Investigating Natural Interaction in Augmented Reality Environments using Motion Qualities

Manuela Chessa and Nicoletta Noceti
Università degli Studi di Genova,
Dept. of Informatics, Bioengineering, Robotics, and Systems Engineering,
Via Dodecaneso 35, Genova, IT-16146, Italy
{*manuela.chessa, nicoletta.noceti*}*@unige.it*

Keywords: Human-computer Interaction, Virtual Reality, Biological Motion Qualities, Reaching Tasks.

Abstract: The evaluation of the users experience when interacting with virtual environments is a challenging task in Human-Machine Interaction. Its relevance is expected to further grow in the near future, when the availability of low-cost portable virtual reality tools will favour a shift – already started – from conventional interaction controllers to a larger use of Natural User Interfaces, where people are asked to use their own body to interact with the device. In this paper, we propose the use of motion qualities to analyze reaching movements, as indicators of the naturalness of the users' actions in augmented reality scenarios. By using such an approach, we compare three different interaction modalities with virtual scenarios, with the goal of identifying the solution that mostly resembles the interaction in a real-world environment.

1 INTRODUCTION

Human-Machine Interaction (HMI) in virtual and augmented reality scenarios will have in the near future an increasing presence in our daily activity. Indeed, in the last year, an growing number of devices for natural interaction have been developed. Those tools have led to a shift from conventional interaction controllers, such as data gloves and motion tracking setup, toward low cost devices. As an example, the Leap Motion[1] is a small sensor device that supports hand and fingers' motions as input and does not requires neither contact nor touching. Supposedly, such kind of sensors become a characterizing trait of applications such as games (Moser and Tscheligi, 2015), immersive 3D modeling (MacAllister et al., 2016), mobile augmented reality application (Kim and Lee, 2016), and rehabilitation (Khademi et al., 2014).

Moreover, virtual reality technology has provided great opportunities for the development of effective assessment techniques for various diseases, because they provide multimodal and highly controllable environments (Iosa et al., 2015). In a virtual world, the patient does not only react to the stimuli, but can actually interact with the computer-generated 3D life-like environment. This provides an entire new realm of possibilities for evaluation and treatments. In order to obtain effective applications of such technologies, work has to be done in order to guarantee an appropriate level of comfort experienced by the end-user. Towards this goal, Natural Human-Machine Interaction (NHMI) aims at creating new interactive frameworks that integrate human language and behavior into technological applications, inspired by the way people work and interact with each other (Baraldi et al., 2009). Such frameworks have to be easy to use, intuitive, entertaining and non-intrusive. As for interaction tasks in the real world, the user should not be asked to use external devices, to wear any kind of tools, or to learn any specific commands or procedures. Therefore, an interesting challenge for NHMI is to make systems self-explanatory by working on their "affordance" (Kaptelinin and Nardi, 2012) and introducing simple and intuitive interaction actions.

From this perspective, one of the main goal of Natural User Interfaces (NUI) is to improve the quality of interaction task within virtual and augmented reality scenarios. As a consequence, there is the need to design qualitative and quantitative evaluations protocols to evaluate the experience as perceived by the user. Most commonly, this task has been addressed in the literature by using qualitative questionnaire (Jaimes and Sebe, 2007), by looking at the absolute positioning error in simple tasks, such as reaching (Solari

[1] www.leapmotion.com

Chessa M. and Noceti N.
Investigating Natural Interaction in Augmented Reality Environments using Motion Qualities.
DOI: 10.5220/0006268401100117
In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017)*, pages 110-117
ISBN: 978-989-758-227-1

et al., 2013), or by looking at physiological measures, such as the variation in the heart rate, when performing the interaction tasks in virtual or augmented scenarios (Chessa et al., 2016). However, if questionnaires or physical measures result in highly subjective feedbacks, the quantitative analysis adopted so far focuses on a very limited part of the interaction, leading to an analysis which only partially evaluate the quality of the interaction itself.

