Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 50 (2016) 541 – 546

www.elsevier.com/locate/procedia

26th CIRP Design Conference

Consumers vs Internet of Things: a systematic evaluation process to drive users in the smart world

Alessandra Papetti*, Andrea Capitanelli, Lorenzo Cavalieri, Silvia Ceccacci, Francesca Gullà, Michele Germani

Università Politecnica delle Marche, Via Brecce Bianche 12, Ancona 60131, Italy

* Corresponding author. Tel.: +39 071 2204880. *E-mail address:* a.papetti@pm.univpm.it

Abstract

Smart Objects (SOs) market offers a wide variety of products apparently similar but characterized by different features that the average users fail to perceive. Consequently, their purchasing is often based on price and brand affection. In this context, users need a tool able to guide them in choosing the most suitable object to satisfy their expectations. To this purpose, this paper proposes a new systematic method to assess SOs in a comprehensive way: it allows to objectively assess and compare products and provides evaluation results tailored on users' needs. A first validation is carried out on three different SO typologies.

Peer-review under responsibility of the organizing committee of the 26th CIRP Design Conference

Keywords: Internet of Things, Systematic evaluation process, Usability, Smart objects

1. Introduction

The world has entered the age of the Internet of Things (IoT), where technologies have reached a maturity that enable each electronic device to be connected [1]. The market of Smart Objects (SOs) is increasing rapidly, from wearable technologies to telemedicine systems and home automation devices; a large amount of new connected devices is coming out in the global market. Despite this trend, there are still barriers that limit the mass dissemination of such technologies. While cost has historically been the most significant barrier to smart systems adoption, in the last years new obstacles have emerged raising consumer concerns. One of the major obstacles is the accessibility of the average user to these technologies.

Turning everyday products into connected products and linking them creating an ecosystem is a complex process. In fact, producers are usually strongly technology-oriented and aim to increase the potentiality of their products and systems neglecting the consumers mistrust towards smart technologies and innovations in general. While technology and logics behind a smart object are complex, the user's experience must be easy and intuitive. In this context, it would be fundamental to adopt a User Centred Design (UCD) approach in order to develop user-friendly products with features understandable by the average user, leading consumers toward the acceptance of these new technologies. Indeed, it has been demonstrated that generally people's perception of system qualities mostly depends on how they interact with it: how easily they understand the way it works, what they feel about it, how much it serves their purposes and fits in the context of use [2].

In addition to this, the constant evolution of Information and Communication Technologies (ICT) and the lack of common technological rules (communication protocols, physical connections, etc.) have led companies operating in the IoT sector to develop and promote its products and services independently. This condition has generated a market that quickly increased the number of companies and devices available to consumers [3]. Consequently, the customer has to choose between a wide variety of products apparently very similar, but each of them hides different features that the average user often fails to perceive.

In this context, it is important to give to the user a tool that can guide him/her in choosing the object able to satisfy his/her

Peer-review under responsibility of the organizing committee of the 26th CIRP Design Conference
doi:10.1016/j.procir.2016.04.128

needs and expectations. In fact, the success of a product or a service is mainly due to its ability to meet the user needs by providing what he/she exactly wants. For this purpose, this paper will present a systematic, flexible and innovative experimental protocol for the evaluation of smart objects that allows assessing and comparing different technologies from a technical and usability point of view. It has been validated by means of an experimental case study focusing on three different SOs belonging to different categories: body scale, blood pressure monitor and IP camera.

2. Research background

Smart Objects (SOs) can be defined as everyday consumer products equipped with sensors, memory and communication capabilities [4, 5], which are able to capture information about their surrounding, communicate with each other and react according to specific rules [6], help users to understand the behaviour and capabilities of their products and allow them to accomplish their tasks in a new intuitive way [7].

Based on our knowledge, although there are some studies that propose methods to assess SOs functionalities according to user needs [8, 9], no studies have been yet conducted with the aim to define a systematic method to objectively evaluate SOs overall quality in order to assist consumer in purchase decision. To help consumer in SOs selection, it is necessary to evaluate and compare them according to their ability to support specific users' goals.

