Review Article

Usability Assessments for Augmented Reality Motor Rehabilitation Solutions: A Systematic Review

Virgínia C. Cavalcanti,1 Maria I. de Santana,2 Alana E. F. Da Gama,2 and Walter F. M. Correia1

1Design Department, Universidade Federal de Pernambuco, Recife 50670-901, Brazil
2Department of Biomedical Engineering, Universidade Federal de Pernambuco, Recife 50670-901, Brazil

Correspondence should be addressed to Virgínia C. Cavalcanti; virginiacarrazzone@gmail.com

Received 31 May 2018; Accepted 2 October 2018; Published 1 November 2018

Academic Editor: Ali Arya

Copyright © 2018 Virgínia C. Cavalcanti et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article aims to understand which methods and user assessment approaches are most commonly used in motor rehabilitation studies that use Augmented Reality (AR) applications. The way in which this was performed and discussed was through a systematic review of the area. Firstly, the different uses of AR in these treatments, and the importance of improving usability studies to evaluate these interfaces and their interactions with their users, were discussed. Then, the systematic review of the literature was structured according to previous studies and covered the period from 2007 to September 2017, using the main scientific journals of the area. Of the 666 results collected in the initial search, 32 articles were selected within the planned requirements and scope. These publications were classified by relevance, using the QualSyst evaluation tool for health technology research, and the type of evaluation, approach, and methods used were catalogued. It was found that most of the studies prioritise methods of data collection, such as task execution and performance analysis. However, through QualSyst, it was verified that the best-evaluated studies chose to use a combination of quantitative and qualitative analysis, through the application of methods such as user preference questionnaires, interviews, and observations. The results presented in this review may benefit designers who intend to design and conduct usability assessments for AR within this context.

1. Introduction

The public health industry has a great relevance in today’s social context and a permanent demand for improvement has driven the development of new assistive technologies. In its sector, motor rehabilitation has provided opportunities for physiotherapists and researchers to explore new technological resources to improve motor recovery of patients suffering from strokes or traumatic accidents, or even elderly people in search of better maintenance of their functionalities.

Between these aspects, Augmented Reality (AR) has been one of the technologies that has been widely studied in the health industry. Through the inclusion of virtual objects in the real world, AR promotes the immersion of the user, while still maintaining their sense of presence with the environment. Using virtual elements that simulate movement behaviour, such as collision, reaction, and physical simulation, for example, AR allows credible and immersive interactions for the user to be created [1]. Some examples of situations for the positive use of this technology are the possibility of training for health professionals and students, since it allows the understanding of spatial relationships and concepts, providing substantial, contextual, and situational learning experiences [2], in addition to providing motivation for patients during therapies [3], and to reduce risks during the exercise learning process [4].

AR makes games and applications an attractive alternative for healthcare, because it provides a variety of applications with immersion, interaction, and involvement characteristics according to its purpose [5]. Games have been incorporated into health solutions to facilitate user’s engagement, by using these interactive technologies to move
affected body areas (such as arms or hands) and encourage the repetitive practice of motor tasks necessary to stimulate the neuroplastic changes responsible for the recovery. It also encourages the motivational aspects and interactions that allow cognitive engagement and challenge [6]. In physiotherapy, the use of motion tracking devices to create gaming applications has provided an improvement in the patient's commitment, motivating the movement sequences established by the professional during the treatment sessions [7].

While researching, it appears that there are some difficulties that arise for the use of game therapies. Burke et al. (2009) describe some pertinent user problems; namely, they have paralysis in one or more areas of their body, have difficulty communicating, and are not familiar with the equipment and it is not uncommon to suffer from depression and therefore they may find it difficult to focus on a therapy program and not get involved with gambling. To encourage these users, it is suggested that games designed for rehab should encourage engagement and reward success. When users deal with failure positively, they are more likely to remain engaged and not feel that failure in the game results from their more physical abilities [6].

Thus, it is necessary to work on the elements of the interface so that they are clear to understand and provide the appropriate feedback to the patients. To achieve this objective, it is necessary to carry out investigations and accurate analysis in usability evaluations. However, there is no consensus in the literature of a methodological approach that is adequate to the diversity of patient's conditions. In the previous evaluation of mixed reality [8], a number of problems were analysed and no usability methods specifically designed for these systems were found. At the time of the survey [8], three categories of methods were identified that were sufficiently general in their approach and that could be appropriate for these systems: questionnaires and interviews, inspection methods, and user testing.

Questionnaires and interviews would be useful resources for gathering subjective data, user preferences, and comparing performance data. On the other hand, the methods of inspection would be more limited since there were not yet guidelines of ergonomics and consolidated design for specific systems of mixed reality. And the user test was the most used method in this area [9].

This review proposes to investigate how user tests on rehabilitation AR applications are being planned and conducted. It aims to understand what the most frequent user profile is, what methods have been used, and how the collected data is analysed. In the next section, this paper describes the reviews used to support the survey and states the objectives of this research (Section 2). The research methodology will follow, where the structure for item and filter classification will be explained (Section 3). Thereafter, the results are presented (Section 4) including study selection through the Prism Flow, review analysis, and synthesis. To finish, a brief discussion about review finding is performed and is followed by the paper's conclusion (Sections 5 and 6).

2. Literature

Swan et al. (2005) [10] surveyed 1104 papers published in the most relevant studies between 1992 and 2004, where only 266 described research for AR. 38 of them addressed HCI (human-computer interaction) aspects and only 21 described a formal user-based experience. At the time, they identified and classified user-based experimentation along three lines: “human perception and cognition,” “user task performance,” and “generic interaction and multi-user communication (collaboration).”