To overcome this limits, in this paper we propose to analyse how actions are performed in simple augmented reality scenarios, by adopting motion qualities to provide an investigation of the overall level of comfort during the interaction based on indicators of the naturalness of the movements. More specifically, we address the problem by considering motion features able to capture on the one hand the *geometrical properties* of the movements – i.e. how the action evolves over time from the spatial point of view – on the other the *dynamic* of the motion – in terms of the hand velocity. We draw inspiration from well known regularities of biological motion, which are the outcome of the fact, as human beings, our movements are constrained by our physicalness. Different, yet related, theories have been formulated, see for instance the *minimun-jerk* and *isochrony* principles (Viviani and Flash, 1995). Among them, we specifically refer to the *Two-Thirds Power Law* – a well-known invariant of upper limb human movements (Viviani and Stucchi, 1992) – which provides a mathematical model of the mutual influence between shape and kinematics of human movements (Greene, 1972).
More specifically, the law describes an exponential relation between velocity and curvature of the motion caused by a physical event, and in the case of human motion (end-point movement in particular) it has been shown that the exponent is very close to the reference value of 1/3. In our work, we adopt the empirical formulation of the law proposed in (Noceti et al., 2015) and and successfully applied to recognise human motion from visual data. Here, we use the obtained statistics to compare the quality of interaction in augmented reality scenarios with a comparable real world scenario. We consider interaction sessions in such scenarios in terms of repetitions of reaching actions towards a common reference target, presented following different visualization strategies. More specifically, we compare a classical 2D and stereoscopic visualization; as for the interaction, the use of a virtual avatar of a hand is compared with more natural the use of the real hand. It is worth noting that we take into account a very simple augmented reality scenario and interaction tasks, in order to focus on the way the action is performed, by limiting the degree of freedoms and the perceptual cues that, of course, influence the movements.

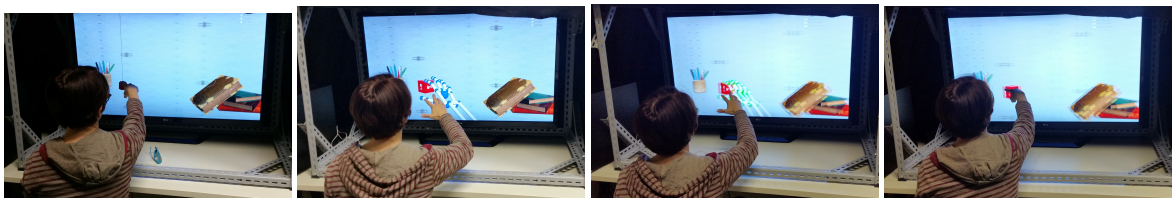
The remainder of the paper is organized as follow. In Sec. 2 we present the augmented reality setup and provide details of the different visualization modalities, which are thoroughly compared in Sec. 3 – where we describe the data collection and the experimental analysis. Sec. 4 is left to conclusions and future lines of research.

2 MATERIAL AND METHODS

2.1 The Augmented Reality Setup

The experimental evaluation described in this paper has been performed by using a setup composed of the following modules:

- Visualization of the virtual scene on a large field of view display, which can be used in both stereoscopic and non-stereoscopic mode. In such a situation, virtual objects appear overlaid onto the real scene (e.g. the desktop, the surroundings of the room). Thus, this setup does not represent an implementation of immersive virtual reality, but an augmented or mixed reality setup. Indeed, virtual and real stimuli coexist at the same time. The virtual scene is designed and rendered by using Unity3D engine.
- Acquisition of the position of the user's hands, by using Leap Motion controller, a small USB peripheral device designed to be placed near a physical desktop. The device is able to track the fine movements of the hands and the fingers in a roughly hemispherical volume of about 1 meter above the sensor at an acquisition frequency of about 120 Hz. The accuracy claimed for the detection of the fingertips is approximately 0.01 mm. However, in (Weichert et al., 2013) the average accuracy of the controller was showed to be about 0.7 mm that allows us to effectively use the Leap Motion in our setup, for the purposes of the experiments. The 3D positions for the 5 fingers, and for the palm center, of each hand, are available. Such information is used both inside the augmented reality environments to render the avatars of the hands in the modalities in which they are required, and saved onto files, for the quantitative evaluation which will be explained in this paper.



