UNIVERSITÀ DEGLI STUDI DI NAPOLI "FEDERICO II"

FACOLTÀ DI INGEGNERIA

Dipartimento di Ingegneria Aerospaziale

**Dottorato di Ricerca in Ingegneria Aerospaziale,
Navale e della Qualità**

Indirizzo "Gestione della Qualità Totale"
XXIV Ciclo di Dottorato

A STATISTICAL APPROACH FOR USABILITY ASSESSMENT AND COMFORT IMPROVEMENT IN PRODUCT DESIGN

TUTOR
Prof. Antonio Lanzotti

CANDIDATA
Mariangela Trotta

DIRETTORE DELLA
SCUOLA DI DOTTORATO
Prof. Antonio Moccia

Ai miei amici dottorandi

...Passati, presenti e futuri...

A chi mi ha insegnato che Siamo cio’ che in Potenza possiamo essere

Ai fiumi e ai mari

Alla neve… Che cadrà a breve

Contributo alla statistica

Su cento persone:
che ne sanno sempre più degli altri
- cinquantadue;
insicuri a ogni passo
- quasi tutti gli altri;
pronti ad aiutare,
purché la cosa non duri molto
- ben quarantanove;
buoni sempre,
perché non sanno fare altrimenti
- quattro, be', forse cinque;
propensi ad ammirare senza invidia
- diciotto;
viventi con la continua paura
di qualcuno o qualcosa
- settantasette;
dotati per la felicità,
- al massimo poco più di venti;
innocui singolarmente,
che imbarbariscono nella folla
- di sicuro più della metà;
crudeli,
se costretti dalle circostanze
- è meglio non saperlo
neppure approssimativamente;
quelli col senno di poi
- non molti di più
di quelli col senno di prima;
che dalla vita prendono solo cose
- quaranta,
anche se vorrei sbagliarmi;
ripiegati, dolenti
e senza torcia nel buio
- ottantatré
prima o poi;
degni di compassione
- novantanove;
mortali
- cento su cento.
Numero al momento invariato
*(**Wislawa Szymborska**)*

"Complicare è facile, semplificare è difficile. Per complicare basta aggiungere, tutto quello che si vuole: colori, forme, azioni, decorazioni, personaggi, ambienti pieni di cose. Tutti sono capaci di complicare. Pochi sono capaci di semplificare. La semplificazione è il segno dell'intelligenza, un antico detto cinese dice: quello che non si può dire in poche parole non si può dirlo neanche in molte"
*(**Munari**)*

Summary

Chapter 1 : Usability and quality in use ………………………………………………………………………………… 11

1.1. Usability definition in ISO 9241-11/1998 ………………………………………………………………….....13

1.1.1. The usability framework …………………………………………………………………………………….. 14

1.1.2. Usability Inspection Methods………………………………………………………………………………. 15

1.2. Comfort ………....16

1.2.1. Seat Comfort ……………………………………………………………………………………………… 17

1.2.2. Plantar comfort…………………………………………………………………………………………………....18

Bibliography……....20

Chapter II: Main issues in participatory usability testing ……………………………………………………….. 25

1.1. Estimation of the number of users…………………………………………………………………………….. 26

1.1.1. The model of Virzi………………………………………………………………………………………….. 27

1.1.1. The Nielsen's approach: five users are enough……………………………………………………... 28

1.1.2. The model of Lewis…………………………………………………………………………………………. 31

1.1.3. The "Evaluator effect" of Hertzum………………………………………………………………….. 31

1.1.4. The Spool's model ………………………………………………………………………………………….. 33

1.1.5. The Caulton's model ……………………………………………………………………………………….. 34

1.1.6. The Turner's model………………………………………………………………………………………….. 34

1.1.7. The Kanis approach…………………………………………………………………………………………. 35

1.2. The level of expertise ……………………………………………………………………………………………….. 36

1.3. Prototipe fidelity……………………………………………………………………………………………………….. 41

Bibliography……….....44

Chapter III: Usability metrics... 47

3.1. Measures of effectiveness.. 48

3.1.1. Task Completion.. 48

3.1.2. Number of errors.. 49

3.1.3. Spatial accuracy.. 49

3.1.4. 3.1.4. Other measures of effectiveness..49

3.2. Measure of efficiency..51

3.2.1. Time..51

3.2.2. Input rate..51

3.2.3. Mental effort..51

3.2.4. Communication effort.. 52

3.2.5. Other measures.. 53

3.3. Learnability: performance metrics over time..53

3.4. Measures of satisfaction and questionnaires for satisfaction's data collection...................54

3.4.1. Post- task ratings..54

3.4.1.1. Ease of use Questionnaire..55

3.4.1.2. After Scenario questionnaire (ASQ) 55

3.4.1.3. Printer Scenario Questionnaire (PSQ) 56

3.4.1.4. Expectation Measure..57

3.4.1.5. Usability Magnitude Estimation..58

3.4.2. Post- session ratings 58

3.4.2.1. System Usability Scale (SUS) 58

3.4.2.2. Post- Study System Usability questionnaire (PSSUQ)...................................60

3.4.2.3. Computer System Usability Questionnaire (CSUQ).......................................60

3.4.2.4. Questionnaire for User Interface Satisfaction (QUIS)....................................61

Bibliography..64

Chapter IV: Combined metrics for usability assessment.. 67

4.1. Summated Usability Index (SUM) 67

4.1.1. The model 67

4.1.2. The methodological approach...68

4.1.3. Main issues in SUM model and applications.. 70

4.2. The Integrated Usability Index.. 71

4.2.1. The model 71

4.2.2. Main issues in Integrate Usability model and applications.. 75

Bibliography...76

Chapter V: A new approach for usability assessment.. *77*

5.1 The case study: a wheelchair mounted manipulator.. 78

5.2 The methodological approach.. 80

5.2.1 The model.. 82

5.2.2 The Usability index.. 86

5.2.2.1 Normalization techniques... 87

5.2.2.2 The weight's assignement: the AHP.. 88

5.3 The experiments.. 90

5.3.1 The experimental setup.. 91

5.3.2 Experimental protocol.. 92

5.4 Results... 93

5.4.1 Weights assessment.. 93

5.4.2 The usability Index.. 95

5.4.3 Preliminary analysis.. 96

5.5 DoE analysis 98

5.6 Comparison with other indexes in literature.. 100

5.6.1 Adapting the model for Sauro et al. Index evaluation 100

5.6.2 Adapting the model for Kim et al. Index evaluation 103

Bibliography...107

Conclusions and future work...*109*

Appendix..*111*

Appendix A: Descriptive Statistics.. *113*

1. Measures of effectiveness.. 113

1.1. Mouvement Error *113*

1.2. Goals number *115*

1.3. Errors number.. *115*

2. Efficiency measures.. *117*

2.1. Time *117*

2.2. Communication *118*

3. Satisfaction measures.. *120*

3.1. User Preference *121*

3.2. Ease of use *121*

Appendix B: PCA application for SUM evaluation.. *123*

Paper A...*131*

Paper B...*146*

Paper C...*155*

Acknowledgements..*165*

Appended Papers:

Paper A: G. Di Gironimo, A. Lanzotti, G. Matrone, A. Tarallo, M. Trotta, A virtual reality approach for usability evaluation of a wheelchair-mounted robot manipulators, Proceedings of TMCE2010, Ancona, 12 - 16 April 2010, ISBN 978- 90- 5155- 060- 3, pp 749- 762

Paper B: A. Lanzotti, A. Vanacore, M. Trotta, Validation of a new index for seat comfort assessment, based on objective and subjective measures, Proceedings of IMPROVE 2011, Venezia, 15 – 17 june 2011, ISBN 978- 88- 7784- 328- 9, pp 7

Paper C: P. Franciosa , S. Gerbino, A. Lanzotti, S. Patalano, V. Rufrano, M. Trotta, L. Silvestri , L. Frattari Improving comfort of occupational footwears through experiments on virtual prototypes, Proceedings of IMPROVE 2011, Venezia, 15 – 17 june 2011, ISBN 978- 88- 7784- 328- 9, pp 8

Additional Publications:

Journal publications

G. Di Gironimo, A. Lanzotti, G. Matrone, A. Tarallo, M. Trotta, A virtual Reality Approach for Usability Assessment: case study on weelchair mounted robot manipulator, Journal of Enginering with Computers (2011) (formal acceptance by “Journal of Enginering with Computers”, Springer, (envisaged date of publication January 2012), pp25

Conferences publications

A. Lanzotti, S. Patalano, M. Trotta, Teaching product design through a participatory contest experience, Proceedings of TRIZ 2010, Bergamo, 2 – 5 Novembre 2010, ISBN 978- 88- 96333- 59- 4, pp 277-283

A. Lanzotti, G. Matrone, M. Staiano, M. Trotta (2009), Adaptive Design Augmentation Strategies for Visual Product Evaluation, Proceedings of Enbis, Goteborg, 20 – 24 September 2009, pubblicato su cd, pp 10

Chapter 1

Usability and quality in use

There are many definition of usability in literature, provided different authors and international standards. Bevan [1] classified the standards in two categories:

a) Top- down approach which is concerned with usability as a broad quality objective that translates into the ability to use a product for its intended purpose
b) Bottom- Up approach, that concerns to the design of specific attributes, and relates more closely to the needs of the interface designer and the role of usability in software engineering

Following a brief overview about usability standards and their category of reference is provided [2]:

Tab. 1: Usability definition in different standards (Bevan's categorization)

Top- Down Approach	Bottom- Up approach
ISO 9241-11 (1998) "Usability is The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specific context of use"	**ISO/IEC 9126-1 (2001)** "Usability is a set of attributesof software which bear on the effort needed for use and on the individual assessment of such use by a stated or implied set of users " **ISO/IEC 9126-1 (2004)**

Usability is the capability of the software product to enable specified users to achieve specified goals with effectiveness, productivity, safety and satisfaction in specified context of use”

IEEE Standard 610.12 (1990)

“Usability is the ease with which a user can learn to operate, prepares inputs for and interprets outputs of a system or component”

An interesting improvement in usability definition, was made with the integration of ISO/IEC 9126-1 [3] and ISO/IEC 25010 [4] standards, that made it possible to define Usability as a characteristic of “Quality in Use”, with sub-characteristics of Effectiveness, Efficiency and Satisfaction. The complete model of Quality in Use is reported in Tab. 2:

Tab. 2: Quality in Use Model

Usability	**Flexibility**	**Safety**
Effectiveness	Context conformity	Commercial damage
Efficiency	Context extendibility	Trust Operator health and safety
Satisfaction	Accessibility	Public health and safety
Likability		Environmental harm
Pleasure		
Safety		
Comfort		

The novelty of this new approach concerns in specific the satisfaction aspect of usability that is translated in four dimensions:

- Likability (cognitive satisfaction) which relates to the level of satisfaction of the user in the accomplishment of a task, taking into account the ease of use of the product, the achievement of pragmatic goals, and the perceived results of use.
- Trust (satisfaction with security): the extent to which the user is satisfied in relation to the affordance of the product.
- Pleasure (emotional satisfaction): the extent to which the user is satisfied with their perceived achievement of hedonic goals.
- Comfort (physical satisfaction): the extent to which the user is satisfied with physical comfort.

However, the definitions of effectiveness and efficiency are broadly consistent with those of the standard 9241-11/1998.

1.1. Usability definition in ISO 9241-11/1998

ISO 9241-11/1998 is a milestone in the literature about usability. Indeed, this standard, not only provides a definition of usability (Tab. 1), but highlights also the necessary information to take into account for usability assessment:

a) a framework to identify the most relevant aspects of usability
b) the definition of usability dimensions (effectiveness, efficiency and satisfaction), that can be used to assess the user- product interaction in a specific context;
c) the definition of several metrics related to both, performance and satisfaction aspects;
d) the definition of usability as a part of a quality plan.

In this chapter point a) will be deepened, while points b) and c) , are discussed in chapter III.

1.1.1. The usability framework

The framework provided by the ISO standard (Figure 1) is based on several assumptions:

- usability is dependent on the context of use
- the level of usability achieved will depend on the specific circumstances in which a product is used.
- the context of use consists of the users, tasks, equipment (hardware, software and materials), and the physical and organizational environments which may all influence the usability of a product

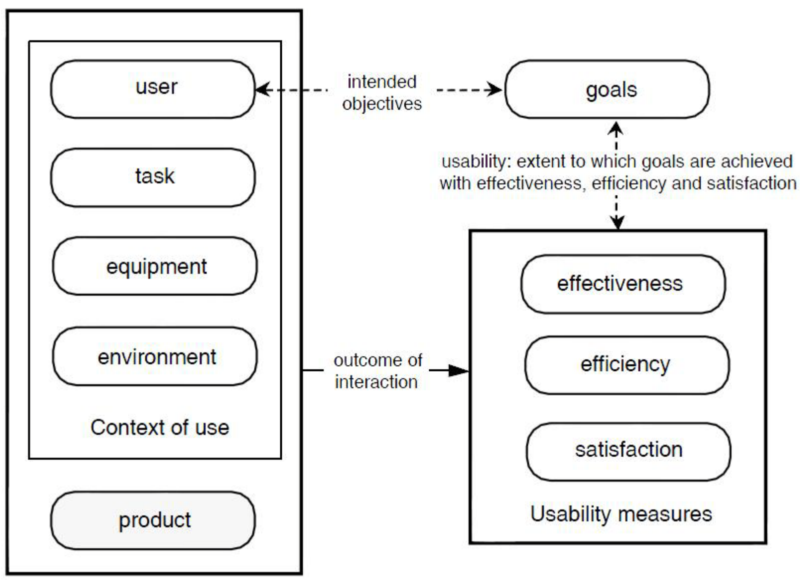


Figure 1: Framework provided by ISO 9241-11/1998 to identify the most relevant aspects of usability

Starting from this framework, at first the goal of the analysis should be defined. Then, the study of the context of use (and its components), the product, the user and the interactions between them, allows to translate usability dimensions (effectiveness, efficiency and satisfaction) in measures that could be collected in an experimental test.

1.1.2. Usability Inspection Methods

The Usability inspection methods[5] are aimed at highlighting the main problems of product interfaces, through the direct inspection of them realized by users. Following the most important Usability inspection methods are briefly described.

– Cognitive Walkthrough (CW) [6]:

this approach [7][8]requires decomposing the task into simpler subtasks that will be subsequently evaluated by a panel of experts. The final aim is to detect potential discrepancies between the actual end-user's cognitive model and the expected one [8].

- Heuristic Evaluation (HE):

this is a usability engineering method for finding the usability problems in a user interface by involving usability specialist. Usually a small set of evaluators examine the interface and judge its compliance with recognized usability principles (the "heuristics").[9]

- Thinking-Aloud Methods (TA) [10]

It is one of the most valuable usability heuristic methods, used to highlight main problem of product interfaces, by direct interaction with them[5]. In the experimental phase, the users verbalizes their thoughts during the accomplishment of a defined tasks, allowing the understanding of the most critical issues.

- Usability Test (UT)

This procedure [8][10] can be a valid alternative to the previous one, since it provides quantitative information about the actual execution of a set of defined tasks. However, the efficiency of this method is limited by the need of physical prototypes and by the impossibility of gathering subjective data.

1.2. Comfort

Specialized literature does not provide a universally recognized definition of comfort, nevertheless in recent years, the assumption that comfort and discomfort are two distinct entities [12] is winning broad respect. In their studies, Zhang and Helander [13] show that sitting discomfort is related to the biomechanical factors associated to the interaction with the seat over time, whereas comfort reflects a perception of instantaneous well-being perceived by the user. Zhang [14] pointed out that poor biomechanics may turn comfort into discomfort even though good biomechanics is not a necessary and sufficient condition for comfort. In other words, good biomechanics can avoid discomfort and thus it can be assumed as a prerequisite for comfort. Being complex concepts, comfort and discomfort are difficult to measure and interpret [15]. A great deal of research has been done to face the problem of sitting comfort/discomfort assessment and several subjective and objective methods have been developed [16]-[20]. Typically comfort assessment is realized on the basis of subjective evaluations or postural analysis. Subjective evaluations are collected by surveying potential seat users who are asked to express their feelings of comfort/discomfort with the seat and/or compare, in terms of perceived comfort/discomfort, different seats belonging to the same class [21][22].

Postural analysis is realized by measuring one or more objective parameters, such as [23]:

- the pattern of muscle activation measured through electromyography (EMG) [24].
- the stress acting on the spine measured through pressure transducer and radio waves [24]
- the postural angles [25] obtained using contact or non contact (like photogrammetric) techniques in real experiments or using virtual manikins in virtual experiments [26]

- the body–seat interface pressure measured through capacitative mats.

Anyway, subjective and objective methods are not alternative since they complement each others. The exclusive use of subjective evaluations can be misleading for several reasons:

- when attention focuses on particular elements of the seat, the response variability is reduced, but the interaction with other neglected features can be a noise factor [24]

- users could not be able to synthesize a subjective perception in a numeric or semantic evaluation causing a partial loss of information [22].

- the perceived differences of ergonomic features are often small and the results from comparisons of different seat concepts are rarely significant;

- the human body is very adaptive and not sensitive to distinguish variations in seats;

- subjective evaluations are costly and time-consuming [27];

- subjective evaluations are rarely applicable early in the design process [15].

On the other hand, the exclusive use of objective measures for comfort assessment, highlights the following criticisms:

- normally, the information provided by objective criteria are complement but not substitute of subjective evaluations related to user's perception of comfort;

- the construction of quantitative measures for comfort assessment cannot disregard from noises often overlooked, such as anthropometric variability.

1.2.1. Seat Comfort

A great deal of research has been performed to find objective measures for predicting seat comfort perception [28]. Research has shown that one of the main factors that affect seat comfort is seat-interface pressure distribution [29]. Moreover, pressure distribution is the objective measure with the clearest correlation with the subjective evaluation methods[15][19]. Human-seat interface pressures have a spread field of application, indeed they have been measured to improve the comfort of office chairs [30], car seats [23], motorcycles saddles [31]and others vehicles seats [32], as well as to pursue product innovation in Kansei Ergonomics [33]. In particular, in office chair design pressure maps have been used to qualitatively verify the effectiveness on seat comfort of product features like, e.g., cushion shape and materials [34]-[36] through correlation studies with the subjective user perceptions. Nevertheless the widespread use of pressure maps, just few authors [37][38] have proposed synthetic indexes for the related multidimensional data, collected by performing real or virtual experiments involving a selected sample of potential users. Furthermore, little effort has been made to highlight the usefulness of these pressure measures for specific purposes defined by designers (e.g. Design for a Target and Design for All).

In order to provide a tool that can be easily used by designers Lanzotti et al. [37][38] proposed the Weighted Pressure Comfort Loss (WPCL) a postural comfort index based on comfort loss due to uneven seat-interface pressure distribution.

1.2.2. Plantar comfort

In [39][40] it was reported that physiological factors, such as plantar pressures, are strongly related to physical parameters such as materials and plantar shape. A first valid scientific contribution to the analysis of correlation was offered by Jordan et al. [41]. They attempted to correlate the subjective perceptions of users with dorsal and plantar pressure distribution through short-term dynamic tests. Perceived comfort was measured by using specific questionnaires, while pressure distributions were monitored through high resolution insole sensors. The correlation analysis was based on the results coming from three different shoes. The study showed a negative correlation between pressures and subjective comfort perception (meaning that a high peak pressure corresponds to a low perceived comfort). Moreover, authors highlighted the need to investigate further other objective parameters that may affect the user perception (see, for example, shear and normal forces, and heat transfer). Witana et al. [42] tried to identify the interactions between comfort and plantar shape. They found substantial differences between the subjective perceptions of users related to the mid-foot for different tested materials, thus confirming that comfort perceptions, for different areas of the plantar foot, are quite different. If on one hand experimental tests, carried out on different product designs, give valuable results, on the other hand, the large number of design parameters would make extremely difficult and expensive to identify the optimal design through tests with real prototypes. In this sense, using virtual simulations and parametric models may be a valid support.

Recently, in order to give a valuable support to experimental investigations, computational methods, based on FE modeling, have been adopted. FE models of human foot have been developed under certain simplifications and assumptions [43]-[46] such as: (i) simplified or partial foot shape, (ii) assumptions of non-linear hyper-elastic material law, (iii) ligaments and plantar fascia modeled as equivalent forces or elastic beams/bars, (iiii) no friction or thermal effect, at plantar foot interface, accounted. In this contest, Cheung and Zhang [14] combined FEM and Taguchi methods to identify the sensitivity of five design factors (arch type, insole and mid-sole thickness, insole and mid-sole stiffness) of footwear on peak plantar pressure. From FEM predictions, the most important design factors, able to reduce the peak plantar pressure, were found-out.

Bibliography:

[1] Bevan, N. Human-Computer Interaction Standards. In Proceedings of the International Conference on Human Computer Interaction, Yokohama, July 1995. Anzai & Ogawa, pp.1-8

[2] Tan, J (2009). FOUUX- A framework for usability & user experience. Master Thesis

[3] ISO/IEC 9126-1: Software engineering – Product quality - Part 1: Quality model. ISO (2001)

[4] ISO/IEC CD 25010.3: Systems and software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Software product quality and system quality in use models. ISO (2009)

[5] Nielsen,J.(1994a).Heuristicevaluation.In J.Nielsen &R.L.Mack (Eds.), Usability inspection methods (pp. 25–62). New York: Wiley

[6] Lewis, C., Polson, P., Wharton, C., & Rieman, J. (1990). Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. In Proceedings of the ACM CHI '90 Conference (pp. 235–242). New York: ACM

[7] Jaspers, M.: A comparison of usability methods for testing interactive health technologies: Methodological aspects and empirical evidence. I. J. Medical Informatics 78(5), 340{353 (2009)

[8] Shackel, B., Richardson, S.: Human Factors for Informatics Usability. Cambridge University Press, New York, NY, USA (2008)

[9] Lewis, C. (1982). Using the "thinking-aloud" method in cognitive interface design (IBM Research Rep.No.RC9265[#40713]).YorktownHeights,NY:IBMThomasJ.WatsonResearchCe nter.

[10] Nielsen, J. (1993). Usability engineering. Boston: Academic

[11] Hertzum M., Jacobsen N.E. (2001). The evaluator Effect: A Chilling Fact About Usability Evaluation Methods. International Journal of Human Computer Interaction 13(4), 421-443

[12] P. Vink. Comfort and design: principles and good practice. CRC Press 2005.

[13] M. G. Helander, L. Z. Zhang. Field studies of comfort and discomfort in sitting. Ergonomics 40, 9 (1997) pp 895-915.

[14] L. Z. Zhang, M. G. Helander, Drury C. G. Identifying factors of comfort and discomfort in sitting. Human Factors 38, 3 (1996) pp 377-389

[15] M. De Looze, L. Kuijt Evers, J. Van Dieen. Sitting comfort and discomfort and the relationships with objective measures. Ergonomics 46, 10 (2003) pp 985-997.

[16] Q. Zeng, Q. Wu, C. Yang, K. S. Wu. An Evolution Method of Driving Seat Comfort Based on Least Squares Support Vector Regression. Proceedings of Fifth International Conference on Natural Computation, August – , 2009, Tianjin, pp 455-459.

[17] Fazlollahtabar. A subjective framework for seat comfort based on a heuristic multi criteria decision making technique and anthropometry. Applied Ergonomics 42, 1 (2010) pp 16-28.

[18] Na, S. Lim, H. Choi, M. Chung. Evaluation of driver's discomfort and postural change using dynamic body pressure distribution. International Journal of Industrial Ergonomics 35,12 (2005) pp.1085-1096.

[19] Kyung, M. A. Nussbaum. Driver sitting comfort and discomfort (part II): Relationships with an prediction from interface pressure. International Journal of Industrial Ergonomics 38, 5-6 (2008) pp 526-538.

[20] K. Fredericks, S. E. Butts. Objectively determining comfortable lumbar support in task seating (Tech. Report). Holland, MI: Haworth. .haworth. (2006).

[21] Shackel, K. D. Chidsey, P. Shipley. The assessment of chair comfort. Ergonomics 20, 1 (1969) pp 269-306.

[22] Reijneveld, M. de Looze, F. Krause, P. Desmet. Measuring the Emotions Elicited by Office Chairs. Proceedings of the International conference on Designing pleasurable products and interfaces, June – , 2003, Pittsburgh, pp 6-10.

[23] Andreoni, G. C. Santambrogio, M. Rabuffetti, A. Pedotti. Method for the analysis of posture and interface pressure of car drivers. Applied Ergonomics 33, 6 (2002) pp 511-522.

[24] K. Lueder. Seat comfort: A review of the Construct in the Office Environment. Human Factors 25, 6 (1986) pp 701-711.

[25] Dreyfuss. The measure of man and woman. Wiley 2002.

[26] Barone, A. Lanzotti. (2009) Robust Ergonomic Virtual Design. In: Statistics for Innovation: statistical design of “continuous” product innovation. Springer 2009, pp 61-83.

[27] F. Tan, F. Delbressine, W. Chen, M. Rauterberg. Subjective and Objective Measurements for Comfortable Truck Driver’s Seat. 9th International Symposium on Advanced Vehicle Control, October – , 2008, Kobe, pp 851-856.

[28] F. M. Kuijt- Evers, T. Bosch, M. A. Huysmans, M. P. de Looze, P. Vink. Association between objective and subjective measurements of comfort and discomfort in hand tools. Applied Ergonomics 34, 3 (2003) pp 207-214.

[29] Stinson, S. Crawford. Optimal Positioning: Wheelchair Seating Comfort and Pressure Mapping. In International Handbook of Occupational Therapy Interventions, Springer 2009, pp 83-90.

[30] Reed, C. Grant. Development of a measurement protocol and analysis techniques for assessment of body pressure distributions on office chairs, TR. Ann Arbor, MI: Univ. of Michigan, Center for Ergonomics(1993).

[31] Carfagni, L. Governi, Y. Volpe. Comfort assessment of motorcycle saddles: a methodology based on virtual prototypes. International Journal on Interactive Design and Manufacturing (IJIDeM) 1,3 (2007) pp 155-167.

[32] Hostens, G. Papaioannou, A. Spaepen, H. Ramon. Buttock and back pressure distribution tests on seats of mobile agricultural machinery. Applied Ergonomics 32 (2002) pp 347-355

[33] Nagamachi. Perspectives and new trend of Kansei/Affective Engineering. The TQM Journal 20, 4 (2008) pp 290-298.

[34] Kamijo, H. Tsujimara, H. Obara, M. Katsumatu. Evaluation of seating comfort, SAE Technical Paper Series 820761. Society of Automotive Engineers, Troy, MI, (1982) pp 1–6

[35] Y. Park, J. Y. Kim, J. H. Shin. Ergonomic design and evaluation of a new VDT workstation chair with keyboard-mouse support. International Journal of Industrial Ergonomics 26, 5 (2000) pp 537-548

[36] Fujimaki, R. Mitsuya. Study of the seated posture for VDT work. Displays, 23, 1-2 (2002) pp 17-24.

[37] Lanzotti, G. Matrone, A. Vanacore. Seat design improvement via a new postural comfort index. Proceedings of ENBIS9 Conference, september – , 2009, Goteborg.

[38] Lanzotti, G. Matrone, A. Vanacore. Robust Ergonomic Design (RED): towards a new index for the evaluation of the seat comfort Proceedings of the Congreso Internacional Conjunto XXI INGEGRAF - XVII ADMME, June – , 2009, Lugo.

[39] C. Leber, P. M. Evanski. A comparison of shoe insole materials in plantar pressure relief. Prosthetics and Orthotics International 10, 3 (1986) pp 135-138.

[40] Frederick, E.C. Physiological and ergonomics factors in running shoe design. Applied Ergonomics 15, 4 (1984) pp 281-287.

[41] Jordan and R. Bartleett. Pressure distribution and perceived comfort in casual footwear. Gait & Posture 3, 4 (1995) pp 215-220.

[42] P. Witana, R. S. Goonetilleke, S. Xiong, E. Y. L. Au. Effects on surface characteristics on the plantar shape of feet and subject's perceived sensation. Applied Ergonomics 40, 2 (2009) pp 267-279.

[43] T. M. Cheung, M. Zhang, A. Kam-Lun Leung. Three-dimensional Finite Element Analysis of the foot during standing: a material sensitivity study. Journal of Biomechanics 38, (2005) pp 1045–1054.

[44] Lemmon, TY. Shiang, A. Hashmi, JS. Ulbrecht, PR. Cavanagh. The Effect of shoe insoles in therapeutic footwear: a finite element approach. Journal of Biomechanics 30, (1997) pp 615-620.

[45] PJ. Antunes, GR. Dias, AT. Coelho, F. Reselo, T. Pereira, Non-Linear Finite Element Modelling of anatomically detailed 3D foot model, technical report (2008).

[46] J. T. M. Cheung, M. Zhang. A 3D Finite Element Model of the human foot and ankle for insole design. Archives of Physical Medicine and Rehabilitation 86, (2005) pp 353-358.

[47] T. M. Cheung, M. Zhang. Parametric design of pressure- relieving foot orthosis using statistics- based finite element method. Medical Engineering & Physics 30, 3 (2008) pp 269-277.

Chapter II

Main issues in participatory usability testing

Both in computer science, and in industrial design, the usability evaluation methods (UEMs) resulted in considerable benefits in identifying critical aspects of product development, allowing a significant reduction in terms of time and costs. [1]. However, in order to obtain significant improvements a huge number of critical factors characterizing user- product interaction, must be taken into account [2] Furthermore, inconsistencies related to some of them, could impact on the reliability of the obtained results, regardless of the adopted methodology [3]. The design of a usability test, cannot disregard these aspects (*Figure **1***) and the variability induced from them on the experimental results. The main factors to consider are:

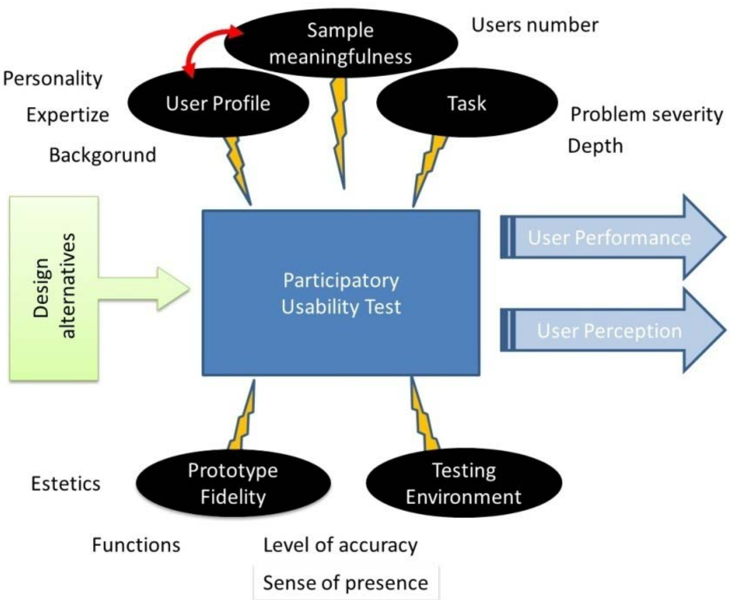


Figure 1: Critical factors in usability testing

- Sample size and representativeness of the selected sample compared to the target population. Generally, these prerequisites are partially neglected, mainly for economic reasons. Moreover, in some cases, the direct involvement of the users is not possible (e.g. disable users or user with limited cognitive skills). In these cases a valid alternative could be the involvement of indirect users (familiars,) of the product or expert users (medical staff, designers) [4], [5].
- Prototipe fidelity. Even in this case, all the choices, at an experimental stage, are strongly conditioned by the constraints of cost and timing. Moreover a low fidelity prototype, could affect the user- product interaction, due to the limited functionalities.
- Task definition. This factor is strictly related to the goal of the survey, particularly in complex studies in which both, performance aspects and subjective measures, should be carried out..
- Testing environment. A very crucial issue is the definition of main differences between testing environment and real environment, especially for remote usability evaluation or virtual experiments.

Several of the aforementioned aspects, will be detailed later in this chapter.

1.1. Estimation of the number of users

Estimation of the number of users for a usability test is actually an unresolved problem [6], [7], [8]. In spite of the goal of the experiments, the analysis carried out must go together with the adopted methodology and the target of users. Indeed, the right selection of the sample strongly affects the validation of experimental results and their significance level (par.2.2). Many studies in the literature, are related to the usability evaluation of interfaces; therefore, the proposed models cannot be applied to industrial products, without prior experimental validation.

O avoid confounding, It is important to clarify the difference between "usability problem" and '"user error in the experimental phase". The first ones are all the factors that affect the use of the product, causing the failure in the achievement of the task, or repeated errors before being able to perform the task itself. In this sense, the number of errors could be considered as an indicator of the level of severity of the usability problems.

Nei successivi paragrafi saranno illustrati i principali modelli presenti in letteratura, le loro evoluzioni e i loro limiti (Figure 2).

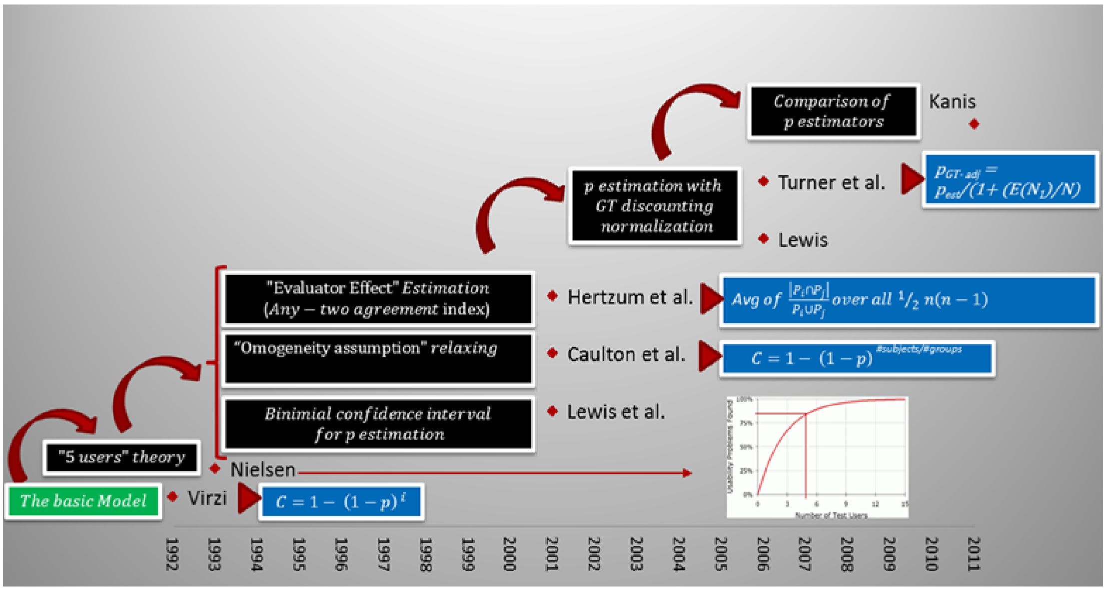


Figure 2: State of the art of the models to estimate the number of usability problems based on the sample size of evaluators

1.1.1. The model of Virzi

The study of Virzi, attempts to estimate the number of users necessary for a meaningful evaluation of usability, through a probabilistic approach. [9]. In three experiments, Virzi tries to answer three fundamental questions:

1. How to determine the statistical link between the sample size and the number of errors identified by users for a single interface.
2. How to assess how the error rate varies according to the level of severity of the identified problems.
3. How to define the level of severity of the interface problems in sample size estimation.

In relation to point 1, the model proposed by Virzi, relates *N*, the rate of usability problems identified with the the sample *i*, according to eq. (1):

(1) $$N = 1-(1-p)^i$$

That is an application of the binomial model assessing the probability that n users are able to find at least one error, being:

- *p* the probability of detecting a particular usability problem
- *i* number of subjects run in the evaluation

The results obtained by comparing the model output with experimental simulations generated with the Monte Carlo method, show that the model seems to overestimate the number of users required for usability evaluations (Figure 3). With regard to point 2, the study shows that problems with high level of severity are identified very quickly even from a limited sample of users. The classification of usability problems, in this case, was carried out by users themselves (Figure 4). Finally, the author proposes an expert-based method (point 3), in order to identify the severity level of usability problems, to ensure that the classification of problems by users, is unaffected by number of errors identified in the experimental phase.

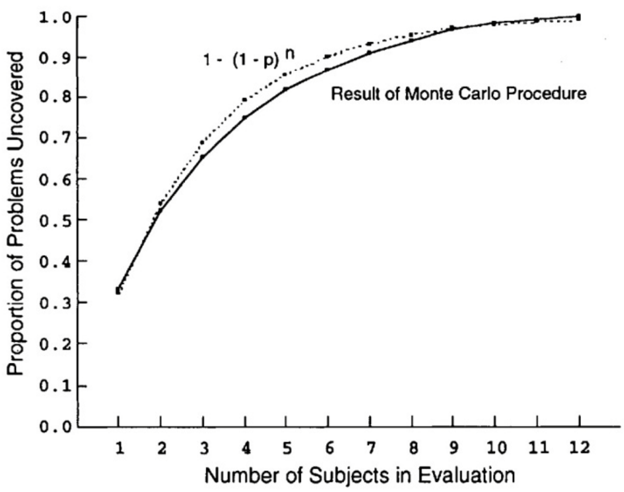


Figure 3: Proportion of usability problems uncovered as a function of the experimental sample size. Source: Virzi R.A. (1992), "Refining the test phase of usability evaluation: how many subjets is enough?"

1.1.1. The Nielsen's approach: five users are enough

Jacob Nielsen in [1] states that a sample of five users seems to be enough to reveal an error rate of at least 75%. [1] (Figure 5):

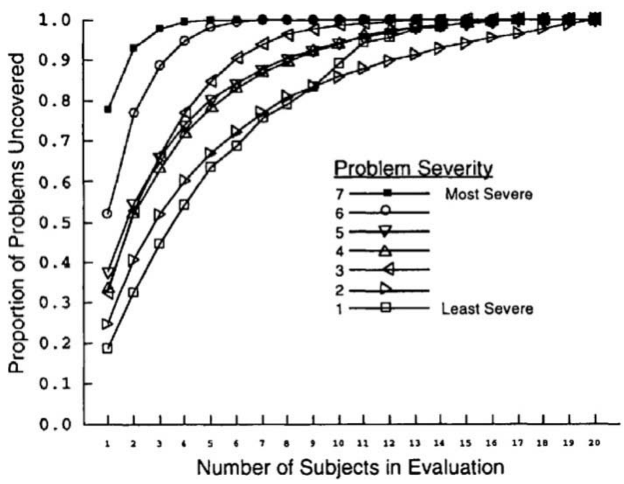


Figure 4 : Proportion of usability problems uncovered as a function of the experimental sample size at a given level of severity. Source: Virzi R.A. (1992), "Refining the test phase of usability evaluation: how many subjets is enough?"

Nielsen considerations relate specifically to the heuristic evaluation of interfaces, taking into account that:

- On average, the error rate for a single user is around 35%
- A cost- benefit analysis and the definition of main issues in product use, must go together in the definition of the sample size.

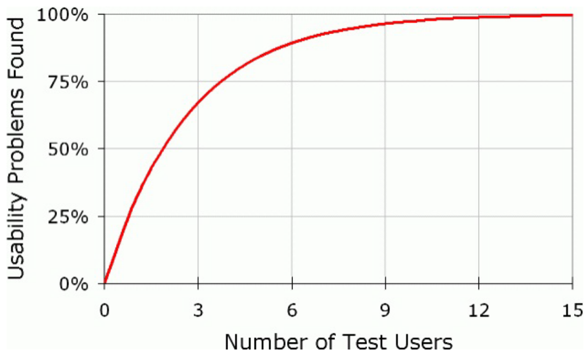


Figure 5 : Usability problems found by heuristic evauation as a function of the number of evaluators. Source: Nielsen J.(1993), "Usability Engineering"

Moreover, the analysis conducted by Nielsen refers to the evaluation of a specific metric (number of errors in the first use of an interface), and in specific experimental conditions. Therefore, the large-scale application of the theory "5 users is enough", must be subject to preliminary statistical analysis. In this perspective, the author proposes a probabilistic model [10] for the identification of usability problems, according to the number of users involved in the experimental phase. Starting from the stochastic process of Poisson, the final formulation is reported in the equation (2):

$$Found(i) = N\left[1-1(1-\lambda)^{i}\right] \quad (2)$$

where λ is the problem discovery rate (equivalent to the value of *p in the* Virzi's model), *N* is the total number of problems in the interface, and n is the number of subjects. λ is dependent from several factors:

- properties of the system;
- stage in the product lifecycle in which the product is tested;
- prototype's fidelity;
- type and quality of the methodology used to to conduct the test;
- complexity of the task;
- user expertise;
- representativeness of the sample of users.

The model is based on the assumption that the problems identified in each test are independent of those found in previous tests, by other users. This hypothesis is quite acceptable in the case of heuristic evaluations, as it becomes stronger (thus generating a greater approximation), in the case of usability testing. One of the main limitations of the model is undoubtedly that all the usability problems have the same probability to be identified (λ) [10]. A more reliable model should replace the fixed value (typically set equal to 0.31) with a probability density function that recognizes the different possibility of detection of usability problems. In conclusion, the claim "five users are enough", is strongly affected by the selected value λ. A λ value greater than or equal to 0.31 (determined by Nielsen), confirming the this statement. However, this result does not take into account the variability induced by the composition of the sample, which, being equal "lambda", can lead to much worse results in terms of usability errors identified, as demonstrated by the study of Faulkner [12].