As SOs are interactive devices, this means, on one hand, to consider their technical and functional features (e.g., material, connection technology, interoperability, reliability, accuracy, price, etc.) and evaluate them with respect to customer requirements. To this purpose, Quality Functional Deployment (QFD) is the best approach, among those proposed in literature. QFD is a systematic methodology for quality improvement and product development, originally defined in 1972 at Mitsubishi's Kobe shipyard site [10]. Several studies report its effectiveness in benchmark analysis [11].

On the other hand, it is necessary to assess SOs human-machine interaction quality. If we consider a product only from an instrumental point of view, the quality of a product perceived by a user during interaction can be measured by assessing usability. 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 specified context of use" [12]. In literature, there are many usability assessment methods that require users involvement or that are based on expert judgments. Among expert evaluation methods, evaluation based on Nielsen's heuristics [13] is the most common one. Meguire [14] provide a review of main methods for test usability with users.

However, usability is not sufficient to cover all the relevant aspects to assess the overall user-product interaction quality. According to Norman [15], product design affects users on three levels of information processing: visceral, behavioural and reflective. In particular, there is a dependency between aesthetic impression of a user interface and its perceived usability [16]. According to the Standard ISO 9241-210 [17], User Experience (UX) is "a person's perceptions and responses that result from the use and/or anticipated use of a product, system or service".

Consequently, UX can be considered as the result of all user emotions, beliefs, preferences, perceptions, physical and psychological responses, and behaviours that users experienced during interaction with a product (i.e., before, during and after use). Several methods can be used to evaluate UX: they range from extensive observation studies to more quick and dirty methods, such as interviews and questionnaires [18]. Among the second ones, the AttrakDiff questionnaire [19] and User Experience Questionnaire [20] are widely used for quick assessment.

Based on our knowledge, no systematic approaches have been defined to assess SOs quality in a comprehensive manner.

In this context, the present research aims proposing an approach to assess SOs from all perspectives.

3. Methodology

In order to increase the consumers' awareness about the IoT world and provide them an efficient tool to discover and compare products, an objective protocol to evaluate SOs has been developed.

Such result has been obtained by defining and adopting an approach that consists in the following six main steps:

1. Market analysis and definition of common patterns of connected devices;
2. Literature and standards analysis to identify what aspects should be taken into account in an evaluation process;
3. Creation of the experimental protocol structure in terms of aspects to assess and criteria for the score assignment, according to the results of the previous steps;
4. Definition of specification documents and procedures that describe how to perform the UX analysis and the tests aimed at the technical evaluation of a specific parameter;
5. Analysis of users' needs and correlation with the evaluation protocol items in order to ensure the protocol adaptability and satisfy the users expectations;
6. Experimental protocol validation by means of the involvement of experts.

The main evidence, originated from the Step 1, is the importance of the data collected and elaborated by SOs. Indeed, it is a peculiarity of smart devices as well as the connectivity requirements and specifications. In addition, interoperability and the offered services are two other significant issues to be considered. Furthermore, it is worth to specify that standard aspects such as the quality, reliability, price, and usability, which are common to the majority of commercial products, cannot be neglected.

Step 2 highlighted the existence of several guidelines to verify if products respect the quality standards, to assess the user experience and to test their reliability. However, evaluation protocols focused on other parameters such as products features, technical specifications, etc., are still missing. This is often due to the specificity of these aspects. However, feedback about them are more appreciate by consumers, as emerged from an analysis of the most important review websites and blogs.

The results of these analysis allowed developing the protocol architecture.

In particular, the goal of Step 3 is the definition of an architecture able to ensure the development of an evaluation protocol targeted to a specific products class and, at the same time, general in terms of review items. From such consideration, it derives the necessity to create a structure with a fix part to make the user familiar with the results and a customized one to adapt the evaluation process according to the product peculiarities.

To this purpose, a new evaluation method, based on a QFD approach, has been defined and several customers' requirements have been identified. They can be grouped in two main categories (i.e., the technical one and the UX one) and both of them have been organized according to a more-level structure (Fig. 1). This choice allows making the protocol adaptable, modular, and scalable. Indeed, the items i of the first level (L1) are common to all devices and represents the most important aspects that have to be considered during the product evaluation process. From a technical point of view, they are: connectivity, cost, installation & configuration, compatibility & services, reliability and characteristics. As far as the user experience is concerned, the evaluation attributes has been grouped in three main classes such as the design, the ease of use and the ease of installation.