Years later, Bai and Blackwell (2012) [11] analysed 71 articles, from 2001 to 2010 that deal specifically with usability assessments. They proposed to group the evaluation types according to 4 categories of usability research, 3 of them previously worked by Swan et al. (2005) [10], “performance,” “perception and cognition,” and “collaboration” and included “user experience,” as did Dünser et al. (2008) [9]. Both studies included the fourth category by detecting that some publications intended to analyse usability with a more subjective approach, in order to discover more individual feelings and experiences of users while testing the system or prototype.

Dünser et al. (2008) [9] selected 161 articles, on a basis of 6071 publications, related to user ratings and also included a rating on the following categories from approach to evaluation: “objective measures,” “subjective measures,” “qualitative analysis,” “usability assessment techniques,” and “informal assessments.” They found that over the years the proportion of formal user evaluations compared to informal evaluations had increased, and between 1995 and 2001 there was an average of 57% of formal evaluations, yet between 2002 and 2007 this percentage increased to 76%. They concluded that there is a growing need to formalise the evaluation process and conduct properly designed user studies.

Dey et al. (2018) [12] selected 291 works, in the 9 years from 2005 to 2014, that were reviewed and classified based on their areas of application. Their main contribution was to show how the scenario changed over the years, identifying areas where there were few studies with users and opportunities for future research. The health area was identified as the most relevant.

These user tests were concentrated within the period of 2002 and 2016 where the majority of the usability tests (83%) were performed with healthy people and the others (17%) involved patients or a combination of users with differing abilities [13]. Demographic data collected from the articles conducting these tests showed that 62% of them tested young participants, mostly university students with an average age of 30 years [12]. This could lead to deficiencies in the analysis and contradict studies [14] that recommend applying user-centred design to the development of interactive technologies in health: including the biased sampling, lack of user experience studies, and noninclusion of patients in the tests. This user-centred design could improve the functionality for systems and usability for patients, increasing the probability of promoting the desired behaviours and health outcomes [14].

These previous reviews [9–13] provided a starting point and a reference for structuring this study. Through the
3. Methodology

This systematic review aims to investigate how user testing with AR applications for rehabilitation is being planned and conducted. Through the results and analysis, it will be possible to reach conclusions on the knowledge collected from these publications [15], which may serve to define the methods of evaluation or the profile of the sampling used in the tests.

To achieve this, papers were searched, selected, evaluated, analysed, and synthesized according to the protocol described below. To organise the data, a checklist of recommendations for systematic PRISMA reviews were used [16], and the publications were counted and catalogued at each phase.

3.1. Search Question. To help define the purpose of the review PICO’s approach to structure and define the research question was used [17]. The question was formulated to fit the four components that make up the acronym PICO: population (P); intervention (I); control (C); outcome (O). The details for each are as follows:

(i) Population: the intended audience required the inclusion of users who needed some treatment or follow-up treatment for motor rehabilitation, whose disabilities included accidents, neurological problems, surgical recovery, and motor limitations, or involving profiles related to the development of applications such as physiotherapists, healthy adults, and developers.

(ii) Intervention: make the use of AR to assist the treatments.

(iii) Control: studies that have control groups for tests performed with users, such as performance tests, comparisons between different profiles or groups, a comparison between different applications or previous experiments, etc.

(iv) Outcome: when the results of the evaluations were supported in the data for greater reliability, but not excluding empirical analysis.

In order to identify trends in this area and compare them with evidence found in previous reviews, the following research question was defined: “What and how methodological approaches have been used in usability assessment research for motor rehabilitation applications, with the use of Augmented Reality?”

3.2. Research Protocol. This review prioritises peer-reviewed journals in the scientific community that are mentioned in other review articles [3, 9, 18] or are chosen due to their popularity in AR-related research. There are four databases that were used: IEEE Xplorer, Springer Link, Science Direct, and ACM Digital Library.

The keywords included in the searches were written in English and concatenated with “AND”: “Augmented Reality” AND “rehabilitation”. The search was performed by two evaluators and the inclusion criteria were separated into two phases: “Screening” and “Eligibility”, which are detailed below (Sections 3.2.1 and 3.2.2). For complete reading and classification, articles that had conflicting approval among researchers were read and selected by a third reviewer.

The final publications were thoroughly examined to provide a score for the relevance of the methodological processes and for the description of the evidence found. QualSyst’s [19] evaluation criteria that were developed specifically to evaluate health items were used to generate these scores (Section 3.2.3). They were catalogue according to the classification (Section 3.2.3).

3.2.1. Screening. The initial screening phase was performed by focusing on paper characteristics and through the reading of the titles and abstract. The inclusion criteria were put in place to investigate an application in Augmented Reality; to perform user tests about the application use and benefits; to be written in English, Portuguese, Spanish, and French; and to be published in the last 10 years (2007 to 2017). The exclusion criteria were short papers (articles with 4 pages or less) and systematic reviews of the literature. The identification process of these publications was accomplished through reviewing their title and abstract.

After being selected, each evaluator downloaded and catalogued articles in folders using the Mendeley Desktop© [20]. The program allowed the evaluators to work individually, checking for duplicate articles and including “select” and “out” tags to identify the articles that would be selected or excluded for the next phase.

3.2.2. Eligibility. This phase identified articles by evaluating research methodology and conclusions. It used the following criteria: to be an experimental research, that is, to involve some type of experiment, where the researcher participates actively in conducting the phenomenon, process, or fact evaluated, acting on the cause, modifying it, and evaluating the changes in the outcome [21]; to be a quantitative research and may simultaneously use a qualitative approach; to have a control group in the sampling or to use criteria for validation of the data found. The exclusion criterion was whether the research experiments were exclusively performance-tested or technical evaluation of system/software/robot/tracking.