(a) RT (b) 2D (c) 3D AV (d) 3D HAND

Figure 1: The augmented reality setup in the 4 modalities (a) Reaching of a real target; (b) Reaching of a virtual cube, visualized monoscopically by using the hand's avatar; (c) Reaching of a virtual cube, visualized stereoscopically by using the hand's avatar; (d) same task of the previous, but by using the users' real hands.

2.2 Visualization and Interaction Modalities

We have devised 4 different scenarios of interaction, with a common virtual background displayed on the monitor and representing a standard office environment, with books and a penholder (see Figure 1). The target to be reached has the form of a small red cube. The scenario has been intentionally kept very simple, in order to better focus on the reaching task towards the cube. Indeed, the aim of the paper is to analyze how a simple action (i.e. to reach an object) is performed, by looking at the 3D trajectories, velocities, and curvatures of some salient points of the users' arms. To analyze human motion in such simple scenario is a good starting point to better understand interaction modalities in more complex AR enviroments.

In the following we provide some details on each visualization strategy:

- *Real target* (henceforth referred to as RT, Figure 1(a)): this modality is expected to provide a baseline, and it is the reference scenario against which the interaction with virtual scenes is compared to. A real cube, whose dimension are $3cm \times 3cm \times 3cm$ is placed in a known position in front of the user.
- *2D target with avatar* (2D, Figure 1(b)): A virtual cube, whose dimensions and 3D positions are the same of the real one is displayed between the user and the display. When the user acts in the scene to perform the required task the avatars of his/her hands are shown. The scene is rendered monoscopically, and only one camera is used inside the game engine. The camera is positioned at the same distance of the user's position (computed as the mean of the distance of his/her eyes) with respect to the monitor, whose position corresponds to the focal plane of the virtual camera.
- *3D target with avatar* (3D AV, Figure 1(c)): differently from the previous case, the scene is rendered stereoscopically. Two cameras with asymmetric frustums to correctly implement the off-axis rendering techniques (Solari et al., 2013) are positioned inside the virtual scene. it is worth noting that the off-axis technique has been ad-hoc implemented, since it is not native in the Unity3D engine. When the user acts in the scene to perform the required task, the avatars of his/her hands are shown.
- *3D target with real hand* (3D HAND,Figure 1(d)): the virtual scene and the stereoscopic rendering are as in 3D AV, but no avatar is shown. The interaction is instead performed with the real user's hands. Conversely to the second and the third scenarios, this modality represents a true "mixed reality" environment, in which a user is asked to interact with virtual objects, by using his/her own body.

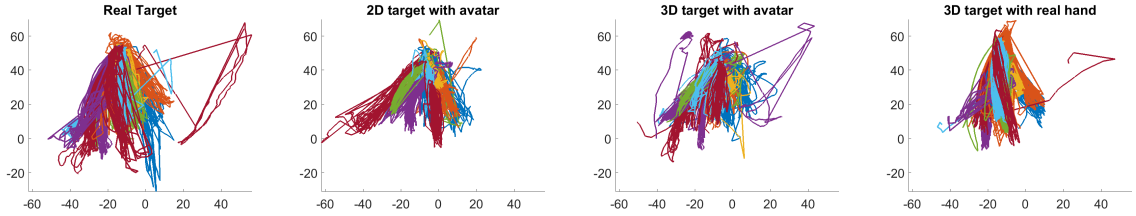
3 EXPERIMENTAL ANALYSIS

In this section we discuss in details the experimental comparison between the different modalities of interaction. After having introduced the data we collected, we will discuss kinematics and dynamic properties of the reaching movements.