Also the second level of the structure (L2) includes parameters (j) common to the majority of products. In this way, consumers can more easily become familiar with the review method and have a reference point.

On the other hand, the third level (L3) allows considering parameters (z) targeted to the specific devices category and, consequently, tailoring the relative scores assignment criteria. For this aim, a 1-3-9 rating scale has been used, which is then adjusted to a 10-point scale to calculate the final score. To evaluate possible plus of the object that cannot be assessed by the defined scale (e.g., a specific certification, backlighted display, etc.), a value of 0,5 has been assigned.

The considered criteria have been selected in order to evaluate the product from different perspectives and take into account all the items usually treated in the review process.

In particular, the scores can be assigned according to the product specifications (e.g., display size, settable languages, weight, etc.), as shown in the examples A and B of Fig. 1, or according to the tests results (e.g., connection quality, synchronization time, resistance to liquids, etc.), as shown in the example C of Fig. 1. In the second case, specific procedures have to be followed in order to ensure objective results. For this reason, Step 4 consisted in the definition of standard procedures to execute the performances and UX analysis. In particular, for the first one, the test typologies have been defined as well as their respective experimental set-up (e.g., operating system, instruments specifications, etc.), test procedure (e.g., steps, times, etc.) and thresholds. For the second one, in addition to these parameters, the user sample characteristics (e.g., age, gender, specific disease, etc.) and the tasks that have to be executed by them have been defined.