3.2.3. Classification. To catalogue the publications by relevance, QualSyst [19], an evaluation tool for health technology researches, was used. It consists of 14 specific questions which compose a note that reflects the quality of the evidence found. QualSyst evaluation criteria looked at many
aspects, for example, (i) research objective description, (ii) adequate study design, (iii) analytical methods which are well described and justified, (iv) sample size, (v) randomization or use of control groups, and (vi) sufficiently detailed results. To perform QualSyst grade composition, the reviewers give points to each question according to their own evaluation and understanding, being “yes” (2 points), “partially” (1 point), “no” (0), and “N/A” when the question does not apply to the study analysed. The score was calculated by summing the total points obtained across relevant items and dividing by the total possible score. The final score is given as a percentage based on the 3 scores given.

After evaluating research quality with QualSyst, the next step was to classify the publications into the “Classification Table.” Among the information collected the following categories were recorded: part of the body or movement worked on the studies; user’s disabilities; the number of individuals who participated and their profile (patients, physiotherapists, developers, and healthy people); type of usability evaluation and methods used.

(I) Type of Assessment Classification. User-based experimentation progresses along three lines of enquiry: (1) those that aimed to understand how human perception and cognition operate in AR contexts; (2) those that examine the performance of the user’s task within specific AR applications; and (3) those that examine the generic user interaction and communication between multiple collaborating users [10].

User experience (UX) was included as a category in other systematic reviews [9, 11] as some studies did not necessarily involve the measurement of performance of user tasks, but instead used other ways of identifying problems with system usability and were identified by using two groups of documents addressing UX. In the first, the main focus of the evaluation was in the user experience itself, with the purpose of evaluating their attitudes and acceptance of the system. In the second, the main focus of the evaluation was on perception, performance, or collaboration, using UX as a supplemental evaluation measure [11]. In this way, the classification of the publications found was made in four categories:

(i) User task performance studies: these studies measure the performance of tasks with users in specific AR applications in order to gain an understanding of how technology can affect the underlying tasks;

(ii) Perception and cognition studies: experiments that study low-level tasks, with the aim of understanding how human perception and cognition operate in AR contexts;

(iii) Collaboration studies: experiments that evaluate the generic interactions between the user and the communication among multiple participating users;

(iv) User experience: study the subjective feelings and experiences of users. It can be presented in two ways: formal and informal. Formal evaluations involve controlled experiences with a fixed sample of voluntary users and collect participants’ experiences with structured surveys or questionnaires. Informal evaluations involve unstructured interviews or observations with a random sample of potential users or domain experts.

(2) Evaluation Approach Classification. The research methods used to classify the evaluation approach were grouped into 5 categories: subjective measures, objective measures, qualitative analysis, usability assessment techniques, and informal assessments [9], as described below:

(i) Objective measurements: studies that include objective measures, the most common being task completion times, accuracy, and error rates. In general, they employ a statistical analysis of the measured variables, though only a few include a descriptive analysis of the results.

(ii) Subjective measurements: publications that study user questionnaires, user classification, or subjective judgments. Regarding the analysis, some of these studies may also employ statistical analysis of the results, whereas others only include a descriptive analysis.

(iii) Qualitative analysis: studies with formal user observations, formal interviews, or classification or analysis of user behaviour (e.g., speech or gesture coding).

(iv) Usability evaluation techniques: publications that employ evaluation techniques that are frequently used in interface usability assessments, such as heuristic evaluation, expert-based assessment, task analysis, and Think Aloud.

(v) Informal evaluations: assessments such as informal observations of users or informal collection of user comments.

4. Results

This section offers the results of the review, which along with paper selection, synthesis, and analysis were demonstrated within their categories, starting with evaluation type, methods used in the tests, profile of the users, and lastly, analysis of the QualSyst score.

4.1. Multilevel Filtering of Studies. Following the protocol described, 666 articles were found, of which 32 remained for systematic review after selection. The number of articles in each step can be seen in Figure 1 based on the PRISMA protocol [16].

In the “Identification” phase (Figure 1), 666 publications were found with the “Augmented Reality” AND “rehabilitation” keywords, and 77% of them (557 articles) were excluded because they did not meet the requirements in the “Screening” phase. Thus, a total of 150 publications were selected, of which 41 were selected simultaneously by
the two reviewers and only counted once, resulting in 109 publications.

In the “Eligibility” phase (Figure 1), of the 109 articles, only 76 remained within the requirements of the selection (Section 3.2.2), of which 21 were chosen by the two reviewers, thus passing 55 to the reading phase and classification. At the end of this selection, 21 were selected by both evaluators and 34 were conflicting. To resolve the conflicts, a third reviewer evaluated and approved 11 of these publications. And so, 32 articles were taken forward for reading and classification (Tables 1 and 2).

4.2. Synthesis and Analyses

4.2.1. Type of Evaluation. “Task Performance” was the most common theme that came up from the results, present in 22 of the studies (51%), followed by 15 studies (35%) that undertook “User Experience” (UX) investigations. Figure 2 shows the type of evaluation through the frequency in which they appear. Of these 15 articles, only 5 evaluated UX and its subjective aspects during the use of the solutions, such as ease of use, utility, and attitude [22]; improvement in the frequency of physical activities [23]; levels of effectiveness, efficiency, and satisfaction [24]; motivation [25]; and preference and immersion between different systems and classical therapy [26]. Another 5 studies [27–31] performed data collection and analysis that configured both types of evaluation.

Lastly, 6 studies (14%) set out to investigate the “Perception and cognition” of users in relation to the technologies studied; 5 of them were also associated with UX. None were found that dealt with “Collaboration” solutions.