3.1 Data Collection

Subjects. 15 subjects, males and females, from 24 to 45 years old, participated in the experiments. They are all confident in the use of technology (everyday use of PC and peripheral) but they did not use the Leap Motion before, neither previously tested the setup. All the subjects have normal or corrected-to normal vision.

Task. All the 4 modalities of visualization and interaction have been tested by all the subjects in a pseudo-random order, to avoid bias effects due to the pre-exposure to a given modality. The subjects



(a) (b) (c) (d)

Figure 2: A visual summary of the collected trajectories on the plane X-Z. The temporal series are color-coded according to the user.

were seated in a fixed position, about 60 *cm* from the monitor, the workspace was represented by a volume of about $100cm \times 60cm$, the leap motion was positioned on a desktop between the user and the monitor. For all the 4 modalities each user was asked to perform 20 reaching movements towards the nearest-upper corner of the cube (virtual or real depending on the considered modality). In the case where the hand's avatar was present, the reaching was performed with the virtual index of the preferred hand. In the other cases, the real index of the preferred hand was used. All the participants were right-handed. No specific constraints were given to the participants about the starting and ending point of each reaching movements, or about the trajectory or velocity. The only reccomandation was to move in the manner they felt more comfortable. To simplify the analysis only the first starting position and the last rest position were fixed. Figure 2 provides a glance of the obtained trajectories (shown in the X-Z plane for better interpretation) from which the variability of the data may be appreciated.

3.2 Estimating Motion Features

The collected trajectories are first analysed in order to detect the *dynamic instants* corresponding to temporal points where (i) a reaching instance starts (henceforth referred to as *starting points*), and (ii) the finger reaches the target (*contact points*). A pair of temporally adjacent starting-contact points delimit a reaching action, where a peak in the velocity can be detected thanks to the well-known bell shapes characterizing biological motion (Morasso, 1981). Figure 3 reports an example where dynamic instants and velocity peaks are marked on a trajectory described with the Z coordinate (left) and on the corresponding tangential velocity (right).

Further, we describe the reaching movements performed by the users in terms of tangential velocity and curvature estimated instantaneously (Noceti et al., 2015). Given a certain 3D location $\mathbf{P}_t = (x_t, y_t, z_t)$ acquired at time t, the velocity vector can be easily obtained as

$$\mathbf{V}_t = (V_t^x, V_t^y, V_t^z) = (x_t - x_{t-1}, y_t - y_{t-1}, z_t - z_{t-1}) \quad (1)$$

Similarly, the acceleration can be obtained as difference between consecutive velocity estimates

$$\mathbf{A}_t = (A_t^x, A_t^y, A_t^z) = (V_t - V_{t-1}, V_t - V_{t-1}, V_t - V_{t-1}). \quad (2)$$

An empirical formulation of the curvature is the following

$$C_t = \frac{||\mathbf{V}_t \times \mathbf{A}_t||}{||\mathbf{V}_t||^3} \quad (3)$$

from which the radius of curvature can be computed as the reciprocal.

It is worth noting that, without loosing in generality, the acquired data have been smoothed with a running average filter to remove noise due to the acquisition setup.

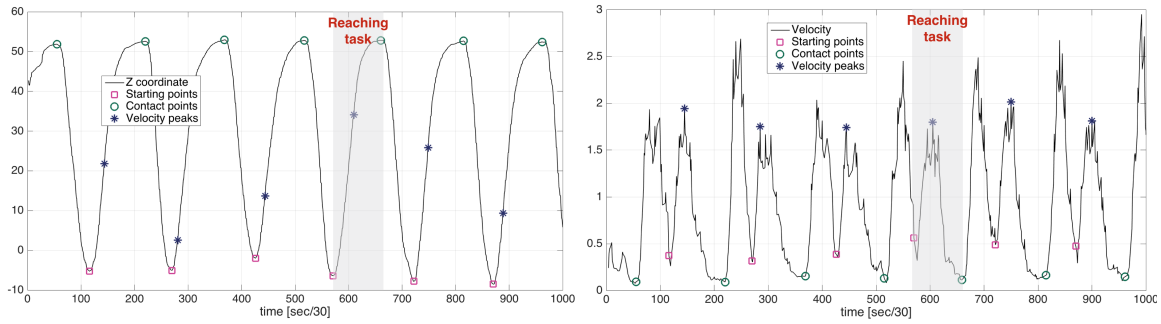
3.3 Motion Analysis

We now present the comparison between different interaction modalities analysing the motion features we introduced in the previous section. Later, we will leverage on the use of the Two-Thirds Power Law as a mean to evaluate motion naturalness.