1.1.2. The model of Lewis

Lewis [13] uses binomial confidence intervals to determine the level of acceptability of the number of errors ,as a function of the number of users, by comparing them with the lower limit of the binomial confidence interval [14]. In a subsequent work Lewis, performs further statistical analysis to find a correct estimate of p in relation to the sample size of users involved in the experiments. [15]. Using data generated by Monte Carlo simulation, the author applies different statistical techniques (discounting, normalization and regression). Finally a combined technique of normalization and Good-Turing discounting is selected as the best for p estimation. Results demonstrates that: “Practitioners can obtain accurate sample size estimates for problem-discovery goals ranging from 70% to 95% by making an initial estimate of the required sample size after running two participants, then adjusting the estimate after obtaining data from another two (total of four) participants”. The work of Lewis, cogently refers to usability errors, rather than usability problems. In this case, the difference in terms of severity level of usability problems, is neglected. On the other hand, the author confirms some of the results already reported by Virzi and Nielsen, which is that the increase in the number of participants allows a decreasing number of errors detected. In any case, the interval estimation of the number of errors is a proposal to deepen.

1.1.3. The “Evaluator effect” of Hertzum

The authors state the Importance of the '"evaluator effect" [16] in the experimental phase. The detections rate of unique usability problems is reported in (3):

(3) $$Detection\ \ rate = Avg\,\frac{P_i}{P_{All}}\,over\ \ all\ \ n\ \ evaluators$$

Being:

- set of problems identified by each evaluator
- total number of problems identified by n evaluators.

Based on what is reported in [17], the "Detection rate" has two fundamental problems:

- Variability in the detection rate, based on samples size. The borderline case of a single evacuato (n=1), in which is obviously the detection rate is 100%, since = . This implies the need to interpret this index with caution, especially in the case of non-high sample size, defining a reasonable range of involved users, with particular attention to the lower limit.
- The basic assumption that the total number of interface problems is coincident with the total number of unique problems encountered by the evaluators is a strong weakness. Indeed, a very small sample may highlight a number of problems lower than the real one, thus affecting the analysis.

In order to overcome this drawback, the authors propose to use a new measure, based on the number of usability problems identified by at least two users, compared to the total number of concordances on the sample analyzed (Eq.):

$$\begin{gathered} Any-two \quad agreement = Avg \quad \frac{p_i \cap p_j}{p_i \cup p_j} \\ over \quad all \frac{1}{2} n(n-1) \quad pairs \quad of \quad evaluators \end{gathered} \tag{4}$$

However, this indicator, which varies from 0 to 100%, , cannot guarantee the detection of the total number of the errors in the interface. With reference to three well-known heuristic methods (CW, HE, TA), the authors evaluate the impact of the so-called "evaluator effect", taking into account various critical aspects in usability assessment (problems severity, complexity of the work- domain, complexity of the product, prototype fidelity, user-expertise). The results reveal a substantial inconsistency in the application of both the indicators (Detection rate and Any-two agreement rate) over the three methodologies. In conclusion, the authors suggest several guidelines for usability tests:

- Definition og the goals of the task
- Esplicitazione e semplificazione dell’obiettivo del task;
- Involvement of a large sample of users, especially for critical evaluations.

1.1.4. The Spool’s model

The study of Spool et al. [18] is based on the assessment of four web- interfaces, by using heuristic evaluation methods. The probability of finding a new problem at the i-th stage is:

(5) $$p_i = (L^{i-1})$$

Where of $^{-1is}$ the expected proportion of usability problems found testing any single user. The probability to find a new usability problem in the first test is (6):

(6) $$(a_i) = new_i \;/\; all_i$$

Being the estimates of L, based on too noisy, the authors used the the cumulative average of the values (7):

(7) $$L = (ia_i - (i-1)a_{i-1})^{1/(i-1)}, \;\; \underline{L} = \sum L(estimated)/i$$

Finally, starting from the eq. (5), the problems that remain to be found are (8):

(8) $$T_{x\%} = Log(x/100)/Log(\underline{L}) + 1$$

The results obtained, for the specific case study demonstrated that five users are allowed to find about the 35% of usability problems, in opposition with Nielsen findings. Instead, it was confirmed that the problems with a greater level of severity, tend to be first identified by users. The study of Spool et al. is strongly affected from the defined task. (purchase products online through Web interfaces) because the wide variety of tested interfaces negatively affected the findings rate [10].

On the other hand, the authors demonstrated the limitations of Nielsen’s theory. The rule "5 users is enough" is valid only if λ is equal to a fixed value (about 0.3), i.e. when all assumptions of the original model are valid, which in some cases may be too restrictive.

1.1.5. The Caulton's model

Caulton [19] develops the model of Virzi, introducing the hypothesis of non-homogeneity 'of the sample of experimenters. Moreover, the author also makes a classification of the usability problems:

- Shared problems, that occur with equal probability in all users;
- Unique problems, which are much more likely to occur in one subgroup than in another.

Based on this assumption, relaxing the homogeneity means that users belonging to different categories, have different probabilities of finding a unique usability problem. In conclusion, the authors propose to use the Virzi's model for shared problems, while for the unique problems, a new model is adopted , as reported in the eq.(9):

(9) $$N = 1-(1-p)^{(\#subjects/\#groups)}$$

This equation reflects that when the number of subgroups in the population increases, the number of unique problems found, decreases. The Caulton's model introducing the relaxation of the homogeneity assumption, allows to consider another important factor in usability testing: the user expertize. However, the identification of the number of subgroups within a population, is an unresolved issue. In addition, the proposed model assumes an equal difference in the level of expertize of users, which is a hypothesis to be tested.

1.1.6. Turner's model

Partendo dal modello di Nielsen e Landauer [10], Turner et. al [20] propose a new criterion for the estimation of p (or λ), in order to ensure a robust estimate of the number of users over a different composition of the sample and the type of the task analyzed. The proposed approach estimates the value of p as the average of the values obtained, respectively, with a normalization procedure(10) and the Good-Touring algorithm(11). The final formulation is reported in the eq. (12):

(10) $$p_{GT-adj} = p_{est} / (1+(E(N_1)/N)$$

$$(11) \qquad p_{norm-adj} = (p_{est} - 1/n)(1 - 1/n)$$

$$(12) \qquad p_{adj} = \tfrac{1}{2}(p_{GT-adj} + p_{norm-adj})$$

The study of Turner certainly represents an important evolution of the Nielsen's model, proposing a criterion for a robust estimation of p and providing guidelines for the application of the model in the experimental stage.

1.1.7. Kanis: *p estimators*

Kanis [22] analyses the mechanisms of biasing in the estimation of the number of usability problems (C), by using four estimators , in several testing methodologies (Think aloud, heuristic evaluation, "one shot" observation).

The author begins with two basic assumptions:

(i) the hypothesis that the number of problems detected in the experimental stage allows a correct estimate of the total number of real problems of the product / interface is illusory, as already highlighted in [17].

(ii) All the estimators proposed in the literature have points of weakness that must be taken into account.

The variables considered are shown in Figure 8:

Legend (variables in the shaded area are only addressed in the Appendix).

C	number of different usability problems
D_j	number of different usability problems discovered after j participants
D_j^*	mean of D_j generated in a number of permutations of the participant order involving a total of n participants ($j < n$)
$\overline{p}_n$	average discovery probability involving D_n different usability problems (after n participants)
p^u	probability of discovering usability problem u by any participant; index $u = 1 \ldots D_n \ldots C$
C_{Av}, C_F, C_D, C_T: estimates of C, see Appendix, including the overview in Table B	
D_j^f	(appears in estimators C_F and C_T) the number of different usability problems encountered by f out of j participants; or, in an alternative phrasing: the frequency of the number of times (f) that distinct usability problems have emerged amongst j participants
D_j^1	number of singletons
$f\|j$	(appears in estimator C_F) discovery probability of a usability problem that has been encountered by f participants out of j
$\overline{D}_1$	(appears in estimator of C_D) average number of usability problems discovered per participant
$\overline{D}_{j-1}$	(appears in estimator C_D) average number of usability problems discovered after $j-1$ participants, with j participants involved ($\overline{D}_{j-1}$ is the mean of J numbers)

Note. The term 'participant' should read 'evaluator' or 'practitioner' in a heuristic evaluation.

Figure 6: Tables of all variables analyzed (source Kanis, 2011)

The four estimators are reported in the eq. 12-16:

(13) $$C_{av} = \frac{D_j}{1-(1-\overline{p_j})^j}$$

(14) $$C_f = \sum_{f=1}^{j} \frac{D_j^f}{1-(1-\frac{f}{j})^j}$$

(15) $$C_D = \frac{\overline{D}_1 \overline{D}_{j-1}}{\overline{D}_1 + \overline{D}_{j-1} - D_j}$$

(16) $$C_T = \frac{\sum_{f=1}^{j} fD_j^f}{\sum_{f=1}^{j} fD_j^f - D_j^1}$$

The analysis shows that is the best estimator of C, although it could be underestimated in several cases. To avoid this problem, the author suggests referring to the maximum number between the two estimators e .

1.2. The level of expertise

One of the most significant factors in the definition of the user profile is the level of competence in the interaction with the product. Sauer et al. state that users can differ each other in several characteristics: [23].

- Competence or expertise: knowledge of the subject a specific context of reference. Based on this factor, a user can be roughly classified as novice or expert;
- Attitude: set of environmental factors that may affect user- product interaction;
- State: temporary conditions that can affect the user's choice;
- Personality: behavioral aspects related to user perceptions.

The level of expertise has been widely considered in literature. One of the aspects in which experts users differs from novice users is the level of proficiency and efficiency in the use of a product[1] . The learning curve for novice users, has a greater slope than that of experienced users, though, the level of efficiency achieved over time is always lower. Thus, designing a product for novice users, means to minimize the learning times, maximizing the affordances of the product [24], that is the product's ability to allow the user to use it with success from its first application.

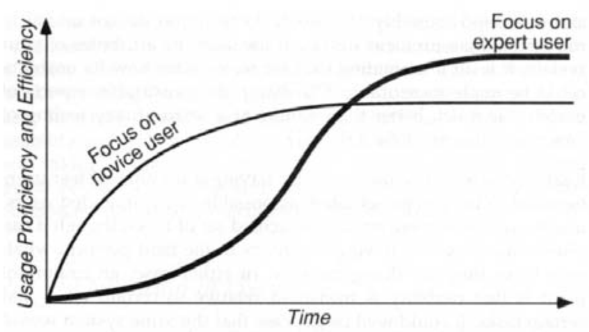


Figure 7: Learning curves for a hypothetical system. Source Nielsen J. "Usability Engineering"(1993)

The level of expertize may refer to[1] [2] as reported in Figure 10:

- a field of application;
- a specific system being evacuate;
- a single task (work domain knowlwdge).

Moreover, it is possible to consider groups of users with different background or individual performance. A good indicator to understand the difference between users in terms of expertise, is the ratio between the 75th and the 25th percentile on performance data, for single tasks. .For many tasks in computer sciences, this ratio is equal to 2. In relation to the field of application and the system under study, the level of expertise can respond to the need for segmentation of the user population, i.e. the product can be intended (and therefore designed) to users with different level of expertise. In this case the product or the interface must be flexible to different user requirements. In relation to the task, however, different considerations can be made, depending on the adopted testing methodology.

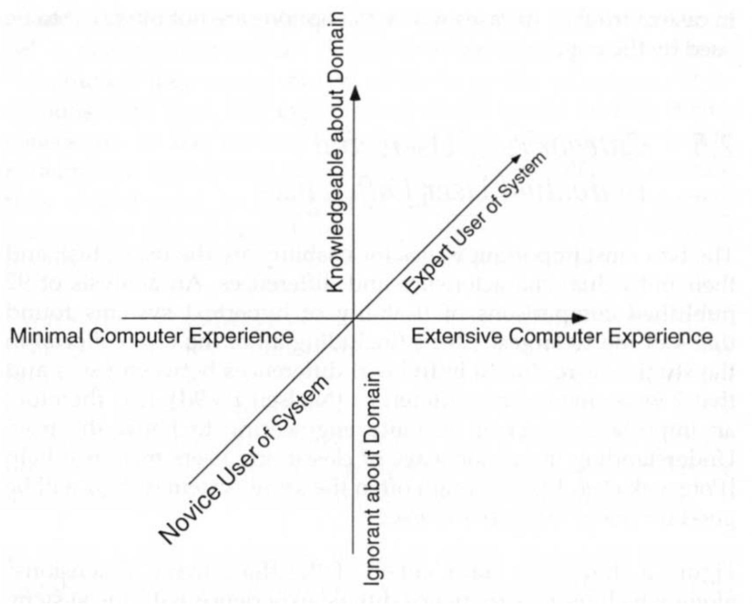


Figure 8: "User cube" of the three main dimensions along which user experience differs. Source Nielsen J. "Usability Engineering"(1993)

In usability testing, the level of expertise, substantially influence the performance of users. Ziefle [25] for instance, compares three models of cellular phones, checking relationships between expertise and measures of effectiveness and efficiency, in the execution of simple tasks. A summary of the main results achieved by Ziefle is shown in the Tab. 1.

The study highlights that level of expertise affects both, the level of success in task accomplishment and the execution time. Moreover the level of proficiency seems to be greater for novice users In some cases, however, the level of expertise can be a noise factor. This occurs, for instance, when it affects the ability of the user to interact with experimental tools, such as in virtual experiments. Indeed, the different familiarity with haptic devices, can completely distort the obtained results. [4]. In heuristic evaluations, however, people with different backgrounds can contribute to the detection of different interface problems. An interesting approach to this problem is that proposed by Caulton [19] (par. 1.1.5), based on the binomial model of Virzi[9]. The mentioned study, refers indirectly to the level of expertise, by considering how the heterogeneity of the sample can affect the rate of usability problems detected by the user In conclusion, several observations can be made:

Tab. 1: Synthesis of results obtained by Ziefle (2002)

Task	**Measure**	**Results**
Calling a number Calling/ phone directory Sending a SMS	Effectiveness measures: - % task	Significant effect of expertize on task success (F(1.58)=32.7;p<0.1)
Hiding the own number Editing a number in the phone directory Call divert	Efficiency measures: - time - # of additional step not required in the execution of the task	Significant effect of expertise on the average time of execution (F(1.58)=47.6;p<000.1) Significant effect of expertize on the # of additional step not required in the execution of the task (F(1.58)=19.1;p<0.0001)

- groups of users with different levels of expertise, highlight different usability problems in relation to the level of detail in the use of the product or interface;
- the existence of different subgroups, in terms of level of expertise, tends to lower the expected proportion of usability problems highlighted.

In carrying out an experiment, it is essential to check the representativeness of the sample analyzed and the relationship between the level of severity of the usability problems and the level of expertise.

In literature, th effectiveness of a usability evaluation is often dealt with the use of quantitative indicators. Hartson et al. [26] propose two metrics reported in the equations (17)(18), which refer, respectively, to the real number of problems, over the total number of problems reported (false positive) and the rate of real problems identified compared the real number of interface problems.

(17) $$Validity = \frac{Number \quad of \quad correct \quad predictions}{Number \quad of \quad problems \quad predictions}$$

(18) $$Throughness = \frac{Number \quad of \quad correct \quad predictions}{Number \quad of \quad real \quad problems}$$

Folstad et al. [27] propose a review of these indices(19), (20), (21), in an empirical study in which they compared the performance of work-domain experts and usability experts.

(19) $$Validity = \frac{ef}{(ef)+(hj)}$$

With:

- f number of real problems
- h number of false positive problems
- j mean of the false positive problems
- e average probability of prediction of the real problems calculated as:

(20) $$Pr\ ediction \quad Pr\ obability = 1 - \frac{n-k}{n} - \frac{n-k-1}{n-1} \ldots \frac{n-k-m+1}{n-m+1}$$

(21) $$Throughness = Mean \quad prediction \quad probability \quad for \quad real \quad problems$$

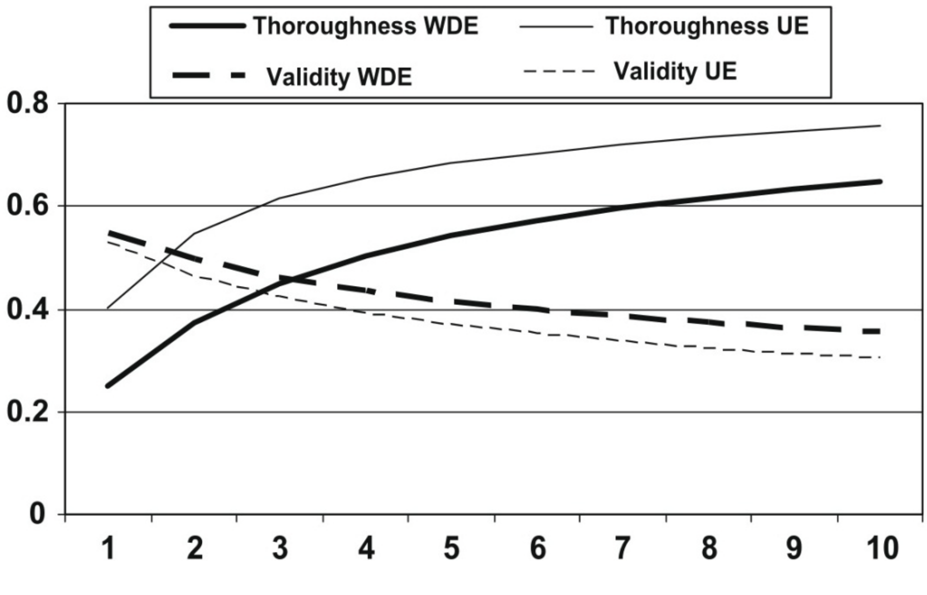


Figure 9: Validity and thoroughness values for nominal groups of work-domain expert evaluators (bold lines) and usability expert evaluators (thin lines). (source: Folstad et al.) [32]

The results obtained confirm that the involvement of domain- experts allow to obtain results comparable to those of usability experts in the term of validity. In terms of thoroughness it is evident but a significant difference has been highlighted that could be balanced by increasing the number of evaluators (work domain experts).

1.3. Prototipe fidelity

The prototype fidelity is one of the factors that most affect the outcome of usability testing. For obvious economic reasons, there is a strong tendency to use low-defined prototypes, which have a lesser impact on the budget. The introduction of virtual prototypes, has strongly influenced this process, due to its flexibility, which is obviously not absolutely comparable to that of a physical prototypes, in subsequent steps of redesign. However, there are, contrasting advices on the effectiveness of the low- fidelity prototypes, [28] because there are strongly limitations in simulating several product functionalities, with subsequent difficulties in usability problems detection. Inoltre, le percezioni dell’utente in termini di soddisfazione possono essere fortemente condizionate dalle differenze nell’interazione con il prodotto e dall’impatto estetico di quest’ultimo. Moreover, user perceptions, in terms of satisfaction, can be strongly affected by product aesthetics. One of the most known about is to Virzi, comparing low and high fidelity prototypes, using as a reference metric, the number of errors identified by users, for two types of electronic products (electronic book, interaction voice response system). The usability test was performed by using the think-aloud methodology.

The analysis of experimental results confirm a substantial equality in the number of usability errors detected with the two prototype. In addition there is a high correlation between the number of subjects identifying a specific usability problem with a single type of prototype. This result must be interpreted. Its validity is related to a specific methodology and a single metric of reference (usability problems). It is not possible, therefore, a generalization without further tests. Sauer et al.[23] they analyze the effects of prototype fidelity and user expertize on the results of a usability test. The tested product is a floor scrubber. The authors use three prototypes at different levels of definition:

Table 1: Synthesis of prototypes used in Sauer et al. (2009)

Level	**Prototype**
Fully operational	High-fidelity 
3D mock-up	Medium-fidelity 
Paper prototype	Low-fidelity 

Moreover the tasks in the study, have been modified in relation to the prototype fidelity.The results obtained confirm that:

- The user overestimates product requirements for low-level prototipes. Thus, the user compensates the absence of feed-back of use with by making cautionary choices.

- There are several limitations on the selection of performance measures , using low-level prototypes, compared to the fully operational prototypes.
- The subjective ratings of satisfaction and aesthetics are not influenced by the productfidelity. The user seems to use some sort of compensatory activities [2], which leads him to consider in the same way low-and high- fidelity prototype [2].

Bibliography

[1] Nielsen, J., (1993), Usability Engineering, Academic Press, Boston, MA, ISBN 1-12-518406-9.

[2] Sauer, J., Sonderegger, A., (2009). The influence of prototype fidelity and aesthetics of design in usability tests: effects on user behaviour, subjective evaluation and emotion. Applied Ergonomics 40, 670-677.

[3] Molich, R., Ede, M.R., Kaasgaard K., Karyukin, B. (2004). Comparative Usability evaluation. Behaviour & Information technology 23 65-74.

[4] G. Di Gironimo, A. Lanzotti, G. Matrone, A. Tarallo, M. Trotta, A virtual reality approach for usability evaluation of a wheelchair-mounted robot manipulators, Proceedings of TMCE2010, Ancona, 12 - 16 April 2010

[5] Patalano S., Galileo F., Lanzotti A., Matrone G., 2010, "Participatory design and virtual concept for postural seating system innovation", CD-Rom Proceedings of TMCE 2010 Symposium, Ancona, 12-16 April 2010, ISBN 9789051550603, pp. 1633-1646.

[6] Barnum, C., Bevan, N., Cockton, G., Nielsen, J., Spool, J., Wixon, D., The "Magic Number 5": Is It Enough for Web Testing? In CHI '03 Extended Abstracts on Human Factors in Computing Systems (Ft. Lauderdale, Florida, USA, April 05 - 10, 2003). CHI '03. ACM, New York, NY, 698-699

[7] Perfetti, C., & Landesman, L. (2002). Eights is not enough. Retrieved April 14, 2003 from ://world.std.com/~uieweb/Articles/eight_is_not_enough..

[8] Macefield, R.How To Specify the Participant Group Size for Usability Studies: A Practitioner's Guide. Journal of Usability Studies 5(1), 34-45 (2009).

[9] Virzi, R. A., Refining the test phase of usability evaluation: how many subjects is enough? Human Factors, 1992, 34(4), 457-468

[10] Nielsen, J., Landauer T.K., (1993), A Mathematical Model of the finding of Usability Problems, in Ashlund, S., Mullet, K., Henderson, A., Hollnagel E., White T. (Eds), Proceedings of ACM INTERCHI'93 Conference of Human Factors in Computing System, ACM, New York, 1993, pp. 206-213

[11] Woolrych, A., and Cockton, G. Why and When Five Test Users aren't Enough, in Proc. IHM-HCI Conference: Volume 2, 105-108, 2001.

[12] Faulkner, L.. (2003). Beyond the five user assumption: benefits of increased sample sizes in usability testing. Behaviour Research Methods, Instruments and Computers, 35(3), 379- 383

[13] Lewis, J.R. Sample sizes for usability studies: Additional considerations. Human Factors 36, 368-378 (1994).

[14] STEELE, R. G. D. and TORRIE, J. H. (1960). Principles and procedures of statistics. (McGraw-Hill, New York, NY).

[15] Lewis, J. R., 2001, Evaluation of procedures for adjusting problem-discovery rates estimated from small samples. International Journal of Human-Computer Interaction, 13, 445-479

[16] Jacobsen, N. E., Hertzum, M., & John, B. E. (1998). The evaluator effect in usability studies: Problem detection and severity judgments. In Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting (pp. 1336–1340). Santa Monica, CA: Human Factors and Ergonomics Society.

[17] Hertzum M., Jacobsen N.E. (2001). The evaluator Effect: A Chilling Fact About Usability Evaluation Methods. International Journal of Human Computer Interaction 13(4), 421-443

[18] Spool, J. and Schroeder, W. "Testing Websites : Five Users is Nowhere Near Enough. In Proc. CHI 2001, Extended Abstracts, ACM 285-286.

[19] Caulton, D. A. (2001). Relaxing the homogeneity assumption in usability testing. Behaviour & Information Technology, 20, 1-7.

[20] Turner, C. W., Lewis J. L., Nielsen J. Determining usability test sample size. International Encyclopedia of Ergonomics an Human Factors, Volume 3, CRC Press.

[21] Hwang , W., Salvendy, G. (2010) Number of people required for usability evaluation: the 10±2 rule. Communication of the ACM. 53(5), 130-133

[22] Kanis, H., (2011). Estimating the number of usability problems. Applied Ergonomics 42,2(2011), 337-347

[23] Sauer, J., Sonderegger, A., 2010. The influence of user expertize and prototype fidelity in usability tests. Applied Ergonomics 41, 130–140.

[24] Norman, D. (1999), Affordance, conventions, and design. Interactions 6(3): 38-43.

[25] Ziefle, M., (2002). The influence of user expertize and phone complexity on performance , ease of use and learnability of different mobile phones. Behaviour & Information Technology. 21(5), 303-311.

[26] Hartson, H.R., Andre, T.S., Williges, R.C., 2001. Criteria for evaluating usability evaluation methods. International Journal of Human–Computer Interaction 13 (4), 373–410.

[27] Folstad, A., Anda, B. C. D., Sjoberg, D. I. K. (2010). The usability inspection performance of work-domain experts: An empirical study. Interacting with Computer 22(): 75-87

[28] Virzi, R.A., Sokolov, J.L., Karis, D., 1996. Usability problem identification using both low- and high-fidelity prototypes. In: Conference Proceedings on Human Factors in Computing Systems: CHI 96, pp. 236–243

[29] J. Zhang, T.R. Johnson, V.L. Patel, D.L. Paige and T. Kubose, Using usability heuristics to evaluate patient safety of medical devices. J. Biomed. Inform. 36 1–2 (2003), pp. 23–24

[30] Medlock, M.C., Wixon, D., Terrano, M., Romero, R., Fulton, B. (2002). Using the RITE method to improve products: a definition and a case study. Proc. Usability Professionals Association (Orlando FL, July 2002).

[31] Turner, C. W., Lewis, J. R., & Nielsen, J. (2002). UPA Panel: How many users is enough? Determining usability test sample size

[32] Nielsen, J. 1992. Finding usability problems through heuristic evaluation. Proceedings ACM CHI'92 Conference (Monterey, CA, May 3-7), 373-380.

Chapter III

Usability metrics

Usability is a multidimensional characteristic, as reported in the ISO 9241[1], that depends from objective and subjective aspects of user- product interaction.

Because of its large extent, it's very complicated to direct measure usability, but it's possible to define several indicators for the indirect measure of that, starting again from user- product interaction. Indeed, measurement of product usability cannot be apart from the analysis of user requirements, the goal of the study, the designed task and the context of use [1]. Moreover, the choice of the right metrics for usability assessment depends also on several issues, such as the technology available for data collection and analysis and the budget.

The ISO 9241 standard shows several examples of usability measures (Tab. 1) and states that there is no general rule for their choice or combination, but it is necessary to provide at least one measure for each usability dimension (effectiveness, efficiency and satisfaction). An effective summary of the appropriate usability metrics related to the most common usability studies, is provided by [2], as reported in *Tab. 2.*

Starting from main literature contributions, the purpose of this chapter is to deepen the most relevant issues in selecting the right metrics for usability assessment. The usability measures will be grouped in measure of effectiveness, measure of efficiency and measures of satisfaction, according to the ISO 9241 standard.

3.1. Measures of effectiveness

The ISO 9241 standard defines effectiveness as the level of "accuracy and completeness with which users achieve a specified goal". Starting from this statement, all the measures of effectiveness could be considered as performance metrics, that estimate the magnitude of specific usability issues, giving information related to the way in which the users behave and interact with the product and also about the use of scenarios and tasks. Following a brief overview of the most important measures of effectiveness used in literature, is provided.

Tab. 1 Examples of measures of usability provided by the ISO 9241 standard

Effectiveness	Efficiency	Satisfaction
- Percentage of goals achieved - Percentage of users successfully completing task - Average accuracy of completed task	- Time to complete a task - Task completed per unit time - Monetary cost of performing the task	- Rating scale for satisfaction - Frequency of discretionary use - Frequency of complaints

3.1.1. Task Completion

The task completion gives refers to whether the user completes a specified task. Usually, this measure is a binary variable (e.g. 1= success, 0= failure) or a discrete variable when the number of correct/ failed tasks is accounted. Especially in the latter case, it is very important to decide beforehand the references to determine the level of completion or level of success in task execution. In [2], six level of completion are reported:

- complete success with assistance
- complete success without assistance
- partial success with assistance
- partial success without assistance

- failure (the user didn't understand that the task is incomplete)
- failure (the user does not complete the task)

Starting from this classification, the role of a moderator and the level of interaction between him and the, change completely the structure and the outcome of the test. The number of correct/ failed tasks can be monitored also in a limited time [3]

3.1.2. Number of errors

The number of errors is a measure of accuracy in the task's completion or in the solution to the task. Based on the goal of the study, it's possible to measure this metric directly (number of errors in task, in a subtask or in a series of tasks) or indirectly (percentage of correct solutions, number of hints to complete a task, task to criterion as, for instance, the number of attempts to complete a given number goals) [3].

3.1.3. Spatial accuracy

The spatial accuracy is another measure of accuracy in product or interface manipulation, during a task accomplishment. This metric could be translated in a distance from a target (point or trajectory) or an error in terms of orientation [3].

3.1.4. Other measures of effectiveness

Other measures used in literature are [3]:

- recall: user's ability to remember specific features of the interface(e.g. button's position) and to recall them in a specific task;
- completeness: user's ability to accomplish the designed task in an exhaustive way. Usually it's measured taking into account the number of secondary tasks done.

Tab. 2 Metrics used in usability assessment. Source Tullis and Albert "Measuring the user experience"(2008)

Usability Study Scenario	Task Success	Task Time	Errors	Efficiency	Learn-ability	Issues-Based Metrics	Self-Reported Metrics	Behavioral and Physiological Metrics	Combined and Comparative Metrics	Live Website Metrics	Card-Sorting Data
1. Completing a transaction	X			X		X	X			X	
2. Comparing products	X			X			X		X		
3. Evaluating frequent use of the same product	X	X		X	X		X				
4. Evaluating navigation and/or information architecture	X		X	X							X
5. Increasing awareness							X	X		X	
6. Problem discovery						X	X				
7. Maximizing usability for a critical product	X		X	X							
8. Creating an overall positive user experience							X	X			
9. Evaluating the impact of subtle changes										X	
10. Comparing alternative designs	X	X				X	X		X		

3.2. Measure of efficiency

Based on ISO 9241 standard, efficiency is “the level of effectiveness achieved to the expenditure of resources”. In this case, measures of effectiveness could be both, performance metrics and human effort that is a subjective aspect in user- product interaction.

3.2.1. Time

The time, usually measure the how long the take the user to complete a specific task. This metric could be taken into account in many different ways:

- time to complete a task or a part of that;
- time for single specific actions;
- time between two actions;
- time in help function;
- reaction’s time to a warning

3.2.2. Input rate

Input rate is an efficiency metric, used in particular for the study of intarfaces’ usability. It could be monitored considering the speed of text entry or the the average number of the correctly entered digits for several input methods.

3.2.3. Mental effort

Mental effort is a measures of the cognitive load of the user in task execution. One of the most effective methods to measure the mental effort is the NASA’s Task Load Index questionnaire [20], based on the six indicators reported in Tab. 2. For each of them a score from o to 100 is assigned. At the same time all the indicators are weighted by using the pair- wise comparison. Finally the mental workload is obtained as a weighted sum of the average scores for each indicator [21].

Other measures user for mental workload assessment are:

- task difficulty (rated by experts)
- physiological measures (heart variability)

3.2.4. Communication effort

The communication effort gives a measure of the amount of resources expended in the communication process [3]. Being related to the cognitive load for the user, this metric could be assessed indirectly, monitoring several indicators during the execution of the task (number of interruptions, number of question asked etc.), or directly with a score assigned by the test's administrator.

Tab. 3 Indicators of the NASA's Task Load Index questionnaire

TITLE	ENDPOINTS	DESCRIPTIONS
MENTAL DEMAND	Low/High	How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	Low/High	How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	Low/High	How much time pressure did you feel due to the rate or pace at which the task or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	Good/Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter? How satisfied were you with your performance in accomplishing these goals?
EFFORT	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	Low/High	How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

3.2.5. Other measures

Other measures of efficiency used in literature are [3]:

- Usage patterns: measure of the level of use of an interface in a specific task (umber of mouse clicks, number of interface actions etc.)

3.3. Learnability: performance metrics over time

The Learnability gives an idea of the proficiency of the user in using a product. It strictly related to all the features of the product that allow users to understand easily how to handle a specifc device, improving the performance level quickly [1]. Thus this metric could be defined as the change of effectiveness and efficiency measures over time [4], as shown in *Figure 1*.

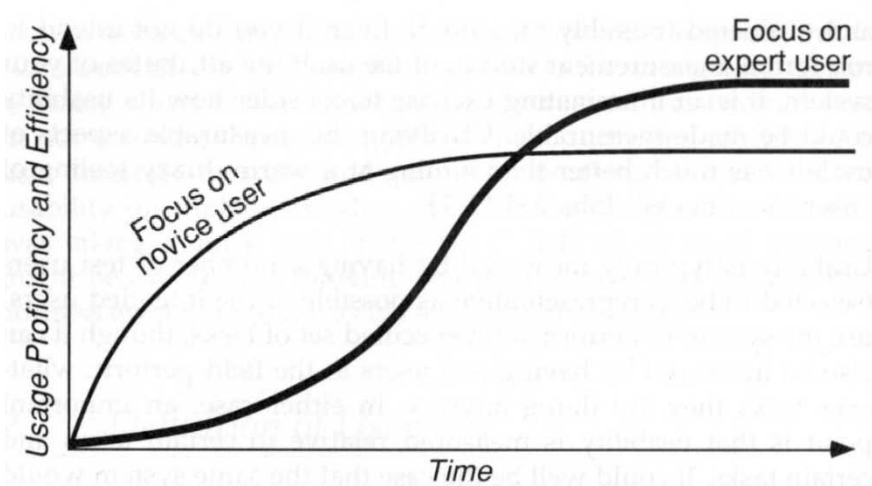


Figure 1: Learning curves for a hypothetical system. Source Nielsen J. "Usability Engineering"(1993)

Starting from this definitions, collecting learnability data, means to collect performance data at multiple times. Thus the main steps in learnability assessment are:

- Selection of performance metrics to track: many studies in literature focus on efficiency metrics (e.g. time) [2]. In some case, also effectiveness metrics have been considered (e.g. percentage of tasks solved)[1].

- Definition of the time to allow between trials (trials within the same session with or without breaks between tasks, trials between sessions) [2].

3.4. Measures of satisfaction and questionnaires for satisfaction's data collection

Starting from ISO 9241, usability cannot be assessed without taking into account also users' perceptions in user- product interaction. Moe specifically, the standard define the satisfaction , the third dimension of usability, as "the condition of freedom from discomfort and positive attitude towards the use of the product". The most used measures of satisfaction are [3]:

- preference: ranking of the interfaces (or products), obtained forcing users to elicit their level of satisfaction;
- ease- of- use: general level of satisfaction related to a specific product or interface;
- specific attitudes: perception of connection between the user and other persons;
- perception of the outcomes: perception of the final result of the interaction
- perception of interaction: perceptions related to the interaction (reliable, natural etc.)

Being related to subjective aspects of user-product interaction, these metrics could be defined also as self- report metrics [2]. Usually this kind of data are collected by using standards questionnaires, that could be administrated at the end of each task (post- task ratings) or at the end of the entire session (post- session ratings). Following the most important questionnaires are reported.

3.4.1. Post- task ratings

The main aim of the post- task rating is to achieve the usability assessment of a product, or of a subsystem of that, in terms of user perceptions. The definition of tested tasks reflects the components of usability that designers consider important for the specific case study. Usually the collection of these kind of data is achieved by using post- task questionnaires, in order to avoid from one side problems of concentration of the user during the test and, on the other side, the bias due to the interaction between users and test's administrator (desirability bias). Following a brief review of the techniques most frequently used, is presented.

3.4.1.1. Ease of use Questionnaire

This tool is used to assess the level of difficulty with which the user interact with a product in a specific task [2]. Usually, the ease of use for the tested product, is assessed by using 5/ 7-point numeric or semantic scales (Likert). In the latter case, the respondents answer to the statement "This task was easy to complete" with their level of agreement as reported in *Tab. 4*.

Tab. 4: Ease of use Questionnaire

This task was easy to complete						
	1	2	3	4	5	
strongly disagree						*strongly agree*

3.4.1.2. After Scenario questionnaire (ASQ)

The "After Scenario Questionnaire" [6] is a three-item questionnaires related to the three "Usability dimensions" provided by the ISO 9241 standard [1]:

- Level of completion of the task in a specific scenario (Effectiveness, satisfaction)
- Time required to complete the task (Efficiency, Satisfaction)
- Adequacy of support information provided to the user before the test administration (satisfaction)

The respondents answer to the statements with their level of agreement by using the 7-points scale anchored at the end with the terms "Strongly agree" and "Strongly disagree" and a "Not applicable" point outside the scale *Tab. 5*.

This questionnaire should be administrated immediately following a scenario- based usability study, where “scenario” means a collection of tasks related to a specific products [9].

Tab. 5: ASQ Questionnaire developed by Lewis (1991)

<table>
<tr><td colspan="7">I am satisfied with the ease of completing the tasks in this scenario</td><td rowspan="3">Not applicable
N/A</td></tr>
<tr><td></td><td>1</td><td>2</td><td>…</td><td>6</td><td>7</td><td></td></tr>
<tr><td>strongly disagree</td><td></td><td></td><td></td><td></td><td></td><td>strongly agree</td></tr>
<tr><td colspan="7">I am satisfied with the amount of time it took to complete the task in this scenario</td><td rowspan="3">Not applicable
N/A</td></tr>
<tr><td></td><td>1</td><td>2</td><td>…</td><td>6</td><td>7</td><td></td></tr>
<tr><td>strongly disagree</td><td></td><td></td><td></td><td></td><td></td><td>strongly agree</td></tr>
<tr><td colspan="7">I am satisfied with the support information</td><td rowspan="3">Not applicable
N/A</td></tr>
<tr><td></td><td>1</td><td>2</td><td>…</td><td>6</td><td>7</td><td></td></tr>
<tr><td>strongly disagree</td><td></td><td></td><td></td><td></td><td></td><td>strongly agree</td></tr>
</table>

3.4.1.3. Printer Scenario Questionnaire (PSQ)

The “Printer Scenario Questionnaire”[6] is the early version of te ASQ. The structure of the two questionnaires is very similar, but the first one uses a 5- point scale, instead of a 7- point scale. Several studies in literature[10] demonstrated that the results of the ASQ and PSQ are broadly comparable. The only difference is in terms of internal consistency. Indeed, the PSQ shows a lower value of alpha, due to the use of a 5-point scale, instead of 7- points scale.

3.4.1.4. Expectation Measure

This method compares, for each user, the perceived level of difficulty (experience rating) and the expected level of difficulty, based on task description (expectation rating)[2].

Usually, a 7- point rating scale anchored at the end with the terms “Very difficult”=1 and “Very easy”=7 is used for both ratings. Analyzing data, it’s possible to define the four scenarios represented in Figure 2:

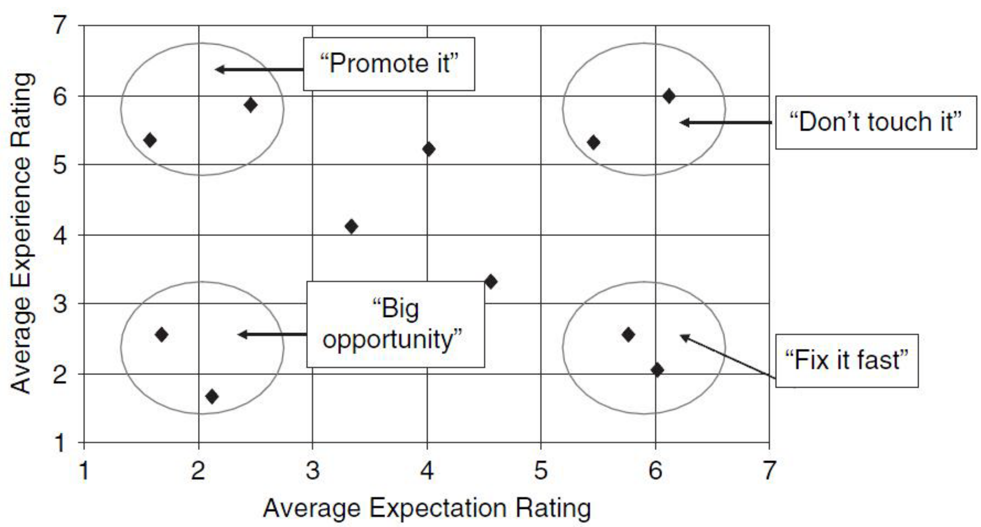


Figure 2: Comparison between average and expectation rating. Source: Tullis and Albert Measuring the user experience, adapted from Albert and Dixon(2003)

- “Fix it fast” scenario, which corresponds to an strong level of dissatisfaction of the users (level of difficulty higher than expected) showing high criticalities of the product that must be promptly solved;
- “Don’t touch it” scenario, in which there is complete consistency between expectations and perceptions of the user in terms of high level of difficulty in interacting with the product. It is therefore an optimal condition;
- “Promote it” scenario, which corresponds to an strong level of satisfaction of the users (level of difficulty lower than expected) showing features that distinguish the product from competitors and that must be improved;

- “Big opportunity” scenario”, in which there is complete consistency between expectations and perceptions of the user in terms of low level of difficulty in interacting with the product. It is therefore a clear opportunity to make improvements.

3.4.1.5. Usability Magnitude Estimation

The Usability Magnitude Estimation approach [11] is based on users self- reported measures. According to classical psychophysics methods, the procedure starts providing to participants the extremes reference designs (examples of “good” and bad “designs”) and asking a rating for both of them. Then, the user should rate the accomplished task, taking into account the scored provided for the extremes design as a reference. The comparison between several studies, is allowed by using the Master Usability Scaling technique, through the creation of a universal usability continuum[13].

3.4.2. Post- session ratings

The post- session metrics are always self- report metrics, that are administrated after the whole usability test (instead that after a specific task), in order to allow comparison between multiple design alternatives or score record of the global usability of a product over time [2].