4.2.2. Evaluation Approach. An evaluation can use quantitative or qualitative data and often includes both. Both methods provide important information for evaluation, and both can improve community engagement. These methods are rarely used alone; combined, they generally provide the best overview of the project.

The “Evaluation Approach” was performed in 23 articles using objective measures (46%) and 18 with subjective measures (36%) (Figure 3). Of these, 10 made use of both
<table>
<thead>
<tr>
<th>n°</th>
<th>Title</th>
<th>Author</th>
<th>Evaluation Type</th>
<th>Evaluation Approach</th>
<th>Method used</th>
<th>Subjects</th>
<th>Body part and Movements used to control application</th>
<th>Type of pathologies</th>
<th>QualSyst Quantitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Visual manipulations for motor rehabilitation</td>
<td>Regenbrecht, Holger</td>
<td>perception and cognition; UX</td>
<td>subjective measures; qualitative analysis; informal evaluations</td>
<td>task execution; questionnaire</td>
<td>(A) 24 participants; (B) 29 participants; (C) 43 participants</td>
<td>treatment of upper limb impairments</td>
<td>stroke</td>
<td>97.92%</td>
</tr>
<tr>
<td>2</td>
<td>Out of reach? - A novel AR interface approach for motor rehabilitation</td>
<td>Regenbrecht, Holger</td>
<td>UX</td>
<td>subjective measures</td>
<td>questionnaire; interview; observation</td>
<td>43 non-clinical participants</td>
<td>Hand and finger</td>
<td>stroke</td>
<td>95.83%</td>
</tr>
<tr>
<td>3</td>
<td>Humans Can Integrate Augmented Reality Feedback in Their Sensorimotor Control of a Robotic Hand</td>
<td>Clemente, Francesco</td>
<td>task performance</td>
<td>objective measures</td>
<td>task execution; performance analysis</td>
<td>8 participants</td>
<td>hand, arm, prosthesis users</td>
<td>stroke</td>
<td>91.67%</td>
</tr>
<tr>
<td>4</td>
<td>AR-based serious game framework for post-stroke rehabilitation</td>
<td>Hossain, M. Shamim</td>
<td>task performance</td>
<td>objective measures</td>
<td>task execution; performance analysis; questionnaire</td>
<td>11 patients</td>
<td>Motor function of the upper extremity of stroke patients, Stroke and Guillain–Barré syndrome</td>
<td>Stroke and Guillain–Barré syndrome</td>
<td>91.67%</td>
</tr>
<tr>
<td>5</td>
<td>Effect of a mixed reality-based intervention on arm, hand, and finger function on chronic stroke</td>
<td>Colomer, Carolina</td>
<td>task performance; UX</td>
<td>objective and subjective measures</td>
<td>task execution; performance analysis</td>
<td>30 patients</td>
<td>upper limb; flexion and extension of the elbow, wrist, and fingers</td>
<td>stroke</td>
<td>89.58%</td>
</tr>
<tr>
<td>6</td>
<td>MirrARbilitation: A clinically-related gesture recognition interactive tool for an AR rehabilitation system</td>
<td>Da Gama, Alana Elza Fontes</td>
<td>task performance; UX</td>
<td>objective and subjective measures</td>
<td>performance analysis; questionnaire</td>
<td>33 users (11 patients, 11 physiotherapists, and 11 software developers)</td>
<td>shoulder abduction movement</td>
<td>Stroke and Mastectomy</td>
<td>89.58%</td>
</tr>
<tr>
<td>7</td>
<td>Touch and hand gesture-based interactions for directly manipulating 3D virtual objects in mobile augmented reality</td>
<td>Kim, Minseok</td>
<td>task performance; UX</td>
<td>objective and subjective measures</td>
<td>task execution; performance analysis; questionnaire</td>
<td>20 participants</td>
<td>Hands, fingers</td>
<td>Was not informed</td>
<td>88.64%</td>
</tr>
<tr>
<td>8</td>
<td>Fractal Fluctuations in Human Walking: Comparison Between Auditory and Visually Guided Stepping</td>
<td>Terrier, Philippe</td>
<td>task performance</td>
<td>objective measures</td>
<td>task execution; performance analysis</td>
<td>36 healthy individuals</td>
<td>Human walking; the muscles of the lower limbs</td>
<td>Fractal fluctuations in human locomotion</td>
<td>87.50%</td>
</tr>
<tr>
<td>n°</td>
<td>Title</td>
<td>Author</td>
<td>Evaluation Type</td>
<td>Evaluation Approach</td>
<td>Method used</td>
<td>Subjects</td>
<td>Body part and Movements used to control application</td>
<td>Type of pathologies</td>
<td>QualSyst Quantitative</td>
</tr>
<tr>
<td>----</td>
<td>-----------------------------------------------------------------------</td>
<td>-------------------------------</td>
<td>-----------------</td>
<td>---------------------</td>
<td>------------------------</td>
<td>--------------------------------------------------</td>
<td>-----------------------------------------------------</td>
<td>--------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>9</td>
<td>Augmented reality in occupational therapy</td>
<td>Dionisio Correa, A.G.</td>
<td>UX</td>
<td>subjective measures</td>
<td>questionnaire</td>
<td>16 patients and 2 therapists.</td>
<td>upper limb rehabilitation</td>
<td>Muscular Dystrophy</td>
<td>82.21%</td>
</tr>
<tr>
<td>10</td>
<td>Computerized Mirror Therapy with Augmented Reflection Technology for Stroke Rehabilitation</td>
<td>Hoermann, S.</td>
<td>perception and cognition; UX</td>
<td>subjective measures; informal evaluations</td>
<td>questionnaire; observation</td>
<td>5 patients</td>
<td>mirror hand (&quot;mirror box&quot;)</td>
<td>Stroke</td>
<td>81.25%</td>
</tr>
<tr>
<td>11</td>
<td>Guidance and movement correction based on therapeutics movements for motor rehabilitation support systems</td>
<td>Gama, Alana Da</td>
<td>task performance, UX</td>
<td>objective and subjective measures</td>
<td>task execution, performance analysis, questionnaire</td>
<td>3 physiotherapist, 4 adults and 3 elderly subjects</td>
<td>Shoulder Abduction</td>
<td>shoulder injuries, neurologic and geriatric therapies</td>
<td>80.68%</td>
</tr>
<tr>
<td>12</td>
<td>Augmented reality-based postural control training improves gait function in patients with stroke: Randomized controlled trial</td>
<td>Lee, Chi-Ho</td>
<td>task performance</td>
<td>objective and subjective measures</td>
<td>task execution, performance analysis</td>
<td>21 patients</td>
<td>stroke</td>
<td>79.97%</td>
<td></td>
</tr>
</tbody>
</table>