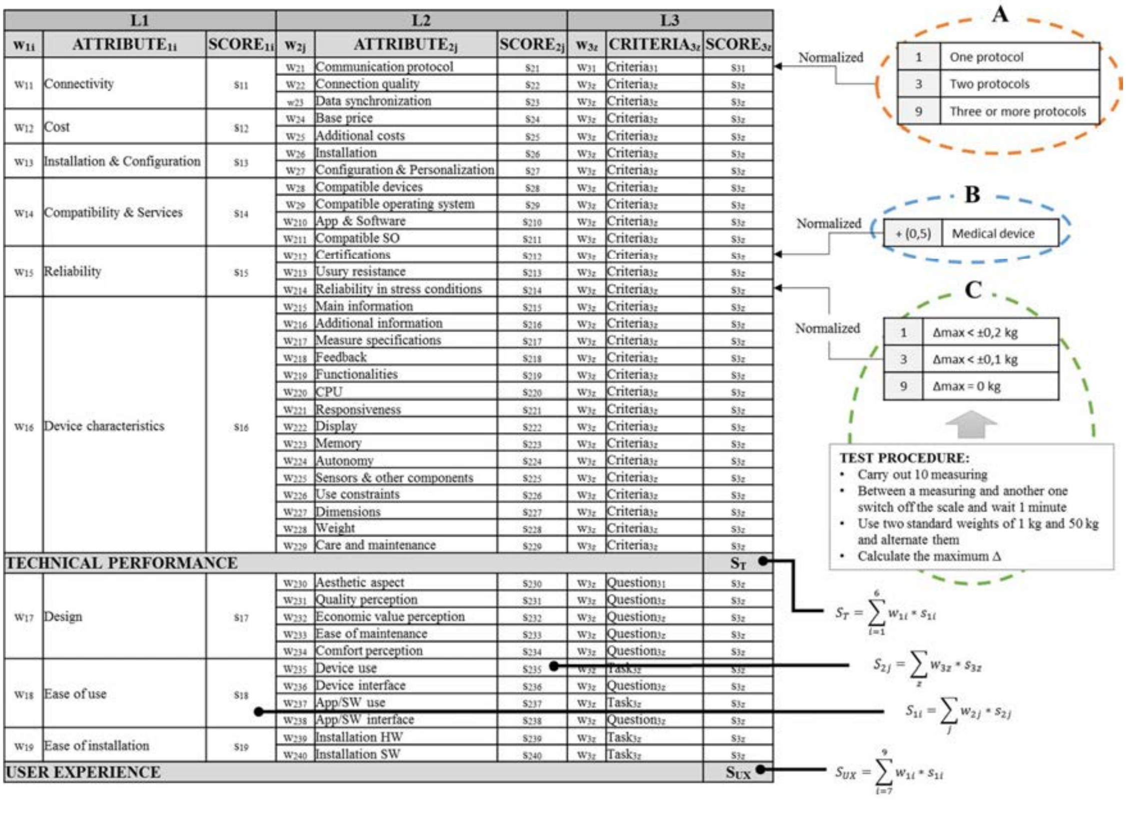


Fig. 1. The evaluation protocol architecture

To better tailor the architecture to the several products, it has been assigned a weight to each item of each level that describes its importance in relation to the specific products category. For this aim, several brainstorming sessions with experts have been organized in order to reduce the subjectivity level in the weights assignment. According to this approach, the technical and UX performances (i.e., S_T and S_{UX}) are obtained as a weighted sum of all scores of the L1-attributes (s_{1i}) and the final score can be calculated as suggested by (1):

$$S = w_T * S_T + w_{UX} * S_{UX} \quad (1)$$

In the same way, the score related to each L1-attribute (s_{1i}) is obtained as a weighted sum of the all relative scores of the L2-attributes (s_{2j}) and so on.

According to the aim of the Step 5, the most common users' needs have been defined and correlated to the L2-attributes. In this way, consumers can select some filters to express their preferences. Such filters enable or disable a specific item in order to tailor the evaluation report according to the users' expectations.

Finally, a group of five experts for each products category has been involved to validate the protocol (Step 6). In particular, it has been asked them to judge a specific product by assigning a weight and a score to each L1-attribute according to their know-how.

Going into more detail, the protocol structure allows increasing the evaluation detail by navigating through the levels (i.e., from L1 to L3). For example, economic aspects is evaluated by two attributes. The first one is the base price of the object, which is compared with the average one of the specific product category. In particular, different percentage gaps are defined to assign the relative score, according to the 1-3-9 rating scale. The second one is related to additional costs. In this case, the score assignment criteria refer to the impact of these charges on the product use (i.e., trial versions, limited features, etc.).

Furthermore, it is worth to specify that the user experience is evaluated through the following procedure:

- Users can interact with the product for few minutes and then are interviewed about their subjective judgment related to product design aesthetics and perceived quality;
- Users has to execute several tasks such as the installation HW and SW of the device, the use of the main function of the product and the consultation of specific data by means of the dedicated application;
- Users are interviewed about the ease and satisfactory of the tasks.

4. Experimental case study

Once defined the evaluation protocol, a case study has been carried out to validate it. In particular, two evaluations process has been performed on three real consumer products: the first one is related to the application of the proposed protocol and the second one is a heuristic evaluation with experts. The goal of the validation process is to check the soundness of the evaluation metrics and the weighing system, which have been attributed empirically. In order to measure the robustness of the proposed method, the results of the two described evaluation processes are compared: a low difference between the scores would allow to demonstrate the goodness of the proposed protocol. In order to explain the adaptability and versatility of this protocol, it has been tested with smart objects belonging to different contexts such as home automation, health and daily life. For this aim, three of the most common smart objects on the market have been selected, which are shown in Fig. 2. They are a smart body scale (a), a blood pressure monitor (b) and an IP camera (c). The choice fell on mid-range products with common functionalities. As far as the application of the proposed protocol is concerned, the technical assessment is carried out by an operator that took steps to assign the metrics according to the functional characteristics of the product and the proposed criteria. On the other hand, a sample of 5 target users is involved to evaluate the User Experience (UX).

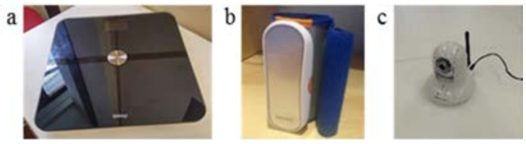


Fig. 2. The three SOs subject of the evaluation

Regarding the heuristic evaluation, a sample of 5 experts is involved. They have an engineering background (electrical, biomedical engineering and computer science disciplines) with a wide knowledge of the interconnected devices world. In particular, they are asked to evaluate the three products by considering the main aspects illustrated in L1 and assigning them a specific weight for each product category. A brief description of these attributes is given to the evaluators in order to avoid misunderstandings about their meaning.