3.4.2.1. System Usability Scale (SUS)

The “System Usability Scale” is a ten-item questionnaires related to user- system interaction, with odd-numbered items worded positively and even-numbered items worded negatively (*Tab. 6*). The respondents answer to the ten statements with their level of agreement by using the 5-points Likert scale, anchored at the end with the terms “Strongly agree” and “Strongly disagree”. After the task execution, the ratings are combined in a overall score with a given technique. The score contribution for each task, range from 0 to 4. For positively-worded items (1, 3, 5, 7 and 9), the score contribution is the scale position minus 1. For negatively-worded items (2, 4, 6, 8 and 10), it is 5 minus the scale position. To get the overall SUS score, the sum of the item score contributions is multiplied by 2.5. Thus, SUS scores range from 0 to 100 in 2.5-point increments [14]. Despite the practitioners describe this tool as a “quick and dirty” usability scale, recent studies demonstrate that SUS has a level of reliability (alpha coefficient of 0.85) higher than typical minimum reliability goal for questionnaires used in usability assessment (about 0.70) [17]. The use of SUS presents many advantages related to [16]: easy of use, minimal training required, immediately comprehensible output, applicability in various domains, easy comparison of different products, easy to use in conjunction with other UEMs, quick application. A review of the SUS is proposed by Findstad [18], who demonstrated that the original version of the SUS could be not suitable for non- native English speakers.

Tab. 6: SUS Questionnaire. Source Tullis and Albert “Measuring the user experience”(2008)

	1 (Strongly disagree)	2	3	4	5 (Strongly agree)	
1. I think that I would like to use this system frquently.					√	4
2. I found the system unnecessarily complex.				√		1
3. I thought the system was easy to use.		√				1
4. I think I would need the support of a technical person to be able to use this system.	√					4
5. I found the various functions in this system were well integrated.		√				1
6. I thought this system was too inconsistent.			√			2
7. I would imagine that most people would learn to use this system very quickly.		√				1
8. I found the system very cumbersome to use.				√		1
9. I felt very confident using the system.					√	4
10. I needed to learn a lot of things before I could get going with this system.		√				3

Total = 22 **SUS Score = 22 × 2.5 = 55**

3.4.2.2. Post- Study System Usability questionnaire (PSSUQ)

The Post- Study System Usability questionnaire (PSSUQ) [6] is a post-study questionnaire, developed to be administrated in person in order to provide to participants an overall evaluation on the product/ system in terms of usability. It consists of 19 items selected from a group of evaluators and related to ease of use, ease of learning simplicity, effectiveness, information and user interface. The psychometric assessment conducted by Lewis, revealed that that basic items, through the principal factor analysis, could be grouped in three human engineering factors (system usefulness, information quality and Interface quality) which account for the 87% of the variance.

3.4.2.3. Computer System Usability Questionnaire (CSUQ)

The Computer System Usability Questionnaire (CSUQ) [6] is a post-study questionnaire, developed to be administrated online, strating from the PSSUQ. It consists of 19 items that could be grouped in four categories:

- System usefluness
- Information quality
- Interface quality
- Overall satisfaction

The respondents answer to the statements with their level of agreement by using the 7-points scale anchored at the end with the terms "Strongly agree" and "Strongly disagree" and a "Not applicable" point outside the scale. The psychometric assessment conducted by Lewis, revealed that that the factor structure of the CSUQ is very similar to that of the PSSUQ. The basic items could be grouped in three human engineering factors (system usefulness, information quality and Interface quality) which account for the 98.6% of the variance.

Tab. 7: CSUQ Questionnaire. Source Tullis and Albert "Measuring the user experience"(2008)

		1	2	3	4	5	6	7		NA
1. Overall, I am satisfied with how easy it is to use this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
2. It was simple to use this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
3. I can effectively complete my work using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
4. I am able to complete my work quickly using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
5. I am able to efficiently complete my work using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
6. I feel comfortable using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
7. It was easy to learn to use this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
8. I believe I became productive quickly using this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
9. The system gives error messages that clearly tell me how to fix problems	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
10. Whenever I make a mistake using the system, I recover easily and quickly	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
11. The information (such as online help, on-screen messages, and other documentation) provided with this system is clear	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
12. It is easy to find the information I needed	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
13. The information provided for the system is easy to understand	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
14. The information is effective in helping me complete the tasks and scenarios	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
15. The organization of information on the system screens is clear	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
16. The interface of this system is pleasant	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
17. I like using the interface of this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
18. This system has all the functions and capabilities I expect it to have	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
19. Overall, I am satisfied with this system	strongly disagree	○	○	○	○	○	○	○	strongly agree	○
		1	2	3	4	5	6	7		NA

3.1.1.1. Questionnaire for User Interface Satisfaction (QUIS)

The Questionnaire for User Interface Satisfaction (QUIS), was developed by the University of Maryland in order to assess customer satisfaction in user- product interfaces. In their first paper [19], Chin et al. developed five version of the questionnaire. For all of them, the aim is the elicitation of subjective user opinions on all usability aspect related to user- product interaction (ease of use, system capability, consistency, learning). The questionnaire consists of 27 rating scales, grouped in five categories. In the first one (overall reaction to the software), the users rate directly the interface without any statement, by using a semantic differential scale with polar opposites.

In the others sections the respondents answer to the 21 statements with their level of agreement by using the 9-points scale anchored at the end again with polar opposites terms and a “Not applicable” point outside the scale [19].

Tab. 8: QUIS Questionnaire. Source Tullis and Albert “Measuring the user experience”(2008)

OVERALL REACTION TO THE SOFTWARE		0	1	2	3	4	5	6	7	8	9		NA
1.	terrible	○	○	○	○	○	○	○	○	○	○	wonderful	○
2.	difficult	○	○	○	○	○	○	○	○	○	○	easy	○
3.	frustrating	○	○	○	○	○	○	○	○	○	○	satisfying	○
4.	inadequate power	○	○	○	○	○	○	○	○	○	○	adequate power	○
5.	dull	○	○	○	○	○	○	○	○	○	○	stimulating	○
6.	rigid	○	○	○	○	○	○	○	○	○	○	flexible	○
SCREEN		**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**		**NA**
7. Reading characters on the screen	hard	○	○	○	○	○	○	○	○	○	○	easy	○
8. Highlighting simplifies task	not at all	○	○	○	○	○	○	○	○	○	○	very much	○
9. Organization of information	confusing	○	○	○	○	○	○	○	○	○	○	very clear	○
10. Sequence of screens	confusing	○	○	○	○	○	○	○	○	○	○	very clear	○
TERMINOLOGY AND SYSTEM INFORMATION		**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**		**NA**
11. Use of terms throughout system	inconsistent	○	○	○	○	○	○	○	○	○	○	consistent	○
12. Terminology related to task	never	○	○	○	○	○	○	○	○	○	○	always	○
13. Position of messages on screen	inconsistent	○	○	○	○	○	○	○	○	○	○	consistent	○
14. Prompts for input	confusing	○	○	○	○	○	○	○	○	○	○	clear	○
15. Computer informs about its progress	never	○	○	○	○	○	○	○	○	○	○	always	○
16. Error messages	unhelpful	○	○	○	○	○	○	○	○	○	○	helpful	○
LEARNING		**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**		**NA**
17. Learning to operate the system	difficult	○	○	○	○	○	○	○	○	○	○	easy	○
18. Exploring new features by trial and error	difficult	○	○	○	○	○	○	○	○	○	○	easy	○
19. Remembering names and use of commands	difficult	○	○	○	○	○	○	○	○	○	○	easy	○
20. Performing tasks is straightforward	never	○	○	○	○	○	○	○	○	○	○	always	○
21. Help messages on the screen	unhelpful	○	○	○	○	○	○	○	○	○	○	helpful	○
22. Supplemental reference materials	confusing	○	○	○	○	○	○	○	○	○	○	clear	○
SYSTEM CAPABILITIES		**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**		**NA**
23. System speed	too slow	○	○	○	○	○	○	○	○	○	○	fast enough	○
24. System reliability	unreliable	○	○	○	○	○	○	○	○	○	○	reliable	○
25. System tends to be	noisy	○	○	○	○	○	○	○	○	○	○	quiet	○
26. Correcting your mistakes	difficult	○	○	○	○	○	○	○	○	○	○	easy	○
27. Designed for all levels of users	never	○	○	○	○	○	○	○	○	○	○	always	○
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**		**NA**

The preliminary study conducted by chin et al. revealed an higher reliability of the questionnaire (Cronbach’s alpha of the QUIS 5.0 equal to 0.94). The use of SUS presents many advantages related to [16]: easy of use, minimal training required, immediately comprehensible output, reliability of the output, modifiability in relation to the requirements, effectiveness also for small sample size.

Bibliography

[1] ISO/IEC standard: Ergonomic requirements for office work with visual display terminals-guidance on usability. ISO 9241-11:1998 (1998)

[2] Tullis, T., Albert, B., (2008).Measuring the user Experience. Morgan Kaufman Publisher

[3] Hornbaek, K., Current practice in measuring usability: challenges to usability studies and research. Int. J. of Human- Computer Studies. 64 (2006), 79- 102.

[4] Nielsen, J., (1993), Usability Engineering, Academic Press, Boston, MA, ISBN 1-12-518406-9.

[5] Ziefle, M., (2002). The influence of user expertize and phone complexity on performance , ease of use and learnability of different mobile phones. Behaviour& Information Technology. 21(5), 303-311

[6] Lewis, J. R. (1991). Psychometric evaluation of an after-scenario questionnaire for computer

[7] usability studies: The ASQ. SIGCHI Bulletin, 23(1), 78–81. Also see http://www.acm.org~perlman/question.cgi?form=ASQ.

[8] Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. International Journal of Human-Computer Interaction, 7(1), 57–78. Also see http://www.acm.org/~perlman/question.cgi?form=CSUQ.

[9] Kirakowski, J.(2007).The Use of Questionnaire Methods for Usability Assessment, Assessment Volume: 2008, Issue: January 20, Publisher: University College Cork, Pages: 1-7 ISBN: 9738120985

[10] Lewis, J. R. An after- scenario questionnaire for usability studies: psychometric evaluation over three trials. SIGCHI Bulletin 23 (4), 79.

[11] Albert, W., & Dixon, E. (2003). Is this what you expected? The use of expectation measures in usability testing. Proceedings of the Usability Professionals Association 2003 Conference, Scottsdale, AZ.

[12] McGee, M. (2003). Usability magnitude estimation. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Denver.

[13] McGee (2004). Master usability scaling: Magnitude estimation and master scaling applied to usability measurement. Proceedings of CHI Conference on Human Factors in Computer Systems, Vienna. New York: ACM Press.

[14] Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), Usability evaluation in industry. London: Taylor & Francis. Also see ://www.usability.serco.com/trump/documents/Suschapt. .

[15] James R. Lewis, J.R., Sauro, J. (2009) The Factor Structure of the System Usability Scale, Human Centered Design . Lecture Notes in Computer Science 5619, © Springer-Verlag Berlin Heidelberg, pp 94-103

[16] Stanton, A. N., Salmon, P. M., Walker, G. H., Baber, C., Jenkins, D. P. (2005) Human Factors Methods. AshgatePublisher

[17] Kirakowski, J.: The Use of Questionnaire Methods for Usability Assessment (1994), ://sumi.ucc.ie/sumipapp.

[18] Finstad, K., The system usability scale and non-native English speakers. Journal of Usability Studie. 1 (4), 185-188

[19] Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. ACM CHI '88 proceedings, 213–218.

[20] Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), Human mental workload (pp. 139–183). Amsterdam: North-Holland.

[21] Rubio, S., Díaz, E., Martín, J., Puente, J.M. (2004). Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods 2004. Applied Psychology 53 (1), 61–86.

Chapter IV

Combined metrics for usability assessment

The collection of different kind of metrics in usability tests is an important source of information for designers, in order to improve product's usability. Therefore, sometimes, starting from single measures, it's complicated to select the best design alternative, enhancing both performance and satisfaction of the users. Thus, the assessment of the global usability of a product is a challenge for many reasons:

- the outcome of user- product interaction depends from factors related to subjective and objective measures that pertain to completely different fields;
- the usability metrics have different measurement scales and magnitude;
- based on product use, metrics could have a different level of importance;
- people involved in product life- cycle process , but not experts in usability studies (designers, managers), are often not able to interpret and use data from a usability test.

All these issues highlight the necessity to provide simplest tools in usability assessment, summarizing contributions (in terms of metrics) of different nature. Several authors tried to assess usability, combining usability metrics in a single score. Following, the main contributions in literature are reported.

4.1. Summated Usability Index (SUM)

4.1.1. The model

In [1], Sauro et al. developed a quantitative model to summarize usability metrics in a single score, starting from ISO 9241 standard definition of usability [2]. The assessment of usability has been realized, starting from four metrics widely used in literature:

- Time (measure of efficiency
- # of errors (measure of effectiveness)
- Completion (measure of effectiveness)
- Average of satisfaction (measure of satisfaction)

The general structure of the model is reported in Figure 1.

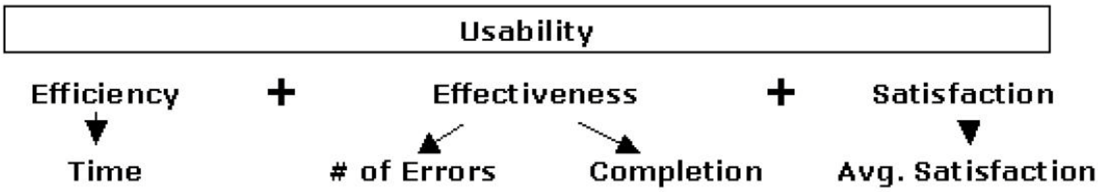


Figure 1: Quantitative Model of Usability. Source: Sauro, J., Kindlund, E. (2008). A method to standardize usability metrics into a single score.

Usability test were conducted to assess three Windows- based interfaces and a web-based application. Then usability metrics were combined in a single index named Summated Usability Metric (SUM), using the Principal Component Analysis (PCA) [3]. The purpose of this technique is the reduction of the original set of observed variables in a reduced set of latent variables. This is done through a linear transformation, which projects the original variables in a new Cartesian system in descending order of variance. Using this methodology, the authors built a better model of usability aimed at remove redundant data from the overlapping variables .

4.1.2. The methodological approach

The main steps of the proposed methodology are:

1. *Examination of relationships between metrics.*

The application of PCA is based on the hypothesis of collinearity in the set of variables analysed. Otherwise, the principal components coincide with the observed variables, except for a rearrangement according to the variance. In order to verify the relationships between variables, the correlation matrix for the four datasets was analysed. Results of this step show a moderated correlation between metrics ranged from 0.3 to 0.5. Moreover satisfaction level seems to be positively affected by performance metrics.

2. *Application of PCA and application of components to retain.*

Once the correlation matrices and their eigenvalues were obtained, the principal components were defined for each test, according to the following heuristic evaluation criteria[1][3]:

- Cumulative variance
- Kaiser's rule
- Scree plot test

The results of this phase show that the first PC accounted for more than the 50% of the variance for all the tests. Moreover, all the variables resulted significant (each variable added new information not contained in the others). Thus the first PC is a linear combination, obtained from the original set of experimental variables, or rather, the four collected metrics (time, errors, completion and satisfaction). The coefficients in this linear combination define the weight of each variable in terms of variance. Since all four variable have roughly the same coefficients, the authors concluded that all metrics had the same relevance.

Finally, the interpretation of the coefficients revealed that the level of completion and the satisfaction tended to increase when the time and number of errors decreased .

3. *Standardization and final definition of the index SUM*

The first PC, so defined, was assumed as a single score for usability assessment. In order to allow different components scores across data sets, all variables were standardized, using techniques reported in [4][5]. Then a single, standardized and summated usability metric (SUM) for each task was obtained by averaging together the standardized values of the variables (time, errors, completion and satisfaction), based on the equal weighting of the coefficients from the PCA. The coherence between the so defined index and the first PC was verified with a regression analysis, which confirmed a strong positive correlation between them Figure 2.

4.1.3. Main issues in SUM model and applications

The main issue in the model to proposed by Sauro et al, is represented by its the lack of generality. When from the application of the PCA, more than one PC must be retained, it is not possible to define a single score that summarize the original variable's set. Moreover, the same results imply that the input variables have different weights (in terms of variance). It is clear that the interpretation of more than one principal component is strongly affected by the designers' experience and sensitivity in evaluating their correlations with the input variables. Furthermore, the PCA needs a huge number of experimental data, which go far beyond the possibilities of participatory tests.

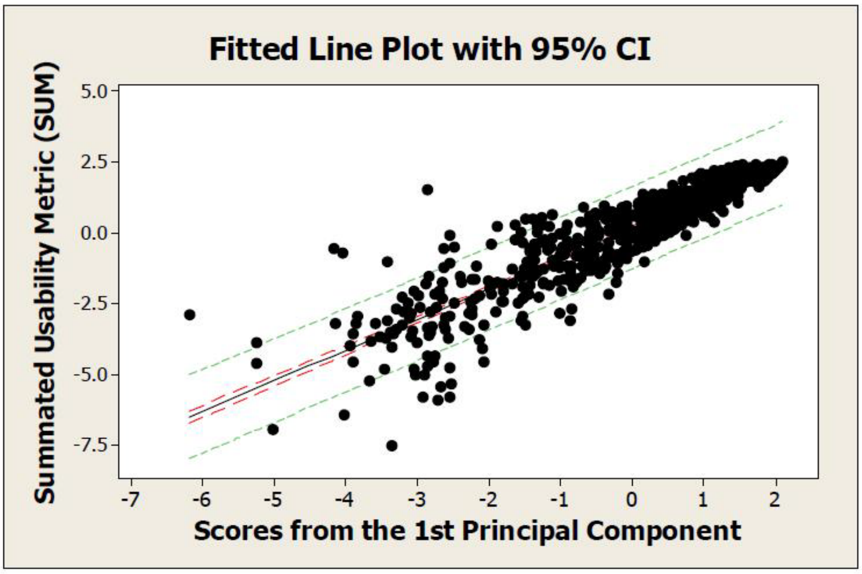


Figure 2: Regression Plot of PCA Score and SUM. Source: Sauro, J., Kindlund, E. (2008). A method to standardize usability metrics into a single score.

4.2. Il modello di Kim

4.2.1. The model

In [6], Kim et al. provided a new approach for the usability assessment of industrial products through the definition of a synthetic index (Integrated usability index). The quantitative model of usability, is based, once again on a hierarchic structure (Figure 3). The lowest level of the model is represented by the usability dimensions, translated during the experimental phase, in measurable functions (usability measures). Based on the field of application, all the usability measures could be grouped. For each so- defined subgroup, a synthetic usability index could be defined (Individual Usability Index). Then, the linear combination of all these indices, is the aforementioned Integrated Usability Index.

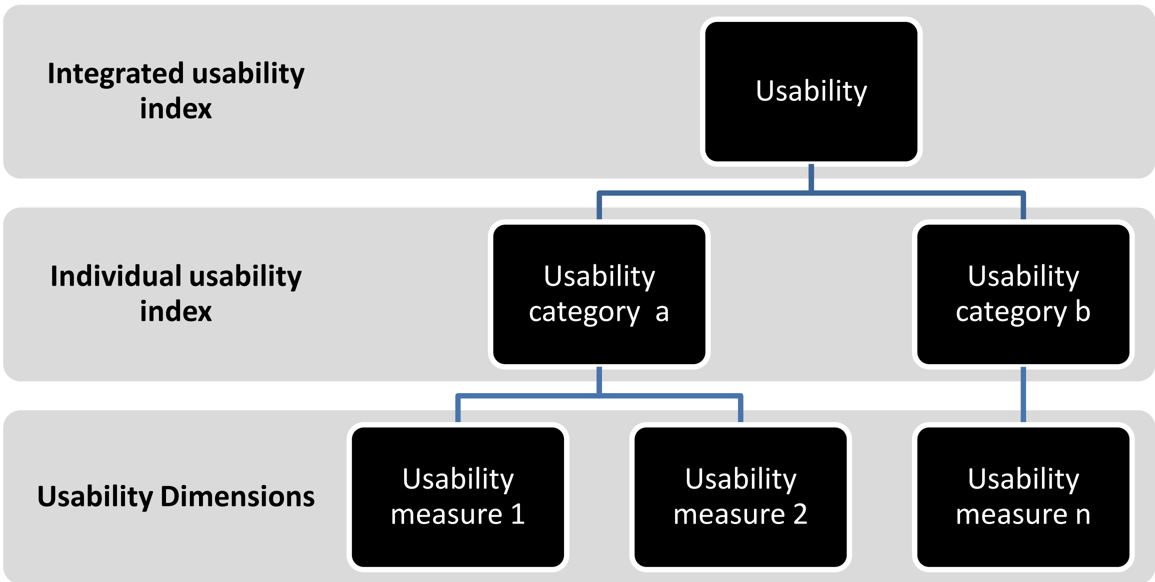


Figure 3: Quantitative model of usability. Adapted from Kim, J., Han, S.H. (2008). A methodology for developing a usability index of consumer electronic products

4.2.2. The methodological approach

The adopted methodology is based on four steps (Figure 4):

1. Classification of Usability dimensions

The evaluation of consumer electronic products starts from the analysis of relationships between product, user and task in a potential context of use, which allows the appropriate measures for the assessment of product usability. More specifically, the authors conducted an in- depth survey, reviewing literature.

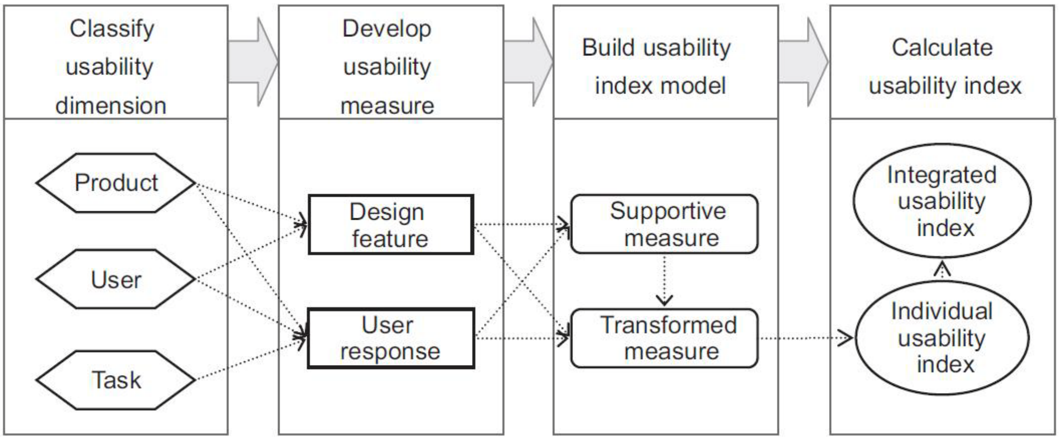


Figure 4: Procedure to calculate the Usability Index. Source: Kim, J., Han, S.H. (2008). A methodology for developing a usability index of consumer electronic products

Finally 50 usability dimensions were collected and then reduced to 18 (Tab. 1).

Tab. 1: Usability dimensions for electronic producs. Source: Kim, J., Han, S.H. (2008). A methodology for developing a usability index of consumer electronic products

Usability Dimensions	**Explanation**
Simplicity	The user interfaces and interaction methods of a product should be simple, plain, and intuitively recognizable
Consistency	The user interfaces and the interaction methods should be consistent within a product and between the same product family
Modelessness	Each user interface and interaction method should have only one designated meaning and behavior
Locus of control	Authority to control all the functions and the appearance of user interfaces should be given to a user
Directness	Any operations should be designed to give a user the feeling

	of direct manipulation
Feedback	The status of a product and the consequences of any user operations should be immediately and clearly provided
Helpfulness	Any helpful information that a user may refer to should be provided whenever a user needs
Forgiveness	When an error is recognized, ability for a user to take corrective actions should be given to a user
Error prevention	The user interfaces and the interaction methods should be designed to prevent a user from making any mistakes or errors
Adaptability	Modification of user interfaces should fit different users and conditions according to users' experience, knowledge and preference
Accessibility	Any functions and user interfaces should be easily accessible when a user wants
Learnability	Efforts required to learn the user interfaces and the interaction methods should be small
Memorability	The user interfaces and the interaction methods should be easy to recall
Familiarity	Familiar user interfaces and the interaction methods should be adopted to make users apply their previous experience
Predictability	The interaction method and the meanings of user interfaces should conform with user's expectations
Informativeness	User interfaces presented to user should be easy and clear to understand
Effectiveness	Every function users want should be implemented in a product
Efficiency	A product should be designed to allow a user to perform functions in a quick, easy, and economical way

Then, all the usability dimensions were classified, in three groups (Figure 5):

- Product based dimensions, related to product features that could be assessed in the early stage of product design;
- Product- user based dimensions, affected by user's control the product, affordance of the interface across different user's profiles.
- Product- user- task based dimensions, influenced by the context. In order to collect these metrics, it is crucial to take into account cognitive aspects of user-product interaction.

2. *Development of usability measures*

Starting from product analysis and literature review, all usability dimension were translated in usability measures that could be directly collected in the experimental phase.

3. *Usability index definition*

The usability measures ( ), collected during experiments were normalized in order to allows the comparison between them. The outcome of normalization procedure are the transformed measures ( ), ranged from 0 to 1. Then, for each subgroup of usability measures, the individual usability index (*IUI*) is defined as in (1).

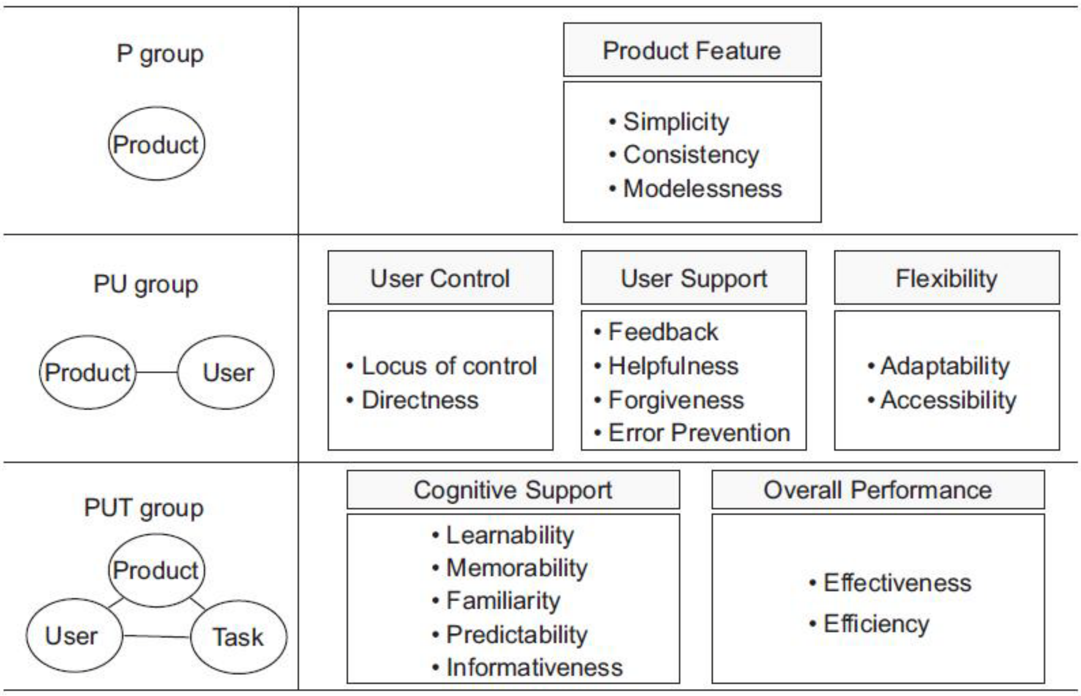


Figure 5: Classification of usability Dimensions. Source: Kim, J., Han, S.H. (2008). A methodology for developing a usability index of consumer electronic products

(1) $$IUI = \sum_{i=1}^{n} w_i \times tm_i$$

Being the weights of each transformed measure, that could be different, based on the level of priority of usability measures in the specific application. Finally, the Integrated Usability Index (UI), calculated across all the usability measures, is the linear combination of all the Individual Usability Index (2):

(2) $$UI = \sum_{i=1}^{n} d_i \times IUI_i$$

Being the weights of each Individual Usability Index, that could be different, based on the specific application. The application of a real case study revealed a high correlation between the index and the subjective score. Thus the proposed model seems to be appropriate to estimate user preference.

4.1.1. Main issues in Integrate Usability model and applications

The proposed approach is a very simple index for the assessment of usability based on a hierarchic model. Being the model defined for consumer electronic products, it could be difficult to use the Integrate usability index for a generic case study, without substantially changing the nature of the metrics. Then, although the conceptual definition of the weights was provided, the authors assumed that the same relevance for all the usability measures an all the Individual usability indices. In order to obtain a more effective assessment of usability with the index, several criteria for the calculation of the weights should be highlighted.

Bibliography

[1] Sauro, J., Kindlund, E.: A method to standardize usability metrics into a single score. In: CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 401{409. ACM, New York, NY, USA (2005). DOI http://doi.acm.org/10.1145/1054972.1055028

[2] ISO/IEC standard: Ergonomic requirements for office work with visual display terminals- guidance on usability. ISO 9241-11:1998 (1998)

[3] Jolliffe, I. T.(2002) Principal Component Analysis, Secaucus, NJ, USA: Springer-Verlag, ISBN 0- 387- 95442- 2.

[4] Sauro, J., Kindlund, E. How Long Should a Task Take? Identifing Specification limits for Task Times in Usability Tests, Retrieved September 13, 2004, from Measuring Usability Web site http://measuringusability.com/time_specs.htm

[5] Sauro, J & Kindlund E. Making Sense of Usability Metrics: Usability and Six Sigma, in Proceedings of the 14th Annual Conference of the Usability Professionals Association, Montreal, Canada

[6] Kim, J., Han, S.: A methodology for developing a usability index of consumer electronic products. International Journal of Industrial Ergonomics 38(3-4), 333{345 (2008). DOI 10.1016/j.ergon.2007.12.002. URL http://dx.doi.org/10.1016/j.ergon.2007.12.002

Chapter V

A new approach for usability assessment

The literature review conducted in previous chapter highlighted several important topic in usability assessment. Starting from these results, the aim of these thesis was the development of participatory design methodologies by using statistical techniques in order to support designers in product development. More specifically a new approach for usability assessment is proposed (*Figure* ***1***), with two main objectives.

– The design of participatory experiments to collect objective and subjective data related to user- product interaction. More specifically, the proposed experimental protocol is related to experiments in virtual reality (VR). Indeed, the use of VR can be a valuable tool for usability assessment in the early stages of product design. A proper experimental setup may in fact allow a significant reduction in time and costs of product development.
– The development of a model for usability assessment. Designers are often not able to interpret and use data from a usability test. In order to help them in designing better products, taking into account the most important aspects of user- product interaction, a single index was defined. This index is a summated metric that synthesizes performance data and satisfaction scores.

Being this study in the exploratory phase, also a first validation of the model is proposed. The approach is tested on a real case study involves the design of an integrated system aimed at assisting disabled people (a powered wheelchair equipped with a robotic arm), in which the usability of two control devices has been evaluated. For this purpose, an user-centered approach, which involves expert users early in the design process, has turned out essential.

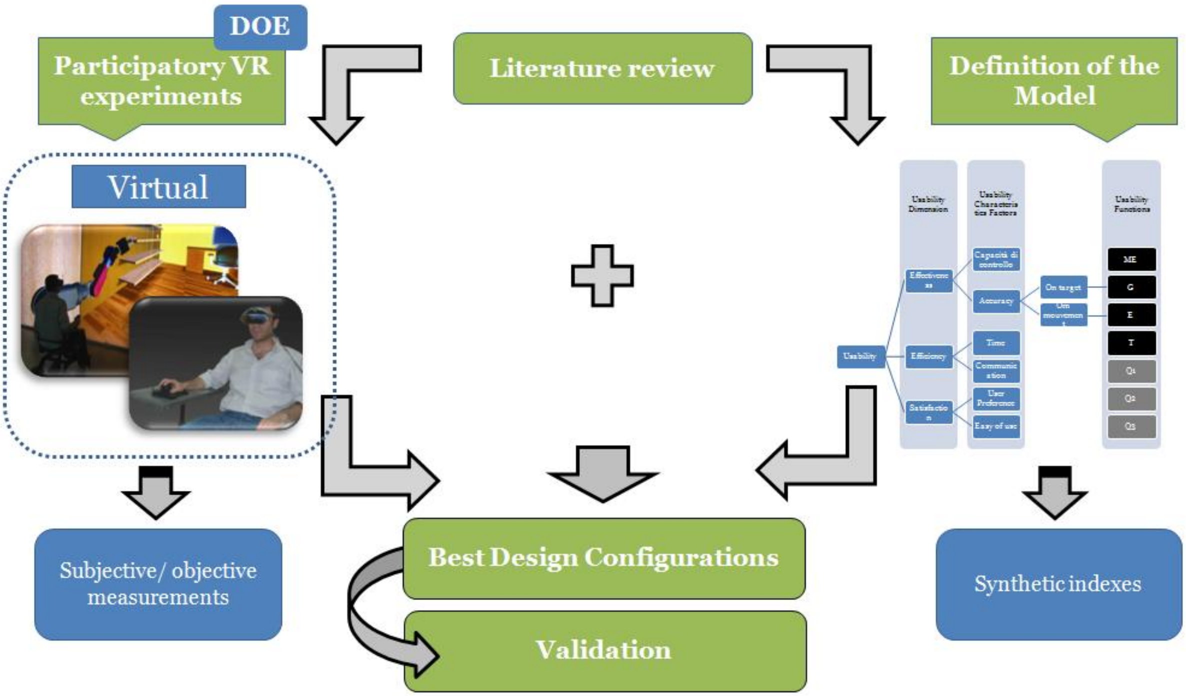


Figure 1: A new approach for usability assessment

5.1 The case study: a wheelchair mounted manipulator

The case study concerns a powered wheelchair equipped with a robotic arm. Starting from two existing products (the powered wheelchair Indoor 2003 by Neatech and the robotic arm KUKA Light Weight Robot) (*Figure* **2**), the virtual model of the integrated system (*Figure* **3**) has been conceived [1]. This is an innovative concept, designed in order to guarantee the maximum usability for disable users in deambulation and handling objects. The robotic arm can move around the wheelchair by sliding along a rail. The so conceived concept allows rotations around an horizontal axis and changes of inclination, widely increasing the robot workspace. Such characteristic strongly improve the interaction by adapting the workspace to user's needs.

Figure 2: The real powered wheelchair and the robotic arm KUKA

It is important to notice that, currently, a real prototype of the product does not exist. All tests, therefore, have only been performed on the virtual prototype of the integrated system.

Figure 3: The wheelchair mounted manipulator

More precisely, this study presents a methodology for assessing the usability of two control devices for such a product. As mentioned, not only the functional requirements needs to be considered, but also the subjective needs of the target user, which are not necessarily obvious. These have been approached with Virtual Reality (VR) technologies. The use of Virtual Reality as a tool to collect experimental data provided significant benefits in terms of performance and repeatability of the tests, ensuring controlled experimental conditions. A Virtual Environment (VE) also shields the user from any risk potentially related to the physical interaction with actual robot prototypes.

5.2 The methodological approach

Hence, the usability evaluation must go together with the analysis of both objective and subjective aspects, that are closer to the emotional sphere of the individual. In this sense, the involvement of the user into the design process is crucial (User Centred Design). The proposed methodology can be summarized in five steps (Figure 4), briefly described following [2]:

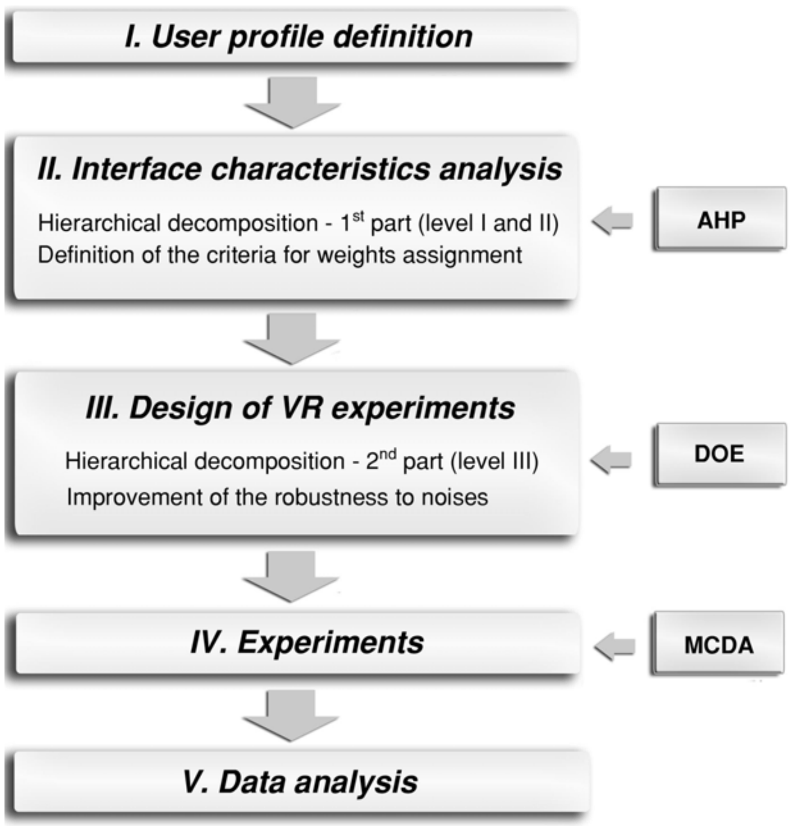


Figure 4: Main steps of the proposed methodology

I. *User profile definition*

During the first phase, both the main characteristics of the user and the potential issues coming from the interaction with the product and its context of use are properly identified. For our case study, the product at issue is a powered wheelchair equipped with a robotic arm, while our standard user is a person suffering total disability of the lower limbs. The analysis of the user profile has highlighted following basic requirements for the control devices

- robot control: the robot arm has been intended to support the standard user in interacting with object allowing them to perform simple daily activities (grasping, handling etc.);
- wheelchair control: the device should allow the normal deambulation, minimizing the efforts of the user;
- cognitive load: It is well understood that an intuitive interface and ergonomic controls greatly facilitate the user in controlling the device. Moreover, since the user-product interface should be consistent with the impairments of the user that imply the inability to perform complex movements, the interface must meet the user needs with no cognitive overload.

II. *Interface characteristics analysis*

During the second phase, the global Usability is broken down into two levels according to the Saaty's Analytic Hierarchy Process (AHP). The first level is made of Usability Dimensions, in compliance with ISO reference standard 9241-11:1998 . The second level contains the so-called Usability Characteristics Factors. The mutual importance of the elements inside each level is scored with proper weights.

III. *Design of VR experiments*

A proper task is defined, according to the requirements coming from the User profile definition. This task allows the first goal of this step, translating the Usability Characteristic Factors into measurable functions (Usability Functions), that are the last level of the hierarchy. The second goal consists in reducing the noise related to the skill of the user in approaching the virtual reality technologies in general.

More specifically, this implies:

- a preliminary selective questionnaire;
- a user training phase;
- some preliminary simulations.

IV. *Experiments*

The fourth step concerns the processing of the experimental data. Basically, a multi-criteria analysis allows combining the values of the individual usability functions into a single index of usability.

V. *Data analysis and conclusions*

- Definition of the last level of the hierarchy
- Enhancing the robustness to VR-related noise

Finally, the last step of the proposed methodology, is the data analysis (see section 5). During that phase, initially, the weights for each level of the hierarchy are defined with a bottom-up approach. Then, the mean effects of each control factor related both to the global usability index and to the usability functions (defined at the lowest level of the hierarchy) are investigated through descriptive statistics, following a DOE approach.

5.2.1 The model

With respect to the case study, the user-product interface actually is the control system of both the robotic manipulator and the powered wheelchair. As aforementioned, for purely research purposes, the authors have chosen to compare two typical control devices: the space-mouse and the joystick (Figure 5). The comparison of the latter devices in terms of usability has been approached with the Saaty's Analitic Herarchy Process (AHP) [3]. The first step of this methodology implies the decomposition of the problem into several levels and factors.

Figure 5: The tested input devices: the space-mouse and the joystick

The first decomposition has been made, according to ISO 9241-11:1998 standard [4] in usability dimensions (UD) (crf. Chapter III). Starting from literature review and the analysis of the case study, a further level of the hierarchical model has been defined, translating usability dimensions in “Usability Characteristics Factors” (UCF). The aim at this stage, was to consider critical aspects in the usability assessment of the devices, object of study, without neglecting the main design characteristics, already defined in the analysis of the product’s interface (robot control, wheelchair control, cognitive load). Finally the six UCF reported in

*Figure **6*** have been identified.

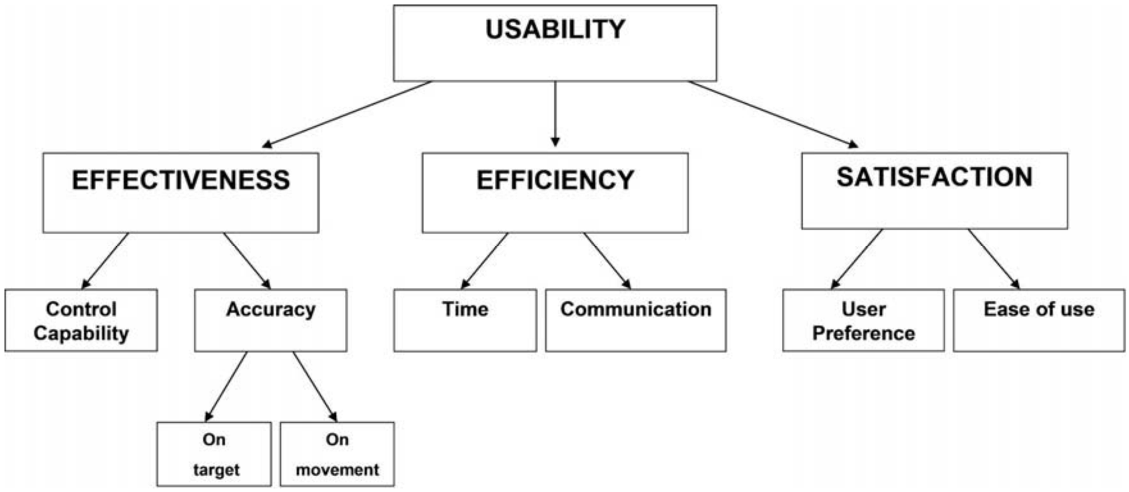


Figure 6: Usability hierarchical decomposition (level I and II)

The last step in the definition of the model, has been the definition of the “Usability Function”(UF) that is strictly related to the definition of the experimental task. In order to correctly assess the usability functions, several performance indicators were measured through a proper VR simulation. The goal of this final test (simulation III) consists in moving a virtual ball between two defined positions along a straight path. That task is accomplished when the manipulator's end-effector reaches the desired position (Figure 7).