Many positions: The first stage includes six subordinated exercise programs that were conducted without the use of any tool in a lying position. The second stage involved four subordinated exercise programs performed while sitting. The third stage consisted of six subordinated exercise programs in the standing position performed using a therapeutic ball or a foothold.
<table>
<thead>
<tr>
<th>n°</th>
<th>Title</th>
<th>Author</th>
<th>Evaluation Type</th>
<th>Evaluation Approach</th>
<th>Method used</th>
<th>Subjects</th>
<th>Body part and Movements used to control application</th>
<th>Type of pathologies</th>
<th>QualSyst</th>
<th>Quantitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>Mixed Reality Serious Games: The Therapist Perspective</td>
<td>Ines, Di Loreto</td>
<td>UX</td>
<td>subjective measures</td>
<td>task execution, questionnaire</td>
<td>3 therapists</td>
<td>upper/bottom limb</td>
<td>Could be applied to post stroke patients.</td>
<td>77.27%</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Visual Occlusion in an Augmented Reality Post-Stroke Therapy Scenario</td>
<td>Allen, Max</td>
<td>perception and cognition, UX</td>
<td>subjective measures, usability evaluation techniques</td>
<td>task execution, questionnaire, observation, think aloud</td>
<td>30 participants</td>
<td>Hand, upper limb</td>
<td>stroke</td>
<td>77.08%</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Augmented reality technology combined with three dimensional holography to train the mental rotation ability of older adults</td>
<td>Lee, I-Jui</td>
<td>perception and cognition, UX</td>
<td>objective and subjective measures, usability evaluation techniques</td>
<td>task execution, performance analysis, questionnaire, interview, observation, think aloud</td>
<td>28 elderly but healthy users.</td>
<td>Mental spatial rotation</td>
<td>Neurobiological research associates it with the human ability to navigate and with other spatially related functions</td>
<td>75.00%</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Interactive augmented reality using Scratch 2.0 to improve physical activities for children with developmental disabilities</td>
<td>Lin, Chien-Yu</td>
<td>UX</td>
<td>qualitative analysis</td>
<td>performance analysis, observation</td>
<td>3 children (2 females: 4 and 6 years; and 1 male 3 years)</td>
<td>A body motion interactive game to enhance the body strength of children with disabilities.</td>
<td>mildly intellectually retarded and seldom talks; severe cerebral palsy; has moderate multiple disabilities, weak legs and low vision</td>
<td>74.05%</td>
<td></td>
</tr>
<tr>
<td>n°</td>
<td>Title</td>
<td>Author</td>
<td>Evaluation Type</td>
<td>Evaluation Approach</td>
<td>Method used</td>
<td>Subjects</td>
<td>Body part and Movements used to control application</td>
<td>Type of pathologies</td>
<td>QualSyst Quantitative</td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>-----------------------------------------------------------------------</td>
<td>-------------------------------</td>
<td>------------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td>---------------------------------</td>
<td>------------------------------------------------------</td>
<td>-----------------------------------------------------</td>
<td>-----------------------</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Functional gait rehabilitation in elderly people following a fall-related hip fracture using a treadmill with visual context: design of a randomized controlled trial.</td>
<td>MW, van Ooijen</td>
<td>task performance</td>
<td>objective measures</td>
<td>performance analysis</td>
<td>Estimated: 126 older adults</td>
<td>walking ability and reducing fall incidence</td>
<td>injuries</td>
<td>74.04%</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>A Kinect-Based System for Lower Limb Rehabilitation in Parkinson’s Disease Patients: a Pilot Study</td>
<td>Palacios-Navarro, Guillermo</td>
<td>task performance</td>
<td>objective measures</td>
<td>task execution, performance analysis</td>
<td>(A) 7 healthy subjects; (B) 7 patients</td>
<td>Lower Limb; lateral leg movements</td>
<td>Parkinson's Disease Patients with idiopathic</td>
<td>67.42%</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>SleeveAR: Augmented reality for rehabilitation using real-time feedback</td>
<td>Sousa, M et al</td>
<td>task performance, perception and cognition</td>
<td>objective and subjective measures, informal evaluations</td>
<td>task execution, performance analysis, questionnaire, interview</td>
<td>18: 14 male and 4 female. Average 26 years old</td>
<td>Upper arm and forearm</td>
<td>injuries</td>
<td>66.67%</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Visually-guided gait training in paretic patients during the first rehabilitation phase: study protocol for a randomized controlled trial</td>
<td>Rossano, Cathia</td>
<td>task performance</td>
<td>objective measures</td>
<td>performance analysis</td>
<td>The current estimation is that a total of 70 to 100 participants will be recruited.</td>
<td>Human walking; the muscles of the lower limbs</td>
<td>Stroke</td>
<td>66.67%</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>An augmented reality home-training system based on the mirror training and imagery approach</td>
<td>Trojan, Jörg</td>
<td>task performance</td>
<td>objective measures</td>
<td>task execution, performance analysis</td>
<td>7 healthy participants</td>
<td>hand training; treatment for phantom limb pain</td>
<td>Several clinical conditions, such as phantom limb pain, stroke and complex regional pain syndrome.</td>
<td>63.64%</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Walking adaptability therapy after stroke: study protocol for a randomized controlled trial</td>