At this point, the weight of each attribute in relation to the target product is defined by evaluators through a focus group. Instead, the evaluation process is performed in a separated way: each expert assigns his vote on a continuous scale from 1 to 10 to each evaluation item.

At the end of the two evaluation processes, the analysis of the differences is carried out. In the following paragraphs the protocol application and the final results are described.

4.1. Smart body scale

The body scale is a very common and simple object, however, the potentialities of the smart ones are often unknown to the average user. For example, the model selected for this study (i.e., *Withings Smart Body Analyzer WS-50*) is able to measure the heart rate, the air quality, the room temperature as well as the standard parameters such as weight, Body Mass Index (BMI), and body fat. Such data are then automatically synchronized to the smartphone/tablet thanks to the Bluetooth or Wi-Fi connection. In addition, it offers several services and functionalities (e.g., dedicated community, weather forecast, automatic user recognition, etc.) that enrich the product value. All these aspects have been evaluated thanks to the technical protocol. The main criteria, which have been added to better evaluate the category "smart scale", are related to the data sharing with a nutritionist, the training programs feature, and the accuracy of the measuring. The reliability tests aim to replay the use conditions (e.g., use of shoes, cleaning with an abrasive sponge, etc.), therefore, the resistance to a prolonged load, liquids, starch, and impacts have been also evaluated. The target user profile is an average user that is between 20 and 35 years old and takes care to his/her body shape. The tasks executed by the users are the HW and SW installation of the scale, the weight measuring and the consultation of historical data by means of the dedicated application.

4.2. Blood pressure monitor

A blood pressure monitor (BPM) is a device used to measure blood pressure (BR) values, the pressure generated by the blood into the body's vessel pushed by the heart rate (HR), and the heartbeat frequency rate, which is the measurement of the number of the heart contractions per unit of time, typically beats per minute. Such data are then synchronized to smartphone or tablet thanks to the Bluetooth or Wi-Fi connection. In the present case study, the model *MyTensio Armband bewell-connect* has been considered. Given the role that the device takes in medicine field and the importance related to vital signs monitoring, in the evaluation protocol, three basic attributes are highlighted: reliability, device characteristics and connectivity. Reliability is measured by three criteria: the presence/absence of medical certificates (e.g. FDA), wear resistance and repeatability of the measure in stress conditions. The device characteristics is measured by eleven main criteria; in this context, the measurement characteristics, referred to the device parameters, BR and HR, are the most important ones. In particular, the percentage gaps of blood pressure and heart rate are evaluated through the comparison of the measurements made with the selected device and those executed with a medical certificate device. Finally, the connectivity is very important in all devices without the physical display use to read the measurement, as the device evaluated in this paper. These assessments are carried out by means of technical tests. The target user profile is an average user that is between 35 and 65 years old with bradycardia or hypertension disorders and low attitude and skill with technology. The tasks executed by the users are the blood pressure monitor HW and SW installation, the pressure and heart rate measuring and the consultation of historical data by means of specific application.

4.3. IP camera

An Internet Protocol (IP) camera is a product generally employed for surveillance that can receive and send data using an internet connection (TCP/IP standard). Unlike the closed circuit television (CCTV) cameras, the IP camera can be connected with many common devices (smartphone, tablet and PC), it is cheaper and generally easy to install and use. Consequently, it is useful for the average user that can easily monitor his/her home; in fact, the market of this kind of products is growing up rapidly. Some key technical parameters must be taken into account to evaluate and choose the IP camera. The image quality is one of the most important characteristic (resolution, focal length, aperture, etc.), but the IP camera at the same time has to establish a fast and stable connection. Other important features evaluated are the presence of alarms in case of motion or audio detection, the indoor or outdoor use, the PIR motion sensor quality, the connection typology (wired or wireless) and many other software features. In this case study, the *TP-LINK TL-SC4171G* IP camera is evaluated by the presented protocol. The target user profile is an average user between 30 and 50 years old with family that wants to improve home security and family safety. The tasks executed by the users are the IP camera HW and SW installation, the ambient monitoring and the image settings regulations.