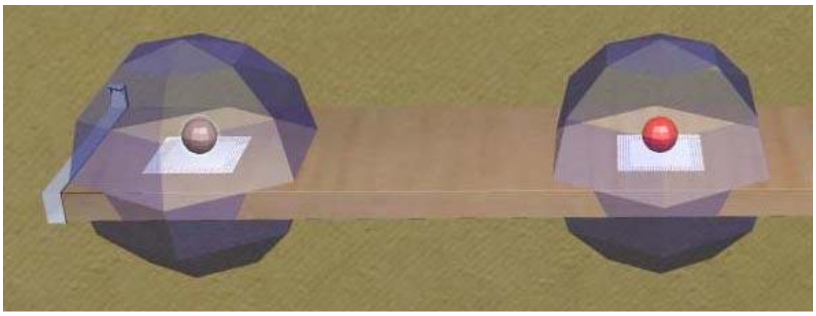


Figure 7: The task: moving a virtual ball between two defined positions along a straight path

According to the hierarchical decomposition above described, the final test provides the following usability functions (UFs):

- Movement Error (ME) (measure of control capability) is the deviation of the real path from the reference one (Figure 8);
- Number of Goals (G) (measure of accuracy on target) is the number of times the user reaches the goal;
- Number of Errors (E) (measure of accuracy on movement) is the number of penalties that the user scores during a single performance, when going beyond the error plans that limit the test area (Figure 9);
- Time (T) (measure of efficiency) is the time needed to accomplish the test;
- Communication effort (Q_1) (measure of efficiency) is a score assigned by the administrator after the test a 5-points scale. It measures the effort made by the user to clarify all his doubts about the functionality of the control devices;
- User Preference (Q_2) (measure of satisfaction) is a score which expresses the preferences of users about the control devices;

  More specifically, it is the average of two different scores:

  a) "Difficulty in use" score, prvided by the user by using a 3-point scale;
  b) "Behaviour score", assigned by the administrator of the test, by using once again, a 3-point scale

- Q3 (ease of use) is a score assigned by the user through a questionnaire (*Tab.* ***1***), according to literature (crf. Par. 3.4.1.1). It defines the ease of use of the system, intended as the ease perceived by the tester about the response of the integrated system compared to the initial training phase (measure of satisfaction).

Tab. 1: Ease of use questionnaire

This task was easy to complete						
	1	**2**	**3**	**4**	**5**	
strongly disagree						***strongly agree***

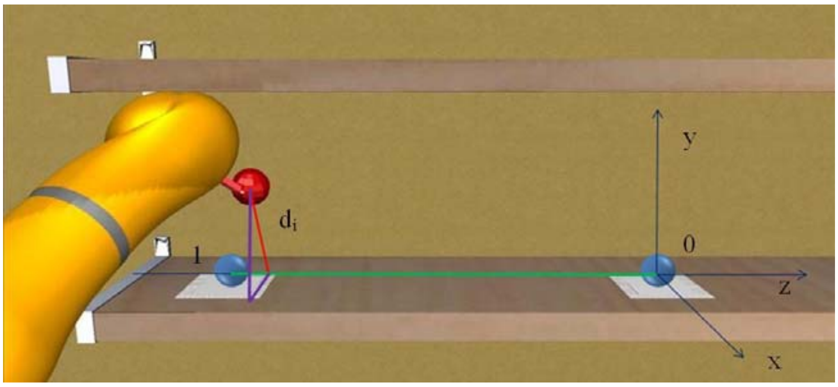


Figure 8: Reference path for Movement Error measurement.

These Usability Functions (UF) define the lowest level of the hierarchical model (Figure 10).

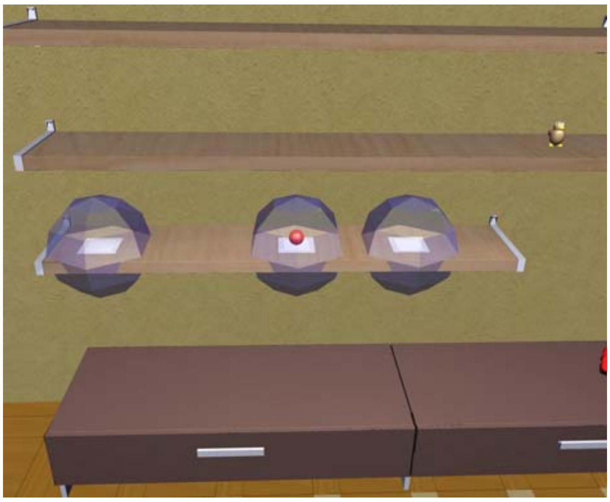


Figure 9: Reference path for # Error measurement.

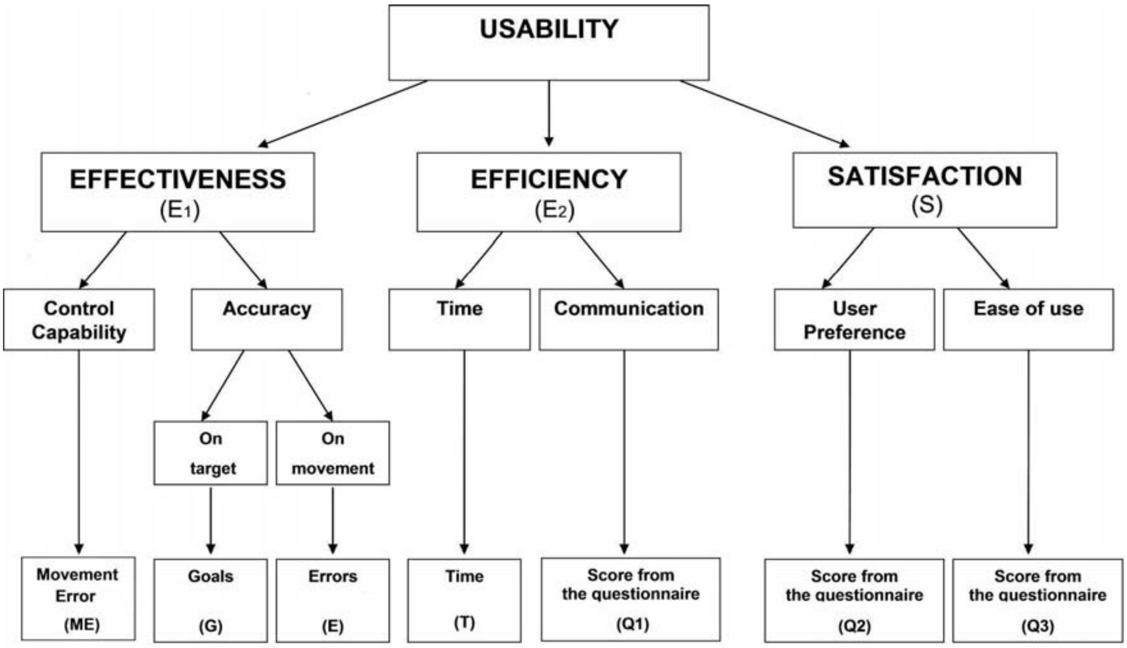


Figure 10: Usability hierarchical decomposition (level III)

5.2.2 The Usability index

Starting from the assumption that all the factors of the hierarchy, for each level are preferentially independent each other, then a simple linear additive evaluation model could be applied to combine all the measures corresponding to the factors of the model into one overall value by means of Multi-Criteria Decision Analysis (MCDA). This is done by multiplying the measure of each factor by a weight based on a specific criterion, and then adding all those weighted scores together. The calculation of the index starts from the usability functions (UF), by using data collected during experiments. Being data of different nature and magnitude, a preliminary normalization is required in order to allows the comparison between them. The normalization techniques adopted for the specific usability functions, are reported in the paragraph 5.2.2.1. The outcome of the normalization procedure are the usability measures (um_i), that range from 0 to 1. Then, for each subgroup of usability measures, the usability dimension index (*UDI*) are defined as in the (1).

(1) $$UDI_i = \sum_{i=1}^{n} w_i \times um_i$$

Being w_i the weights of each usability measure, that could be different, based on the level of priority of usability measures in the specific application. The three usability dimension index are in specific:

- the Index of effectiveness
- the Index of efficiency
- the Index of satisfaction

Finally the weighted sum of these three index provides the overall results for the usability index (2):

(2) $$UI = \sum_{i=1}^{n} w_i \times UDI_i$$

5.2.2.1 Normalization techniques

To avoid the effect of different measurement units, the data should be normalized, that is to transform them within a smaller and common range (usually [-1; 1] or [0,1]).

Following, the adopted normalization techniques are briefly described:

- Min- Max normalization performs a linear transformation of the original data. The considered value e_{ij}, is transformed in a new value e_{ij}^{I} ranged in the interval [0,1] using the formula (3):

(3) $$e_{ij}^{I} = \frac{e_{ij} - min_i}{max_i - min_i}$$

where min_i and max_i are the extremes values in the *i* dimension (column dimension).

- 0- Max normalization performs, once again, a linear transformation of the original data (4). This is a particular case of the min- max standardization, that occurs when the lower limit of the interval of original values is equal to 0 (min_i = 0).

(4) $$e^I_{ij} = \frac{e_{ij}}{max_i}$$

The normalization techniques adopted for each usability function are reported in *Tab.* ***2***.

Tab. 2: Normalization techniques adopted for each usability function

Normalization technique	**Usability Function**
Min- Max	ME, E, T
0-Max	G, Q_1, Q_2, Q_3

5.2.2.2 The weight's assignement: the AHP

The second phase of Saaty's methodology deal with the scoring of all the factors of the hierarchy [3]. The AHP is applied in order to evaluate the relevance of the factors in the hierarchy, taking into account the analysis of user-product interaction. Starting from the hierarchical structure of the model, all the weights are assigned. All the elements of the same cluster are compared in pairs by adopting he Saaty's scale, that is a 9-points scale anchored at the end with the terms "Equivalent alternatives" and "The chosen alternative is absolutely better than the other one"(*Tab.* ***3***)

Tab. 3: The Saaty's questionnaire

1	3	5	7	9
Equivalent importance	Weak importance	Essential importance	Demonstrated importance	Absolute importance

For each cluster, a total of $n(n-1)/2$ pair-wise comparison are evaluated, where n is the number of factors of the hierarchy for each cluster. Let A denote the generic matrix of the pair- wise comparison (5):

(5)
$$\begin{bmatrix} 1 & a_{12} & \dots & a_{1n} \\ a_{21} & 1 & \dots & a_{2n} \\ \dots & \dots & \dots & \\ a_{n1} & \dots & \dots & 1 \end{bmatrix}$$

The generic matrix element aij is the result of the pair- wise comparison between the attribute of the row i and the column j, with respect to a certain task, using the Saaty's scale .Thus, the main diagonal of the matrix consists of unit elements only (self-compared attributes), while the values of other cells are always positive, according to the reciprocity property (6):

(6)
$$a_{ij} = \frac{1}{a_{ji}}$$

Once the pairs comparison matrix has been defined, the weight of each element is assumed as (7):

(7)
$$w_i = \frac{\left(\prod_{j=1}^{n} a_{ij}\right)^{1/n}}{\sum_{i=1}^{n}\left(\prod_{j=1}^{n} a_{ij}\right)^{1/n}}, \quad i, j = 1, 2, \dots, n$$

Where n is the dimension of the metrics related to the element at issue. In particular, the weights are allocated with a bottom-up logic, starting from the lowest level of the hierarchy (Usability Functions) and ending with the highest one (Usability).

5.3 The experiments

The familiarity of the users with VR technologies and interfaces can be intended as a noise factor for the experiments because the potential user population may include individuals with different skill levels. In order to limit their impact on the final results, these factors have been considered by means of several steps:

- a selective preliminary questionnaire has been administrated in ordet to select only users with a minimum level of experience. More specifically, the familiarity of the users with very common computer gaming interfaces and control devices has been considered a relevant factor for the skill level assessment.
- a preparatory phase in which the users are introduced to the tests;
- the administration of two preliminary VR simulations to train the user.

The two preliminary simulations are:

- Simulation I This simulation is intended to train the user on the navigation through the virtual environment. In this simulation the interactions with objects are not allowed; the user can only control the powered wheelchair moving it through the virtual flat.
- Simulation II The user can move through the virtual environment, interacting with objects. During this simulation, the user has to accomplish a specific task: to move a book between two shelves of a library, from a lower shelf to a higher one. The task is achieved when the book collides with a predefined control volume (*Figure* ***11***).

The test has three replications. The administrator collects the individual execution time. It is worth noticing that Simulation II is intentionally more difficult than the final test (simulation III) that is actually used for the data collection.

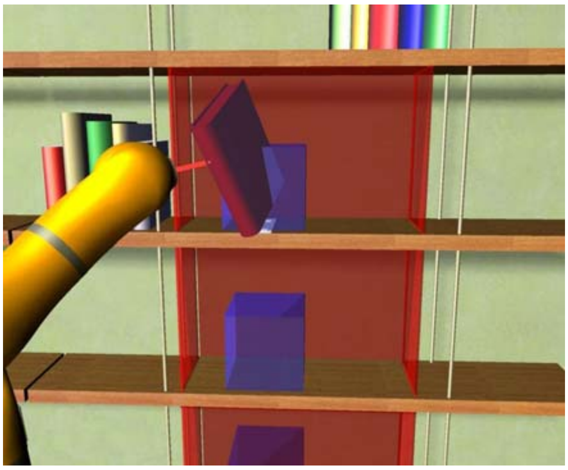


Figure 11: Simulation II: moving a book between two shelves of a library

Indeed, Simulation II is intended not only to train the user with the control interfaces, but also to assess its familiarity level with VR technologies. This familiarity level is assumed inversely proportional to the average execution time measured during the three administrations of the test. Those who have completed the test in an average time of less than 60 seconds, have been considered “expert users" that are particularly skilled in the use of VR technologies. Finally the simulation III. was accomplished.

5.3.1 The experimental setup

The interface characteristics analysis and the definition of the user profile have suggested that the most important design features are:

- the ability to control the robot (R);
- the ability to control the powered wheelchair (K);
- the logical and cognitive load of the user (C).

These design features have been used as factors of the Design Of Experiments (DOE). On the other hand, the level of expertise (L) in using the two input devices (Skill level) has been assumed as a noise factor with two levels (*Tab.* ***4***). All the latter factors have been summarized in the cross array shown in *Tab.* ***5***. Because of the high complexity of the test, starting from the above identified control factors, a fractionated factorial design,

2^{3-1}, has been developed as inner array and a two-skill-level design has been adopted as outer array.

Tab. 4: Control Factors and Noise Factors

Control Factors		**0**	**1**
Robot Control	R	Spacemouse	Joystick
Wheelchair Control	K	Joystick	Spacemouse
Cognitive Load	C	one hand	two hands
Noise Factor		**0**	**1**
Skill Level	L	low	high

Tab. 5 The cross array planned for the experimental phase

				0	1	**L**
	R	**K**	**C**			**Mean**
TEST I	0	0	1	$X_{I,0}$	$X_{I,1}$	M_I
TEST II	0	1	0	$X_{II,0}$	$X_{II,1}$	M_{II}
TEST III	1	0	0	$X_{III,0}$	$X_{III,1}$	M_{III}
TEST IV	1	1	1	$X_{IV,0}$	$X_{IV,1}$	M_{IV}

5.3.2 Experimental protocol

VR experiments have been conducted according to the following experimental protocol:

1. administration of a questionnaire for the selective collection of information on the cultural background of the user and its familiarity with the control interfaces. The questionnaire is a selective tool to recruit testers with appropriate skills;
2. briefing to explain the contents of the tests;
3. user training with simulation I and II;
4. viewing a video tutorial about the final test;
5. administration of the final test (Simulation III).

The final test has been administrated three times to each user. During the test execution, performance measures (ME, G, E, T) have been collected. The administrator has also recorded his impression about the user to determine his communication effort while interacting with the virtual environment (Q1). Moreover, after the simulation, a questionnaire has been administrated to each user, in order to assess the preference (Q2) and Ease of use (Q3) factors. Finally, a further VRSART questionnaire for assessing the sense of presence has been administrated.

5.4 Results

5.4.1 Weights assessment

The weights for all levels of the hierarchic model of usability have been assigned by using a pair- wise comparison method. In particular, a questionnaire has been administrated to seven experienced designers, who already knew the case study and its main characteristics. The assignment of the weights has then followed a bottom-up approach. More precisely, starting from the lowest level of the hierarchy, all weights were calculated by comparing in pairs all usability function stemming from the same usability characteristic (*Figure 10: Usability hierarchical decomposition (level III)Figure **12***). For instance, with reference to the accuracy (i.e. an usability characteristic factor), the two usability functions number of goals and number of errors have been defined. Depending on the task and the above mentioned interface requirements, each expert assigned a preference score to the best between the two usability functions of each pair, by using Saaty's scale (crf. par. 5.2.2.2).Once the respondents selected the best usability function, they answer to the statements "Taking into account product functionalities and the application field, how much the selected usability function is better than the other one?". Finally, the weights were obtained from (7). Tab. 6 shows an example of weights calculation based on the scores assigned by the expert 1.

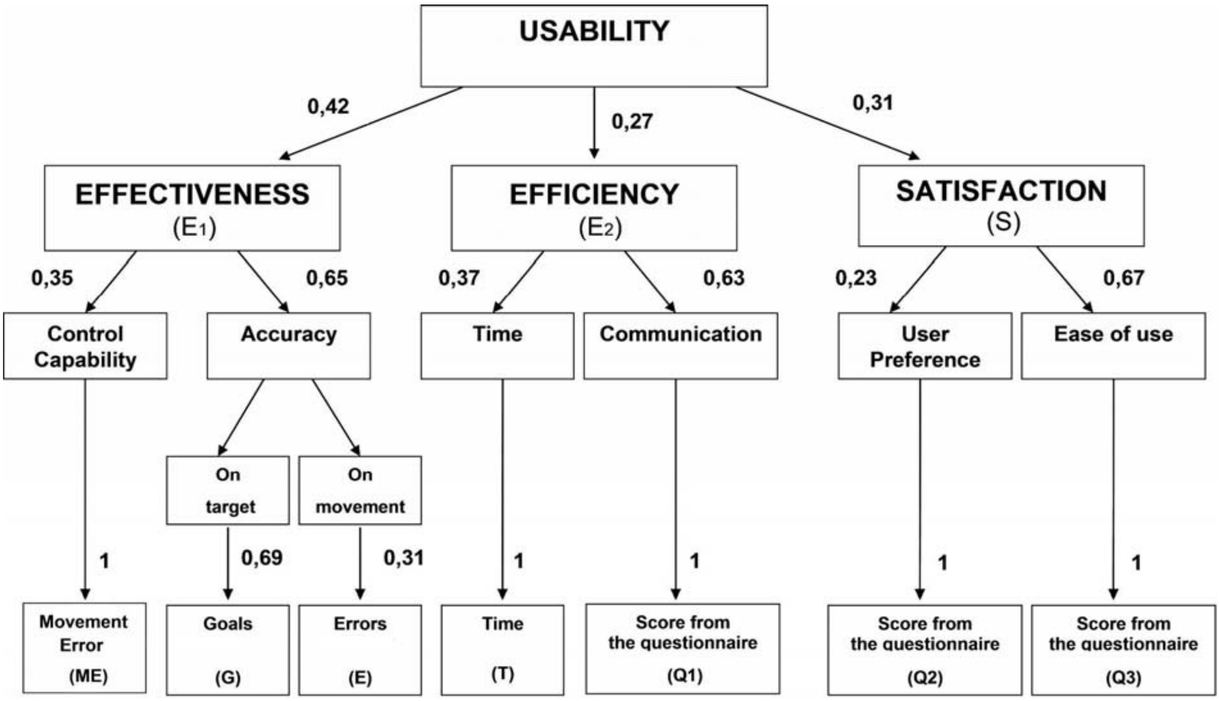


Figure 12: The weighted model of usability

Tab. ***7*** summarizes the weights obtained for G and E functions, that depend on the scores assigned by the whole panel of experts. Further, moving to the second level of the hierarchy, all the weights for the Usability Characteristic Factors (UCF) have been found following a bottom- up approach. Finally, the vector p of the weights for the Usability Dimensions (UD) respect to the global Usability (U) has been defined as (8):

(8) $$p = [0.42, 0.27, 0.31]$$

Tab. 6: Example of weights calculation

	Target (G)	**Movement (E)**
Target (G)	1.0	0.5
Movement (E)	0.25	1.0

$\left(\prod_{j=1}^{n} a_{ij}\right)^{1/n}$	0.25	0.5
w_i		

Tab. 7: Weight assigned for the usability functions G and E

	Ex_1	Ex_2	Ex_3	Ex_4	Ex_5	Ex_6	Ex_7	Mean
***w_i* (Target)**	0.80	0.33	0.75	0.75	0.88	0.80	0.72	0.80
***w_i* (Movement)**	0.20	0.67	0.25	0.25	0.12	0.20	0.28	0.20

5.4.2 The usability Index

Starting from the Usability Functions, through the combination of AHP and MCDA, the Usability Dimensions Indices (UDI_i) have been assessed for each experiment (*Tab.* ***8***).

Tab. 8: Usability Dimensions Indices (UDIi)

	E_1	E_2	S
Test I	0.45	0.75	0.71
Test II	0.80	1	1
Test III	0.57	0.50	0.79
Test IV	0.75	0.95	0.78

At this level of the hierarchy, Test II achieves the best results in terms of effectiveness, efficiency and satisfaction. For this, regardless of the choice for the weights vector at the last level, the best simulation in terms of usability will always be the second one. In fact, the weighted sum of the Usability Dimensions Indices provides the overall results for the usability index as reported in the equation (2).

All the results for the four test are summarized in *Tab.* ***9***:

Tab. 9: Usability Index for all the tests

	UI
Test I	0,609
Test II	0,915
Test III	0,621
Test IV	0,812

As expected, the best solution belongs to Test II, that obtains an usability value of 0,915, which is related to (R0;K1;C0) configuration, that is:

- Space-mouse for the robot control;
- Space-mouse for the wheelchair control
- Only one hand to handle the interface (minor cognitive load)

This allows to evaluate even the remaining tests that were not able to be ranked because of the different values of E1, E2and S at the previous level. Nevertheless, Test IV obtains a good UI value, while UI values coming from Test I and Test III are quite far from the best one.

5.4.3 Preliminary analysis

The purpose of these preliminary analysis has been to evaluate the differences between novice and experts users, based on measures of effectiveness efficiency and satisfaction, in a task's execution. The experimental protocol, with the two preliminary training stages, has been designed for the elimination of the gap between users with different confidence level with VR technologies ("expertise effect"). Indeed, the level of expertise could influence both, performance measures and subjective perceptions, in terms of satisfaction. The descriptive statistics seems to show a better performance of expert users in task execution in terms of ME (*Figure* ***13***), while there is no evidence of differences between novice and expert users for others objective measures. The analysis of subjective measures, also highlights an higher communication effort (*Figure* ***14***) for novice users, in task execution, while it’s not possible to provide information on the two analysed samples, in terms of satisfaction (ease of use and user preference).

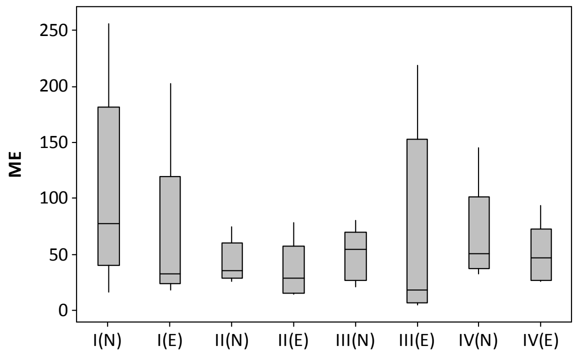


Figure 13: Comparison between Novice and expert- users in term of ME using histogram chart and box-plot chart

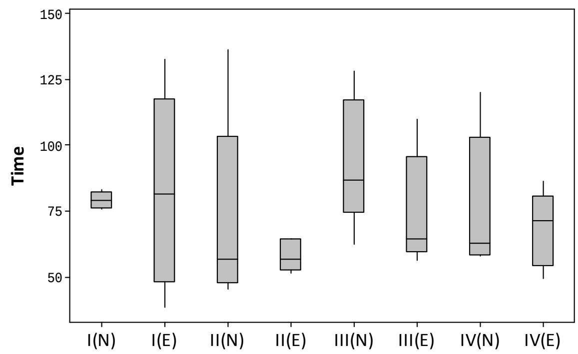


Figure 14: Comparison between Novice and expert- users in term of Q1 using histogram chart and box-plot chart

Finally, the significance of the difference between the two samples has been assessed for each usability measure, by using the nonparametric Mann-Withney test (α=0.05)[5], that verified the null hypothesis of equality of medians for the considered samples. (novices vs experts). Finally, for the four tests it’s impossible to reject the null hypothesis of medians equality. All the obtained results, for the preliminary analysis, are reported in the appendix A of this chapter.

Starting from obtained results, the difference between expert users and novice users seems to be not significant for all the tests. This result partly contrasts with the literature [6]. On the other hand, the result confirms the original intent of the proposed approach, aimed at the reduction of the noise induced from the level of expertise, through the two preliminary sessions of training (crf par 5.2.).

Moreover the obtained results could be considered coherent with the learning curves of Nielsen [7]. Indeed the increasing of the confidence level of the users over time in the two preliminary experimental sessions of training, makes comparable proficiency and efficiency of expert and novice users in product use (*Figure* ***15***). This result should be deepened with further experiments involving a larger sample of users.

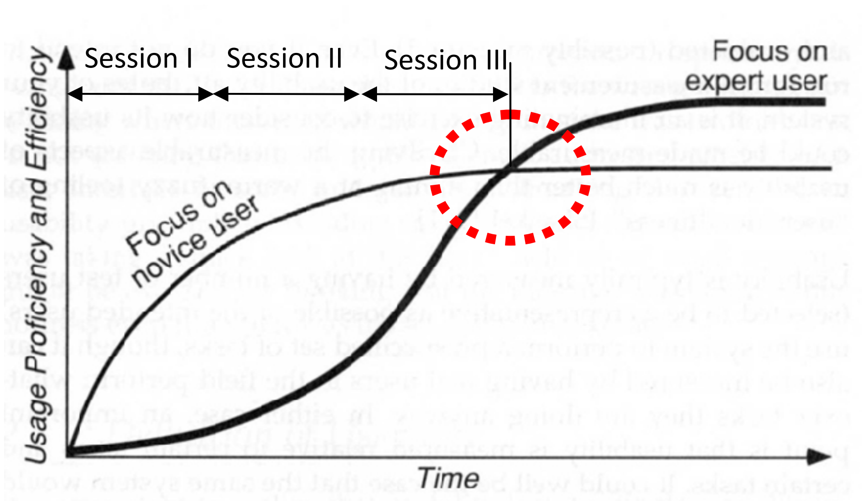


Figure 15: Learning curves for a hypothetical system. Adapted from Nielsen "Usability Engineering"(1993)

5.5 DoE analysis

Ten users have been involved in the experiments, five for each of the two skill levels. The seven usability function have been considered as response functions. Using data collected in the experimental phase, for each of them, the main effects analysis has been performed in order to define the impact of control factors on the individual responses. Thus, the expected optimal combination has been defined and then it has been compared with the experimental one. For instance, in the case of ME function, the charts in *Figure **16*** show a strong effect of K and C compared to that of R.

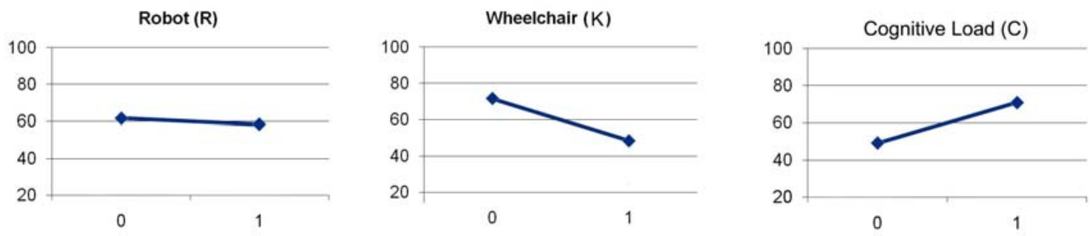


Figure 16: Main effects of the control factors on the Movement Error (ME).

In this case, the experimental optimal combination (R0, K1, C0) is different from the expected one (R1, K1, C0) (*Tab. **10***). However, given the quite marginal impact of R, which has been derived with Pareto-ANOVA analysis [8], the two combinations can be considered substantially equivalent.

Tab. 10: Experimental and expected Movement Error.

		Experimental	Expected
Movement Error (ME)	**R**	0	1
	K	1	1
	C	0	0

Whenever the expected optimal combination is not coincided with the experimental one, it has been verified that the difference has been related to the factor R with no significant effects. Thus, at a first approximation, we can consider the two configurations virtually identical. However, it would be better to repeat the experiment in order to confirm expected results. Finally, the analysis of the main effects has been also conducted for each UD and for the UI, by considering them as response functions. In *Figure **17*** the plots of the effects of the control factors on the usability index are shown. They highlight a noticeable impact of K, if compared with that of R or C. This behavior is similar to "ease of use" function, that most affects the global index in terms of weight.

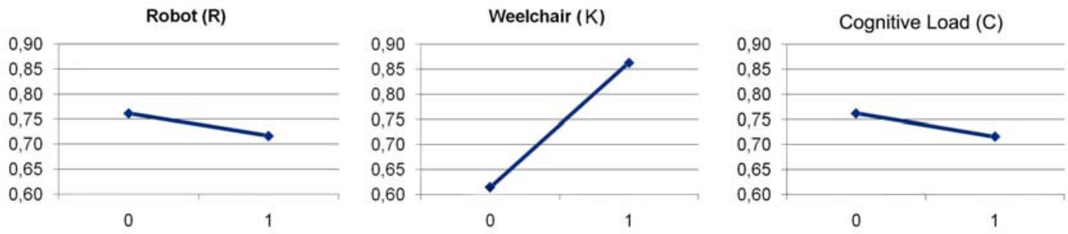


Figure 17: Main effects of the control factors on the Usability Index (UI).

5.6 Comparison with other indexes in literature

In order to compare and validate our results, in this section the proposed index is compared with the other indices already mentioned in chapter IV:

- the Summated Usability Metric (SUM) by Sauro et al.107[9][10]
- the Integrated Usability Index by Kim and Han [11]

Both of these approach tried summarize usability metrics (subjective an objective measures) in a single score.

5.6.1 Adapting the model for Sauro et al. Index evaluation

The Sauro's methodology has been applied to the lowest level of the hierarchical model of usability (usability functions). In order to evaluate the relationships between each usability function, the correlation matrix [10] has been defined for each test shown in table 2. The results confirmed a clear correlation between Communication (Q1) and Ease of use (Q3) usability functions, specifically for Test II that achieved the higher values of UI.

The moderate correlation between the subjective and the objective usability functions is consistent with Frøkjær 's work [12], that founding a weak correlation between the usability dimensions, suggested to consider the three dimensions of usability as independent aspects. Once the eigenvalues (λ_i) and the eigenvectors of the correlation matrices were obtained, the definition of the number of principal components to retain(Y_i) is allowed by the following heuristic evaluation criteria:

- Kaiser's rule: all the principal components with eigenvalues greater than 1 are retained;
- Cumulative variance: the number of principal components depends on the level of the cumulative variance (70%- 90%). When that level is reached the retaining of principal components is stopped;
- Scree plot test: the eigenvalues are plotted in descending order. Then, if the so defined plot presents a change in the sign of the slope, all the principal components corresponding to those eigenvalues that are at the bottom of the point of “slope inversion”, are retained.

The three heuristic criteria above described, have been applied to the experimental data in order to define the minimum number of principal components to retain. For instance, by considering the test I, following results have been carried out:

- Kaiser's rule: the vector of the eigenvalues λ (9), highlights two values greater than 1. Thus two principal components are retained.

(9) $$\lambda = \left[3{,}89 \;\; 1{,}88 \;\; 0{,}61 \;\; 0{,}35 \;\; 0{,}16 \;\; 0{,}10 \;\; 0{,}00 \right]$$

- Cumulative variance: With reference to a level of 90% of the cumulative variance, three principal components are retained, as reported in *Tab. **11***.

Tab. 11: Eigenanalysis of the Correlation Matrix

λ_i	3.89	1.88	0.61	0.35	0.63	0.10	0.00
Proportion	0.56	0.27	0.09	0.05	0.02	0.02	0.00
Cumulative	0.56	0.82	0.91	0.96	0.98	1.00	1.00

- Scree plot test: the diagram shows that two principal components are retained.

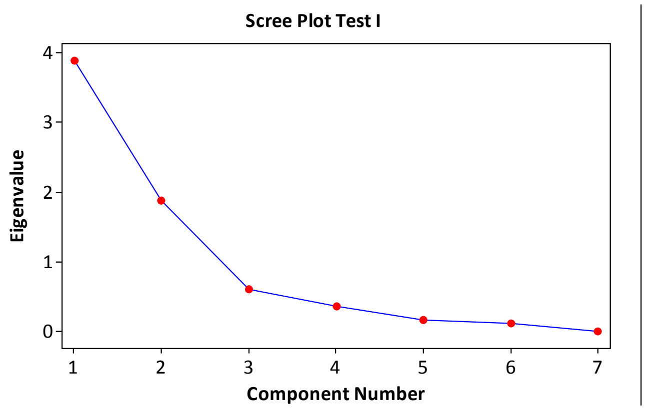


Figure 18: Scree plot diagram for the test I

Finally, for test I, the three heuristic criteria retain always more than one principal components. A similar result was achieved also for other tests, as reported in *Tab.* ***12***:

Tab. 12: Principal components to retain.

Test	Kaiser's rule	Cumulative variance	Scree plot test
I	Y_1, Y_2	Y_1, $Y_{2,}$ Y_3	Y_1, Y_2
II	Y_1, Y_2	Y_1, $Y_{2,}$	Y_1, Y_2
III	Y_1, Y_2	Y_1, $Y_{2,}$ Y_3	Y_1, Y_2
IV	Y_1	Y_1, Y_2	Y_1, $Y_{2,}$ Y_3

The results in Table 9 show that the number of input variables (usability dimensions) cannot be summarized in a single principal component (except in the test IV, but only for the heuristic criterion of the Kaiser's rule). Indeed, each test needs at least a two-dimensional information, therefore it is not possible to define a single model in which each variable adds only informations not contained in other variables. Moreover, the same results imply that the input variables must have different weights, differently than in [9].

Tab. 13:Eigenvectors of the Correlation Matrix test I

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
ME	0,40	0,04	0,73	-0,33	0,20	-0,24	-0,31
T	0,41	0,24	-0,56	0,04	0,43	-0,50	-0,16
G	0,01	-0,69	-0,24	-0,28	-0,40	-0,31	-0,36
E	0,26	-0,52	0,19	0,75	0,22	-0,05	0,11
Q1	-0,50	0,06	0,01	0,24	0,30	0,13	-0,77
Q2	0,40	0,38	0,00	0,37	-0,65	0,09	-0,36
Q3	-0,46	0,20	0,22	0,23	-0,22	-0,76	0,17

For istance, starting from the eigenvectors matrix (*Tab.* ***13***), the two principal components for the test I are:

$$Y_1 = 0.40ME + 0.41T + 0.01G + 0.26E - 0.50Q_1 + 0.40Q_2 - 0.46Q_3 \quad (10)$$

$$Y_2 = 0.04ME + 0.24T - 0.69G - 0.52E + 0.06Q_1 + 0.38Q_2 - 0.20Q_3 \quad (11)$$

- Y1strongly depends from ME on the performance side. All the usability functions related to user perception, seem to be relevant for this principal component.
- Y2strongly depends from G and E that have the highest coefficients in absolute value.ME and Q1seem to be not relevant for this principal component.

It is clear that the interpretation of more than one principal components is strongly affected by the designers' experience and sensitivity in evaluating their correlations with the input variables. Furthermore, the analysis of the principal components needs a huge number of experimental data, which go far beyond the possibilities of participatory tests. In conclusion, Sauro's model is not suitable for the present case study because it needs large sample sizes that are not commonly available in product design.

5.6.2 Adapting the model for Kim et al. Index evaluation

In order to adapt the proposed hierarchical model of usability, the following assumptions were made:

- Usability dimensions: Efficacy, efficiency, Satisfaction

- Usability measures: ME, G, E, T, Q1, Q2, Q3

Usabilility Measures were normalized and corrected, obtaining the *transformed measures*, used for the individual usability indices calculation. Normalization techniques are the same described in Section 5.2.2.1.

The values of transformed usability measures for each test are reported in *Tab.* ***14***..

Tab. 14: Trasformed usability measures

Test	Efficacy			Efficiency		Satisfaction	
	ME	G	E	T	Q1	Q2	Q3
I	0,00	1,00	0,00	0,55	0,87	0,79	0,68
II	1,00	0,63	0,80	1,00	1,00	1,00	1,00
III	0,56	0,57	0,60	0,00	0,80	0,79	0,79
IV	0,59	0,76	1,00	0,86	1,00	0,74	0,79
ω_i	0,33	0,33	0,33	0,50	0,50	0,50	0,50

The last row shows the weights ω_i related to each usability measures, which are calculated with the formula (5):

(12) $$\omega_i = \frac{1}{n_i}$$

where n_i is the number of usability measures that help to define the Individual Usability Index. Then, for each subgroup of usability measures, the individual usability index (IUI) is defined (crf. 4.2.2). Results obtained for the three individual usability index are shown in Table 8: Finally, it is possible to evaluate the integrated usability index as the weighted sum of the individual usability index reported above, for each test.

Tab. ***16***, shows that the best solution belongs to Test II, according to the results obtained using the UI.

Tab. ***17*** compares the results achieved by the proposed methods.The last row shows the correlation coefficients between usability dimension scores and usability index and satisfaction usability measures and usability index.

Results obtained shows an higher correlation for the proposed index, compared to the integated usability index of Kim et al. Although the best solution (Test II) is confirmed by both methods analyzed, it is clear that the analytical definition of the weights has a positive influence on the index, leading to higher correlation on both measures of satisfaction and usability dimensions.

Tab. 15: Individual Usability Index and Usability Dimensions weights

Test	Efficacy UI	Efficiency UI	Satisfaction UI
I	0,33	0,71	0,73
II	0,81	1,00	1,00
III	0,58	0,40	0,79
IV	0,78	0,93	0,76
d_i	0,33	0,33	0,33

Tab. 16: Integrated Usability Index

Test	**Integrated UI**
I	0,59
II	0,94
III	0,59
IV	0,82

Tab. 17: Individual Usability Index (IUIi), Integrated Usability Index (Kim et al.) and the Usability Index (UI)

Test	E1	E2	S	E1 (Kim)	E2 (Kim)	S (Kim)	Q2	Q3	UI	UI (Kim)
I	0,45	0,75	0,71	0,33	0,71	0,73	0,79	0,68	0,61	0,59
II	0,80	1,00	1,00	0,81	1,00	1,00	1,00	1,00	0,91	0,94
III	0,57	0,50	0,79	0,58	0,40	0,79	0,79	0,79	0,62	0,59
IV	0,75	0,95	0,78	0,78	0,93	0,76	0,74	0,79	0,81	0,82
Pearson Coefficient	0,95	0,87	0,83				0,63	0,85		
				0,87	0,88	0,76	0,62	0,83		

Bibliography

[1] Di Gironimo, G., Matrone G., Tarallo A., Trotta M. A virtual reality approach for usability assessment oo a wheelchair- mounted robot manipulator. In TMCE 2010 Symposium (Ancona, Italy April 12-16, 2010), Organizing Committee of TMCE 2010 Symposium, ISBN 978-90-5155-060-3, pp. 749- 762

[2] G. Di Gironimo, A. Lanzotti, G. Matrone, A. Tarallo, M. Trotta, A virtual Reality Approach for Usability Assessment: a case study on weelchair mounted robot manipulator, Journal of enginering with Computers (2011) (in press)

[3] Saaty, L.: Decision making with the analytic hierarchy process. International Journal of Services Sciences 1(1), 83 { 98 (2008). DOI http://dx.doi.10.1504/IJSSCI.2008.017590

[4] ISO/IEC standard: Ergonomic requirements for office work with visual display terminals guidance on usability. ISO 9241-11:1998 (1998)

[5] Piccolo, D. “Statistica”, Ed. Il Mulino, 1998

[6] Ziefle, M., (2002). The influence of user expertize and phone complexity on performance , ease of use and learnability of different mobile phones. Behaviour& Information Technology. 21(5), 303-311

[7] Nielsen, J., (1993), Usability Engineering, Academic Press, Boston, MA, ISBN 1-12-518406-9.

[8] Park, S.: Robust design and analysis for quality engineering. Kluwer Academic Publishers, London, UK (1996)

[9] Sauro, J., Kindlund, E. A method to standardize usability metrics into a single score. In CHI ’05: Proceedings of the SIGCHI conference on Human factors in computing systems (New York, NY, USA, 2005), ACM, pp. 401– 409.