<td>Timmermans, Celine</td>
<td>task performance</td>
<td>objective measures</td>
<td>task execution, performance analysis</td>
<td>40 persons after stroke (≥3 months)</td>
<td>Human walking</td>
<td>Stroke</td>
<td>62.05%</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Data modeling mobile augmented reality: integrated mind and body rehabilitation</td>
<td>Hsiao, Kuei Fang</td>
<td>UX</td>
<td>subjective measures</td>
<td>questionnaire, interview</td>
<td>51 senior citizens</td>
<td>Wrist &amp; Arm, Neck and Knee</td>
<td>Alzheimer</td>
<td>60.42%</td>
<td></td>
</tr>
<tr>
<td>n°</td>
<td>Title</td>
<td>Author</td>
<td>Evaluation Type</td>
<td>Evaluation Approach</td>
<td>Method used</td>
<td>Subjects</td>
<td>Body part and Movements used to control application</td>
<td>Type of pathologies</td>
<td>QualSyst</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>----------------------------------------------------------------------</td>
<td>--------------</td>
<td>-----------------</td>
<td>--------------------</td>
<td>-----------------------------------------------</td>
<td>--------------------</td>
<td>-------------------------------------------------</td>
<td>---------------------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Augmented reality-based training system for hand rehabilitation</td>
<td>Liu, Jia</td>
<td>task</td>
<td>objective and subjective measures</td>
<td>performance analysis, questionnaire</td>
<td>20 healthy subjects</td>
<td>hand rehabilitation</td>
<td>Stroke</td>
<td>58.33%</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Hand tracking and trajectory analysis for physical rehabilitation</td>
<td>Boato, G.</td>
<td>task</td>
<td>objective measures</td>
<td>task execution, performance analysis</td>
<td>was not informed</td>
<td>Hand in injuries or cognitive disabilities.</td>
<td></td>
<td>47.73%</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Design and Assessment of a Remote Vibrotactile Biofeedback System for Neuromotor Rehabilitation Using Active Markers</td>
<td>Montaño-Murillo, R.</td>
<td>task</td>
<td>objective measures</td>
<td>performance analysis</td>
<td>8 patients</td>
<td>Upper limbs (shoulders, elbows or wrists)</td>
<td>injuries</td>
<td>45.08%</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Evaluation the Post-Stroke Patients Progress Using an Augmented Reality Rehabilitation System</td>
<td>Alamri, Atif</td>
<td>task</td>
<td>objective measures</td>
<td>task execution, performance analysis</td>
<td>15 healthy subjects</td>
<td>Hand, upper limb</td>
<td>Stroke</td>
<td>43.75%</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Manipulating the experience of reality for rehabilitation applications</td>
<td>Regenbrecht, Holger</td>
<td>perception and cognition, UX</td>
<td>subjective measures, qualitative analysis, informal evaluations</td>
<td>questionnaire</td>
<td>(A) 24 participants; (B) 30 participants; (C) 43 participants; (D) 2 tests: 23 participants and 30 participants; (E) 100 physiotherapists, who evaluated the system in the role as a therapist and as a patient</td>
<td>Hands and Fingers</td>
<td>Stroke</td>
<td>43.18%</td>
<td></td>
</tr>
<tr>
<td>n°</td>
<td>Title</td>
<td>Author</td>
<td>Evaluation Type</td>
<td>Evaluation Approach</td>
<td>Method used</td>
<td>Subjects</td>
<td>Body part and Movements used to control application</td>
<td>Type of pathologies</td>
<td>QualSyst</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>----------------------------------------------------------------------</td>
<td>-------------------</td>
<td>------------------</td>
<td>---------------------</td>
<td>-----------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------</td>
<td>--------------------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>A Gesture Control System to Support Rehabilitation Exercises</td>
<td>Sousa, Kleber A.</td>
<td>task performance</td>
<td>objective measures</td>
<td>task execution, performance analysis</td>
<td>10 participants (19 to 24 years, 8 men e 2 women)</td>
<td>15 joints injuries</td>
<td></td>
<td>40.91%</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>AR-Rehab: An augmented reality framework for post stroke-patient rehabilitation</td>
<td>Alamri, Atif</td>
<td>task performance</td>
<td>objective and subjective measures</td>
<td>task execution, performance analysis, questionnaire, observation</td>
<td>15 subjects poststroke-patient rehabilitation of his / her hands and arms</td>
<td>Stroke</td>
<td></td>
<td>37.50%</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Cloud-Based Rehabilitation Exergames System</td>
<td>Hoda, Mohamad</td>
<td>task performance</td>
<td>objective and subjective measures</td>
<td>performance analysis, questionnaire</td>
<td>6 subjects upper limb</td>
<td>Stroke</td>
<td></td>
<td>37.50%</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Rehabilitation exercise with real-time muscle simulation based EMG and AR</td>
<td>Aung, Yee Mon</td>
<td>task performance</td>
<td>objective measures</td>
<td>questionnaire</td>
<td>5 healthy subjects Shoulder</td>
<td>Traumatic Brain Injury (TBI), Spinal Cord Injury (SCI) and Stroke or Cerebrovascular Accident (CVA)</td>
<td></td>
<td>25.00%</td>
<td></td>
</tr>
</tbody>
</table>
4.2.3. Evaluation Methods. The evaluation methods are the resource the researchers use to collect data in their experiments. They can be used alone or in combination to provide better evidence for the analysis. Among the most used methods, 91% of the selected papers chose to make a combination of them, the most common being "Execution of Tasks" and "Performance Analysis," appearing in 16 articles (Figure 4(a)).