4.4. Results discussion

The expert evaluation has been conducted through the process described above. The results show a good homogeneity in the scores: the standard deviations in the three examined cases varies from 0,08 to 0,20. These values allow determining the reliability of the chosen sample. To compare the experts scores with the protocol scores, the average values of experts evaluations have been calculated for each product. In the comparative analysis, the section total scores underline the strong matching between the two evaluation methods: the scores gaps are minimized (from a min value of 0,03 to a max value of 0,82). The comparison between the two evaluation processes is shown in Fig. 3. The qualitative results allow making some interesting considerations. First of all, the trends of two evaluations are very similar: both in the case of the IP camera and in that one of the BPM, they follow a similar and consistent pattern. However, there are some differences in the Smart Body Scale evaluations. In particular, a consistent gap is present in the "Installation & Configuration" category: the lower score is due to the low level of personalization offered to the user, which is an attribute of the device configuration.

These results show a qualitative evaluation of the proposed method: they can be help the authors to improve the evaluation methodology planning a revision of the weighting values to reach an assessment protocol able to be used regardless of the operator expertise.

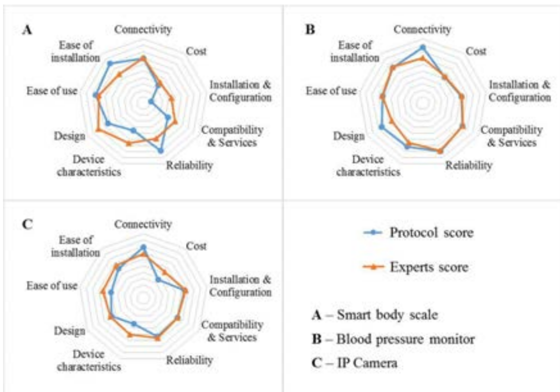
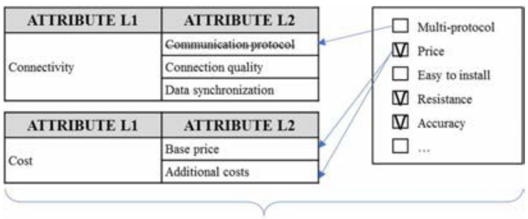


Fig. 3. Comparison of the evaluation results

In order to demonstrate the results adaptability to the users needs, it has been asked to a possible user of the proposed tool to choose some criteria that better represent his/her expectations about a smart body scale.

According to the rules defined in the Step 5, the significant items of L2 has been defined and the new results has been calculated. In particular, the attribute of L2 has been disenabled (e.g., communication protocol) or enabled according to the flag set by the user (e.g., base price, additional costs, etc.), as shown in Fig. 3. It is worth to specify that the absence of arrows means that there is no correlation between the user preferences and the attribute (e.g., multi-protocol and connection quality, etc.). In the specific case, the user does not matter about the following features: multi-protocol, installation, maintenance and device interface. The new results show as the scores of the selected smart scale increased, therefore, its performances are closer to the preferences of the specific user.



ATTRIBUTE	PRE	POST
Connectivity	6,9	8,1
Cost	3,7	3,7
Installation & Configuration	1,2	0,8
Compatibility & Services	4,4	5,5
Reliability	7,9	7,9
Device characteristics	4,5	6,2
TECHNICAL PERFORMANCE	**5,0**	**5,8**
Design	6,4	6,9
Ease of use	7,5	7,7
Ease of installation	8,0	N/A
USER EXPERIENCE	**7,1**	**7,4**

Fig. 4. Adaptability of smart scale results according to the users' expectations

5. Conclusions

The rapid spread of the IoT market and its peculiarities are highlighting the necessity to develop and offer to consumers tools able to guide them in knowing and comparing SOs and, consequently, selecting the better one in relation to their needs.

For this reason, the present research work proposes an objective evaluation protocol aimed to assign an understandable, complete and accurate score that represent the overall quality of the product. Indeed, it takes into account both technical features and the user experience.

In order to validate it, an experimental case study has been carried out on three commercial SOs belonging to different products categories. It allows comparing the results obtained by means of the proposed protocol and those ones resulted from a heuristic evaluation with experts. In this way, the goodness of the defined evaluation criteria has been demonstrated.