[10] Tullis, T., Albert, B. “Measuring the user experience”, Morgan Kauffmann Publishers, ISBN 978- 0- 12- 373558-4

[11] Kim, J., Han, S.: A methodology for developing a usability index of consumer electronic products. International Journal of Industrial Ergonomics 38(3-4), 333{345 (2008). DOI 10.1016/j.ergon.2007.12.002. URL http://dx.doi.org/10.1016/j.ergon.2007.12.002

[12] Frøkjær, E., Hertzum, M., and Hornbæk, K. (2000) Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? In Proc. CHI 2000, (pp.345-352). Washington, D.C.: ACM Press.

Conclusions and future work

The proposed methodology provides a valuable tool for comparing different product design alternatives in terms of usability during the design phase. Further, it is suitable to be used with small-size groups of testers.

Moreover, a set of minimum requirements so that the user feels involved "and present" inside the VE has been defined, while possible external noises have been reduced. This is particularly relevant for semi-immersive experimental set-up that may raise some problems in terms of sense of presence, especially with respect to simulations that involve both real input and virtual outputs. Moreover, the proposed methodology takes into account the possible mix-up between product usability and VR usability. Indeed, the critical analysis of the experimental set-up has been fundamental to guarantee that the satisfaction feeling of the user was actually related to the product, rather than the experimental modalities.

The results obtained on the specific case study have been also validated through the use of Kim's methodology, that is well known in the related literature. Other approaches, like the Sauro's Single Usability Measure (SUM), need a huge number of experimental data and therefore have been considered not suitable for product design.

On the other hand, a weakness of our approach could be the assignment of the weights through the MCDA analysis. In order to ensure more reliable results, which better reject the scale of user priorities, the experts team should include not only designers as in our case, but also medical specialists. Also, it would be worth studying more deeply how the familiarity of the testers with

VR simulation tools can affect the simulation results, even if the published literature has highlighted only a partial relevance of this aspect . It is worth emphasizing that composition and sample size may have affected the final results. Specifically, the selected sample consisted of able-bodied people properly informed on experiments and their purpose. However, the authors believe that an expert-based approach, such as the one described above, would make easy the administration of the tests even to disabled people. In that sense, further confirmatory tests need to be also conducted on disabled users.

Thus, future research will focus on alternative data collection tools that can reduce the noise introduced by the subjective feelings of the users.

Finally, although this work has mainly focused on the usability assessment of an assistive technology for disabled people, the described approach can be extended to other fields too. For instance, in recent years, research in robotics is focusing on applications where the human being is free to interact with the machine by means of di_erent modes (namely, the so-called Physical Human- Robot Interaction. Thus, the developed model could be tested for the usability assessment of robotic tools designed to assist and support the human operators during their working, such as power extenders, robots for microsurgery and other manipulators for adverse environments (e.g. space or undersea manipulators, nuclear plants service robots, etc.). It is understood that the development of such tools, which involves the study of both the robots

control algorithms and their control interfaces, requires a design methodology that can consider not only the functional requirements of the product itself, but also the problems arising from its interaction with human beings.

Appendix

Appendix A

Descriptive Statistics

The purpose of these preliminary analysis has been to evaluate the differences between novice and experts users in a task's execution. More specifically several descriptive statistics have been conducted for each usability functions, taking into account the classification between usability dimension (effectiveness, efficiency and satisfaction). Following, the main results for all the usability functions, are reported, starting from several assumptions:

- all the usability functions have been assessed in four tests, with three replications;
- all the analysis are related to the average of the measures carried out in the single replications, for the single users.

1. Measures of effectiveness

1.1. Mouvement Error (ME)

Experimental results carried out for the usability function ME are reported in Tab. 1:

Tab. 1: Experimental results for the usability function ME

N	I	II	III	IV	E	I	II	III	IV
U2	128.2	26.0	14.7	32.5	**U1**	5.3	35.5	8.3	39.9
U4	38.4	26.0	35.3	50.5	**U3**	2.6	65.2	24.5	47.2
U5	23.3	55.5	28.3	58.7	**U6**	33.4	18.0	12.0	24.7
U9	48.8	153.6	205.5	145.0	**U7**	579.5	84.0	122.6	86.9
U10	30.1	31.7	49.6	42.3	**U8**	37.3	25.8	251.2	8.5
Mean	**53.7**	**58.5**	**66.6**	**65.8**	**Mean**	**131.6**	**45.7**	**83.7**	**41.4**
St.Dev	**42.7**	**54.5**	**78.6**	**45.3**	**St.Dev**	**250.8**	**27.9**	**104.8**	**29.5**
Median	**38.4**	**31.7**	**35.3**	**50.5**	**Median**	**33.4**	**35.5**	**24.5**	**39.9**

The comparison of the average values of ME highlights a better performance of expert users, except for the test III, in which the result seems to be the opposite(Figure 1). The worst result has been achieved in test I. Data present an higher variability, as shown by box-plot in, which provided also further information related to the anomalous result obtained for test III. It is clear that the median values of ME are actually lower for expert users, but the variability of the data is significantly higher for the sample of novice users.

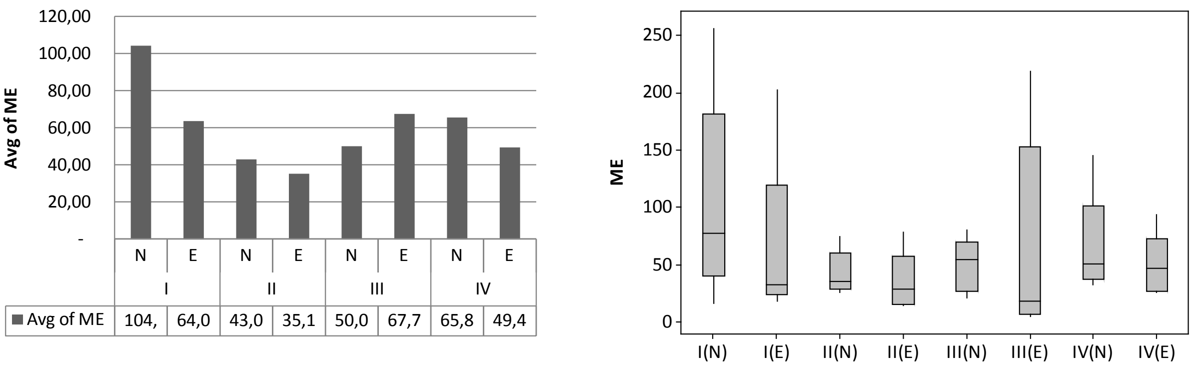


Figure 1: Comparison between Novice and expert- users in term of ME using histogram chart and box-plot chart

Indeed, checking the data, there is an abnormal performance of the user 9. Based on descriptive statistics, the performances of expert users seems to be better of them of novice users.

Tab. 2: Nonparametric Mann-Withney test for ME data, related to novice and experts performance (α= 0.05)

Test		Median	CI	W	p
I	N	77.40	(-124.8;223.2)	32.0	0.4034
	E	32.70			
II	N	35.78	(-42.67; 46.03)	31.0	0.5309
	E	28.67			
III	N	55.10	(-163.3; 62.4)	30.0	0.6761
	E	18.50			
IV	N	50.53	(-43.14; 97.86)	31.0	0.5309
	E	47.11			

The significance of the difference between the two samples has been assessed using the nonparametric Mann-Withney test that verified the null hypothesis of equality of medians for the considered samples. (novices vs experts). Finally, for the four tests it's impossible to reject the null hypothesis of medians equality, as reported in Tab. 2.

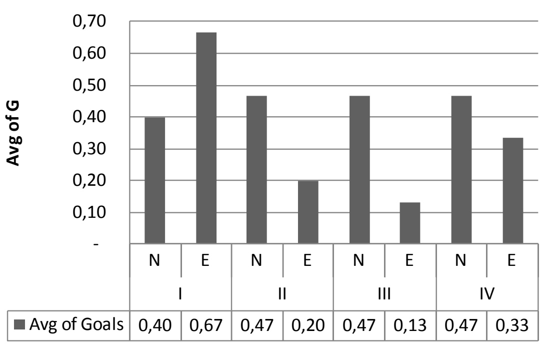
1.2. Goals number (G)

Experimental results carried out for the usability function goals number are reported in Tab. 3:

Tab. 3: Experimental results for the usability function G

N	I	II	III	IV	E	I	II	III	IV
U2	0.3	0.3	0.3	1.0	**U1**	0.7	0.0	0.3	0.0
U4	0.7	1.0	1.0	1.0	**U3**	0.7	0.0	0.3	0.0
U5	0.3	0.0	0.3	0.0	**U6**	0.7	0.3	0.0	0.3
U9	0.3	0.7	0.3	0.3	**U7**	0.7	0.7	0.0	1.0
U10	0.3	0.3	0.3	0.0	**U8**	0.7	0.0	0.0	0.3
Mean	**0.4**	**0.5**	**0.5**	**0.5**	**Mean**	**0.7**	**0.2**	**0.1**	**0.3**
St.Dev	**0.2**	**0.4**	**0.3**	**0.5**	**St.Dev**	**0.0**	**0.3**	**0.2**	**0.4**
Median	**0.3**	**0.3**	**0.3**	**0.3**	**Median**	**0.7**	**0.0**	**0.0**	**0.3**

The comparison of the average values of goal's number highlights, surprisingly, a better performance of novice users, except for the test I, in which the result seems to be the opposite(Figure 2). In this case the nature of data does not allow the application of the nonparametric test of Mann- Withney.



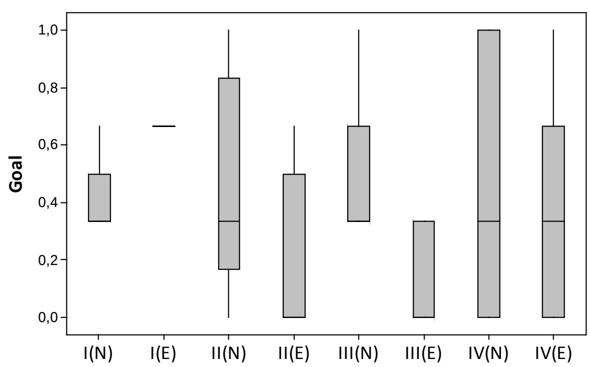


Figure 2: Comparison between Novice and expert- users in term of G using histogram chart and box-plot chart

1.3. Errors number (E)

Experimental results carried out for the usability function errors number are reported in Tab. 4. The comparison of the average values of errors number highlights a better performance of expert users, except for the test II, in which the result seems to be the opposite (Figure 3).

Tab. 4: Experimental results for the usability function E

N	I	II	III	IV	E	I	II	III	IV
U2	0.7	0.3	0.3	1.0	**U1**	0.3	0.0	0.3	0.0
U4	1.3	0.3	1.3	1.7	**U3**	0.7	0.3	0.0	0.3
U5	1.0	0.0	0.3	0.3	**U6**	0.7	0.7	0.3	1.0
U9	0.3	1.3	0.3	0.3	**U7**	1.3	2.0	1.7	1.0
U10	0.7	0.7	2.0	0.0	**U8**	0.7	1.0	0.0	0.3
Mean	**0.8**	**0.5**	**0.9**	**0.7**	**Mean**	**0.7**	**0.8**	**0.5**	**0.5**
St.Dev	**0.4**	**0.5**	**0.8**	**0.7**	**St.Dev**	**0.4**	**0.8**	**0.7**	**0.5**
Median	**0.7**	**0.3**	**0.3**	**0.3**	**Median**	**0.7**	**0.7**	**0.3**	**0.3**

The box-plot diagrams highlight a substantial equality of medians for analyzed samples, except, once again, for the test II.

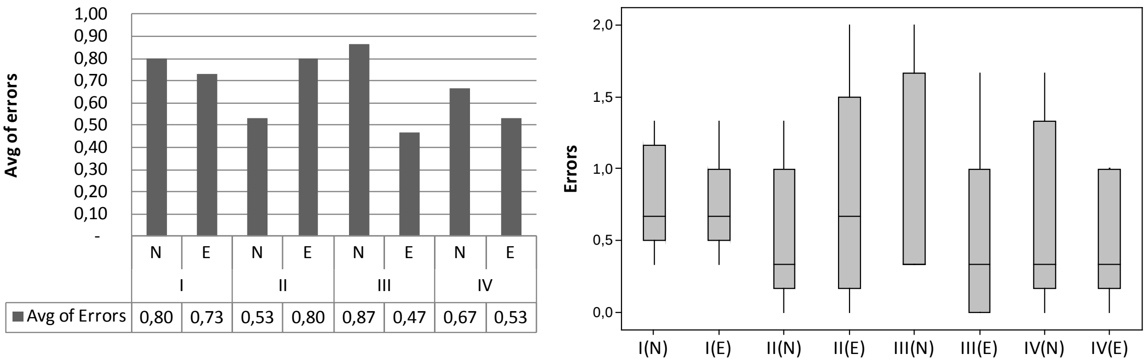


Figure 3: Comparison between Novice and expert- users in term of E using histogram chart and box-plot chart

The application of the Mann- Withney test, confirmed the impossibility to reject the null hypothesis of medians equality for all the tests (*Tab.* ***5***)

Tab. 5: Nonparametric Mann-Withney test for E data, related to novice and experts performance (α= 0.05)

Test		Median	CI	W	p
I	N	0.67	(-0.67;0,67)	29.0	0.8345
	E	0.67			
II	N	0.33	(-1.67;0.67)	25.0	0.6761
	E	0.67			
III	N	0.33	(-1.33; 1.67)	33.0	0.2963
	E	0.33			
IV	N	0.33	(-0.67; 1.33)	28.5	0,9168
	E	0.33			

2. Efficiency measures

2.1. Time (T)

Data related to the time of task execution (*Tab.* ***6***), revealed, once again, a better performance of expert users, except for test I. The difference between the two samples is not substantial, as shown by box-plot diagrams (*Figure* ***4***).

Tab. 6: Experimental results for the usability function T

N	**I**	**II**	**III**	**IV**	**E**	**I**	**II**	**III**	**IV**
U2	83.3	45.3	87.0	58.0	**U1**	57.5	51.7	82.0	75.0
U4	79.0	70.7	86.3	58.7	**U3**	38.7	53.7	56.5	49.7
U5	81.7	56.7	62.7	86.3	**U6**	81.7	56.7	62.7	86.3
U9	76.0	13.,7	12.,5	120.3	**U7**	133.0	64.0	110.0	58.7
U10	76.3	50.0	106.0	63.0	**U8**	102.3	64.7	64.5	71.3
Mean	**79.3**	**71.9**	**0.87**	**0.67**	**Mean**	**82.6**	**58.1**	**75.1**	**68.2**
St.Dev	**3.2**	**37.5**	**24.6**	**26.7**	**St.Dev**	**37.0**	**5.9**	**21.7**	**14.3**
Median	**79.0**	**56.7**	**87.0**	**63.0**	**Median**	**81.7**	**56.7**	**64.5**	**71.3**

The Mann-Withney test reject for all the tests the null hypothesis of medians equality (*Tab.* ***7***).

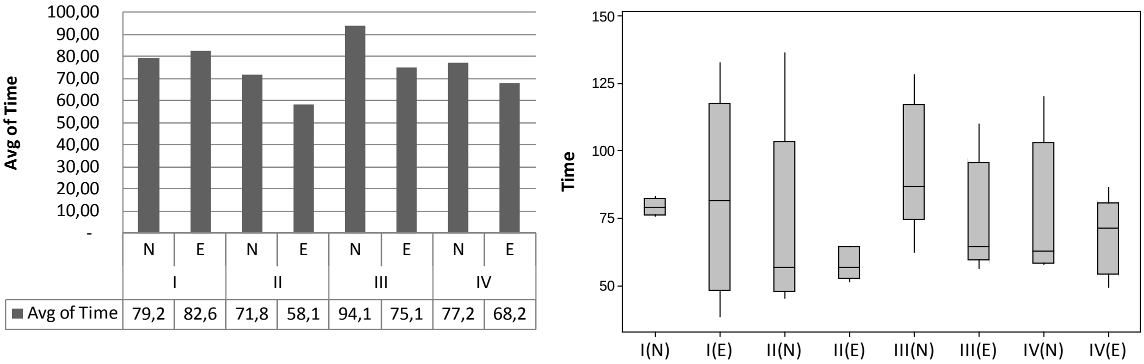


Figure 4: Comparison between Novice and expert- users in term of T using histogram chart and box-plot chart

Tab. 7: Nonparametric Mann-Withney test for T data, related to novice and experts performance (α= 0.05)

Test		**Median**	**CI**	**W**	**p**
I	N	79.0	(-54.0; 40.4)	26.5	0.9168
	E	81.7			
II	N	56.7	(-14.7; 80.0)	27.5	1.0000
	E	56.7			
III	N	87.0	(--23.0;64.0)	33.5	0.2506
	E	64.5			
IV	N	63.0	(-23.3;49.0)	29.0	0.8345
	E	71.3			

2.2. Communication ( )

Data of communication effort highlight higher scores for novice users in all tests as reported in *Tab.* ***8***. This result seems to be confirmed by box- plot chart. although differences between samples with different level of expertise are less evident in several tests (e.g. test IV). Finally, the Mann-Withney test, once again, does not reject the null hypothesis of equality of the medians (*Tab.* ***9***).

Tab. 8: Experimental results for the usability function Q1

N	I	II	III	IV	E	I	II	III	IV
U2	4.0	2.0	5.0	3.0	**U1**	2.0	2.0	2.0	2.0
U4	4.0	3.0	4.0	3.0	**U3**	1.0	3.0	2.0	3.0
U5	3.0	4.0	1.0	3.0	**U6**	1.0	4.0	1.0	3.0
U9	2.0	5.0	3.0	3.0	**U7**	4.0	3.0	3.0	4.0
U10	3.0	3.0	3.0	4.0	**U8**	2.0	1.0	3.0	3.0
Mean	**3.2**	**3.4**	**3.2**	**3.2**	**Mean**	**2.0**	**2.6**	**2.2**	**3.0**
St.Dev	**0.8**	**1.1**	**1.5**	**0.5**	**St.Dev**	**1.2**	**1.1**	**0.8**	**0.7**
Median	**0.7**	**1.0**	**1.3**	**0.4**	**Median**	**1.1**	**1.0**	**0.7**	**0.6**

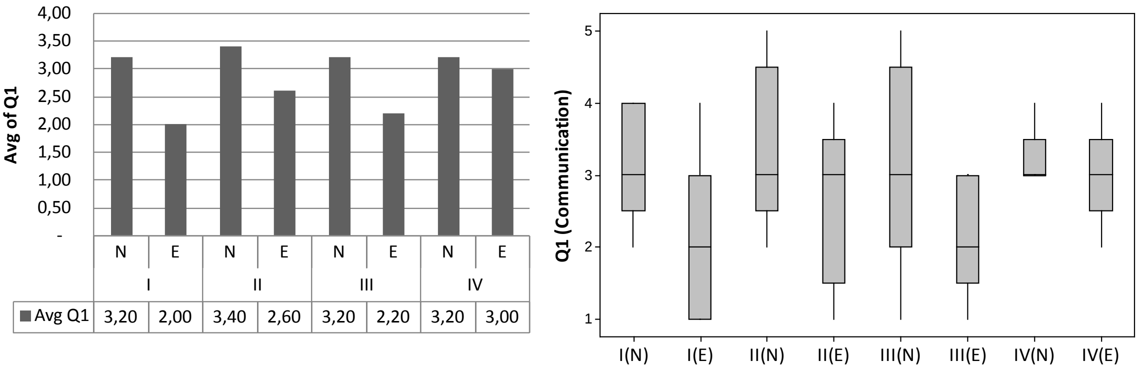


Figure 5: Comparison between Novice and expert- users in term of Q1 using histogram chart and box-plot chart

Tab. 9: Nonparametric Mann-Withney test for communication effort (Q1) data, related to novice and experts performance (α= 0.05)

Test		Median	CI	W	p
I	N	3.00	(-1.00; 3.00)	35.0	0.1437
	E	2.00			
II	N	3.00	(-1.00; 3.00)	32.0	0.4034
	E	3.00			
III	N	3.00	(-1.00; 3.00)	33.5	0.2506
	E	2.00			
IV	N	3.00	(-1.00; 1.00)	29.5	0.7540
	E	3.00			

3. Satisfaction measures

3.1. User Preference ( )

Experimental results carried out for the usability function errors number are reported in *Tab.* ***10***.

Tab. 10: Experimental results for the usability function Q2

N	**I**	**II**	**III**	**IV**	**E**	**I**	**II**	**III**	**IV**
U2	3.0	4.0	2.0	2.0	**U1**	3.0	4.0	4.0	3.0
U4	4.0	5.0	4.0	2.0	**U3**	1.0	4.0	4.0	2.0
U5	3.0	4.0	5.0	2.0	**U6**	5.0	5.0	3.0	3.0
U9	3.0	1.0	2.0	4.0	**U7**	3.0	4.0	3.0	2.0
U10	2.0	2.0	1.0	5.0	**U8**	3.0	5.0	2.0	3.0
Mean	**3.0**	**3.2**	**2.8**	**3.0**	**Mean**	**3.0**	**4.4**	**3.2**	**2.6**
St.Dev	**0.7**	**1.6**	**1.6**	**1.4**	**St.Dev**	**1.4**	**0.6**	**0.8**	**0.6**
Median	**3.0**	**4.0**	**2.0**	**2.0**	**Median**	**3.0**	**4.0**	**3.0**	**3.0**

The analysis of average values and box plot charts for user preference data does not allow to highlight a different trend in perceptions of users with different level of expertize.

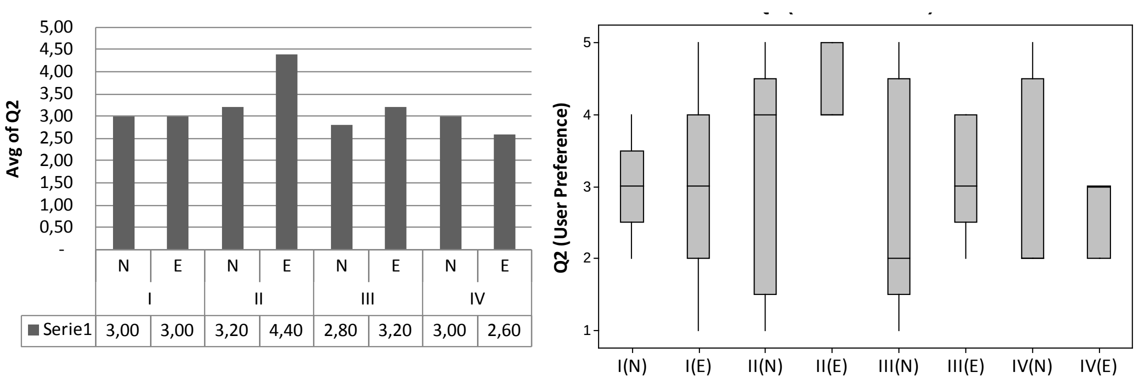


Figure 6: Comparison between Novice and expert- users in term of Q2 using histogram chart and box-plot chart

Also the Mann-Withney test reject for all the tests the null hypothesis of medians equality (*Tab.* ***11***).

Tab. 11: Nonparametric Mann-Withney test for user preference (Q2) data, related to novice and experts performance (α= 0.05)

Test		Median	CI	W	p
I	N	3.00	(-2.00; 2.00)	27.5	1.0000
	E	3.00			
II	N	4.00	(-3.00; 1.00)	22.0	0.2963
	E	4.00			
III	N	3.00	(2.00; 3.00)	25.0	0.6761
	E	2.00			
IV	N	2.00	(-1.00; 2.00)	28.0	1.0000
	E	3.00			

3.2. Ease of use ( )

Starting from the analysis of the average values, scores related to ease of use seem to be higher for expert users (*Tab.* ***12***). This result is not confirmed from box-plot charts in which is impossible to identify a common difference between the samples analyzed, over all tests (*Figure* ***7***).

Tab. 12: Experimental results for the usability function Q3

N	**I**	**II**	**III**	**IV**	**E**	**I**	**II**	**III**	**IV**
U2	1.0	3.0	2.0	1.0	**U1**	2.0	3.0	2.0	3.0
U4	2.0	3.0	3.0	3.0	**U3**	1.0	3.0	2.0	2.0
U5	3.0	3.0	3.0	2.0	**U6**	3.0	3.0	3.0	3.0
U9	2.0	2.0	2.0	2.0	**U7**	2.0	3.0	2.0	2.0
U10	1.0	3.0	2.0	3.0	**U8**	2.0	3.0	2.0	2.0
Mean	**1.8**	**2.8**	**2.4**	**2.2**	**Mean**	**2.0**	**3.0**	**2.2**	**2.4**
St.Dev	**0.8**	**0.4**	**0.5**	**0.8**	**St.Dev**	**0.7**	**0.0**	**0.4**	**0.5**
Median	**2.0**	**3.0**	**2.0**	**2.0**	**Median**	**2.0**	**3.0**	**2.0**	**2.0**

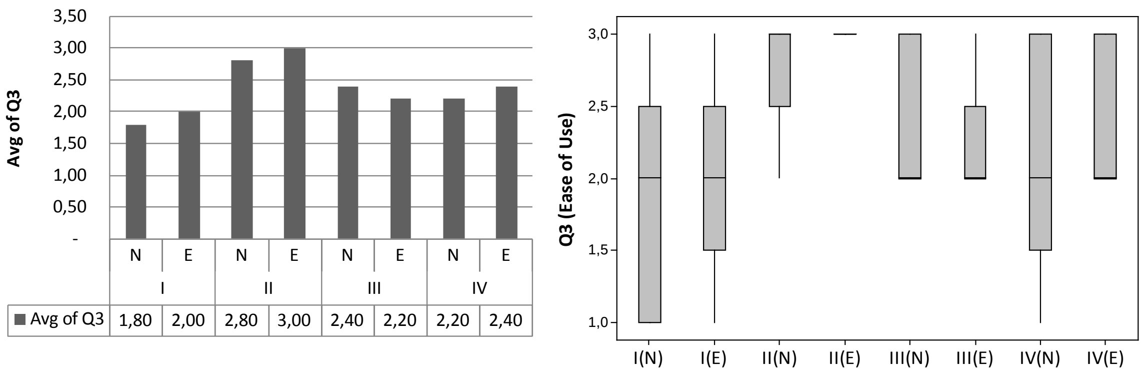


Figure 7: Comparison between Novice and expert- users in term of Q3 using histogram chart and box-plot chart

Tab. 13: Nonparametric Mann-Withney test for ease of use (Q3) data, related to novice and experts performance (α= 0.05)

Test		Median	CI	W	p
I	N	2.00	(-1.00; 1.00)	25.5	0.7540
	E	2.00			
II	N	3.00	(-- .; --)	--	--
	E	3.00			
III	N	2.00	(-1.00; 1.00)	30.0	0.6761
	E	2.00			
IV	N	2.00	(-1.00; 1.00)	26.0	0.8345
	E	2.00			

Appendix B

PCA application for SUM evaluation

Tab. 1: Correlation Matrix (Test I)

Prova I	**ME**	**T**	**G**	**E**	**Q1**	**Q2**	**Q3**
ME	1,00	0,28	0,10	0,38	-0,25	0,40	0,04
T	0,28	1,00	0,08	0,26	-0,18	0,54	-0,03
G	0,10	0,08	1,00	0,50	0,02	-0,10	0,07
E	0,38	0,26	0,50	1,00	0,04	0,28	0,09
Q1	-0,25	-0,18	0,02	0,04	1,00	-0,23	0,70
Q2	0,40	0,54	-0,10	0,28	-0,23	1,00	0,09
Q3	0,04	-0,03	0,07	0,09	0,70	0,09	1,00

Tab. 2: Correlation Matrix (Test II)

Prova II	**ME**	**T**	**G**	**E**	**Q1**	**Q2**	**Q3**
ME	1,00	0,08	0,35	0,69	-0,30	0,32	-0,39
T	0,08	1,00	0,15	0,06	-0,58	0,58	-0,45
G	0,35	0,15	1,00	0,65	-0,31	0,26	-0,11
E	0,69	0,06	0,65	1,00	-0,11	0,09	-0,39
Q1	-0,30	-0,58	-0,31	-0,11	1,00	-0,51	0,77
Q2	0,32	0,58	0,26	0,09	-0,51	1,00	-0,51

Q3	-0,39	-0,45	-0,11	-0,16	0,77	-0,51	1,00

Tab. 3: Correlation Matrix (Test III)

Prova III	**ME**	**T**	**G**	**E**	**Q1**	**Q2**	**Q3**
ME	1,00	-0,09	-0,11	0,01	-0,36	0,21	-0,31
T	-0,09	1,00	0,36	0,72	-0,21	0,28	-0,38
G	-0,11	0,36	1,00	0,48	0,15	0,23	0,18
E	0,01	0,72	0,48	1,00	-0,05	0,16	-0,16
Q1	-0,36	-0,21	0,15	-0,05	1,00	-0,66	0,55
Q2	0,21	0,28	0,23	0,16	-0,66	1,00	-0,50
Q3	-0,31	-0,38	0,18	-0,16	0,55	-0,50	1,00

Tab. 4: Correlation Matrix (Test III)

Prova IV	**ME**	**T**	**G**	**E**	**Q1**	**Q2**	**Q3**
ME	1,00	0,34	0,07	0,08	-0,02	0,24	0,16
T	0,34	1,00	-0,10	-0,21	0,05	-0,17	0,30
G	0,07	-0,10	1,00	0,70	-0,21	0,23	-0,39
E	0,08	-0,21	0,70	1,00	0,03	0,10	-0,38
Q1	-0,02	0,05	-0,21	0,03	1,00	-0,14	0,48
Q2	0,24	-0,17	0,23	0,10	-0,14	1,00	0,23
Q3	0,16	0,30	-0,39	-0,38	0,48	0,23	1,00

TEST I:

Tab. 5: Eigenanalysis of the Correlation Matrix test I

	3.89	1.88	0.61	0.35	0.63	0.10	0.00
Proportion	0.56	0.27	0.09	0.05	0.02	0.02	0.00
Cumulative	0.56	0.82	0.91	0.96	0.98	1.00	1.00

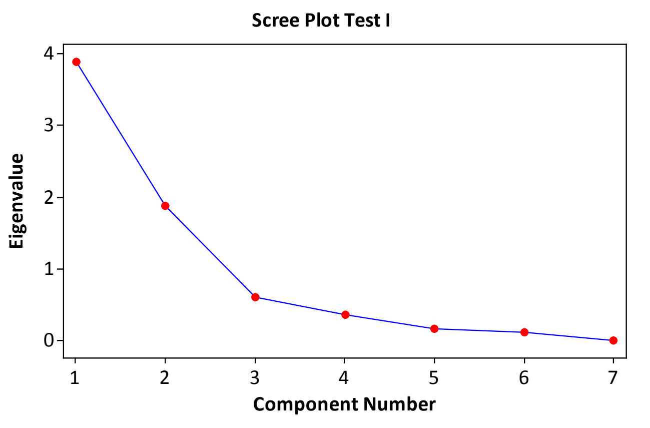


Figure 1: Scree plot diagram for the test I

Tab. 6: Principal components to retain test I

Test	Kaiser's rule	Cumulative variance	Scree plot test
I	,	, ,	,

Tab.7: Eigenvectors of the Correlation Matrix test I

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
ME	0,40	0,04	0,73	-0,33	0,20	-0,24	-0,31
T	0,41	0,24	-0,56	0,04	0,43	-0,50	-0,16
G	0,01	-0,69	-0,24	-0,28	-0,40	-0,31	-0,36
E	0,26	-0,52	0,19	0,75	0,22	-0,05	0,11
Q1	-0,50	0,06	0,01	0,24	0,30	0,13	-0,77
Q2	0,40	0,38	0,00	0,37	-0,65	0,09	-0,36
Q3	-0,46	0,20	0,22	0,23	-0,22	-0,76	0,17

TEST II:

Tab.8: Eigenanalysis of the Correlation Matrix test II

	4.94	1.53	0.37	0.12	0.04	0.00	0.00
Proportion	0.70	0.22	0.05	0.02	0.01	0.00	0.00
Cumulative	0.70	0.92	0.97	0.99	0.98	1.00	1.00

Figure 2: Scree plot diagram for the test II

Tab.9: Principal components to retain test II

Test	Kaiser's rule	Cumulative variance	Scree plot test
I	,	, ,	,

Tab.10: Eigenvectors of the Correlation Matrix test I

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
ME	0,37	-0,35	0,59	-0,17	0,48	-0,19	-0,33
T	0,35	0,47	-0,16	0,58	0,14	0,06	-0,52
G	0,34	-0,37	-0,75	-0,24	0,06	-0,31	-0,17
E	0,31	-0,58	0,04	0,36	-0,34	0,57	0,04
Q1	-0,43	-0,19	0,11	-0,03	-0,49	-0,20	-0,69
Q2	0,39	0,37	0,06	-0,65	-0,32	0,39	-0,20
Q3	-0,44	-0,09	-0,23	-0,16	0,54	0,59	-0,29

TEST III:

Tab. 11: Eigenanalysis of the Correlation Matrix for the test III

	3,89	2,14	0,53	0,31	0,09	0,04	0,00
Proportion	0,56	0,31	0,08	0,04	0,01	0,01	0,00
Cumulative	0,56	0,86	0,94	0,98	1,00	1,00	1,00

Figure 3: Scree plot diagram for the test III

Tab. 12: Principal components to retain. for the test III

Test	Kaiser's rule	Cumulative variance	Scree plot test
I	,	, ,	,

Tab.13: Eigenvectors of the Correlation Matrix test III

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
ME	0,25	0,52	0,07	-0,75	-0,01	-0,13	0,29
T	0,40	-0,36	0,36	0,16	-0,07	-0,65	0,35
G	0,11	-0,56	-0,64	-0,44	-0,19	-0,14	-0,15
E	0,34	-0,44	0,41	-0,27	0,36	0,56	-0,02
Q1	-0,46	-0,23	0,20	-0,13	-0,57	0,27	0,52
Q2	0,46	0,12	-0,47	0,35	0,01	0,34	0,57
Q3	-0,48	-0,14	-0,17	-0,09	0,71	-0,18	0,43

TEST IV:

Tab. 54: Eigenanalysis of the Correlation Matrix for the test IV

	3,68	1,57	1,23	0,44	0,04	0,04	0,00
Proportion	0,53	0,23	0,18	0,06	0,01	0,01	0,00
Cumulative	0,53	0,75	0,93	0,99	1,00	1,00	1,00

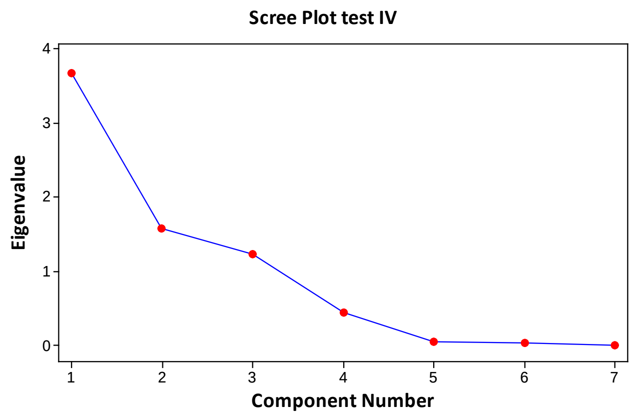


Figure 4: Scree plot diagram for the test IV

Tab. 15: Principal components to retain for the test IV

Test	Kaiser's rule	Cumulative variance	Scree plot test
I		,	, ,

Tab.16: Eigenvectors of the Correlation Matrix test IV

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
ME	0,09	0,69	-0,13	0,70	-0,09	0,08	-0,09
T	0,35	0,31	-0,49	-0,46	0,28	-0,07	-0,50
G	-0,51	-0,01	-0,14	-0,14	-0,40	0,61	-0,42
E	-0,49	-0,18	-0,18	0,23	-0,02	-0,69	-0,41
Q1	0,34	-0,55	0,02	0,46	0,32	0,30	-0,43
Q2	-0,18	0,31	0,77	-0,13	0,37	0,00	-0,36
Q3	0,48	-0,04	0,32	-0,06	-0,72	-0,23	-0,31

Papers

Paper A

G. Di Gironimo, A. Lanzotti, G. Matrone, A. Tarallo, M. Trotta, A virtual reality approach for usability evaluation of a wheelchair-mounted robot manipulators, Proceedings of TMCE2010, Ancona, 12 - 16 April 2010, ISBN 978- 90- 5155- 060- 3, pp 749- 762

Proceedings of TMCE 2010, April 12-16, 2010, Ancona, Italy, edited by I. Horváth, F. Mandorli and Z. Rusák

A VIRTUAL REALITY APPROACH FOR USABILITY ASSESSMENT OF A WHEELCHAIR-MOUNTED ROBOT MANIPULATOR

Giuseppe Di Gironimo [1]
giuseppe.digironimo@unina.it

Giovanna Matrone [2]
giovanna.matrone@unina.it

Andrea Tarallo [1]
andrea.tarallo@unina.it

Mariangela Trotta [2]
mariangela.trotta@unina.it

Antonio Lanzotti [2]
antonio.lanzotti@unina.it

[1] Department of Mechanics and Energetics (DIME)
[2] Department of Aerospace Engineering (DIAS)
PRO-I Research Group
University of Naples Federico II, Italy

Abstract

This work concerns the usability assessment of two control devices for a wheelchair-mounted robot manipulator aimed at assisting physical disabled people. The assessment of the usability is a crucial issue for the design of such products, since they communicate with their users not only through their shape, but especially through their control interfaces. In a first phase, the study focuses on defining a synthetic usability index on the basis of the methodologies currently in use. In a second phase, some experiments in Virtual Reality (VR) have been carried out. The use of VR technologies for the collection of the experimental data has been fundamental in terms of safety, costs and repeatability of the tests. Another important result has been the reduction of the sources of noise, thanks to preliminary simulations in VR and non-invasive questionnaires and interviews for capturing the subjective perceptions of users. Finally, it is worth noticing that the developed model may show its validity also in evaluating the usability of other products. Indeed, it provides a basis for a more extensive use of VR experiments for evaluating different design solutions in terms of global usability requirements.

KEYWORDS

Participatory Design, Virtual Reality Experiments, Assistive Robotics, Usability index, Analytic Hierarchy Process (AHP)

1. INTRODUCTION

The success of a product is strongly influenced not only by its ability to be used for a specific purpose, but also by users perception of it. In short, the goal of the designers is to develop products that satisfy specific needs, assuring at same time a positive feeling to the end-user [30]. ISO reference standard 9241-11:1998 [16] summarizes these aspects in a more general concept of *usability*, defined as the *extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use*. The rigorous assessment of these subjective aspects is certainly one of the major challenges proposed by the reference standard. In particular, the concept of *satisfaction* deals both with the subjective perception about the performance of the product and with a more instinctive feeling of "pleasure" in using the product itself [13]. Nielsen [29] also stresses the importance of the user satisfaction as a measure of the degree of pleasure related to the use of the system. Bevan [2] proposes a framework for product usability measurement inside a more general concept of *quality in use*. In any case, it is understood that the usability evaluation must go together with the assessment of the subjective aspects of the user-product interaction, even if they are extremely difficult to be evaluated systematically [27]. At present, this issue is generally faced with heuristics or hybrid approaches [21, 20] based both on deterministic methods for the assessment of product performance and on heuristic ones for evaluating more subjective aspects. Interesting applications of these methodologies can be found mainly in the medical field, since complex medical equipments often require a careful study of user-product interfaces [37]. In this field, Liljegren [24] notes the inadequacy of the *Usability Questionnaires* (UQ), because of their lack in assessing the subjective unconscious user feelings and his familiarity with the devices under test. Differently, the so-called *Cognitive Walkthrough* (CW) approach [37, 17] is based on the decomposition of the tasks into simpler operations and on their subsequent evaluation by a panel of experts. This methodology helps to detect some problems related to the discrepancies between the actual cognitive model of the end-user and the one expected by the system designer [37]. Similarly, the so-called *think aloud* method [29, 17] evaluates the quality of user-interface interaction, by means of a step-by-step verbalization of the activities of the tester during the experimental phase.

Overall, the principal limitation of these approaches stands in their heuristic nature, which cannot consider the objective performance of the user. In this sense, the so-called *Usability Test* (UT) [37, 29] can be a valid alternative, because it provides quantitative informations about the actual execution of a set of defined tasks. However, the efficiency of this method is limited by the need of physical prototypes and by the impossibility of gathering subjective data. As a result, it is understood that only a combined use of the above discussed methods could provide an objective and subjective assessment of product usability. However, such an approach would require a number of physical prototypes during the test phase, at least one for each re-design process, with obvious consequences in terms of cost.

An effective way to simulate the interaction with a product during its development, limiting the needs of physical prototypes, is the use of Virtual Reality (VR) technologies [4]. Nowadays these technologies are very sophisticated and suitable for different fields of application. Moreover, the diffusion of the Internet has made easier to involve common users, with their subjective feelings and needs, in the evaluation processes, also in the field of product/service design [4]. This approach, focused on end-user involvement all long the product design process, is called *Participatory Design* [29]. VR technologies have helped the spread of Participatory Design in several industries [28, 5, 26]. In fact, virtual and mixed reality environment can be used to evaluate the usability of the final product, even during the concept phase, taking into account both cognitive and physical aspects [19, 22]. These issues become particularly important when the usability tests involve physically disabled people. Indeed, standardized tests could be not suitable for people with varying degrees of disability and, more generally, could be simply hard to administrate [32]. There are many examples in literature of participatory approaches aimed at designing for disabled people. For instance, already more than ten years ago, Eriksson and Johansson [10] developed a computer-based tool to evaluate the design adaptations and the usability of some architectural solutions for physically disabled people. In [38] an iterative participatory approach for designing a *Wheelchair Convoy System* aimed at assisting disabled people is shown. The described methodology allows the designer to collect the feedback of the testers in each phase of product development. Lanzotti et al. [23] have already proposed a participative approach for continuous product innovation based on the identification of users needs through human-product interaction simulations and VR experiments. Furthermore, Wallergard et al. [39] have used a VR-based approach to help people with cognitive disabilities to communicate their feelings about a public transport system. This study has shown the effectiveness of VR-based participative experiments in order to assess even the cognitive and emotional feelings of the testers. Finally, several VR-based simulations about wheelchair-mounted manipulators are described in [33].