There were only 3 that chose to use solely questionnaires [25, 34, 35]. Two of these received low scores in the QualSyst analysis [34, 35], because they did not sufficiently describe the sampling, and the results were not well defined or robust to the research proposal. Of the 6 articles that made "Observations," half of them (3) also had "Performance analysis" for data collection [23, 33, 36], and only 1 did not apply a questionnaire [23].

"Performance analysis" was the most applied method, appearing in 23 of the 32 studies, accounting for 32% of the methods found (Figure 4(b)). "Execution of tasks" appeared in 19 publications (26%) and questionnaires in 18 (25%), rounded off by "Observations" (8%), "Interviews" (6%), and "Think Aloud" (3%). No heuristic evaluations with specialists were performed in the selected papers.

4.2.4. Participants Profile. Within the 32 selected articles, there were a total of 806 participants with an average age of 39.89 years, distributed through a proportion of 37% men, 23% women, and 40% unknown (Figure 5). Of the 32 articles, 14 did not report the participants’ ages, 2 reported within an age range (20-30 years and 27-35 years) [26, 37], and 3 articles reported the mean ages per group [22, 28, 29]. From the included papers, 3 presented study protocols [38–40] to be completed during the research.

User profile investigations included a mixture of healthy users (59%), patients (34%), and studies that investigated both profiles (7%) (Figure 6). Of the 10 studies that reported the mean age, 6 used individuals under 30 years old and the remaining 4 were over the age of 58. Studies which included older users investigated health seniors to study the negative effects of aging [33], Parkinson’s disease patients [41], and those who had a stroke in the last 3 months and were in a long-term rehabilitation program [42].

In relation to the researchers conducting the experiments, 44% reported being conducted by the researchers themselves, 37% were physiotherapists, 16% were undefined, and 3% were both physiotherapists and researchers (Figure 7). If the situation allowed the remote monitoring or configuration by physiotherapist professionals, 20 of them reported "yes", 8 reported "no," and 4 were undefined (Figure 8).

4.2.5. Disabilities and Limbs Studied. Among the conditions addressed in the studies, those that suffered a stroke were the most frequent, appearing in 17 publications (48%), and included both stroke sufferers, and patients with motor or cognitive limitations. On the other hand, 8 of the studies (23%) aimed to investigate users who suffered injuries due to traumatic accidents, such as cerebral and spinal cord [43] and hip fracture in the elderly caused by fall [40]. Others involved several disabilities and included unique occurrences, such as muscular dystrophy fluctuations in human locomotion [44], motor coordination [33], mastectomy and stroke [29], Alzheimer [22], prosthesis users [37], Parkinson’s [41], phantom limb pain [45], and different developmental difficulties (mildly intellectually retarded; cerebral palsy; moderate multiple disabilities, weak legs, and low vision) [23], and only one study [31] did not indicate the disability altogether (Figure 9).

The upper body for the movement was the most commonly used to control applications, appearing in 14 studies...
Task Execution
Performance Analysis
Questionnaire
Interview
Observation
Think Aloud

(a)

(b)

Figure 4: Method used in articles numbers (a) and percentage (b).

Proportion by sex between the participating users

Male
Female
unknown

40%
37%
23%

Figure 5: Proportion by sex.

Type of users who did the tests

Patients
Healthy users
Patients and Healthy users

34%
59%
7%

Figure 6: Profile of the users who did the experiments.

Monitoring application tests

Therapists
Researchers
Both
undefined

37%
44%
16%
3%

Figure 7: Monitoring application tests.

4.3. Evidence Quality according to QualSyst Criteria. The processing of data, use of comparison groups, sample size, description of analytical methods, outcome well defined,
The system has the possibility of remote monitoring.

![Bar chart showing yes, no, and undefined responses with 20 yes, 8 no, and 4 undefined.]

The system has the possibility of remote monitoring.

The grades were assigned with the averages given by scales for each evaluated item. Correa et al. (2013) [44] used a blind evaluator who applied a semistructured questionnaire to each group, developed by the authors themselves. The alternatives sought to measure the degree of satisfaction, the motivation, and the perception of the effectiveness of the exercise. The percentage responses served to compare the results of the therapies with and without the technology.

5. Discussion

This review focused on applications with AR for motor rehabilitation and the techniques used to evaluate the usability and found a predominance of structured tests for task execution and performance analysis, which is very common in AR usability studies [9, 11, 12]. In general, studies that planned performance analysis tests were satisfied with the data they acquired to answer their research questions. However, the results of this review still demonstrate the use of few observational and empirical methods to validate or compare with analytical data generated by the tools. These results are reinforced by the QualSys evaluation, which demonstrates that the best articles (Table 1) performed a combination of methods and provided evidence based on quantitative data collection and subjective measures, like questionnaires. The empirical analysis, such as observation and interviews, is still a little-used practice in the area (Figure 4(a)) and should be further explored in an attempt to understand users’ behavior and expectations more deeply.