In the next future, the experimental protocol will be tested with a wider range of products to improve its structure and consistency. Furthermore, a tool that supports the operator in its application will be developed in order to reduce the level of expertise required. Finally, artificial intelligence algorithms will be implemented to make the correlation between users' needs and evaluation attributes more accurate.

References

[1] L. Wang, M. Törngren and M. Onori, "Current Status and Advancement of Cyber-Physical Systems in Manufacturing," Journal of Manufacturing Systems, Vol.37, Part 2, pp.517-527, 2015.

[2] Hornbæk, F. Current practice in measuring usability: Challenges to usability studies and research. INT J HUM-COMPUT INT 2005. 64, 79-102.

[3] INTERNET OF THINGS (IOT) MARKET FORECAST 2015-2020, Visiogain, March 2015.

[4] Ferguson, G.T.: Have your objects call my objects. Harvard business review (2003), vol. 80, no. 6, pp. 138-143.

[5] Gellersen, H.W., Schmidt, A. and Beigl, M.: Adding Some Smartness to Devices and Everyday Things, Proc. of WMCSA 2000 – 3rd IEEE Workshop on Mobile Computing Systems and Appli-cations, December 2000, IEEE Computer Society 2000, Monterrey, USA.

[6] Ziefle, M.; Rocker, C.,: Acceptance of pervasive healthcare systems: A comparison of different implementation concepts, Pervasive Computing Technologies for Healthcare (PervasiveHealth), 4th International Conference on-NO PERMISSIONS, (2010), vol., no., pp.1,6, 22-25 March 2010.

[7] Bohn, J., Coroama, V., Langheinrich, M., Mattern, F. and Rohs, M.: Living in a World of Smart Everyday Objects—Social, Economic, and Ethical Implications, Human and Ecological Risk Assessment, vol.10, pp. 763–785, Taylor&Francis, 2004.

[8] Bevilacqua, R., Ceccacci, S., Germani, M., Iualè, M., Mengoni, M., & Papetti, A. (2014). Smart Object for AAL: A Review. In Ambient Assisted Living (pp. 313-324). Springer International Publishing.

[9] Papetti, A., Iualé, M., Ceccacci, S., Bevilacqua, R., Germani, M., & Mengoni, M. (2014). Smart Objects: An Evaluation of the Present State Based on User Needs. In Distributed, Ambient, and Pervasive Interactions (pp. 359-368). Springer International Publishing.

[10] Park, T. & Kim, K., 1998. Technical note Determination of an optimal set of design requirements using house of quality. Journal of Operations Management, Volume 16, pp. 569-581.

[11]Chan, L. K., & Wu, M. L. (2002). Quality function deployment: A literature review. European Journal of Operational Research, 143(3), 463-497.

[12]ISO 9241-10: Ergonomic requirements for office work with visual display terminals (VDTs) - Part 10: Dialogue principles. Beuth, Berlin, Germany (1996)

[13]Nielsen, J. & Molich, R., 1990. Heuristic evaluation of user interfaces. Proc. ACM CHI'90 Conf. (Seattle, WA, 1–5 April), 249-256.

[14]Maguire, M., 2001. Methods to support human-centred design. Int. J. Human-Computer Studies, Issue 55, pp. 587-634.

[15]Norman, D.: Emotional Design: Why We Love (or Hate) Everyday Things. Basic Books, New York (2004)

[16]Tractinsky, N.: Aesthetics and Apparent Usability: Empirical Assessing Cultural and Methodological Issues. In: CHI 1997. Electronic Publications (1997)

[17]ISO 9241-210 (2010) Human-centred design process for interactive systems. ISO.

[18] Bevan, N. (2009). What is the difference between the purpose of usability and user experience evaluation methods. In Proceedings of the Workshop UXEM (Vol. 9, pp. 1-4).

[19] Hassenzahl, M. (2004). The interplay of beauty, goodness, and usability in interactive products. Human-Computer Interaction, 19(4), 319-349.

[20]Rauschenberger, M., Schrepp, M., Cota, M. P., Olschner, S., & Thomaschewski, J. (2013). Efficient measurement of the user experience of interactive products. How to use the user experience questionnaire (ueq). example: spanish language version. IJIMAI, 2(1), 39-45.