

NAVMETRO®: PRELIMINARY STUDY APPLICATIONS OF USABILITY ASSESSMENT METHODS

Eveline FERREIRA¹, e Diamantino FREITAS²

¹ Faculdade de Engenharia da Universidade do Porto

² Faculdade de Engenharia da Universidade do Porto

ABSTRACT

NAVMETRO® is a spoken dialogue system that provides information and a guidance function to the people with visual disabilities on Trindade's metro station, in Porto, Portugal. The system usability in addition to assure the quality dialogue for all costumers should ensure the success in the guidance on the desirable way in the metro station. This article presents the results of the preliminary assessment developed on NAVMETRO® and that represents an important stage of the usability assessment of this system.

KEY-WORDS

NAVMETRO®, spoken dialogue system, usability evaluation.

1. INTRODUCTION

The voice recognition technology has been widely used in recent years in dynamic contexts applications of usage, where the users have to perform multiple simultaneous actions, like the usage of mobile communication devices (Hooper, 2011), cars, and home controls systems (Möller, 2005). Nowadays this technology has been enabling visually impairment people to interact with many devices using their own voice through applications with low cost, making it an important tool for accessibility.

The design of a spoken dialogue system should take into consideration the user at all the stages of development by ensuring that his/her needs are respected during this process.

NAVMETRO®, like any other user-centered interactive system, should be subjected to usability evaluations in order to ensure the quality of dialogue provided by the system. The dialogue should be understandable and pleasant, so as to provide a good experience to the user. In the same way the guidance provided by the system must ensure to the user a safe mobility, free from constraints within the metro station.

The preliminary results of the usability evaluation performed on this system are presented in this article.

1.1 NAVMETRO®

NAVMETRO®, in operation since December 2009 in Trindade's metro station in Porto – Portugal aims to provide free of charge to visually impairment costumers, access to information relating to urban public transport, as well as a service to aid personal guidance within the station.

The information service is available also through the *Olá Metro* phone line and contains information to support trips, information regarding the acquisition of the transport ticket, a support to “lost and found section/office” and an area for suggestions and complaints.

The personal guidance orientation service, relying on studies regarding the acoustic orientation human capability (Freitas, 2008), allows visually impairment costumers to be conducted within the station through mobile telephone equipment alerts provided by an automatic dialogue system based on VOIP (Voice Over Internet Protocol) and supported by buoys sounds strategically placed throughout the station.

The information service can be used simultaneously by a large number of customers; for the personal guidance service there is a waiting line to use the acoustic space (Freitas, et. al, 2008). Gradually the system has been improved to allow greater flexibility in the shared usage of the service.

The main objective of the system is to provide to the visually impairment people, autonomy and consequently, quality in the usage of the metro.

1.2. System Operation

The start personal guidance service at the station, the client should call the free of charge number and follow the steps provided by the self service center.

Initially, the client selects the desired destination, among them: the metro lines, the three exits of the station, and the other resources within the station, including: bathroom, pharmacy for disabled, bar, tickets store, ticket machine, and food machine.

Subsequently, the system locates the user on the floor that he/she is: last, middle and ground/underground, and then in the area corresponding to the previously selected. At this time, the sound buoys are activated, in numerical order, so that the user can choose what he hears highest¹

From then the client begins his movement to the desired destination by following the instructions trough a mobile telephone equipment, and directing their movements through bird's sounds that they keep listening sequentially inside the station. The user can also change direction when he/she feels appropriate by making new location.

According to the technical document describing the architecture of the system (Freitas, 2008) in terms of hardware the system consists of a compartment that contains the telecommunications sub-systems, servers LAN, the sub-system information servers, the sound server, and the safety equipment supplier. In terms of software, the system has proprietary modules that are supported

¹ The system operation can be better understood through the video available on this website:
<http://paginas.fe.up.pt/~mdi11018/wordpress>

on databases which has partly rely on third party software, which facilitates the use of voice devices upgraded versions (Figure 1).