The present work, starting from the ISO reference 9241-11:1998 [16], shows the effectiveness of Multi-Criteria Decision Analysis (MCDA) and Saaty's Analytic Hierarchy Process (AHP) [34] in defining a single index of usability for a product. The main novelty stands in the analytical nature of the approach and in the completeness in collecting the experimental data using virtual prototypes. The case study involves the design of an integrated system aimed at assisting disabled people (a powered wheelchair equipped with a robotic arm), in which the usability of two control devices has been evaluated. For this purpose, an user-centered approach, which involves expert users early in the design process, has turned out essential.

2. USABILITY OF A WHEELCHAIR-MOUNTED MANIPULATOR

The development of systems aimed at assisting disabled people makes the analysis of the usability particularly important, both for the limitations of disabled users in interacting with the interface and for the product itself, that significantly affects their quality of life. In particular, this work concerns the usability assessment of two input devices for controlling a powered wheelchair equipped with a robotic arm (Figure 1).

Figure 1 The wheelchair with the robot manipulator.

Indeed, while assistive robots are becoming quite common [14, 8], realistic simulation tools and methods for studying their usability are still required. The present study aims to provide a tool to easy recognize the weaknesses of such a product, through the evaluation of its usability, taking into account not only the functional requirements, but also the subjective needs of the target user, which are not necessarily obvious. This objective is pursued through the identification of a metric for a quantitative assessment of the usability in order to compare different design alternatives. The main issues the authors have faced in this study are:

- identification of a single index of usability starting from many different objective and subjective contributions;
- assignment of a numerical value to characteristics that are not easily quantifiable.

The proposed approach responds to such issues by performing different tests in a Virtual Environment (VE). These experiments have allowed the authors to quantify the usability functions, that then have been merged into a single index by means of the Multi-criteria Decision Analysis (MCDA). The use of Virtual Reality as a tool aimed to the measure of experimental data provides significant benefits in terms of performance and repeatability of the tests, ensuring controlled experimental conditions. The interaction with a virtual product also shields the user from any risk eventually related to the interaction with real prototypes.

2.1. Experimental set-up

In this work, the authors have used Virtual Reality technologies to give the user the impression of moving a robotic arm attached to an ordinary powered wheelchair for physical disabled people. In particular, the case study refers to a powered wheelchair (*Indoor 2003* by Neatech srl) equipped with a *kuka* light-weight robot [6]. The main goal has been the development of a three-dimensional virtual environment in which the user was able to control a robot manipulator attached to a wheelchair, in 1:1 scale and from his own point of view.

The experimental activity has been mainly carried out at “VRoom”, that is a low-cost VR laboratory equipped with two LCD projectors and polarized glasses for *passive* stereoscopic view [4]. Further tests have been also carried out at *VRTest*, that is a high-end laboratory with three DLP projectors and shutter-glasses for *active* stereoscopic view [3]. In order to enhance the impression of moving a real appendix of a wheelchair, a physical wheelchair has been placed in the laboratory in such a way that the user viewpoint coincided with the virtual wheelchair starting position. Moreover, the glasses are endowed with optical targets, and the user can also adjust the point of view on the virtual scene by moving the head. In this way, the authors have set up a semi-immersive VE, where the user can move and control both the wheelchair and the virtual robotic arm by means of different devices (Figure 2).

The first step in order to carry out the virtual simulations has been the design of the VE. The authors have designed a “virtual flat” with all the common furnishing. In particular, it is completely unstructured with respect to the robotic manipulator (Figure 3).

The realism of the VE has been particularly consid-

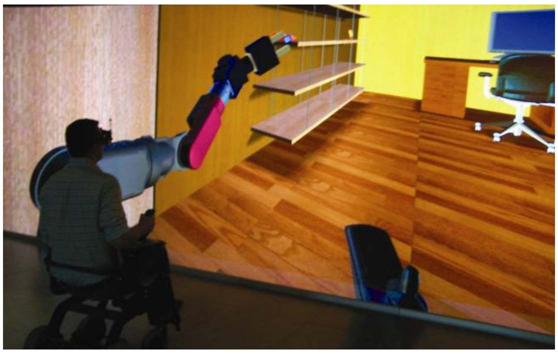


Figure 2 The semi-immersive set-up at *VRTest*.

Figure 3 The "virtual flat" with all the common furnishing.

ered, because a semi-immersive experimental set-up may raise some problems in terms of sense of presence, especially with respect to experiments that involve both real input and virtual outputs. In fact, a low sense of presence of the user may undermine the validity of test results.

The second phase has concerned the programming of the VE, that means, essentially, defining its behaviour in response to the user interaction. The software platform that has been used as Simulation Manager for this work is Virtual Design 2 (VD2), by vrcom GmbH. In particular, the VE can be programmed with a complete set of commands that essentially describe actions that operate on the objects in the VE. In short, the programmer defines certain events that will trigger some action (Figure 4). For instance, a collision between two objects in the VE can cause a warning message as well as the increasing of an error counter, etc.

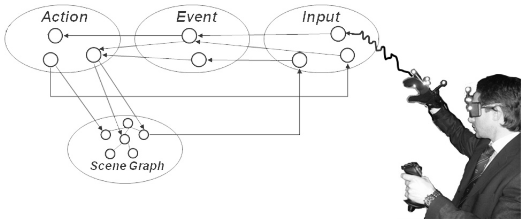


Figure 4 The "input-event-action" paradigm.

In order to achieve this goal, the VE can be programmed through a scripting language. However, the Software Development Kit (SDK) allows the programmer to enhance the basic functionalities of the system by developing external modules that interface with the software kernel. In this way, the programmer can define new classes of actions and events, such as the ones we have used in order to control the virtual robotic arm. The software application that has been developed [7] allows the user to move a kinematic chain in the virtual environment by means of a multidimensional input device, such as a *joystick* or a *space-mouse* (Figure 5).

Figure 5 Input devices.

The *space-mouse* is an input device with 6 Degrees of Freedom (DOF). It has a round "puck" or a "ball" that can be manipulated out of its quiescent position in order to apply rotations as well as translations.

The *joystick* is a very common input device, generally consisting of a stick that pivots on a base and reports its vectorial direction. Moreover a lever controls the "vertical elevation". Thus, the joystick is a 4-DOF input device.

Although the space-mouse and the joystick have different degrees of freedom, in this work only three DOF have been used, in order to control only the position of the end-effector, but not its orientation.

However, both the space-mouse and the joystick are equipped with several buttons that can be used to trigger user-defined actions. For instance, the user can control both the wheelchair and the robot with the same interface (e.g. the space-mouse). This is achieved by simply pressing a button, that switches the active control between the wheelchair and the robotic arm and vice versa.

Finally, it is worth noticing that the user can even move the powered wheelchair in the virtual space with the joystick while he is controlling the robot with the space-mouse and that other kinds of input devices can be tested.

3. THE METHODOLOGICAL APPROACH

The traditional design process tends to favour the functional aspects of an object at the expense of the cognitive-emotional ones, not considering that an object can even have only an emotional function, as in the case of objects of style, figment of the artists' imagination [30].

Hence, it is clear that the usability evaluation can not be separated from the analysis of both the objective and subjective aspects, that are closer to the emotional sphere of the individual. In this sense, the contribution of the user to the design process is crucial (User Centred Design). Starting from these considerations, the proposed approach requires the involvement of potential users during all phases of usability evaluation. In particular, the logical flow chart of this approach is shown in Figure 6.

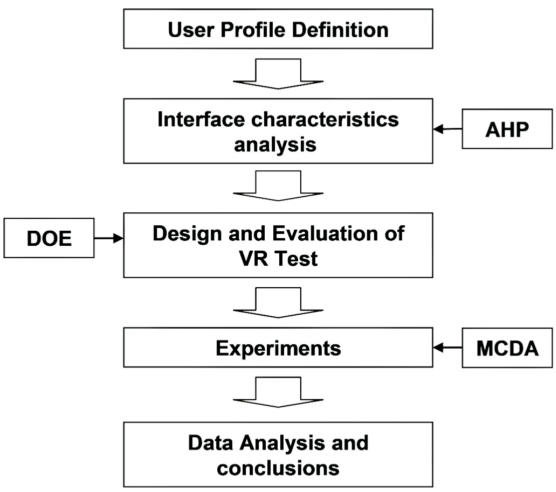


Figure 6 Methodological approach.

3.1. Definition of the user profile

The product-user interaction depends not only on the design elements of the products, but also on the kind of user and the context of use. All attributes that identify specific needs, desires and interests [1], and even behaviours, contexts of use and personal preferences [12], define a specific user profile. The identification of a user profile requires the analysis of the following information:

- product-related user needs;
- cultural background and familiarity of the user with VR technologies;
- context of use;
- identification of user-product spheres (who does what);
- purpose of the interface.

In the present case study, the product at issue is intended for people suffering from total disability of the lower limbs and partial disability of the upper limb, with good cognitive ability, absence of severe disturbances of memory and any delays in perceptual processes. The standard user is intended to be able to interact with an integrated system supporting him in walking and handling objects. In particular, the user-product interface should be consistent with the impairments of the user that imply the inability to perform complex movements. Hence, the interface should satisfy the user needs easily and with no cognitive overload.

3.2. Analysis of the interface characteristics

With respect to the case study, the user-product interface essentially is the control system both of the robotic manipulator and the powered wheelchair. Considering the impact of the interface on the performance of the user, the problem of product usability has been faced by evaluating different control systems. For purely research purposes, the authors have chosen to compare two typical VR input devices: the space-mouse and the joystick (Figure 5).

The usability evaluation of the input devices chosen to control the integrated system has been carried out using Saaty's AHP [34], that is essentially based on the decomposition of the problem into several levels of factors and then on the scoring of the factors of each level, by comparing them in pairs. In our case, the top level of the hierarchy is the usability of the product. The first decomposition can be made, according to ISO 9241-11:1998 [16] and [15], in "*usability dimensions*" [36], namely:

Effectiveness (E_1): The measurement of the effectiveness relates the targets with the accuracy and completeness of the results achieved [29]. The effectiveness value can be assessed in terms of overall device control capability and in terms of the accuracy related to the two main tasks of the integrated device (handling and manipulation).

Efficiency (E_2): ratio between the effectiveness level and the use of resources, meant as physical (*time*) or even cognitive (*communication*) [36].

Satisfaction (S): user-perceived benefit and level of comfort felt during the use of the product. This dimension is strongly related to the subjective perception of user performance.
The assessment of the satisfaction usually requires the evaluation of some of the following parameters:
Preference: choice made by the user;
Ease of use: degree of satisfaction about the final performance in relation to the user expectations;

Starting from these considerations, a preliminary decomposition of the usability is shown in Figure 7. At the first level, there is the Usability (U) of the product, that is decomposed in Usability Dimensions at the second level. In turn, these are broken down at the next level in Usability Characteristics Factors.

However, the numerical assessment of the usability requires a further level to be added to the aforementioned hierarchy in order to translate the Usability Characteristic Factors in Usability Functions that can be quantified during the experiments. These functions have been determined with precision in the experimental phase, in relation to the structure of the tests.

Once the hierarchical decomposition has been completed, the matrix of weights has been defined. This matrix is constructed for each level of the hierarchy and for each group (namely, the set of elements that are children of the same father in the upper level of the hierarchy) by placing all the elements of the group both on the rows and on the columns of the matrix, that is therefore a square matrix.

The generic matrix element a_{ij} is the result of the pairwise comparison between the attribute of the row i and the column j, with respect to a certain task, using the Saaty scale (from 1 to 9) [34].

Thus, the main diagonal of the matrix consists of unit elements only, while the values of other cells are always positive, according to the reciprocity property:

$$a_{ij} = \frac{1}{a_{ji}} \tag{1}$$

Once the pairs comparison matrix has been defined, the weight of each element is assumed as [25]:

$$w_i = \frac{\left(\prod_{j=1}^{n} a_{ij}\right)^{1/n}}{\sum_{j=1}^{n}\left(\prod_{j=1}^{n} a_{ij}\right)^{1/n}} \tag{2}$$

Where n is the dimension of the metrics related to the element at issue. In particular, the allocation of weights is done with a bottom-up logic, from the lowest level of the hierarchy (Usability Functions) to the highest (Usability).

3.3. Design and analysis of experiments in Virtual Reality

The analysis of the product and the user profile has suggested that the most important design features are:

- ability to control the robot;
- ability to control the powered wheelchair;
- logical and cognitive load of the user.

The ability to control the robot refers to the movement of the robotic arm and the manipulation of objects, while the ability to control the wheelchair refers to the movement of the whole integrated system. Finally, the logical and cognitive load refers to the mental workload that the user has to bear while using the device. It is obvious that an intuitive interface and ergonomic controls greatly facilitate the user in controlling the device. The aforementioned features have been used as factors of the Design Of Experiments (DOE), in order to find the best solution between the two input devices (space-mouse and joystick) in terms of usability. The design of the experiments has been directed to achieve two fundamental objectives:

- definition and evaluation of the response functions;
- minimization of the effects of noise factors.

Definition and evaluation of response functions

The response functions are evaluated through a proper VR simulation (*final test*), during which sev-

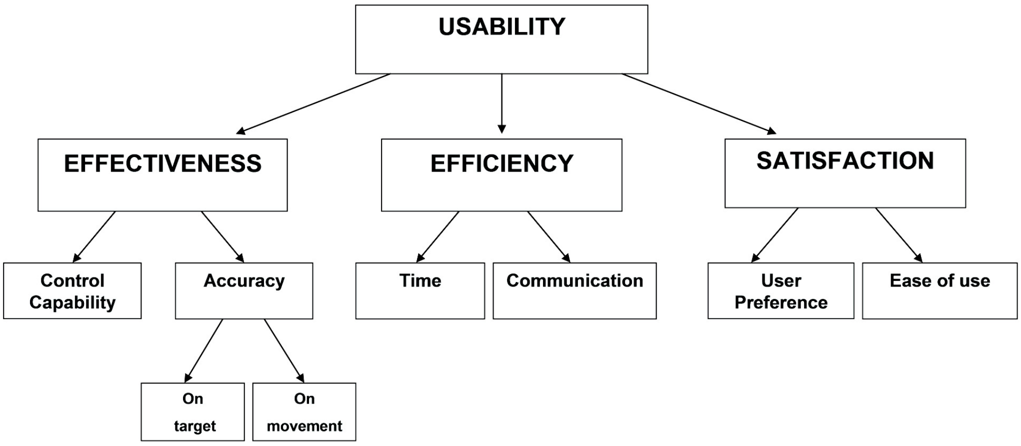


Figure 7 Usability hierarchical decomposition.

eral performance factors are measured. The goal consists in moving a virtual ball between two fixed positions along a straight path. The final test is considered valid only if a certain position is achieved (Figure 8).

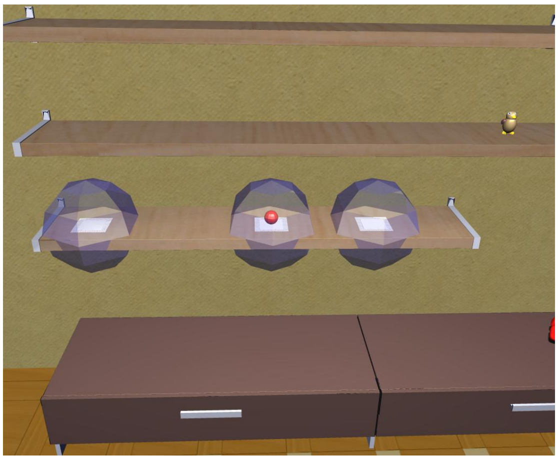


Figure 8 Moving a virtual ball between two fixed positions.

The test has been designed to provide the following response functions according to the hierarchical decomposition above described:

- *Movement Error (ME)* (measure of control capability) is defined as the deviation of the real path from the reference one (Figure 9);
- *Goal number (G)* (measure of accuracy on target) is the number of times the user reaches the goal, that consists in moving a ball between two predefined positions of the test area;
- *Error number (E)* (measure of accuracy on movement) is the number of penalties that the user scores during a single performance, when going beyond the error plans that limit the test area (Figure 10);
- *Time (T)* (measure of efficiency) is the time needed to accomplish the test.

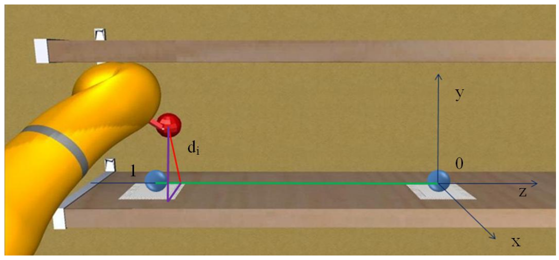


Figure 9 Reference path for Movement Error measurement.

With regard to the subjective response functions, their evaluation has been carried out by means of questionnaires [36] that have been administered to users at the end of the test sessions. The results of these surveys have been classified in three categories of scoring:

- Q1 is a score assigned by the administrator after the test. It measures the *communication effort* of the user (measure of efficiency), intended as the effort made by the user to clarify all his doubts about the functionality of the control devices;
- Q2 is a score assigned by the user by means of a questionnaire. It expresses his *preferences* about the control devices used (measure of satisfaction);
- Q3 represents the *ease of use* of the system. It is a score assigned by the user through a question-

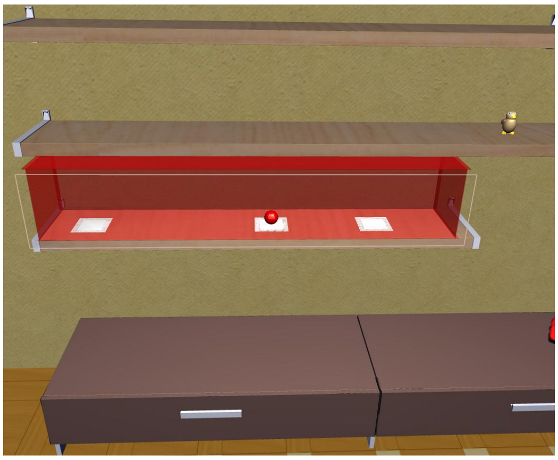


Figure 10 Error plans that limit the test area.

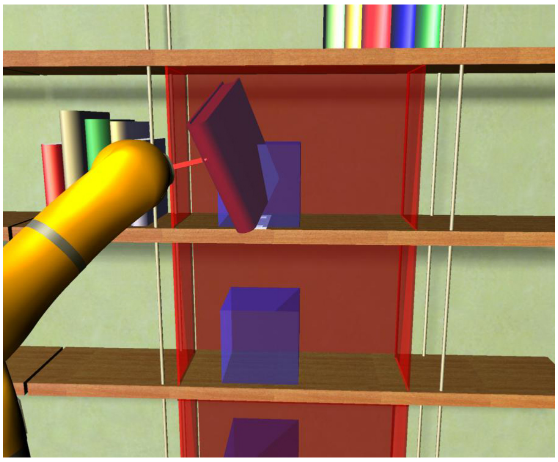


Figure 11 Moving a book between two shelves (task in Simulation II).

naire and defines the ease perceived and the expectations of the tester about the response of the integrated system compared to the initial training phase (measure of satisfaction).

These response functions are the Usability Functions that define the lowest level of the hierarchical model (Figure 12).

Minimization of effects of noise factors related to Virtual Reality

The familiarity of the users with VR technologies and interfaces can be intended as a noise factor for the experiments because the potential users population may include individuals with different skill levels. In order to limit its impact on the final results, the tests has been designed considering these factors by means of:

- a selective preliminary questionnaire to evaluating the skill level of the users, in order to select only those with a minimum level of experience;
- a preparatory phase in which the users are introduced to the tests;
- the administration of two preliminary VR simulations to train the user.

Specifically, the two preliminary simulations are:

Simulation I This simulation is intended to train the user on the navigation through the virtual environment. There are no interactions with objects, the user can only control the powered wheelchair moving it through the virtual flat.

Simulation II The user can move through the virtual environment, but now he can interact with objects in order to move them. In particular, the user has to move a book between two shelves of a library, from a lower shelf to a higher one (Figure 11). The test should be administrated three times. The administrator collects the individual execution time.

It is worth noticing that Simulation II is intentionally more difficult than the final test described in section 3.3 and used for the data collection. Indeed, Simulation II is aimed not only at training the user in interacting with VR devices, but also at defining its familiarity level with VR technologies. This familiarity level is assumed proportional to the average execution time measured during the three administrations of the test. Those who have completed the test in an average time of less than 60 seconds, have been considered “confident users” that are particularly skilled in the use of a specific interface.

3.4. Analysis of the virtual environment

The user may not be at ease in the use of Virtual Reality. In fact, a low sense of presence may induce him to assume postures and behaviors (patterns) that are not those that he would actually assume in everyday life. In this way the user may nullify the whole experiment in Virtual Reality and its results. Therefore, a comprehensive evaluation of the experimental setup is essential, taking into account both the objective characteristics, and the users subjective sphere. The objective aspects of the scenario are evaluated during a preliminary technical review. This phase includes the exploration of the virtual environment and the analysis of the problems by means of some heuristics. Usually, the relevant aspects are due to:

Table 1 Control Factors and Noise Factor.

Control Factors		0	1
Robot Control	R	Spacemouse	Joystick
Wheelchair Control	K	Spacemouse	Joystick
Cognitive Load	C	one hand	two hands
Noise Factor			
Skill Level	L	low	high

Table 2 The cross array planned for the experimental phase.

				0	1	L
	R	K	C			Mean
Test I	0	0	1	$X_{I,0}$	$X_{I,1}$	M_I
Test II	0	1	0	$X_{II,0}$	$X_{II,1}$	M_{II}
Test III	1	0	0	$X_{III,0}$	$X_{III,1}$	M_{III}
Test IV	1	1	1	$X_{IV,0}$	$X_{IV,1}$	M_{IV}

Table 3 Impact matrix.

	ME	G	E	T	Q1	Q2	Q3
Test I	$V_{1,1}$	...	...	...	...	...	...
Test II	$V_{2,1}$	...	...	...	...	...	...
Test III	$V_{3,1}$	...	...	...	...	...	...
Test IV	$V_{4,1}$	...	...	...	...	...	...

- quality of graphics;
- some lacks of the sensory feedback;
- the handling of the devices.

The severity of the problem is evaluated in a scale from 1 to 4, which measures the impact of the problem on the achievement of the task. The subjective factors related to the interaction between the user and the virtual scene are evaluated with a VRSART test. The purposes of the diagnostic tool are [18]:

- help to detect the factors that may impact on the sense of the presence of the user;
- provide a structured method to evaluate the actual impact of the sense of presence on the experimental results;
- classify the sense of presence in specific categories;
- provide an indication about the weaknesses of the user interface;
- provide an immediate feedback of the users performance.

A negative outcome during the technical review or even during the subsequent involvement of the user in the experiments may imply a revision of the whole experimental set-up.

With respect to the case study, the test has been administrated to the user as a 14-items questionnaire related to the above mentioned issues, where the user could express an agreement opinion as a value ranging from 1 to 5. The questionnaire has been administrated only after the last simulation with the dual purpose of being unobtrusive with respect to the test itself and to provide a validation of the experimental scenario.

The problems were primarily due to the structure of the experimental set-up, which involves real inputs and virtual outputs. In any case, the Virtual Reality technology has made it possible to achieve a good sense of presence in spite of the semi-immersive virtual environment. Thus, no changes to the experimental scenario have been considered necessary.

4. EXPERIMENTAL PHASE

4.1. Usability index

In order to consider each factor involved in the experiment, a cross array with three control factors and one noise factor has been used. Depending on the study of the critical design elements, the robot control, the wheelchair control and the cognitive load have been chosen as control factors, each with two levels. The different aptitudes of the users for using the input devices has been chosen as noise factor with two levels (Table 1).

Because of the high complexity of the test, starting from the above identified control factors, a fractionated factorial design, 2^{3-1}, has been developed as inner array and a two-skill-level design has been adopted as outer array (Table 2).

The mean values of the response functions are the elements of the so-called *impacts matrix*, starting from which a single index of usability is defined, by means of the MCDA (Table 3).

Because of the lack of homogeneity of the response functions, all the values have been set to a common base through the normalization of the impacts matrix.

- In order to normalize the collected values of *G*, *Q1*, *Q2* and *Q3*, the Zero-Max normalization has been chosen [11]:

$$e^{'}(i,j) = \frac{e(i,j)}{e_{max}(j)} \tag{3}$$

- In order to normalize *ME*, *E* and *T* response functions, the Min-Max normalization has been used [11]:

$$e^{'}(i,j) = \frac{e(i,j) - e_{min}(j)}{e_{max}(j) - e_{min}(j)} \tag{4}$$

The impacts matrix has been further normalized with respect to the functions with negative effects on the usability. In these cases the reference value has been replaced with its one's complement. The final value of usability has been achieved from the product of the normalized impacts matrix with the vector of weights coming from AHP. As for the weights, the final value of usability has been achieved with a bottom-up approach, starting from the response functions and climbing the Saaty's hierarchy until the final value of usability.

4.2. Experimental Protocol

The four tests of the experimental design have been carried out on the basis of an experimental protocol previously defined as follows:

1. administration of a questionnaire for the selective collection of information on the cultural background of the user and its familiarity with the control interfaces. The questionnaire is a selective tool to recruit testers with appropriate skills;
2. briefing to explain the contents of the tests;
3. user training with simulation I and II;
4. viewing a video tutorial about the *final test*;
5. administration of the *final test*.

Three executions of the *final test* (see section 3.3) have been carried out for each user, during which objective data have been collected (ME, G, E, T). Moreover, during the test execution, the administrator has recorded his impressions about the user in order to determine his communication effort while interacting with the virtual environment (Q1). After the simulation, a questionnaire on a scale from 1 to 5 has been administrated to each user, in order to assess the *Preference* (Q2) and *Ease of use* (Q3) factors. Moreover, a further VRSART questionnaire for assessing the *sense of presence* has been administrated.

5. RESULTS

In preliminary experiments, after the setting of the virtual scene and the design of the experiments, the weights for all levels of the hierarchical structure have been assigned. In particular, a questionnaire based on the method of comparison in pairs has been administrated to seven experienced designers. The weights vectors have been derived from the mean values of the collected data (Figure 12) for each level of the hierarchy. Then, the vector **p** of the weights of the Usability Dimensions on U at the highest level has been defined as:

$$\mathbf{p} = [0,42 \;\; 0,27 \;\; 0,31]^T \tag{5}$$

Table 4 Experimental and expected Movement Error.

		Experimental	Expected
Movement Error (ME)	R	0	1
	K	1	1
	C	0	0

Ten users have been involved in the experiments, five for each of the two skill levels. The above described seven response functions have been collected for each experiment, for each of which the analysis of the main effects has been performed in order to define the impact of control factors on the individual responses.

Thus, the expected optimal combination has been defined and then it has been compared with the experimental one. For instance, in the case of ME function, the charts in Figure 13 show a strong effect of K and C compared to that of R. In this case, the experimental optimal combination (R_0, K_1, C_0) is different from the expected one (R_1, K_1, C_0) (Table 4). However, given the quite marginal impact of R, which has been derived with Pareto-ANOVA analysis [31], the two combinations can be considered substantially equivalent.

Whenever the expected optimal combination has not coincided with the experimental one, it has been verified that the difference has been related to the factor R with no significant effects. Thus, at a first approximation, we can consider the two configurations virtually identical. However, it would be better to repeat the experiment in order to confirm expected results. Starting from the Usability Functions, through the combination of AHP and MCDA, the Usability Dimensions have been assessed for each experiment (Table 5). Table 5 can be summarized in the follow-

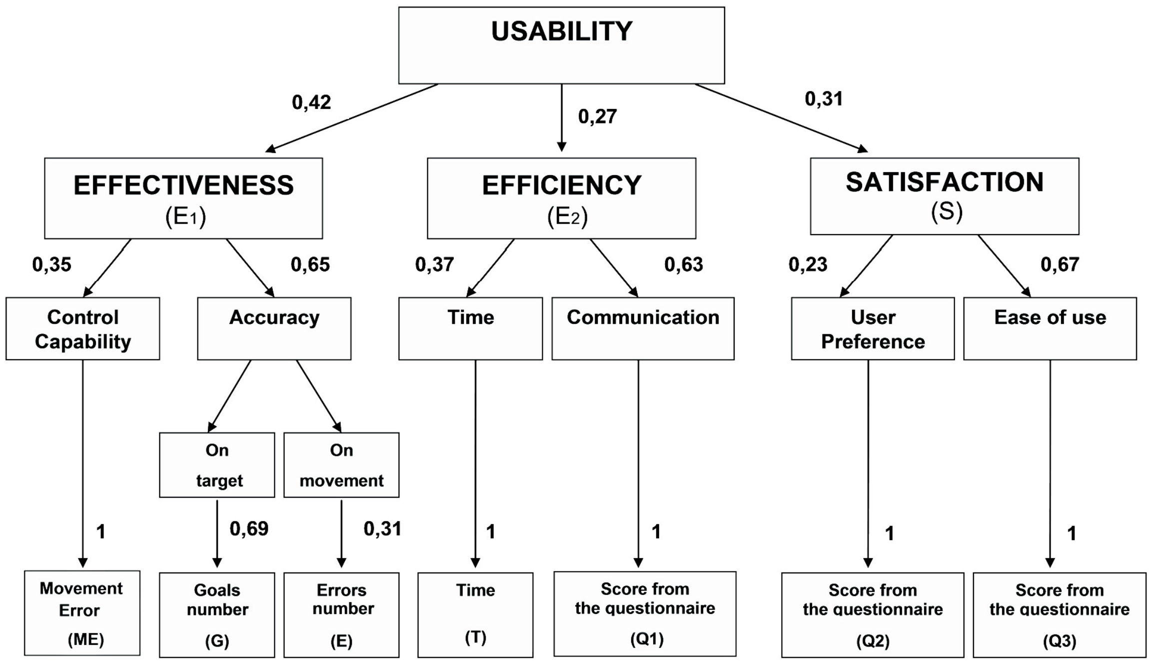


Figure 12 Complete hierarchical model of Usability with weights.

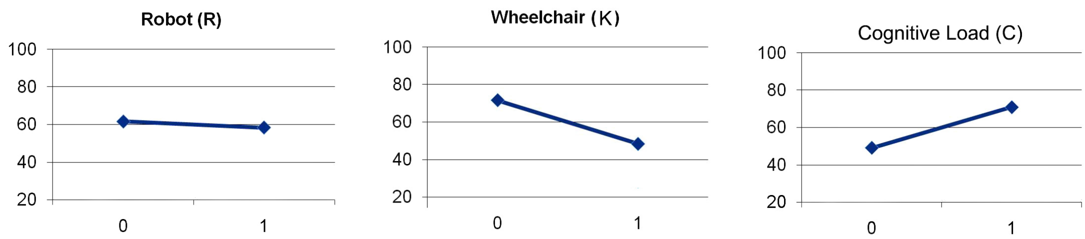


Figure 13 Main effects of the control factors on the Movement Error (ME).

Table 5 Characteristic dimensions of the usability.

	E_1	E_2	S
Test I	0,45	0,75	0,71
Test II	0,8	1	1
Test III	0,57	0,5	0,79
Test IV	0,75	0,95	0,78

Table 6 Usability assessments for each Test.

	Usability Index (UI)
Test I	0,609
Test II	0,915
Test III	0,621
Test IV	0,812

ing *Usability Dimensions Matrix*:

$$\mathbf{U_D} = \begin{bmatrix} 0,45 & 0,75 & 0,71 \\ 0,8 & 1 & 1 \\ 0,57 & 0,5 & 0,79 \\ 0,75 & 0,95 & 0,78 \end{bmatrix} \tag{6}$$

At this level of the hierarchy, Test II achieves the best results in terms of both effectiveness and efficiency and satisfaction. For this, regardless of the choice for the weights vector at the last level, the best simulation in terms of usability will always be the second one. In fact, the weighted sum of the Usability Dimensions values provides the overall results for the usability index (Table 6):

$$\mathbf{u} = \mathbf{U_D} \cdot \mathbf{p} \tag{7}$$

As expected, the best solution belongs to Test II, that obtains an usability value of 0,915, which is related

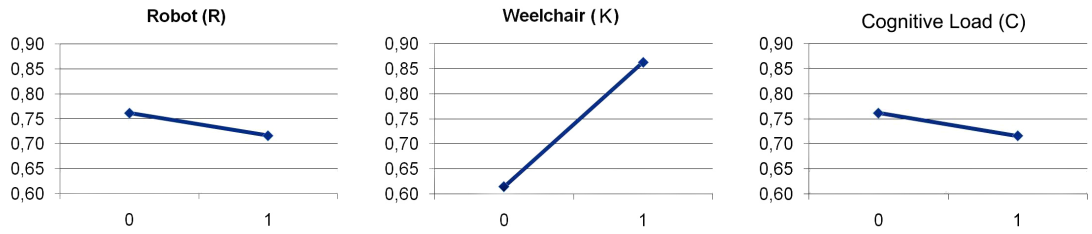


Figure 14 Effects of the control factors on the Usability Index (UI).

to (R_0, K_1, C_0) configuration, that is:

- Space-mouse for the robot control;
- Space-mouse for the wheelchair control;
- Only one hand to handle the interface (minor cognitive load).

This allows the authors to evaluate even the remaining tests that were not able to be ranked because of the different values of E_1, E_2 and S at the previous level. In any case, Test IV obtains a good UI value, while UI values coming from Test I and Test III are far from the best one.

The analysis of the main effects has been carried out also for each UD and for the UI, by considering them as response functions. In Figure 14 the plots of the effects of the control factors on the usability index are shown. They highlight a noticeable impact of K, if compared with that of R or C. This is a behaviour similar to "ease of use" function, that most affects in terms of weight the global index.

6. CONCLUSIONS AND FUTURE WORK

A new approach to usability assessment has been developed through a non-heuristic but analytic methodology, based on both AHP and MCDA. A single usability index has been assessed through several VR experiments that have considered both objective and subjective aspects of user-product interaction. A weakness of this approach could be the evaluation of the weights for MCDA analysis. In order to ensure reliable results, which better reflect the users scale of priorities, the experts team should include not only designers as in our case, but also medical specialists. Furthermore, it would be worth studying more deeply how the familiarity of the testers with VR simulation tools can affect the simulation results, even if the literature has highlighted only a partial relevance of this aspect [35]. Finally, a further weakness can be the eventual mix-up between "product usability" and "VR usability". For this, the experimental set-up is fundamental to guarantee that the satisfaction feeling of the user is actually related to the product, rather than the experimental modalities. The case study has focused on the usability assessment of an integrated system for disabled people. Even during the earliest design stages, the described VR approach could help the designers to select and validate the best architecture for an assistive robotic system. It is worth emphasizing that the final result is likely to have been influenced by the composition and the size of the users sample. Indeed, the selected sample has consisted of able-bodied people, properly informed about the experiments and their purpose. However, the authors believe that an expert-based approach, such as the one here described, makes easier the administration of the tests even to disabled people [32]. On this basis, future confirmatory tests can be carried out on disabled users, in order to verify the consistency of the data already collected. In this case, in order to minimize noise related to subjective feelings assessment, future research trends [36, 9] will focus on the study of data collection tools alternative to questionnaires.

ACKNOWLEDGMENT

The authors thank Dr. *Agostino De Santis* for his interesting discussions about assistive robotics and Human-Robot Interaction and *Amedeo Sgura* for his precious technical support. We have to mention also *Neatech s.r.l.* for having provided the physical wheelchair with the corresponding CAD model used for VR experiments.

References

[1] Abbattista, F., Degemmis, M., Licchelli, O., Lops, P., Semeraro, G., and Zambetta, F. Improving the usability of an e-commerce web site through personalization. In *Recommendation and Personalization in Ecommerce, 2nd In-*

ternational Conference on Adaptive Hypermedia and Adaptive Web Based Systems (2002), F. Ricci and B. Smith, Eds., pp. 20–29.

[2] Bevan, N. Extending quality in use to provide a framework for usability measurement. In *HCD 09: Proceedings of the 1st International Conference on Human Centered Design* (Berlin, Heidelberg, 2009), Springer-Verlag, pp. 13–22.

[3] Caputo, F., and Di Gironimo, G. Vrtest: a Virtual Reality sysTEm for tranSporTation design. *Centre of Competence for Transport Systems of the Campania Region, An experience of innovation and training.* (2007).

[4] Caputo, F., Di Gironimo, G., and Papa, S. A virtual reality system for ergonomics and usability validation of equipment controls. *ANALES DE INGENIERÌA GRÀFICA 18* (Sept. 2006), 47–64.

[5] Davies, R.C. Adapting virtual reality for the participatory design of workenvironments. *Comput. Supported Coop. Work 13*, 1 (2004), 1–33.

[6] De Santis, A., Di Gironimo, G., Marzano, A., Siciliano, B., and Tarallo, A. A virtual-reality-based evaluation environment for wheelchair-mounted manipulators. In *Proceedings of the 6th EUROGRAPHICS Italian Chapter 2008 conference* (Salerno, Italy, 2008), pp. 1–8.

[7] Di Gironimo, G., Marzano, A., and Tarallo, A. Human robot interaction in virtual reality. In *5th EUROGRAPHICS Italian Chapter Conference* (Trento, Italy, 2007), pp. 1–8.

[8] Eftring, H., and Boschian, K. Technical results from manus user trials. In *Proc. International Conference on Rehabilitation Robotics (ICORR)* (Stanford, CA, 1999), pp. 136–141.

[9] Eraslan, E., Akay, D., and Kurt, M. Usability ranking of intercity bus passenger seats using fuzzy axiomatic design theory. In *CDVE* (2006), pp. 141–148.

[10] Eriksson, J., and Johansson, G. Adaptation of workplaces and homes for disabled people using computer aided design. *International Journal of Industrial Ergonomics 17*, 2 (Feb. 1996), 153–162.

[11] Figueira, J., Greco, S., and Ehrgott, M. *Multiple Criteria Decision Analysis: State of the Art Surveys*. Springer, 2005.

[12] Ghosh, R., and Dekhil, M. Discovering user profiles. In *WWW '09: Proceedings of the 18th international conference on World wide web* (New York, NY, USA, 2009), ACM, pp. 1233–1234.

[13] Hassenzahl, M. The thing and i: understanding the relationship between user and product. In *Funology: From Usability to Enjoyment (Human-Computer Interaction Series)* (Norwell, MA, USA, 2003), M. A. Blythe, K. Overbeeke, and A. F. Monk, Eds., Kluwer Academic Publishers, pp. 31–42.

[14] Hillman, M., and Gammie, A. The bath institute of medical engineering assistive robot. In *Proc. ICORR 94* (Wilmington, 1994), pp. 211–212.

[15] Hornbaek, K. Current practice in measuring usability: Challenges to usability studies and research. *Int. J. Hum.-Comput. Stud. 64*, 2 (2006), 79–102.

[16] ISO/IEC standard. Ergonomic requirements for office work with visual display terminals - guidance on usability. *ISO 9241-11:1998* (1998).

[17] Jaspers, M.W.M. A comparison of usability methods for testing interactive health technologies: Methodological aspects and empirical evidence. *I. J. Medical Informatics 78*, 5 (2009), 340–353.

[18] Kalawsky, R.S. New methodologies and techniques for evaluating user performance in advanced 3d virtual interfaces. *IEE Digest. 1998*, 437 (1998), 5/1–5/8.

[19] Kanai, S., Horiuchi, S., Kikuta, Y., Yokoyama, A., and Shiroma, Y. *An Integrated Environment for Testing and Assessing the Usability of Information Appliances Using Digital and Physical Mock-Ups*. Springer Berlin / Heidelberg, Berlin / Heidelberg, 2007.

[20] Kilic Delice, E., and Güngör, Z. The usability analysis with heuristic evaluation and analytic

hierarchy process. *International Journal of Industrial Ergonomics 39*, 6 (November 2009), 934–939.

[21] Kim, J., and Han, S. A methodology for developing a usability index of consumer electronic products. *International Journal of Industrial Ergonomics 38*, 3-4 (March 2008), 333–345.

[22] Kuutti, K., Battarbee, K., Säde, S., Mattelmäki, T., Keinonen, T., Teirikko, T., and Tornberg, A. Virtual prototypes in usability testing. In *HICSS '01: Proceedings of the 34th Annual Hawaii International Conference on System Sciences (HICSS-34)-Volume 5* (Washington, DC, USA, 2001), IEEE Computer Society, p. 5029.

[23] Lanzotti, A., Matrone, G., Tarantino, P., and Vanacore, A. Statistical design for innovation in vr. *Statistics for innovation* (2008).

[24] Liljegren, E. Cognitive engineering methods as usability evaluation tools for medical equipment. *International Journal of Industrial Ergonomics 34*, 1 (July 2004), 49–62.

[25] Lin, M.C., Wang, C.C., Chen, M.S., and Chang, C.A. Using ahp and topsis approaches in customer-driven product design process. *Comput. Ind. 59*, 1 (2008), 17–31.

[26] Mahdjoub, M., Monticolo, D., Gomes, S., and Sagot, J.C. A collaborative design for usability approach supported by virtual reality a multi-agent system embedded in a plm environment. *Computer-Aided Design* (2009).

[27] Montero, F., Lozano, M., and Gonzalez, P. Usability-oriented quality model based on ergonomic criteria. *Handbook of Research on Web Information Systems Quality* (2008).