If not otherwise stated, we adopt Two-way ANOVA methods to compare distributions of measures from the different modalities. The influence of the type of visualization and of the subject are taken into account.

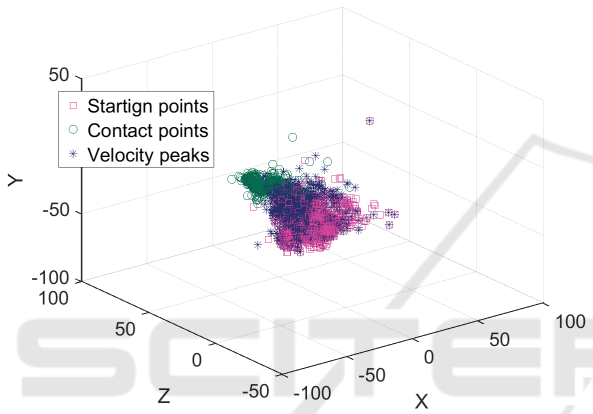
Geometry of the Reaching Task. First, we consider how participants used the space while performing the reaching actions. We report in Figure 4 a visual representation of the distribution in space of dynamic instants. It can be observed how the spread characterizing the area of the starting points (squares) is reduced in the middle of the movement where the velocity peak is achieved (asterisks).

Contact points further shrink in the area corresponding to the target. To provide an overall idea of the spatial extent of the performed movements,



(a) (b)

Figure 3: Dynamic instants and velocity peaks are marked on a trajectory described with the Z coordinate (left) and on the corresponding tangential velocity (right). Starting and contact points are marked with squares and circles, respectively. Peaks velocities are marked with asterisks.



(a)

Figure 4: A visual representation of the distribution in the 3D space of dynamic instants.

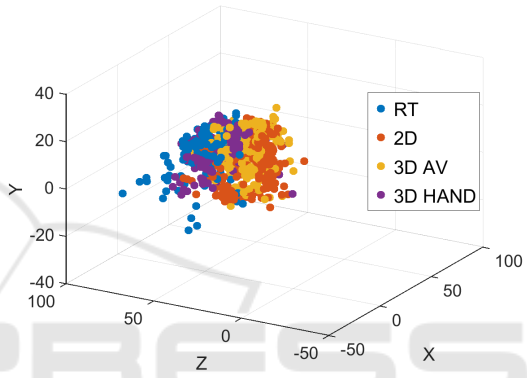


Figure 5: A visualization of vectors connecting starting and contact points (consecutive in time) reveals a clear separation between scenarios where avatar is used and scenarios where the real hand is employed.

we provide a visual representation of the displacements between pairs of temporally adjacent starting and contact instants as 3D points (see Figure 5). It can be noticed a clear separation between observations from RT and 3D HAND with respect to the remaining two scenarios. As a further evidence, we computed the average magnitude for each scenario, obtaining (in *cm*) 51.79 ± 11.30 for RT, 31.77 ± 11.24 for 2D, 34.69 ± 11.30 for 3D AV, and 44.98 ± 10.75 for 3D HAND. Though starting and contact points seem to discriminate between avatar-based modalities with respect to real hand ones, an ANOVA test on the positions of the velocity peaks does not reveal a significant difference between the population of the four scenarios. All the contact points are correctly located in the vicinity of the virtual or real targets.