Studies that evaluated user’s preferences and perceptions or UX predominantly used structured and semistructured questionnaires, with a smaller proportion of them also using observations and interviews. Performance surveys, for the most part, did not evaluate subjective aspects of users’ interactions, such as the experience of use, satisfaction, and motivation when using the tool. When some subjective analysis was performed, it was done through a questionnaire, squandering the opportunity to conduct in-depth interviews with the user that would bring more intrinsic information about the opinions, values, emotions, and perception of the user experience [52].

Results of the “Evaluation methods” showed an absence of heuristic evaluation with specialists (designers, developers, and physiotherapists) (Section 4.2.3), which shows a lack of research with expert users based on AR-adapted guidelines. It would be a good research opportunity for application requirements area and is particularly important because such evaluations focus on enhancing the experience and detecting critical errors, before testing with ordinary users [53].

Furthermore, it is possible to analyse whether it was a recurrent practice to perform research with users who did not suffer from the disability proposed by the applications (as much as 59% of the studies) in selected articles. The mean age of non-young adults (39.89 years) reflects that the number of studies with higher sampling and younger ages lowered the general mean age, since the studies were generally conducted with healthy students within the university environment. This may be a problem because the use of
nonrepresentative users in the development of assistive and rehabilitative technologies, without considering their clinical needs, may result in products with low applicability [54]. The results demonstrate a need to invest in surveys that test their applications with final users to whom the application is intended. Some authors [54] recommend the development of these applications that users need to be involved in the development process, including prototyping and validation to increase and direct the applicability of the solution.

The studies with patients had fewer users and a more advanced age, because the sample contained people with characteristic disabilities that caused the need for motor rehabilitation: stroke (47%) and traumatic accidents (23%). Three of these studies pointed out some difficulties in working with this type of participant. The first of them indicated the difficulty for older people to use new technologies that they are not used to. In addition to low sampling, the characteristics were intrinsically linked to the specialised neurorehabilitation service where the study was performed and may have restricted the generalisation of the results [30]. The second, which studied individuals with Parkinson’s disease, also evaluated an intervention with a small group and did not include a control group. They emphasised the need for a follow-up evaluation to be able to consider possible residual effects and recommended additional research with more subjects, in order to increase the results to a population [55]. The third pointed out the difficulty in recruiting the participants, because the intervention was relatively new for the elderly, or because the participants had limited time for testing because of personal routine. They recommended recruiting and recording larger samples to provide stronger evidence [33].

In a usability study for games with gestural interaction [56], it was also found that it was not evident that there is a standardisation in usability assessments that cover who should be evaluated and how it should be evaluated. They assume that it is because each application intends to investigate different variables and consequently has different objectives. It is recommended that standardisation of usability assessments is implicated; this would be important to guide a validation for future studies and to check whether particular hardware or software is appropriate for intended use [56].

The QualSyst evaluation criteria contributed to classifying the best-structured studies with the most solid and consistent results [19]. It is noticed that the publications with the highest grades used a combination of methods of data collection [24, 25, 29, 50, 51] and sought to quantify these results by validating them with analysis based on statistical calculations. For this, they experimented with different configurations, by either grouping user profiles [29], testing different versions of the system [24, 50], or comparing with and without the use of technology [25], in order to obtain data to make comparisons or detect patterns of changes.

Although the technical and practical success of rehabilitation applications is fundamental to improve them, the researches that make subjective judgments of the users are essential for the usability improvement. Of the 10 studies that received QualSyst’s best scores, 7 used questionnaires (validated or authored) to verify aspects related to user experience, such as motivation [29, 51], satisfaction [24, 25],
perception [42, 50], and ease of use [31]. This analysis contributed to bringing parameters beyond those provided by performance assessments and helped to justify or explain events related to changes in user behaviour patterns. This review may help developers and researchers to have a current overview of the scientific community and the types of assessments that are being conducted when using AR technology for motor rehabilitation. This overview can help to guide the formulation of research protocols in this context and also be a good reference for a methodological model for future revisions.

6. Conclusions

This study sought to provide an overview of the studies that had performed tests with users using AR games and applications for motor rehabilitation purposes. It was verified that the great majority still prioritised data collection methods in the execution of tasks and performance analysis, which certainly generate solid quantitative data. However, it was concluded through an external scoring system (QualSyst) that the studies which generated the best evidence chose to use a combination of quantitative and qualitative analysis, through the application of methods such as user preference questionnaires, interviews, and observations. Through user-based studies, it is possible to understand the evolution of interactions with projected space and gain insight into new approaches to solving user interface design problems identified as part of the design/evaluation process [57]. The perception is that it follows the trend found in another study [9] that seems to be a growing understanding of the need to formalise the evaluation process and conduct properly designed user studies.

The difficulties of performing tests with users who have suffered a stroke or any other type of injury that limits their movements are reflected directly on the sampling found in this selection. It has been noted that the predominance of healthy users, students, and young people can lead to results that do not reflect the needs for the reality of these users. It would be suggested that the researchers included in their protocols at least one group of users that represented this profile.

It was important for this review to measure the publications through the QualSyst score by two reviewers. By using this system, it was guaranteed that the best-evaluated publications were the ones that were more methodologically structured and have strong evidence in their experiments. For future work, it is recommended to use this evaluation system, not only for health studies but also for other areas such as computer science and design, requiring minor adjustments to adapt it to the context of these researches.

In further reviews of this area, it is recommended to include more specific terms that focus on usage analysis, such as “usability” or “user testing.” It might also be interesting to expand searches in health database, more specifically where studies in motor physiotherapies are published. In this way, it would be possible to expand the search in areas that may have a direct correlation with the searched theme.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was carried out due to the support of CNPq, National Council of Scientific and Technological Development, Brazil.

References