[28] Muzzupappa, M., Bruno, F., Mattanò, R. M., and Pina, M. A new approach to participatory design: usability tests in virtual environment. In *Research in Interactive Design, Vol. 1, Cap. X* (New York, NY, USA, 2006), SpringerVerlag., pp. 80–90.

[29] Nielsen, J. *Usability Engineering*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.

[30] Norman, D. *Emotional Design*. Perseus Publishing, New York, 2004.

[31] Park, S.H. *Robust design and analysis for quality engineering*. Kluwer Academic Publishers, London, UK, 1996.

[32] Petrie, H., Hamilton, F., King, N., and Pavan, P. Remote usability evaluations with disabled people. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems* (New York, NY, USA, 2006), ACM, pp. 1133–1141.

[33] Pithon, T., Weiss, T., Richir, S., and Klinger, E. Wheelchair simulators: A review. *Technology and disability 21*, 1 (2009), 1–10.

[34] Saaty, L. Decision making with the analytic hierarchy process. *International Journal of Services Sciences 1*, 1 (2008), 83 – 98.

[35] Sauer, J., Seibel, K., and Rüttinger, B. The influence of user expertise and prototype fidelity in usability tests. *Applied Ergonomics* (July 2009).

[36] Sauro, J., and Kindlund, E. A method to standardize usability metrics into a single score. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 2005), ACM, pp. 401–409.

[37] Shackel, B., and Richardson, S.J. *Human Factors for Informatics Usability*. Cambridge University Press, New York, NY, USA, 2008.

[38] Sharma, V., Simpson, R.and LoPresti, E., Mostowy, C., Olson, J., Puhlman, J., Hayashi, S., Cooper, R., Konarski, E., and Kerley, B. Participatory design in the development of the wheelchair convoy system. *Journal of NeuroEngineering and Rehabilitation 5*, 1 (2008), 1.

[39] Wallergard, M., G., Eriksson, and Johansson, G. A suggested virtual reality methodology allowing people with cognitive disabilities to communicate their knowledge and experiences of public transport systems. *Technology and disability 20*, 1 (2008), 9–24.

Paper B

A. Lanzotti, A. Vanacore, M. Trotta, Validation of a new index for seat comfort assessment, based on objective and subjective measures, Proceedings of IMPROVE 2011, Venezia, 15 – 17 june 2011, ISBN 978-88- 7784- 328- 9, pp 7

Proceedings of the IMProVe
International conference on Innovative Methods in Product Design
June 15^{th} – 17^{th}, 2011, Venice, Italy
ADM – INGEGRAF

Validation of a new index for seat comfort assessment based on objective and subjective measurements

A., Lanzotti, A., Vanacore, M., Trotta

University of Naples Federico II, Faculty of Engineering

Article Information

Keywords:
User centered design
Chair design
Comfort assessment
Comfort Loss
Ordinal logistic regression

Corresponding author:
Antonio Lanzotti
Tel.:+39-0817682506
Fax.:+39-0817682187
e-mail: antonio.lanzotti@unina.it
Address: P. le Tecchio, 80, 80125, Naples (NA), Italy

Abstract

Purpose:

This work aims at validating a new statistical index (Weighted Pressure Comfort Loss, WPCL) for seat comfort assessment. The validation is carried out by deepening the relationship between subjective comfort evaluations and objective measures of seat comfort and comparing, from an engineering standpoint, the new index with the pressure peak which is currently one of the most used indexes for seat comfort assessment.

Method:

In the experimental phase, 22 experimenters evaluated four office chairs, by using different evaluation scales and methods (rating, ranking, comfort degree). The subjective comfort perceptions were collected through questionnaires. At the same time, several objective parameters related to seat comfort were measured by using a capacitive mat. In order to select the objective comfort measures which are significantly associated to the perceived comfort, a logistic regression model was adopted.

Result:

The logistic regression model selected the peak pressure as a significant predictor of perceived comfort whereas, the hypothesis of absence of correlation between the perceived comfort and the WPCL index cannot be rejected. However, from an engineering standpoint, the final seat rating evidences substantial coherence of peak pressure and WPCL index, showing not redundant results useful to design team for seat comfort improvement. Since results were strongly influenced by experimental conditions and anthropometric variability of the experimenters, further investigations should be carried out. On the basis of the first experiments, a refinement of the index and new test conditions could be investigated.

1 Introduction

Specialized literature does not provide a universally recognized definition of comfort, nevertheless in recent years, the assumption that comfort and discomfort are two distinct entities [1] is winning broad respect. In their studies, Zhang and Helander [2] show that sitting discomfort is related to the biomechanical factors associated to the interaction with the seat over time, whereas comfort reflects a perception of instantaneous well-being perceived by the user. Zhang [3] pointed out that poor biomechanics may turn comfort into discomfort even though good biomechanics is not a necessary and sufficient condition for comfort. In other words, good biomechanics can avoid discomfort and thus it can be assumed as a prerequisite for comfort. Being complex concepts, comfort and discomfort are difficult to measure and interpret [4]. A great deal of research has been done to face the problem of sitting comfort/discomfort assessment and several subjective and objective methods have been developed [5]-[9]. Typically comfort assessment is realized on the basis of subjective evaluations or postural analysis. Subjective evaluations are collected by surveying potential seat users who are asked to express their feelings of comfort/discomfort with the seat and/or compare, in terms of perceived comfort/discomfort, different seats belonging to the same class [10][11].

Postural analysis is realized by measuring one or more objective parameters, such as [12]:

- the pattern of muscle activation measured through electromyography (EMG) [13].
- the stress acting on the spine measured through pressure transducer and radio waves [13]
- the postural angles [14] obtained using contact or non contact (like photogrammetric) techniques in real experiments or using virtual manikins in virtual experiments [15]
- the body–seat interface pressure measured through capacitative mats.

Anyway, subjective and objective methods are not alternative since they complement each others.

The exclusive use of subjective evaluations can be misleading for several reasons:

- when attention focuses on particular elements of the seat, the response variability is reduced, but the interaction with other neglected features can be a noise factor [13]
- users could not be able to synthesize a subjective perception in a numeric or semantic evaluation causing a partial loss of information [11].

- the perceived differences of ergonomic features are often small and the results from comparisons of different seat concepts are rarely significant;
- the human body is very adaptive and not sensitive to distinguish variations in seats;
- subjective evaluations are costly and time-consuming [16];
- subjective evaluations are rarely applicable early in the design process [4].

On the other hand, the exclusive use of objective measures for comfort assessment, highlights the following criticisms:

- normally, the information provided by objective criteria are complement but not substitute of subjective evaluations related to user's perception of comfort;
- the construction of quantitative measures for comfort assessment cannot disregard from noises often overlooked, such as anthropometric variability.

In this perspective, a great deal of research has been performed to find objective measures for predicting seat comfort perception [17]. Research has shown that one of the main factors that affect seat comfort is seat-interface pressure distribution [18]. Moreover, pressure distribution is the objective measure with the clearest correlation with the subjective evaluation methods[4][8]. Human-seat interface pressures have a spread field of application, indeed they have been measured to improve the comfort of office chairs [19], car seats [12], motorcycles saddles [20] and others vehicles seats [21], as well as to pursue product innovation in Kansei Ergonomics [22]. In particular, in office chair design pressure maps have been used to qualitatively verify the effectiveness on seat comfort of product features like, *e.g.*, cushion shape and materials [23]-[25] through correlation studies with the subjective user perceptions. Nevertheless the widespread use of pressure maps, just few authors [26][27] have proposed synthetic indexes for the related multidimensional data, collected by performing real or virtual experiments involving a selected sample of potential users. Furthermore, little effort has been made to highlight the usefulness of these pressure measures for specific purposes defined by designers (*e.g.* Design for a Target and Design for All).

In order to provide a tool that can be easily used by designers Lanzotti et al. [26][27] proposed the *Weighted Pressure Comfort Loss* (*WPCL*) a postural comfort index based on comfort loss due to uneven seat-interface pressure distribution. In this paper the *WPCL* index is statistically validated by assessing how its results correlate with comfort perception expressed in short-term experimental sessions. The experiments were planned by using robust design approach, taking into account the noise related to the anthropometric variability of the experimenters.

2 *Identification of the goals of seat comfort assessment*

The results presented in this paper are part of a wider and long-standing research activity carried on at the Department of Aerospace Engineering of University of Naples Federico II and aimed at developing simple and repeatable procedures useful to design teams for the development of more comfortable seats. To this aim, the first research step is the definition of simple quantitative seat comfort measures. These measures can be expressed into synthetic indexes that objectively meets two fundamental requirements:

- the index must be representative of user perceptions and it must be a valid surrogate of information obtained through questionnaires, until now extensively used in this field of study;
- the index must be an usable and interpretable indicator that supports the designer in his design choices.

Further, the second research step is to apply a robust design approach to validate these indexes and to identify and choose optimal levels for seat features (like materials and shapes) that improve contact between the human body and the part of a chair on which one's weight rests directly (the seat). The focus of this paper is on the validation of a new comfort index. The proposed validation procedure consists of four phases:

- Experimental setup design
- Definition of the objective and subjective measurement methods
- Comfort index definition and validation
- Experimental results elaboration

In the first phase, the experimental setup was defined in terms of control factors and noise factors by using robust design approach.

In the second phase, the experimenters, during short-term static sessions, evaluated the comfort of some office chairs expressing their judgments on three different scales (rating, ranking, comfort degree). Simultaneously, a capacitive mat allowed to capture the pressure distribution on seat interface. In this way, for each experimenter, subjective and objective measurements were collected.

In the third phase, the best objective predictors for perceived comfort were selected and validated by adopting the ordinal logistic regression (OLR). This statistical technique was applied in order to investigate the nature of relationships between the objective measurements, obtained from pressure maps and perceived comfort (subjective measurements). So the validation of WPCL index starts with the correlation analysis between objective and subjective measurements.

In the fourth phase, the validation follows an engineering approach based on the comparison of design choices strictly linked to the adoption of objective indexes. Even if the experimental set up is simple and just linked to one design factor, experimental results were analyzed and interpreted in order to verify if and how indexes can condition and help to improve seat design.

2.1 *Previous study*

In previous works [26][27], the authors proposed the index WPCL based on the human-seat interface pressures measured over a bidimensional pressure map obtained by discretizing the whole contact surface between the human body and the seat in a finite number, (N), of equal-area cells. When the user j is seated, (with $\leq N$) cells are *activated* by the effective contact between the human body and the seat. The pressure value reported in correspondence of any activated cell is always positive. The formulation of the WPCL index is coherent with the assumption, supported by literature, that the uniformity of pressure distribution increases the level of perceived comfort [8][25]. Coherently with these assumption, for each user, a target value was defined as the mean pressure over the whole contact area (eq. 1).

$$x_{0j} = \frac{PS_j}{n_j} = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j} \qquad (1)$$

where:

- indicate the overall pressure impressed by the j-th user on the seat ,
- is the number of activated cells in the pressure map for the j-th user,
- is the pressure value measured by the i-th cell when the j-th user is seated.

For each user and for each cell of the map it is possible to identify a pressure comfort loss based on a "Nominal is the Best" (NB) loss function, standardized with respect to the nominal pressure. Starting from the (1), for the *j-th* user the Pressure Comfort Loss Index over the activated cells of the contact surface is defined as:

$$PCL_{ij}(x_{ij}) = k_{ij}\left(\frac{x_{ij} - x_{0j}}{x_{0j}}\right)^2 \qquad (2)$$

where is a coefficient that for each cell measures the loss corresponding to the maximum accepted deviation from the target.

Starting from eq. 2, assuming the hypotheses that the loss coefficient is the same for all the cells and the loss is additive, for the user j, the Pressure Comfort Loss index over the activated cells of the contact surface is:

$$PCL_j(\vec{x}) = k_j \sum_{i=1}^{n_j}\left(\frac{x_{ij} - x_{0j}}{x_{0j}}\right)^2 \qquad (3)$$

being $\vec{x}$ the vector on the pressure variables . Additional information on calculation are reported in the Appendix (eq. A1). The final formulation of the index takes into account the need to design for a specific target population through the introduction of a parameter θ related to the composition of the sample in terms of sex (eq. 4):

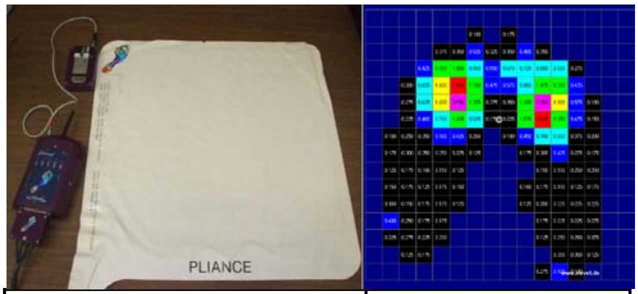
$$WPCL(\theta) = \theta WPCL_f + (1-\theta) WPCL_m \qquad (4)$$

with:

- comfort loss function for the female population obtained by appropriately summing all the of female population.
- comfort loss function for the male population obtained by appropriately summing all the over the male population.
-

2.2 Laboratory and devices

The experiments were performed at the Department of Aerospace Engineering (DIAS) of the University of Naples Federico II. A room, suitably cleared of furnishings, was chosen as scenario for the experiments. In order to collect data on pressure distribution impressed by participants on the seats the Novel Pliance mat by Novel was used fig.1. The mat is made of flexible material, characterized by 16×16 sensors uniformly distributed on its surface. The sensors send the sampled electric signals to the pliance box for converting them into digital data. Then, a dedicated software processes the data and displays them on the screen as a pressure map (fig.1). The map is a scheme of the mat; it is a matrix of 256 cells (24,5 mm x 24,5 mm) respectively corresponding to the 16×16 sensors. Each cell is characterized by a number (pressure value in kPa) and a colour (pressure range).



Number of sensors	16x16 (256)
Data transmission	Bluetooth
Pressure	kPa
Sensors	capacitive

Fig. 1 Equipment and related output.

Thanks to its flexible structure the mat is a minimally invasive instrument, which does not interfere with user perception of seat comfort. Several examples of application involving these devices in comfort assessment are reported in [12][32].

3 Experimental setup design

The experiments were carefully planned to reduce noise in the evaluation of the comfort of sitting [26]. In particular, five office chairs were tested, assessing the effect of the design parameter softness on perceived comfort. Tested chairs, have a five-point base, a backrest and armrests and they differ from each other for shape and materials. The chairs are named with fantasy names (tab. 1) so as to avoid any conditioning of the brand name or the model name on the evaluation.

Denomination	Chairs
OC	Oslo Chair
MC	Madrid Chair
CC	Chicago Chair
TC	Tourin Chair

Tab. 1 Tested chairs.

3.1 Definition of the control factor

The characteristic softness (S) was considered as a qualitative ordinal variable with four levels (from 0 to 3), in order of decreasing rigidity of the seat. In particular, each seat was representative of this control factor's level (tab. 2)

Control Factor				
Softness (S)	0	1	2	3
Chair	OC	MC	CC	TC

Tab. 2 Control Factors.

3.2 Definition of the noise factor

The noise factor taken into account was the anthropometric variability of experimenters (weight) stratified by sex. The random variable (r.v.) weight of the Italian female population and the r.v. weight of the Italian male population are both normally distributed, with parameters [27] reported in tab.3.

Variable	pdf	μ (kg)	σ (kg)

Female weight	Normal	58	9,48
Male Weight	Normal	75	10,05

Tab. 3 Parameters of the Normal r.v. weight (kg) for Italian females and males.

The r.v. weight of the whole Italian population can be modeled as a mixture of two normal distributions, whose probability density function (pdf) is [28]:

$$f(\theta) = \theta \cdot f_f + (1-\theta) \cdot f_m \tag{5}$$

where:

- θ is the mix coefficient representative of the proportion of females in the target population;
- f_f is the pdf of the r.v. weight of females;
- f_m is the pdf of the r.v. weight of males;

3.3 Experimenters

The experimental phase involved 22 experimenters, including 8 females (F) and 14 males (M). Anthropometric data collected from the experimenters included stature and weight. Statistics regarding these variables are reported in tab. 3.

Sex	N		Mean	St. Dev	Min	Max
F	8	stature	164,3	7,5	153,0	178,0
		weight	67,2	13,3	52,8	96,1
M	14	stature	181,6	8,3	170,0	198,0
		weight	79,4	9,3	64,4	93,0

Tab. 4 Anthropometric characteristics of experimenters.

The experimental sample is representative of the reference populations reported in tab. 3. Indeed, the sub-sample consisting of only women, covers the range from 29th to 99th percentile of the female weight distribution (μ=58; σ=9,48), while the sub-sample of the men covers the range from the 14th to the 96th percentile of the male weight distribution (μ=75; σ=10,05). Further details on experimenters, tested chairs and experimental setup are in [26].

3.4 Experimental protocol

More specifically, experimenters tested the seats in four short-term static experimental sessions. During the test, they were asked to read a text on VDT. According to [29], who demonstrates the invariance of global comfort rating over time, the duration of each experimental session was 5 minutes. In order to avoid the noise due to the sequence of the tested seats, the order of the test was randomized for each experimenter. Furthermore, all experimenters were blindfolded before and after each experimental session, to avoid that visual impact with the chair could affect their comfort perceptions [10].

4 Definition of the objective and subjective measurement methods

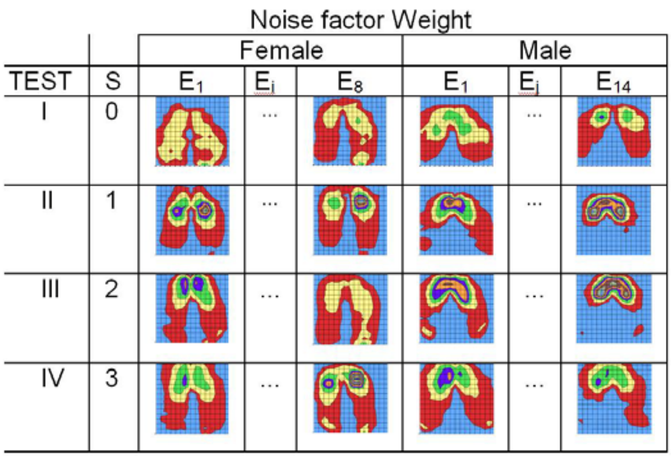
During the experimental session, for each experimenter, two types of data were recorded for each chair: objective data, obtained from pressure maps and subjective data, collected by questionnaires (tab 5). Once design factor, noise factor and responses are defined, the classical cross array showed in tab. 6 was used to plan the experiments.

4.1 Objective measures

With reference to objective data, obtained from pressure maps, many parameters were recorded: the maximum pressure (peak pressure) and the minimum pressure for each map, the sum of pressure values over all activated cells (overall pressure) and the mean of pressure values over all activated cells (mean pressure). Moreover, the total area (map area) and the weight on the mat (download weight), were measured. Finally, known the pressures of individual cells, it was possible to calculate the index PCL for each user and for each seat, using the equation 3.

Type	Label	Source
Objective	Peak pressure [N/]	Pressure maps
	Min pressure [N/]	Pressure maps
	Overall pressure [N/]	Pressure maps
	Mean pressure [N/]	Pressure maps
	Maps area []	Pressure maps
	Download weight [N]	Pressure maps
	PCL	Calculated from pressure data
Subjective	Comfort rating	Questionnaire
	Comfort ranking	Questionnaire
	Comfort degree	Questionnaire

Tab. 5 Typology and sources of recorded data.



Tab. 6 Cross array.

4.2 Subjective measures

After the test, each user expressed his/her subjective perception of comfort using three evaluation scales (rating, ranking, comfort degree) and the data were collected by questionnaires. For the rating evaluation the Borg CR10 scale [30],[31] modified by Kyung et al [32] was used. Rating scores ranged from 0 (no comfort) to 10 (extreme comfort). Every experimenter gave also a ranking of the chairs based on the perceived seat comfort. Finally, the third scale measured the user agreement with the statement "the seat is comfortable" using a four-point semantic scale : "I do not agree at all" (NA), "I scarcely agree" (SA), "I fairly agree" (FA), "I absolutely agree"(AA).

5 Comfort indexes definition and validation

The last step of the presented validation framework was the identification of good objective predictors for perceived comfort. From a statistical standpoint the nature of dependencies between perceived seat comfort and seat pressure variables, collected in the experimental phase, was analysed through a logistic regression model. More specifically, in order to identify a robust response function to use in the regression model, an association analysis was performed on the three evaluation scales. Then an ordinal logistic regression was performed to detect the significant dependencies, if any, of perceived comfort from anthropometric variables (*i.e. sex, weight, stature*) and pressure variables (full model). Finally, starting from parameters that were significant in the full model, a new ordinal logistic regression model was re- fitted to deepen the nature of dependencies previously identified.

5.1 Choice of a robust evaluation scale for perceived comfort

Few studies in literature have dealt with the validation of subjective scales for comfort assessment, although this aspect strongly affects the achieved results. In order to verify the consistency of the subjective data collected, the three evaluation scales adopted to collect the perceived comfort judgment were analyzed to verify their level of association. All three adopted scales are ordinal and polytomous. According to [34] the Goodman and Kruskall's index was applied to all possible combinations of binary association:

$$\gamma = \frac{(S - D)}{(S + D)} \tag{6}$$

where:

- *S* is the total number of pairs of responses on different evaluation scales which verify the condition i>i' and j>j' or both i<i' and j<j'
- *D* is the total number of pairs of responses on different evaluation scales which verify the condition i>i' and j<j' or both i<i' and j>j'

Results obtained, summarized in tab. 7, show a substantial consistency of the three scale. The minimum value calculated (between ranking and rating, equal to 0,653) reveals, however, a medium-high level of association between the scales. It is evident that the responses given on the scale "comfort degree" were highly associated with the other ones. So the comfort degree was selected as a good proxy of perceived comfort and set as response function in the adopted logistic regression model.

	Comfort degree	**Rating**	**Ranking**
Comfort degree	***1,000***	***0,984***	***0,860***
Rating	***0,984***	1,000	0,653
Ranking	***0,860***	0,653	1,000

Tab. 7 Results for the association analysis on the evaluation scales.

5.2 Logistic Regression model

According to both experimental data and results achieved in previous phases of the validation procedure, the full model of logistic regression was built. This model included all variables that were assumed explicative for the response function "comfort degree". Comfort degree was an ordinal response function with four ordered levels : "I do not agree at all " (NA), "I scarcely agree "(SA) " I fairly agree "(FA)" I absolutely agree "(AA). The list and classification of variables in the full model is reported in tab. 8: Quantitative variables are described in par. 4.2. Qualitative variables of the model were:

- Sex, that is a dichotomous variable (0=female, 1=male)
- Softness is a polytomous variable with four modalities (0, 1, 2, 3).

Type	Name
Quantitative	Peak pressure [N/]
	Mean pressure [N/]
	Maps area []
	Download weight [N]
	PCL
	Rate stature/weight of users
Qualitative	Sex
	Softness

Tab. 8 Full-model variables.

The baseline logit model [35] was used to identify significant relationships between the response comfort degree and the explicative variables in tab. 5. The generalized linear predictor equation was:

$$\hat{g}(x)_k = \beta_{0k} + x_i'\beta_k \tag{7}$$

where:

- $\hat{g}(x)_k$ is the generalized linear predictor with K=4 (index of the logits);
- x_i' are all model variables reported in table 5;
- are the parameters of the model.

The significance of all parameters was tested by using a stepwise backward elimination algorithm, that verified the null hypothesis that the model parameters are equal to 0. The results showed that the null hypothesis should be rejected with the conclusion that at least two parameters were significant in the model (PCL, peak pressure). Based on these results, the model could be re-fit. Then the ordinal logistic regression model (OLR) was applied [35][36] by using the comfort degree as a response function and peak and PCL as model variables. Based on the proportional odds approach, the model compares, for each ordinal level of the response function, the probability of an equal or smaller response function Y≤ k, with the probability of a larger response Y>k. The model output is reported in tab. 9. The results indicate that peak pressure significantly affects perceived comfort.

Pred	Coeff	SE Coeff	*z*-val	*p*-val	OR	95 %CI Lower	95 % CI Upper
Const	-6,49	1,29	-5,02	0,00			
Const	-3,83	0,79	-4,85	0,00			
Const	-0,05	0,66	-0,07	0,94			
Peak	2,93	0,79	3,71	0,00	18,79	3,99	88,46
PCL	-0,02	0,01	-1,68	0,09	0,98	0,95	1,00

Tab. 9 Ordinal logistic regression table.

The positive coefficient of 2,93 for peak is the estimated change in the logit of the cumulative comfort degree probability when a set of levels is compared with the others covariates, whereas PCL held constant. Because the *p*-value for estimated coefficient is close to 0, there is evidence to conclude that peak has a significant effect upon comfort degree. The odds ratio value is greater than one (18,79), this indicates that high peak pressures values tend to be associated with low values of comfort degree. The p-value indicates that there is no evidence to conclude that the PCL affects the comfort degree. The value of the odds ratio is approximately equal to 1, this indicates the independence between PCL and comfort degree.

6 Experimental results elaboration

The last step of the proposed framework was aimed at the validation of the index from an engineering point of view. Mean values of peak and WPCL for the four chairs were compared to verify the consistency of information provided by these indexes. Furthermore, the analysis of the pressure maps related to the worst values of peak and WPCL, allowed the identification of chair characteristics which were critical to improve seat comfort. Given the value of k (see Appendix), it is possible to calculate the index WPCL from PCL for a mixed population. For the analyzed sample, it was θ =0.36 and (1-θ) = 0.64 (36% females and 64% males). The results, assuming WPCL as a response function, are shown in tab. 10, for female, mixed and male population.

		Response		
		F	Mix	M
TEST	S		WPCL	
I	0	0,74	0,987	1,125
II	1	0,699	0,949	1,09
III	2	0,395	0,609	0,729
IV	3	0,213	0,342	0,415

Tab. 10 Results from using WPCL as a response function.

Level 3, corresponding to the highest level of cushion softness, was the best one in terms of WPCL, whereas levels 0 and 1 got the worst results, with comparable values of WPCL (fig.2).

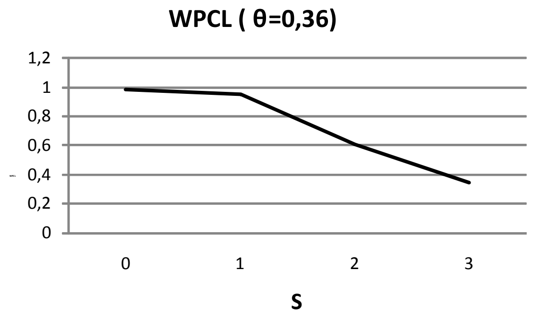


Fig. 2 Mean effects assuming WPCL as response function for a mixed population.

Level 3 seems to be also the most robust one against changes in the composition of the reference sample (fig.3). A minor change in the slope of mean effects diagram, in fact, indicates a minor change in WPCL index over different composition of the sample. The mean effects diagrams for the other levels highlight slightly higher slopes. However, whatever is the softness level, the index WPCL is greater for males than females, since it is influenced by the distribution of body weight.

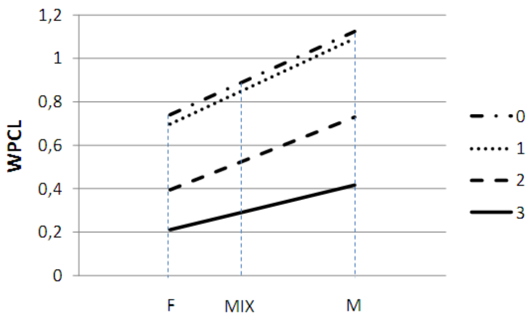


Fig. 3 WPCL Index for different sample compositions.

		Response		
		F	Mix	M
TEST	S		peak	
I	0	1,425	2,284	2,767
II	1	1,488	1,936	2,189
III	2	0,908	1,112	1,227
IV	3	0,688	0,946	1,092

Tab. 11 Results from using peak pressure as a response function.

The same analysis was carried out, assuming the peak pressure as a response function. As shown in fig. 4, the lowest values of peak pressure were recorded for level 3. Level 2 got comparable performance, whereas level 0 and 1 once again resulted to be the worst ones.

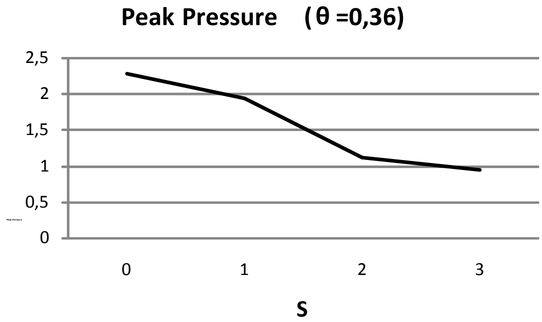


Fig. 4 Mean effects assuming peak pressure as response function for a mixed population.

The diagrams of pressure peaks for different sample compositions (Fig. 5) confirmed that level 3 is the best one, since it presents the lowest peak pressure values for any mix of the population. However, it is evident that, in this case, level 2 is more robust against the anthropometric variability induced by sex, as evidenced by the lower slope of the mean effects diagram; once again, levels 0 and 1 got the worst performance. Assuming that the sample were composed exclusively of women (θ = 0), level 0 would be better than level 1. However, level 0 seems to be less robust against anthropometric variability induced by the composition of the sample, as the highest slope of its main effects diagram highlights.

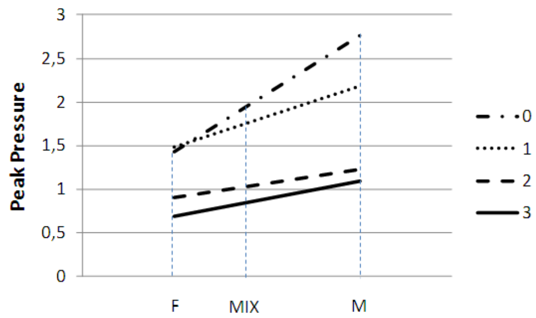


Fig. 5 Peak pressure for different sample compositions.

The ranking of chairs shows substantial coherence of the results provided by peak pressure and PCL.

With regard to level 3 (*i.e.* seat TC), the minimum values of these indexes are related, for each sub-sample, to the same pressure map and thus identify the same experimenter (fig.6).

This coherency in results does not mean that peak pressure and PCL provide the same information.

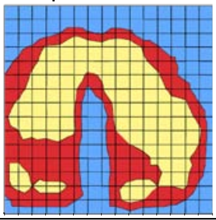
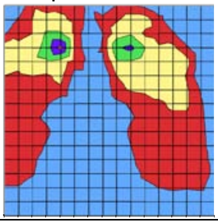
	Female	Male
Min PCL and peak pressure Map	Experimenter 1 	Experimenter 18 

Fig. 6 Pressure maps related to minimum values of PCL and peak pressure for the seat TC.

For instance, fig. 7 show the pressure maps related to the maximum values of peak pressure and PCL for level 0 (*i.e.* seat OC), which resulted to be the worst one in terms of perceived comfort.

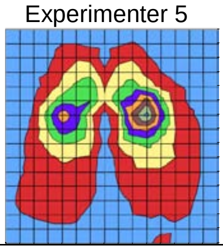
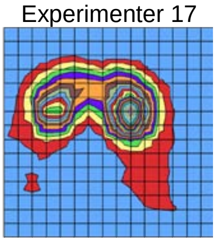
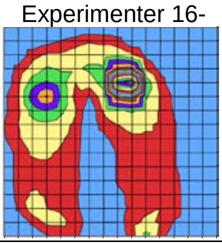
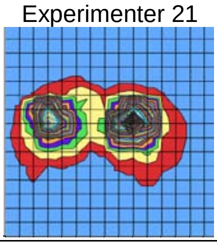
	Female	Male
Maximum PCL Maps	Experimenter 5 	Experimenter 17 
Maximum Peak pressure maps	Experimenter 16- 	Experimenter 21 

Fig. 7 Pressure maps related to maximum values of PCL and peak pressure for the seat OC.

The joint analysis of these indexes allows to obtain important information for the improvement of the seat.

Based on selected maps it is possible to highlight main issues in improving the design of tested chairs. By integrating the information provided on sensitive areas by maximum peak pressure and PCL, it is possible to improve the seat in terms of comfort loss. More specifically, it is important to identify and analyze the most stressed areas, in order to reduce load on bony prominences of the pelvis, taking into account anatomical differences related to the sex of the experimenter [37]. As shown in fig. 6, in fact, the pelvis of women are developed more in width, while in men the sacral and iliac bone is thicker and heavier, generating localized peaks of greater magnitude. The analysis of pressure maps stratified by sex help to take into account variability and redesign the seat's shape and materials. To mitigate the peak loads at the ischial tuberosities, for different anthropometric percentiles, an insertion of material could be expected (*e.g.*, polyurethane foam of assigned density) to reduce significantly the discomfort caused by body compression on the seat.

7 Conclusions

The purpose of this work was the validation of an index for seat comfort assessment, which could be a valuable support in the design phase. More specifically, the WPCL index proposed in a previous work, was compared with both objective and subjective parameters obtained in experimental tests planned to compare office chairs.

From the statistical standpoint, relationships between perceived comfort and objective parameters were investigated through a logistic regression model, assuming as a response function the subjective measure of users' comfort perception (comfort degree). Among others objective measures, OLR identifies peak pressure and PCL as the two parameters that are significantly associated to perceived comfort. The results revealed that comfort degree strongly depends on peak pressure, whereas there is no statistical evidence of dependence on WPCL. The assumption that the high pressure values are predictors of comfort is unsatisfactory. In fact, the peak pressure can be a useful parameter for the designer, only if integrated by information about the position of the peak itself [8].

On the other hand, the failure to identify significant correlation between WPCL index and comfort degree, must be deepened. It could be that subjective evaluation in a short-time session is more related to instantaneous stimuli like the peak pressure. This means that the opinions of users may be misleading and therefore not suitable in an analysis like the one proposed in this paper. Further investigation will concern the following critical issues:

- a refinement of the index so as to take into account variations between neighbouring cells of a pressure map instead of single values;
- An in-depth study of the most significant anthropometric variables is necessary in order to improve the robustness of the seats over different types of users (design for all).
- From an engineering standpoint, the index WPCL and the peak pressure, got consistent results with regard to softness, providing not redundant information that could help designers to improve chair design, taking into account different sensitive areas of the seat.

Acknowledgement

The present work was developed with the contribution of the Italian Ministry of education, University and research (MIUR) within the framework of the PRIN 2008 project "Innovation in service quality management: statistical approach and application in some fields of national interest".

The author thank also "Polo ST UniNa" that partially funded the activity within the framework of "Faro Project".

Finally, the authors deeply thank prof. Monica Carfagni, Director of the Department of Mechanics and Industrial Technologies of the University of Florence, engg. Lapo Governi, Yari Volpe and Domenico Maria Del Giudice for his precious technical support.

Appendix: definition

The calculation of was made on the basis of pressure maps data, assuming that the maximum value of the ratio, expressed in formula (eq. 7) was the maximum tolerable by the user. More specific only maps which had a comfort degree score equal to 4 (completely comfortable) were selected. Identified the maximum of this ratio, the value of (one for all the maps) was calculated as its reciprocal. More specifically, the resulting value was equal to 0.10.

$$\frac{PCL_j(\bar{x})}{k_j} = \sum_{i=1}^{n_j} \left(\frac{x_{ij} - x_{0j}}{x_{0j}} \right)^2 \qquad \text{(A1)}$$

References

[1] P. Vink. Comfort and design: principles and good practice. CRC Press 2005.

[2] M. G. Helander, L. Z. Zhang. *Field studies of comfort and discomfort in sitting*. Ergonomics 40, 9 (1997) pp 895-915.

[3] LL. Z. Zhang, M. G. Helander, Drury C. G. *Identifying factors of comfort and discomfort in sitting*. Human Factors 38, 3 (1996) pp 377-389

[4] M. De Looze, L. Kuijt Evers, J. Van Dieen. *Sitting comfort and discomfort and the relationships with objective measures.* Ergonomics 46, 10 (2003) pp 985-997.

[5] Z. Q. Zeng, Q. Wu, C. Yang, K. S. Wu. *An Evolution Method of Driving Seat Comfort Based on Least Squares Support Vector Regression.* Proceedings of Fifth International Conference on Natural Computation, August – , 2009, Tianjin, pp 455-459.

[6] H. Fazlollahtabar. A subjective framework for seat comfort based on a heuristic multi criteria decision making technique and anthropometry. Applied Ergonomics 42, 1 (2010) pp 16-28.

[7] S. Na, S. Lim, H. Choi, M. Chung. Evaluation of driver's discomfort and postural change using dynamic body pressure distribution. International Journal of Industrial Ergonomics 35,12 (2005) pp.1085-1096.

[8] G. Kyung, M. A. Nussbaum. Driver sitting comfort and discomfort (part II): Relationships with an prediction from interface pressure. International Journal of Industrial Ergonomics 38, 5-6 (2008) pp 526-538.

[9] T. K. Fredericks, S. E. Butts. Objectively determining comfortable lumbar support in task seating (Tech. Report). Holland, MI: Haworth. *.haworth.* (2006).

[10] B. Shackel, K. D. Chidsey, P. Shipley. *The assessment of chair comfort.* Ergonomics 20, 1 (1969) pp 269-306.

[11] K. Reijneveld, M. de Looze, F. Krause, P. Desmet. *Measuring the Emotions Elicited by Office Chairs.* Proceedings of the International conference on Designing pleasurable products and interfaces, June – , 2003, Pittsburgh, pp 6-10.

[12] G. Andreoni, G. C. Santambrogio, M. Rabuffetti, A. Pedotti. *Method for the analysis of posture and interface pressure of car drivers.* Applied Ergonomics 33, 6 (2002) pp 511-522.

[13] R. K. Lueder. Seat comfort: A review of the Construct in the Office Environment. Human Factors 25, 6 (1986) pp 701-711.

[14] H. Dreyfuss. The measure of man and woman. Wiley 2002.

[15] S. Barone, A. Lanzotti. (2009) *Robust Ergonomic Virtual Design*. In: Statistics for Innovation: statistical design of "continuous" product innovation. Springer 2009, pp 61-83.

[16] C. F. Tan, F. Delbressine, W. Chen, M. Rauterberg. *Subjective and Objective Measurements for Comfortable Truck Driver's Seat*. 9th International Symposium on Advanced Vehicle Control, October – , 2008, Kobe, pp 851-856.

[17] L. F. M. Kuijt- Evers, T. Bosch, M. A. Huysmans, M. P. de Looze, P. Vink. *Association between objective and subjective measurements of comfort and discomfort in hand tools.* Applied Ergonomics 34, 3 (2003) pp 207-214.

[18] M. Stinson, S. Crawford. *Optimal Positioning: Wheelchair Seating Comfort and Pressure Mapping.* In International Handbook of Occupational Therapy Interventions, Springer 2009, pp 83-90.

[19] M. Reed, C. Grant. Development of a measurement protocol and analysis techniques for assessment of body pressure distributions on office chairs, TR. Ann Arbor, MI: Univ. of Michigan, Center for Ergonomics(1993).

[20] M. Carfagni, L. Governi, Y. Volpe. *Comfort assessment of motorcycle saddles: a methodology based on virtual prototypes*. International Journal on Interactive Design and Manufacturing (IJIDeM) 1,3 (2007) pp 155-167.

[21] I. Hostens, G. Papaioannou, A. Spaepen, H. Ramon. *Buttock and back pressure distribution tests on seats of mobile agricultural machinery*. Applied Ergonomics 32 (2002) pp 347-355

[22] M. Nagamachi. Perspectives and new trend of Kansei/Affective Engineering. The TQM Journal 20, 4 (2008) pp 290-298.

[23] K. Kamijo, H. Tsujimara, H. Obara, M. Katsumatu. *Evaluation of seating comfort*, SAE Technical Paper Series 820761. Society of Automotive Engineers, Troy, MI, (1982) pp 1–6

[24] M. Y. Park, J. Y. Kim, J. H. Shin. *Ergonomic design and evaluation of a new VDT workstation chair with keyboard-mouse support.* International Journal of Industrial Ergonomics 26, 5 (2000) pp 537-548

[25] G. Fujimaki, R. Mitsuya. *Study of the seated posture for VDT work.* Displays, 23, 1-2 (2002) pp 17-24.

[26] A. Lanzotti, G. Matrone, A. Vanacore. *Seat design improvement via a new postural comfort index.* Proceedings of ENBIS9 Conference, september – , 2009, Goteborg.

[27] A. Lanzotti, G. Matrone, A. Vanacore. *Robust Ergonomic Design (RED): towards a new index for the evaluation of the seat comfort* Proceedings of the Congreso Internacional Conjunto XXI INGEGRAF - XVII ADMME, June – , 2009, Lugo.

[28] A. Lanzotti, A. Vanacore. An Efficient and Easy Discretizing Method for the Treatment of Noise Factors in Robust Design. The Asian Journal on Quality 8, 3 (2008) pp188-197.

[29] R. A. Wachsler, D. B. Learner. *An analysis of some factors influencing seat comfort*. Ergonomics 3, 4 (1960) pp 315-320

[30] W. Shen, K. C. Parsons. *Validity and reliability of rating scales for seated pressure discomfort*. International Journal of Industrial Ergonomics 20, 6 (1997) pp 441-461.

[31] M. Helander. A guide to Human Factors and Ergonomics. CRC Press 2006.

[32] G. Kyung, M.A.Nussbaum, K.B. Reeves. Driver sitting comfort and discomfort (part I): Use of subjective ratings in discriminating car seats and correspondence among ratings. International Journal of Industrial Ergonomics 38, 5-6 (2008) pp 516-525.

[33] M. Kolich. A conceptual framework proposed to formalize the scientific investigation of automobile seat comfort. Applied Ergonomics 39, 1 (2008) pp 15-27.

[34] L. A. Goodman, W. H. Kruskal. *Measures of Association for Cross Classifications. II: Further Discussion and References.* Journal of the American Statistical Association 54, 285 (1959) pp 123-163.

[35] D. W. Hosmer, S. Lemeshow. Applied Logistic Regression, 2nd Edition. Wiley, 2000.