Figure 6 shows the average reaching errors and their standard deviations, estimated as the distance between contact points and exact target positions. As expected, the lowest errors are for the RT modality, followed by 3D HAND scenario.

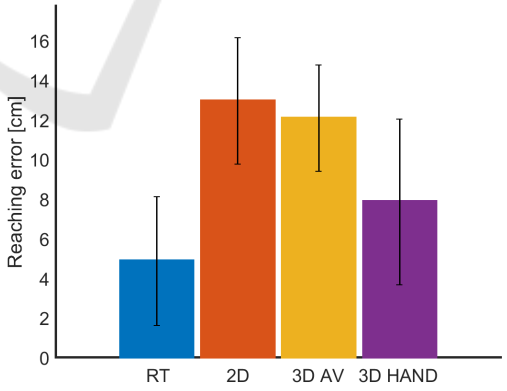


Figure 6: Absolute average errors and associated standard deviation of the target localization (position of the contact point) with respect the known target position.

Such errors are due to different causes: in the real scenario the users could make the cube swing, thus causing an error in the localization of the target. In the steroscopic 3D scenarios, the users'eyes position was not tracked, thus causing a localization error as

explained in (Solari et al., 2013). This does not affect the generality and validity of the experiment.

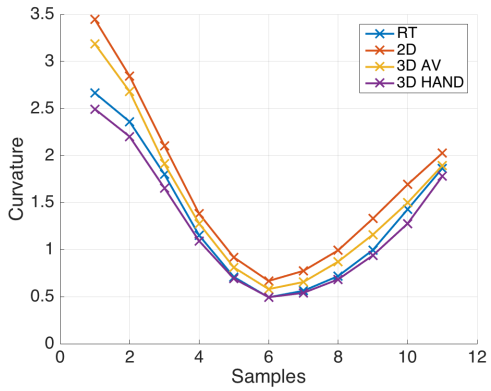
We then consider the populations of curvature (Eq. 3) estimations in the different scenarios. Significant effects of modality $F(3,446.48) = 24.18, P < 0.05$ and subject $F(13,114.55) = 6.2, P < 0.05$ factors have been observed. Also, an interaction between the two factors $F(39,43.25) = 2.34, P < 0.05$ has been detected as statistically significant. More specifically, modalities RT and 3D HAND resulted similar with $P = 0.16$; a similar observation holds for modalities 2D and 3D AV, with $P = 0.96$. An investigation on the statistical distributions of curvature among participants reveals the presence of two subjects whose measures statistically diverge from the rest of the population.

We report in Figure 7(a) the curvature between a pair of starting and contact point, sampled so to have a common number of observations and averaged for each scenario. It can be noticed how modalities RT and 3D HAND are characterized by lower curvature (i.e. more "direct" reaching, related to the minimum-jerk) which may be indicative of higher velocity.

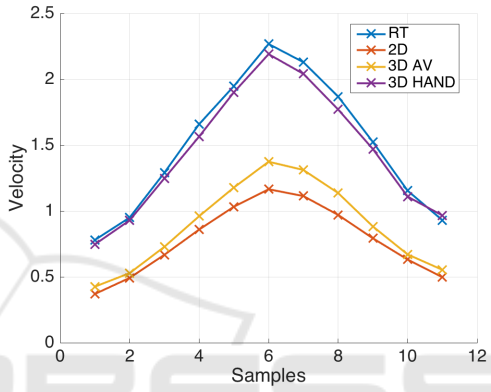
In summary, all the observations consistently indicate that the presence of the avatar tends to spatially constraint the movements of users, providing more compact temporal observations and an expected lower level of naturalness of the motion.