[36] C. Lawson, D. C. Montgomery. *Logistic Regression Analysis of Customer Satisfaction Data.* Quality and Reliability Engineering International 22,8 (2006) pp 971-984.

[37] G. Lambertini. Human Anatomy (in italian). Piccin editore 1977.

Paper C

P. Franciosa , S. Gerbino, A. Lanzotti, S. Patalano, V. Rufrano, M. Trotta, L. Silvestri , L. Frattari Improving comfort of occupational footwears through experiments on virtual prototypes, Proceedings of IMPROVE 2011, Venezia, 15 – 17 june 2011, ISBN 978- 88- 7784- 328- 9, pp 8

Proceedings of the IMProVe
International conference on Innovative Methods in Product Design
June 15th – 17th, 2011, Venice, Italy
ADM – INGEGRAF

Improving comfort of occupational footwears through experiments on virtual prototypes

Pasquale Franciosa [a], Salvatore Gerbino [a], Antonio Lanzotti [b], Stanislao Patalano [b], Vito Rufrano [b], Mariangela Trotta [b], Luca Silvestri [c], Luca Frattari [c]

[a] University of Molise, School of Engineering
[b] University of Naples, Federico II, School of Engineering
[c] SAFE WAY s.r.l

Article Information

Keywords:
occupational footwear,
comfort assessment,
numerical-physical correlation,
virtual prototyping,
pressure map

Corresponding author:
Salvatore Gerbino
Tel.: +39 0874 404593
Fax.: +39 0874 404978
e-mail: salvatore.gerbino@unimol.it
Address: Via Duca degli Abruzzi - 86039 Termoli (CB) - Italy

Abstract

Purpose:
The present paper focuses on the parametric analysis of the sole of occupational footwear in order to improve the perceived human comfort. By combining real experimental tests and virtual simulations, the sensitivity of both geometric and material design factors, on comfort degree, was investigated.

Method:
The correlation among perceived human comfort and physical parameters, such as plantar pressures, was estimated by conducting real tests. Experimenters were asked to wear four commercial shoes and to express their perceived comfort degree. By adopting plantar sensors, plantar pressures were also monitored. Once given such a correlation, a parametric FE model of the footwear was developed. In order to better simulate the contact at plantar surface, a detailed FE model of the foot was also generated starting from CT scan images.
A fractional factorial design array was, finally, used to study the sensitivity of different sets of design factors on comfort degree. In the present study only a static standing-up configuration was analyzed.

Result:
Findings of this research showed that sole thickness and its material highly influence perceived comfort. In particular, softer materials and thicker sole designs contribute to increase comfort degree.

Discussion & Conclusion:
Despite all simplifications and limitations, the proposed methodology may be successfully adopted in other industrial applications, in which the design (or re-design) of new products is driven by the satisfaction or the sensations of users.

1 Introduction

Comfort assessment is a crucial task in product design. This is especially true for certain categories of products characterized by repeated and prolonged usage such footwears. It was reported that the perceived human comfort is strongly related to the footwear design, in terms, for example, of adopted materials, insole and outsole thickness and shape [1].

In this contest, the opinions of users may provide valuable information whether or not a shoe is comfortable. However, this information is often limited to qualitative descriptions, which cannot quantify causes of comfort or discomfort. Therefore, in order to "quantify" what may influence comfort and discomfort, the relationship between the human perceived parameters and measuring parameters should be determined [2, 3].

Over last two decades, researchers, especially in the medical and bio-mechanical fields, have addressed their attention on comfort issues. Some studies were mainly based on questionnaires as an indication of user preferences [4, 5]. However, very few researches have focused on the evaluation of the analytical correlation between subjective and objective parameters.

In [6, 7] it was reported that physiological factors, such as plantar pressures, are strongly related to physical parameters such as materials and plantar shape.

A first valid scientific contribution to the analysis of correlation was offered by Jordan *et al.* [8]. They attempted to correlate the subjective perceptions of users with dorsal and plantar pressure distribution through short-term dynamic tests. Perceived comfort was measured by using specific questionnaires, while pressure distributions were monitored through high resolution insole sensors. The correlation analysis was based on the results coming from three different shoes. The study showed a negative correlation between pressures and subjective comfort perception (meaning that a high peak pressure corresponds to a low perceived comfort). Moreover, authors highlighted the need to investigate further other objective parameters that may affect the user perception (see, for example, shear and normal forces, and heat transfer).

Witana *et al*. [9] tried to identify the interactions between comfort and plantar shape. They found substantial differences between the subjective perceptions of users related to the mid-foot for different tested materials, thus confirming that comfort perceptions, for different areas of the plantar foot, are quite different.

If on one hand experimental tests, carried out on different product designs, give valuable results, on the other hand, the large number of design parameters would make extremely difficult and expensive to identify the optimal design through tests with real prototypes. In this sense, using virtual simulations and parametric models may be a valid support.

Recently, in order to give a valuable support to experimental investigations, computational methods, based on FE modeling, have been adopted. FE models of human foot have been developed under certain simplifications and assumptions [10-13] such as: (i) simplified or partial foot shape, (ii) assumptions of non-linear hyper-elastic material law, (iii) ligaments and plantar fascia modeled as equivalent forces or elastic beams/bars, (iiii) no friction or thermal effect, at plantar foot interface, accounted.

In this contest, Cheung and Zhang [14] combined FEM and Taguchi methods to identify the sensitivity of five design factors (arch type, insole and mid-sole thickness, insole and mid-sole stiffness) of footwear on peak plantar pressure. From FEM predictions, the most important design factors, able to reduce the peak plantar pressure, were found-out.

Starting from the literature review, the present paper focuses on the parametric analysis of the sole of occupational footwear in order to figure out which parameters influence human comfort. In this term, the analysis allows to define the best design of that sole in order to maximize comfort.

Subjective perceived comfort and plantar pressure maps were correlated by using experimental tests on different commercial shoes. Then, a comfort function was estimated. Knowing the comfort function the Taguchi's method was adopted to study the influence of different design settings. A FE model was adopted for this purpose. Finally, by statistically analyzing simulated plantar pressure maps, the most influencing design factors were identified.

2 Methodological overview

Fig. 1 depicts the general methodology adopted in the present paper.

First of all, how to correlate subjective perceived comfort to physical and measurable variables, such as contact pressure maps, was investigated. To do this, 23 users were involved in short-term static tests and for each of them four different footwears were worn. During the experiments, plantar pressure maps were recorded by means of high resolution insole sensors. Then, subjective ratings, related to perceived comfort, were collected by means of questionnaires.

By comparing the perceived comfort rate to the measured pressure maps, a comfort function, depending on the peak pressure, was established.

Once the comfort function was estimated, Taguchi method was used to study the sensitivity of different design settings on plantar peak pressure and then on the comfort rate, with respect to the sole of the occupational footwear, being optimized. A 3D FE model of the footwear was developed for this purpose. In order to simulate as much as possible the contact among the plantar surface and the foot, an anatomical detailed FE human foot model - with soft tissue, bones and cartilages - was created from CT scans.

In accordance with a fractional factorial design, virtual prototypes were developed, selecting combinations of design factors (materials and geometry shape parameters). By statistically analyzing plantar pressure maps, the most influencing design factors were identified.

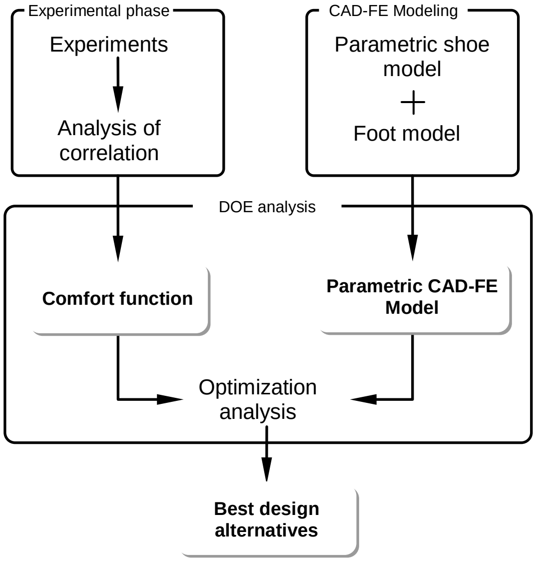
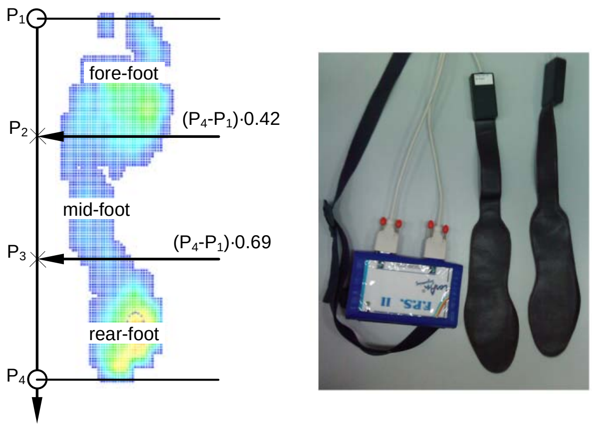


Fig. 1 General work-flow methodology

The following hypotheses were formulated:

- only footwear sole accounted: vamp and upper-sole were not modeled. This means that the interaction effect among dorsal/lateral and plantar pressure was neglected;
- no thermal effect considered: gradient of temperature may influence perceived comfort, but it is here neglected as users have worn shoes for a short time; and,
- short-term static tests considered: when running or walking, perceived human comfort may be influenced by temperature, humidity and interaction among dorsal and plantar pressures. In the present research only a static balanced standing-up configuration was modeled.

The plantar foot was subdivided into three zones, as also suggested in [15, 16]. Fig. 2.a depicts the proposed foot division (three areas are shortly identified: rear-foot, mid-foot and fore-foot).



a - foot sub-division *b - insole sensors*

Fig. 2 Foot sub-division and insole sensors

Plantar pressure maps were recorded by using a high resolution plantar sensor. The adopted equipment device (sensor size: 0.5x0.7 cm - number of sensors: 512 - insole size: 39-41 European - producer: Loran Eng. - Italy) is shown into fig. 2.b. The insole sensor provides a limit set of values: that is, pressures are read on 512x4 points (every sensor provides four pressure values).

3 Experimental phase

The experimental phase deals with the measurement of pressure map distributions in order to carry-out the correlation among subjective and physical variables.

According to [8], the following physical (or objective) parameters were accounted (each of them is related to sub-areas of the foot):

- contact area: number of active sensor points (a sensor point is assumed "active" if the related pressure is not zero);
- peak plantar pressure: maximum pressure value; and,
- weight distribution: net force, calculated as resultant pressure over contact area. The weight distribution may be considered as percentage of body-weight transmitted to the insole for each foot zone.

Whereas, the considered subjective parameters were:

- global comfort degree; and,
- local comfort degree, related to every foot area.

For each experimenter four shoe-configurations were tested (named "A", "B", "C" and "D"). Configuration D corresponds to a bare-foot test.

3.1 Participants

The participants of the experimental session were selected among students of the School of Engineering at University of Naples, Federico II - Italy. Their ages ranged between 20 and 28 years, with 41 shoe size (European size).

a *b* *c*

Fig. 3 Testing procedure

After screening all those subjects with foot pathology and abnormal pressure maps, 23 experimenters (13 males and 10 females) were selected.

In order to avoid the influence of aesthetic qualities of the shoes (notice that it is usual to classify as "comfortable" a running shoe, while an occupational footwear is assumed "un-comfortable"), all experimenters conducted their test blindfolded.

Fig. 3 shows a typical testing procedure: experimenter is firstly blindfolded (fig. 3.a); then, she/he is aided to wear the shoes in which insoles have been previously inserted (fig. 3.b); finally, pressure maps are recorded.

When recording pressure maps, experimenters were asked to stand-up in a balanced configuration and not make sharp movement. Output pressure maps were stored for 10-15 seconds. The average map was then saved.

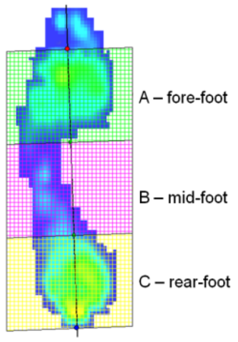
3.2 Experimental protocol

Participant's feet were cleaned with warm water and then dried. Physical information was also collected: height, weight, foot size.

The order of tests was randomly selected, to avoid the effect of noise factors connected to sequence in shoes wearing and experimental conditions [17].

The insole sensor device was re-calibrated before every test. Right and left feet pressures were measured to identify possible pathology unknown to the experimenter, before the test. Subjects with abnormal pathology were, then, discarded from the data analysis.

Output pressure maps were processed within an *ad-hoc* MatLAB® tool, allowing to calculate objective parameters: contact area, peak plantar pressure and weight distribution. The tool assures the repeatability of the subdivision of the foot areas. For each map two extreme points must be manually identified (point and , as depicted into fig. 2.a). Then, the tool automatically provided the breakdown points (and into fig, 2.a).



Perceived comfort	
0	*Absence of comfort*
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	*Extremely comfortable*

Fig. 4 Adopted questionnaire for comfort rating

After measuring pressure, in order to evaluate perceived comfort, a 10-level scale questionnaire (see fig. 4) was asked to be compiled (0 - absence of comfort; 10 - extremely comfortable). More specifically, for each test, experimenters expressed their rating preferences both for the whole plantar (global comfort degree) and for each plantar area (local comfort degree).

3.3 Data analysis

To assess the reliability of questionnaire results, all experimenters repeated unconsciously one of the four tests. The repeatability test (E) was conducted for the shoe-configuration "C". Based on the Mann-Whitney test [18], five experimenters were excluded from the data analysis (meaning that only 18 experimenters were finally available) since their rating degrees did not appear reliable.

It is of interest analyzing trends of both objective and subjective parameters with respect to shoe-configurations (in tab. 1 mean values, calculated over all experimenters, are reported). Analyzing peak pressure data (in KPa), one can note that configuration C exhibits the lowest peak pressure in fore-foot and rear-foot areas. As expected, in test D (bare-foot) the highest peak pressure was achieved in the rear-foot and fore-foot zones.

Objective parameters (contact area, peak pressure and weight distribution, related to fore-foot and rear-foot areas) and subjective parameters were correlated each-other. In particular, as contact area and weight distribution are dependent each-other (notice that the weight distribution is calculated averaging pressures over contact area), only peak pressure and weight distribution will be kept in the following. Tab. 2 shows results of such correlations: objective and subjective parameters are correlated to the global comfort degree. Correlation indexes (ρ) were calculated by using the Pearson's correlation coefficient [19].

Shoe Configuration	Fore-foot	Mid-foot	Rear-foot
Comfort degree			
A	5,94	6,61	6,50
B	4,89	6,50	6,83
C	6,33	6,56	7,22
D	6,00	6,28	6,28
Contact area			
A	779,72	278,22	548,72
B	793,11	316,78	552,22
C	766,06	355,11	562,61
D	712,44	207,78	490,17
Peak pressure			
A	39,78	28,94	48,17
B	41,72	39,28	46,72
C	40,56	33,78	44,44
D	43,33	24,44	52,28
Weight distribution			
A	57,05	5,62	37,33
B	53,74	8,38	37,87
C	50,19	9,81	40,00
D	56,32	3,92	39,77

Tab. 1 Experimental data

Parameter	Correlation coefficient (with respect to the global comfort degree)
Fore-foot comfort degree	0.87
Rear-foot comfort degree	**0.99**
Fore-foot peak pressure	-0.62
Rear-foot peak pressure	**-0.95**
Fore-foot weight distribution	**-0.96**
Rear-foot weight distribution	0.30

Tab. 2 Table of correlations

Correlation analysis between subjective scores showed a very high positive correlation for the fore-foot (ρ = 0.87) and rear-foot (ρ = 0.99) comfort degrees. This means that the perceived comfort in those areas strongly influences the global comfort perception.

Moreover, as expected, peak pressure has a negative impact on perceived comfort. More specifically, a negative high correlation appears both for the rear-foot peak pressure (ρ = -0.95) and for the fore-foot weight distribution (ρ = -0.96).

Finally, the data analysis showed the following key issues:

- perceived comfort in fore-foot and rear-foot areas highly influences the global comfort degree; and,
- high peak pressure values correspond to low perceived comfort degrees.

Based on these preliminary results, next Section will describe how to calculate the comfort function, depending on the objective parameters.

3.4 Comfort function estimation

Comfort function depends, as discussed above, on objective parameters. Let , , and be the peak pressures and the weight distributions, related to the fore-foot and the rear-foot areas. Moreover, let be the comfort degree. Tab. 3 reports normalized mean values of such parameters (peak pressures were normalized with respect to the maximum value).

	- [0-1]	- [0-1]	- [%]	- [%]	- [0-10]
A	0.340	0.412	57.05	37.33	6.33
B	0.346	0.379	50.70	38.95	6.83
C	0.341	0.374	50.19	40.00	6.89
D	0.361	0.435	56.32	39.77	6.17

Tab. 3 Normalized mean values

27.637	-34.923	0.147	0.079

Tab. 4 Comfort constants

Assuming a linear relationship among the comfort function, , and the objective parameters, one can write:

$$C_f = f(P_f, P_r, W_f, W_r) = \ldots$$
$$\ldots C_f = \alpha_1 \cdot P_f + \alpha_2 \cdot P_r + \alpha_3 \cdot W_f + \alpha_4 \cdot W_r \quad (1)$$

where , , and are unknown constants (here called "comfort constants"), which can be now evaluated by assigning all four constraints (derived from the experiments), stated into tab. 3. Tab. 4 reports the so-calculated constant values.

Once comfort constants are known the comfort function is completely defined. This means that for any set of objective parameters, the related comfort degree can be obtained, univocally.

Notice that the present approach may be easily extended when more than four shoe-configurations are available. In that case comfort constants should be evaluated by solving for a least squares problem (that is, the number of unknowns is less than the number of available equations).

The comfort function, , will be adopted in the next to perform the shoe design optimization.

4 CAD-FE Modeling

In order to quickly analyze different design settings, a parametric CAD model of the sole of the occupational footwear was created into SolidWorks® 2010 (by Dassault Systemes) CAD system (see fig. 5.a), according to the design constraints of SAFE WAY s.r.l company - Italy.

The sole is made of two sub-domains: inner-sole and outer-sole. The body weight is transmitted from the foot to the inner-sole and then to outer-sole, which comes into contact with the ground (not modeled here).

The mechanical behavior of the sole was captured through a non-linear incompressible hyper-elastic law, characterized by two material constants, and (Mooney-Rivlin formulation). These constants were extracted from stress-strain experimental curves (experimental tests were conducted at Dept. of Materials and Production Engineering, School of Engineering - Naples).

The physical interaction among the inner-sole and the outer-sole was modeled by defining identity pairs among interfacial surfaces. Identity pairs assures that the displacement fields of both parts at interfacial surfaces are identical each-other.

Since pressure maps are aimed to be calculated, a detailed foot model, previously developed into [20] starting from CT scans, was incorporated into the FE model (fig. 5.b). Contact pairs (see fig. 5.c) were introduced between plantar foot surfaces and the inner-sole. No friction was here accounted.

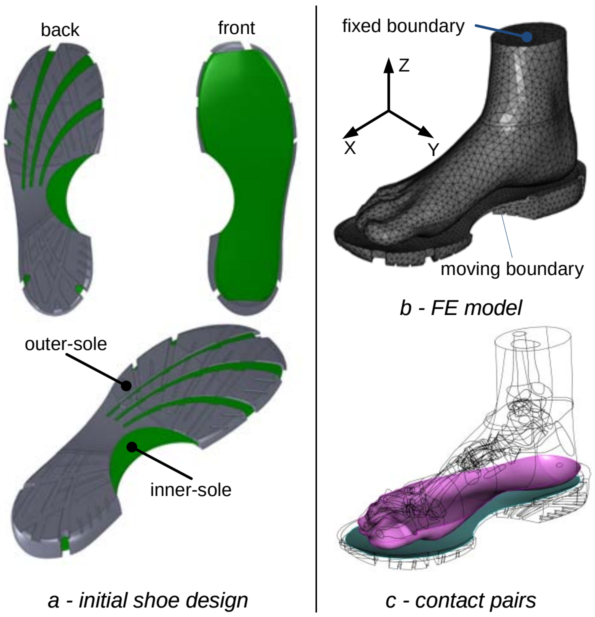


Fig. 5 CAD-FE model

Boundary conditions were applied as depicted into fig. 5.b. The upper surface of the ankle was supposed fixed. The lower boundary of the sole was moved along the Z direction. The maximum displacement was chosen so that the reaction force calculated at the fixed boundary was greater or equal to 650/2 N, that is half of the body weight.

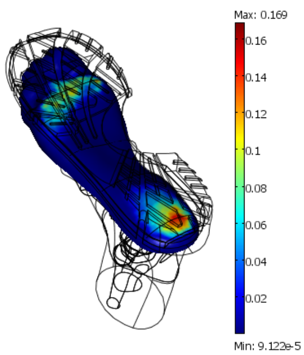


Fig. 6 Pressure distribution (MPa) for the initial shoe design

FEM simulation was performed within Comsol Multiphysics® 3.5a. As hyper-elastic materials and contact pairs were modeled, an iterative non-linear static solver was adopted. The simulation took about 100 min on a DELL Precision T7400 workstation (WinXP 64bit, 16GB RAM, 2 Xeon E5420 quad-core processors).

Fig 6 shows numerical FE results related to the initial shoe design. One should note that the highest peak pressures are located in the rear-foot and fore-foot areas, thus confirming experimental analyses.

5 DOE analysis

This Section discusses how to investigate the most influencing design parameters, based on a DOE (Design Of Experiments) approach.

Generally speaking, when facing out an optimization problem, it is asked to calculate the best set of design parameters, which optimizes (in terms of minimization or maximization) a given objective function.

In the present research, the objective function is the comfort function (see Section 3.4), to be maximized. This function depends on physical variables, such as contact pressure. However, the analytical relationship among design factors (see for example, sole materials, sole thickness) and physical variables is not know. By using a FEM solution, this relationship can be obtained for a given set of design factors.

Therefore, the proposed approach may be summarized as follows: (i) generate a set of combinations of design factors; (ii) calculate the comfort function for every combination by solving a FE model; and, then, (iii) analyze design scenarios, looking for the best combination of design factors.

Here, combinations of design factors were generated by using a factorial design approach.

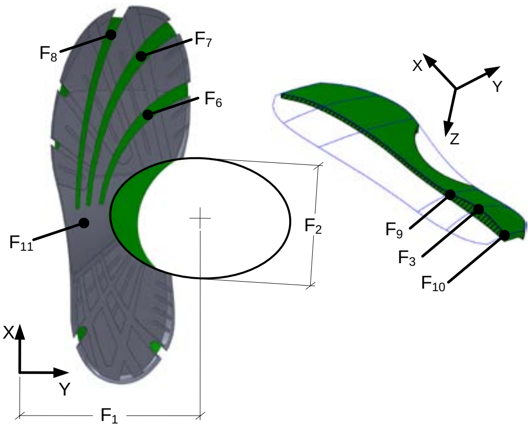


Fig. 7 Geometrical design factors

Factor ID	Level		
	129 mm	149 mm	169 mm
	93 mm	73 mm	53 mm
	2 mm	3 mm	4 mm
	=1.265	=0.969	=1.325
	=-0.416	=-0.314	=-0.314
	=0.408	=0.578	=0.158
	=-0.248	=-0.314	=-0.071
	ON	OFF	
	ON	OFF	
	ON	OFF	
	1 mm	2 mm	
	1 mm	2 mm	
	ON	OFF	

Tab. 5 Design factors and their levels for design optimization

5.1 Design factors

Looking at fig. 7, the following design factors were considered:

- arch shape: its elliptical shape was parameterized in terms of width () and centre position () with respect to the global coordinate frame;
- outer-sole cuttings (, ,);
- inner-sole thickness: parameterized through the Z coordinates of points , and ;
- outer-sole notching (); and,
- sole materials: outer-sole material () and inner-sole material ().

ID												- [0-1]	- [0-1]	- [%]	- [%]	- [0-10]
I	1	1	2	1	1	2	2	2	2	2	1	0.4086	0.5914	43.6398	56.3450	1.1525
II	1	1	2	3	2	1	1	1	1	1	2	0.4492	0.5508	45.2914	54.7064	3.2021
III	1	2	1	1	2	2	1	1	2	2	1	0.4703	0.5297	45.3102	54.6878	4.2259
IV	1	2	2	2	3	2	2	2	1	2	2	0.5409	0.4591	48.6694	51.3289	7.8142
V	1	3	1	1	1	1	1	2	1	1	2	0.4895	0.5105	46.5954	53.4030	5.2216
VI	1	3	3	2	3	1	2	1	2	1	1	0.5457	0.4543	48.5012	51.4942	8.0387
VII	2	1	3	1	3	2	2	1	1	1	2	0.5217	0.4783	47.6650	52.3303	6.8328
VIII	2	2	2	3	3	2	1	2	2	1	1	0.5880	0.4120	46.9760	53.0211	10.0000
IX	2	2	3	2	1	1	1	2	2	2	2	0.5126	0.4874	45.8789	54.1196	6.2995
X	2	3	1	2	2	2	2	2	2	1	2	0.4925	0.5075	47.0003	52.9967	5.3854
XI	2	3	1	3	1	1	2	1	1	2	1	0.4172	0.5828	40.8475	59.1524	1.4264
XII	3	1	1	2	3	1	1	2	1	2	1	0.4329	0.5671	42.2972	57.7027	2.2577
XIII	3	2	1	3	1	1	2	1	2	1	2	0.4898	0.5102	46.1833	53.8151	5.2151
XIV	3	2	3	1	2	1	2	2	1	1	1	0.4548	0.5452	45.7568	54.2424	3.4966
XV	3	3	2	1	3	1	1	1	2	2	2	0.5801	0.4199	47.7451	52.2514	9.6596
XVI	3	3	2	2	1	2	1	1	1	1	1	0.5823	0.4177	46.1013	53.8960	9.6810

Tab. 6 Adopted mixed fractional factorial array and simulated peak pressures and weight distributions

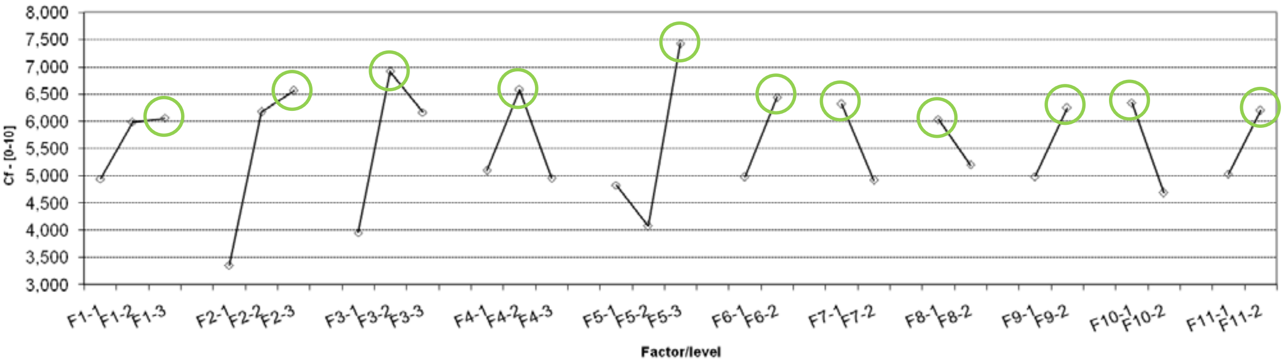


Fig. 8 Mean effects related to the comfort degree. Optimal design levels: 3-3-2-2-3-2-1-1-2-1-2

Tab. 5 reports the adopted design factors and their levels. Factors to had 3 levels (shortly named "1", "2", "3"), while two levels (named "1", "2") were assigned to factors to .

A full factorial design would have required a large amount of tests (· =15552). However, since every test is related to a FEA run (which is very time consuming - about 100 min to solve), a fractional factorial array was adopted (see tab. 6). This array was generated by using the MatLAB® built-in function "rowexch" [21]. Notice that the array has the minimum number of treatments, allowing to capture at least the main effects for every design factor (interactions among design factors are not here accounted).

The results obtained from FEM simulations are given in tab. 6 (columns 13 through 16). Those values were normalized as discussed into Section 3.4. Comfort degree values (last column into tab. 6), for every treatment, were calculated by using the comfort function stated into eq. 1.

5.2 Discussion of results

Looking at tab. 6, notice that treatment VIII gives the maximum comfort degree value. However, as a fractional array was here adopted, treatment VIII corresponds to a "relative" optimal configuration, among the 16 tested configurations. More investigations are then needed.

Fig. 8 shows the mean effects of each level for the eleven design factors on the comfort degree. Levels, for every design factor, which maximize the comfort degree, are marked as circle. Notice that the set of design factors (here called "optimal design"), maximizing the comfort degree, does not match any tested configuration, listed in tab. 6. This is due to the adopted fractional array, which does not contain all combinations among levels of factors. A confirmation experiment is then required [22]. The purpose of the confirmation experiment is to demonstrate the validity of results coming-out from the analysis of the mean effects. Therefore, a new FE model was generated and resolved, according to the optimal design parameters.

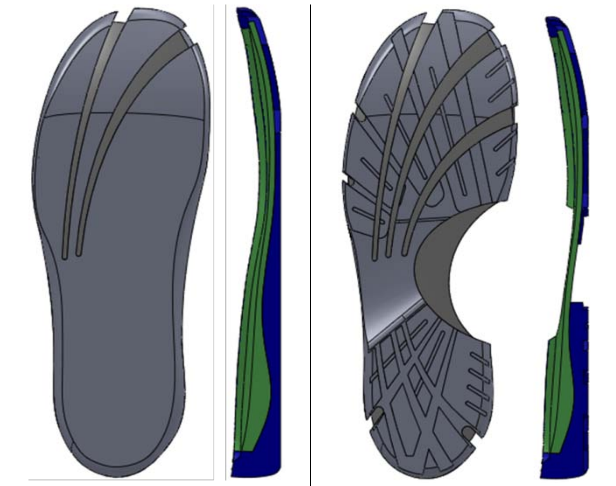
Fig. 9 compares initial shoe design and the optimal design. The estimated comfort degree for the initial shoe design equals 4.79, whereas it becomes 8.28 for the optimal design (with about 42% of comfort improving).

$$\begin{cases} SMQ_j = \sum_{i=1}^{Nl} R_{ji}^2 \\ \Delta_j = \dfrac{SMQ_j}{\sum_{i=j}^{Nf} SMQ_j}, \forall j = 1,\ldots,N_f \end{cases} \qquad (2)$$

By analyzing data through a Pareto ANOVA [22], contribution indexes (Δ) were calculated, as stated into eq. 2, where and are the number of levels and factors, respectively, and is the mean effect response of factor "j" at level "i".

Fig. 10.a depicts the so-calculated contribution indexes.

Looking at fig. 10.b, where cumulative contributions are shown, one should observe that, assuming a significant level of 90%, factors , , , exhibit a poor incidence on comfort degree: this means that variations of those factors slightly influence the comfort degree and then their variation may be neglected. This result says, for example, that the outer-sole notching plays no significant role on improving comfort degree and, then, it may be introduced with no variation of the comfort degree.



Comfort degree	**8.28**	Comfort degree	**4.79**
a - optimal shoe design		*b - initial shoe design*	

Fig. 9 Optimal design vs initial design

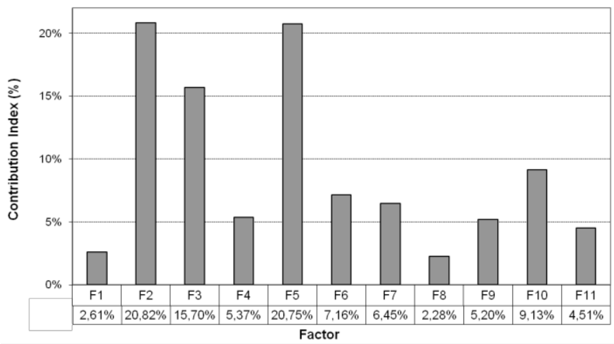


Fig. 10.a Contribution indexes

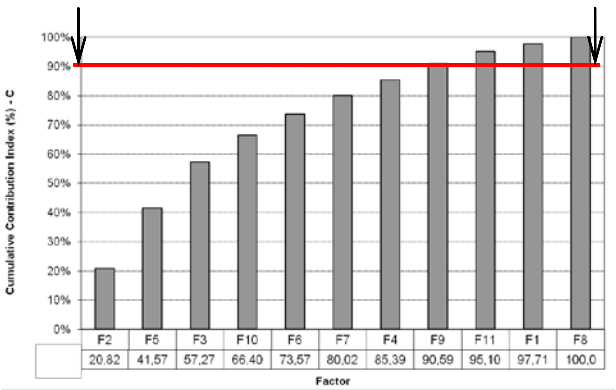


Fig. 10.b Cumulative contribution indexes

From the analysis of fig. 10.a, about 60% of the contribution rate is due to factors ("arch shape"), ("inner-sole material") and ("inner-sole thickness"). By analyzing mean effects of these factors (see fig. 8), one can note that:

- a general increasing in comfort degree is observed when decreasing the width of the arch shape (from level 1 to level 3). However, the increasing rate is less pronounced when moving from level 2 to level 3;
- material stiffness highly influences comfort degree. Except for a minimal reduction on comfort degree when moving from level 1 to level 2, a very considerable improvement of comfort is obtained when adopting a softer material (level 3);
- increasing of inner-sole thickness will improve the comfort degree.

The design of a new product, as footwear, involves many factors and key features to be accounted. When considering also the human perceived comfort, the design stage becomes a very huge task. The present study, despite its limitations, gives some guidelines to choosing and selecting the best design alternatives, by statistically analyzing different design settings.

Designer should consider these results carefully since they give only a preliminary screening in selecting the right design setting. Real experimenters are always required to validate such predicted results.

6 Conclusions and final remarks

In the present research different footwear design scenarios, based on virtual prototypes, were investigated in order to improve the perceived human comfort. Attention was posed on occupational footwears, which are commonly un-comfortable. In particular, only the footwear sole was here taken into account.

In order to "quantify" the perceived comfort, a preliminary experimental session was conducted. Twenty three experimenters were selected and, for each of them, four different footwears were worn. Every experimenter was asked to compile a questionnaire reporting the degree of perceived comfort. Plantar pressures were also monitored. By comparing the perceived comfort degree to the measured pressure maps, a comfort function, mainly depending on the peak pressure, was determined. Results showed that an increase in plantar pressure corresponds to a decrease of perceived comfort.

The comfort function was, then, adopted to study the sensitivity of different design factors. A parametric FE model was developed for this purpose. The sensitivity study was based on a fractional factorial design array. Findings of this study have suggested that the sole material and its thickness may strongly influence perceived comfort. More specifically, softer material and thicker inner-sole may play a significant role in improving comfort. Other specific features, such as sole notching, exhibit a negligible contribution level. This means that when designing such features other criteria, different from comfort assessments, can be adopted. Cost or manufacturing rules can be here evocated.

In the present research a static balanced standing-up configuration was accounted. Moreover, the interaction among the foot and the upper shoe cover was neglected. This hypothesis may be accepted for a static test, whereas when considering walking or running configurations it becomes a strong limitation. In fact, the perceived comfort is often affected by the interaction between plantar pressures and dorsal/lateral pressures. More investigation is, then, required when considering dynamic foot motion.

Apart from the simplifications and the limitations, the proposed methodology for studying and quantifying the comfort function, based on virtual prototyping, may be successfully extended to other engineering applications, involving customer satisfactions and sensations, such as seat design or postural assessments.

Acknowledgement

This research was partially supported by the grant POR xxxxx under the research cooperation with SAFE WAY s.r.l. company and PST Marche.

References

[1] H. Chen, B. M. Nigg, J. De Koning. *Relationship between plantar pressure distribution under the foot and insole comfort*. Clinical Biomechanics 9, 6 (1994) pp 335-341.

[2] P. Vink. Comfort and design: principles and good practice. CRC Press, Inc. 2005.

[3] P. Vink, M. P. De Looze. Crucial elements of designing for comfort. In Product Experience, Elsevier 2008, pp 441-460.

[4] M. Helander. Work posture. In a guide to human factors and Ergonomics, CRC Press 2005, pp 179-183.

[5] G. Kyunga, M. A. Nussbauma, K. Babski-Reevesb. *Driver sitting comfort and discomfort (part I): Use of subjective ratings in discriminating car seats and correspondence among ratings*. International Journal of Industrial Ergonomics 38, 5 (2008) pp 516–525.

[6] C. Leber, P. M. Evanski. *A comparison of shoe insole materials in plantar pressure relief.* Prosthetics and Orthotics International 10, 3 (1986) pp 135-138.

[7] E. C. Frederick. *Physiological and ergonomics factors in running shoe design.* Applied Ergonomics 15, 4 (1984) pp 281-287.

[8] C. Jordan and R. Bartleett. *Pressure distribution and perceived comfort in casual footwear*. Gait & Posture 3, 4 (1995) pp 215-220.

[9] C. P. Witana, R. S. Goonetilleke, S. Xiong, E. Y. L. Au. *Effects on surface characteristics on the plantar shape of feet and subject's perceived sensation.* Applied Ergonomics 40, 2 (2009) pp 267-279.

[10] J. T. M. Cheung, M. Zhang, A. Kam-Lun Leung. *Three-dimensional Finite Element Analysis of the foot during standing: a material sensitivity study*. Journal of Biomechanics 38, (2005) pp 1045–1054.

[11] D. Lemmon, TY. Shiang, A. Hashmi, JS. Ulbrecht, PR. Cavanagh. *The Effect of shoe insoles in therapeutic footwear:* a *finite element approach.* Journal of Biomechanics 30, (1997) pp 615-620.

[12] PJ. Antunes, GR. Dias, AT. Coelho, F. Reselo, T. Pereira, Non-Linear Finite Element Modelling of anatomically detailed 3D foot model*, technical report* (2008).

[13] J. T. M. Cheung, M. Zhang. *A 3D Finite Element Model of the human foot and ankle for insole design.* Archives of Physical Medicine and Rehabilitation 86, (2005) pp 353-358.

[14] J. T. M. Cheung, M. Zhang. *Parametric design of pressure- relieving foot orthosis using statistics- based finite element method*. Medical Engineering & Physics 30, 3 (2008) pp 269-277.

[15] H. R. Ashry, L. A. Lavery, D. P. Murdoch, M. Frolich , D. C. Lavery. *Effectiveness of diabetic insoles to reduce foot pressures.* Journal of Foot Ankle Surgery 36, 4 (1997) pp 268–271.

[16] M. Lord, R. Hosein. *A study of in-shoe plantar shear in patients with diabetic neuropathy*. Clinical Biomechanics 15, 4 (2000) pp 278–283.

[17] A. Lanzotti, G. Matrone, A. Vanacore. *Seat design improvement via a new postural comfort index.* Proceedings of ENBIS9 Conference, september – , 2009, Goteborg.

[18] D. C. Montgomery, G. C. Runger. *Nonparametric statistics*. In Applied Statistics and Probability for Engineers, Wiley 2003, pp 585-588.

[19] P. Erto. Probability and Statistic for Science and Engineering (in Italian), McGraw-Hill, ISBN 978-88-386-6413-7, 2008.

[20] P. Franciosa, S. Gerbino, *From CT Scan to Plantar Pressure Map Distribution of a 3D Anatomic Human Foot*, in Proc. of COMSOL Conference'10, Paris (France), november -, 2010.

[21] MatLAB® 2009b User Guide (Statistical Toolbox - User's Guide).

[22] D. C. Montgomery. Design and Analysis of Experiments. Wiley, Inc. 2001.

Acknowledgments

Acknowledgements

I would like to express my gratitude to my supervisor, Prof. Lanzotti for his suggestions and his support during my Ph. D. Course.

I'd like also to thank all my coauthors and everyone who contributed to my research for their help and for keeping me focused on our goals. I learned many things by everyone.

Finally I would like to thank Professor Kazerounian and Professor Ilies, who gave me the opportunity of an important research experience at the University of Connecticut.

Pensieri (II)

Insensata incoscienza o penuria di adrenalina… Approfitto di questa quiete prima della tempesta per chiudere questo lavoro nel miglior modo che mi venga in mente…

Così… Semplicemente…

Ho iniziato questo dottorato per realizzare un'aspirazione.

Durante questo dottorato ho imparato a difendere il mio lavoro.

Termino questo dottorato con la consapevolezza di aver avuto un'esperienza umanamente impagabile, perché impagabili sono tutti coloro che hanno condiviso con me questi tre anni.

A te. Sei stata inizio, percorso, poesia, filo conduttore a distanza senza saperlo.

Ai miei compagni di equipaggio. Abbiamo remato tanto e sempre insieme sulla stessa barca.

A chi potrei dire con buona pace che abbiamo fatto molto più che armonizzare.

A i tenaci, prima, durante e dopo.

A una delle persone più in gamba che io conosca.

A chi mi ha concesso uno splendido viaggio condiviso in tanti sensi

To Radu, Morad, Elena, Josephina, Mai and all the other fantastic guys that I meet at UCONN. Thanks for ice-cream, ping-pong matchs, pictures on the Moon, rumenian Monday, farsi Wednesday and a lot of other things that it's complicated to write… Three months is a short period but you will be always in my heart.

A coloro che mi hanno reso parte di qualcosa che va oltre, di cui non immaginavo la potenza e l'importanza…

All'anomalia che spero ricordi sé stessa e che non è presenza ma essenza.

A me per averci creduto.

A tutte le persone che condividono questo percorso a distanza e che nonostante vite diverse, case diverse e futuri diversi, continuano a farmi ricordare da dove vengo.

Alla mia famiglia, al voler costruire e al saper resistere.

A chi in ogni caso, in un certo qual modo, allorquando o quantomeno

Per chi capirà… Prometto solennemente che non collezionerò foto, ma solo ricordi nel fagotto, per quello che mi aspetta ora…