Dynamics of the Reaching Task. We computed the tangential velocity in each point as in Eq. 1. A two-way ANOVA reveals significant effect of the modality factor ($F(3,557.52) = 81.10, P < 0.05$) and of the subjects ($F(13,100.42) = 14.61, P < 0.05$), as well as an interaction between them ($F(39,38.76) = 5.64, P < 0.05$). The comparison among modalities shows strong similarities between RT and 3D HAND ($P = 0.40$) and between the avatar-based modalities ($P = 0.79$). 3 subjects out of the 14 show a statistically significant divergence of their velocity profiles with respect to the rest of the population. In Figure 7(b) we report the sampled velocity bells representing the reaching dynamic and corresponding to the curvature plots in Figure 7(a). The average bells show that 2D, 3D AV, 3D HAND and RT are characterized by an increasing overall velocity, which may be in turn interpreted as an indication of the increasing level of comfort experienced by the users in the scenarios.

Combining Curvature and Velocity. We conclude our analysis by evaluating the pertinence of the observed motion to the Two-Thirds Power Law. We consider the following formulation of the law



(a)



(b)

Figure 7: Average curvature and velocity of reaching actions, sampled to obtain comparable representations of equal length.

$$V_t = k\left(\frac{1}{C_t}\right)^{\beta} \tag{4}$$

and adopted PCA (Jolliffe, 2002) to estimate the function parameters k and β in the linear relation

$$\log(V_t) = \log(k) + \beta \log\left(\frac{1}{C_t}\right) \tag{5}$$

While we can not have a prior on the value of k, we expect β to be close to the reference value of 1/3, which characterise human motion. A divergence with respect to such a value may indicate a decrease in the naturalness of the motion. The Two-Way ANOVA test suggests that both modalities and subjects influence the outcome (with, respectively, $P(3,0.18) = 92.90, P < 0.05$, and $P(13,0.75) = 386.70, P < 0.05$). Also, an interaction between factors is observed ($P(39,0.09) = 44.24, P < 0.05$). Figure 8 reports average and standard deviation of the estimated βs. Only the RT scenario is characterized by a population of βs not significantly different from the distribution with $\frac{1}{3}$ mean. The other scenarios – 3D HAND, 3D AV and 2D – progressively approach

the reference exponent, indicating an increasing level of smoothness of the movements.

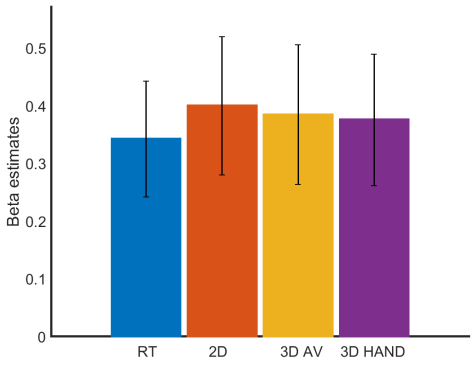


Figure 8: Average and standard deviation of the estimated βs.

3.4 Discussion

In order to provide a further qualitative evaluation of their experience, we asked the participants to provide us with a ranking of their preferences when interacting with the virtual scenarios. Table 1 shows the obtained percentages, from which it is possible to state that the most appreciated modality (in terms of easiness and positive feeling in the interaction) is 3D HAND. Overall, stereoscopic scenarios ranked first in $\sim$ 86% of the cases. The quality of interaction, as self-reported by the users, are thus in line with the outcomes of the analysis we proposed. However, it is worth noting that this last investigation shows a higher variability of the users' acceptance with respect to the considerations resulting from the motion-based quantitative evaluation – where the highest pertinence of 3D HAND to RT is consistently inferred, thus confirming the importance of a qualitative evaluation of interaction to assess AR scenarios.

Table 1: Ranking of preference (in percentages) as self-reported by the users.

	1st	2nd	3rd
3D HAND	**57.14**	21.43	21.43
3D AV	28.57	**57.14**	14.29
2D	14.29	21.43	**64.29**

4 FUTURE DEVELOPMENTS

In this paper, we have used motion qualities to analyze reaching movements, as indicators of the naturalness of the users' actions in augmented reality scenarios. We have considered a simple AR setup, in order to focus on a reaching movement, trough different visualization modalities.

It may be possible that the results are affected by the precision of the Leap Motion. Nevertheless, the use of more precise, and more expensive or invasive, tracking device is out of the scope of the paper and it is far from our idea of obtaining Natural HCI by using low-cost and off-the-shelf devices.

The obtained results show, in a consistent way, that the interaction in a stereoscopic environment by using own real hand is the most similar to the interaction in a real-world scenario. This confirm the validity of the proposed approach.

As a future work we plan the extend the evaluation of interaction naturalness in virtual and augmented scenarios, by considering more complex tasks and richer environments. Moreover, we plan to jointly analyze quantitative motion qualities, subjective users' evaluation, and physiological measures, with the aim of guiding the development of effective Natural Human-Machine Interaction systems.

ACKNOWLEDGEMENTS

The authors would like to thanks all the participants to the data collection.

REFERENCES

Baraldi, S., Del Bimbo, A., Landucci, L., and Torpei, N. (2009). Natural interaction. In *Encyclopedia of Database Systems*, pages 1880–1885. Springer.

Chessa, M., Maiello, G., Borsari, A., and Bex, P. J. (2016). The perceptual quality of the oculus rift for immersive virtual reality. *Human–Computer Interaction*, (in press).

Greene, P. H. (1972). Problems of organization of motor systems. *Progress in theoretical biology*, 2:303–338.

Iosa, M., Morone, G., Fusco, A., Castagnoli, M., Fusco, F. R., Pratesi, L., and Paolucci, S. (2015). Leap motion controlled videogame-based therapy for rehabilitation of elderly patients with subacute stroke: a feasibility pilot study. *Topics in stroke rehabilitation*, 22(4):306–316.

Jaimes, A. and Sebe, N. (2007). Multimodal human–computer interaction: A survey. *Computer vision and image understanding*, 108(1):116–134.

Jolliffe, I. (2002). *Principal component analysis*. Wiley Online Library.

Kaptelinin, V. and Nardi, B. (2012). Affordances in HCI: toward a mediated action perspective. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 967–976. ACM.

Khademi, M., Mousavi Hondori, H., McKenzie, A., Dodakian, L., Lopes, C. V., and Cramer, S. C. (2014). Free-hand interaction with Leap Motion controller for stroke rehabilitation. In *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems*, pages 1663–1668. ACM.

Kim, M. and Lee, J. Y. (2016). Touch and hand gesture-based interactions for directly manipulating 3d virtual objects in mobile augmented reality. *Multimedia Tools and Applications*, pages 1–22.

MacAllister, A., Yeh, T.-P., and Winer, E. (2016). Implementing native support for oculus and leap motion in a commercial engineering visualization and analysis platform. *Electronic Imaging*, 2016(4):1–11.

Morasso, P. (1981). Spatial control of arm movements. *Experimental brain research*, 42(2):223–227.

Moser, C. and Tscheligi, M. (2015). Physics-based gaming: Exploring touch vs. mid-air gesture input. In *Proceedings of the 14th International Conference on Interaction Design and Children*, IDC '15, pages 291–294, New York, NY, USA. ACM.

Noceti, N., Sciutti, A., and Sandini, G. (2015). *Cognition Helps Vision: Recognizing Biological Motion Using Invariant Dynamic Cues*, pages 676–686. Springer International Publishing, Cham.

Solari, F., Chessa, M., Garibotti, M., and Sabatini, S. P. (2013). Natural perception in dynamic stereoscopic augmented reality environments. *Displays*, 34(2):142–152.

Viviani, P. and Flash, T. (1995). Minimum-jerk, two-thirds power law, and isochrony: converging approaches to movement planning. *Journal of Experimental Psychology: Human Perception and Performance*, 21(1):32.

Viviani, P. and Stucchi, N. (1992). Biological movements look uniform: evidence of motor-perceptual interactions. *Journal of experimental psychology: Human perception and performance*, 18(3):603.

Weichert, F., Bachmann, D., Rudak, B., and Fisseler, D. (2013). Analysis of the accuracy and robustness of the Leap Motion controller. *Sensors*, 13(5):6380–6393.