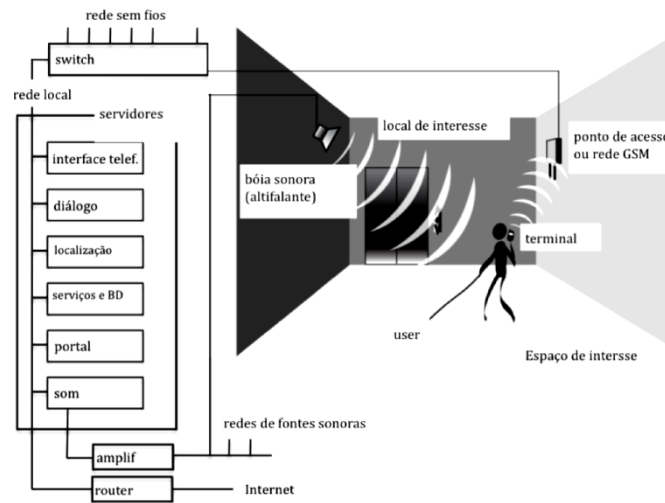


Figure 1: Representation of utilization and system systems architecture (Freitas, 2008)

2. STATE-OF-ART

2.1. Visually Impairment People (VIP) Needs in the Public Transportation

According to the World Health Organization (WHO, 2009), a legally blind person has visual acuity of 20/200 feet with the best correction in the best eye; and a field of vision whose diameter represents an arc not greater than 20 degrees. Therefore the VIP see 1/10 or less than a person with normal vision, thus having great difficulty in performing tasks that require vision.

According to the INE (2001), in Portugal the VIP represent 25.7% of the total population with some type of disability (Figure 2), considering the visual one, the type of disability in which there is a higher relative proportion of the population in the labor market, representing 52.6% of the population between 18 and 64 years old (CRPG, 2007).

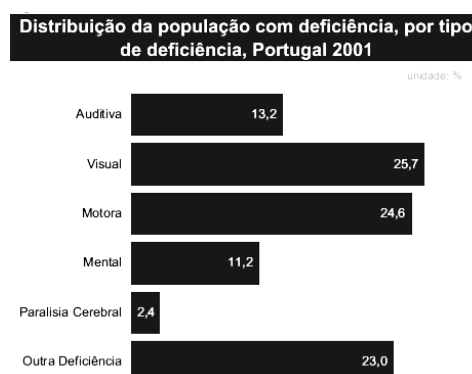


Figure 2: Population census (INE, 2001)

VIP use public transportation daily, and so it should work properly. However, it is common to see them using the help of others to perform simple activities such as obtaining information about the metro itinerary, identifying and moving to the board of the quay inside the metro stations, or purchase a ticket.

Therefore, this justifies the urgency of ensuring full accessibility to VIP in the urban public transport, safeguarding their right to enjoy the benefits of all the required information for them to be guided within the metro stations.

According to the National Plan for the Promotion of Accessibility (Ministers Council Resolution N^o. 09, 2007) the metro stations must *ensure the existence of alternative formats, accessible information used by people with special needs used in orientation and mobility.*

NAVMETRO[®] is a promising technical support system for macro navigation that will increasingly provide to the VIP a quiet and safe circulation within Trindades's metro station, and also within the future stations in the district of Porto, Portugal.

2.2. Usabilidade e Sistemas de Diálogo Falado (SDFs)

According to Weinschenk (2000), spoken dialogue systems allow users to dialogue with the computer in a natural way, using their own voice as input and output to control the system.

In these systems the flow of dialogue occurs through two main technologies: voice recognition, refers to technology that allows the computer to identify the human voice, and voice synthesis, which is the artificial reproduction of the human voice, commonly called text-to-speech. Thus, dialogue occurs when the human speech is recognized and encoded by the computer, which in turn performs voice synthesis that is understood by the user.

The flow of the dialogue occurs in the following way: the users voice is recognized by the Automatic Speech Recognition (ARS) that enables the computer to identify their response accepting it in order to control the system. Then, the dialogue management module processes the user's response – it is in this stage that the sound guidance system reproduces the sounds through the sound buoys installed according to the scheme depicted by figure 3.

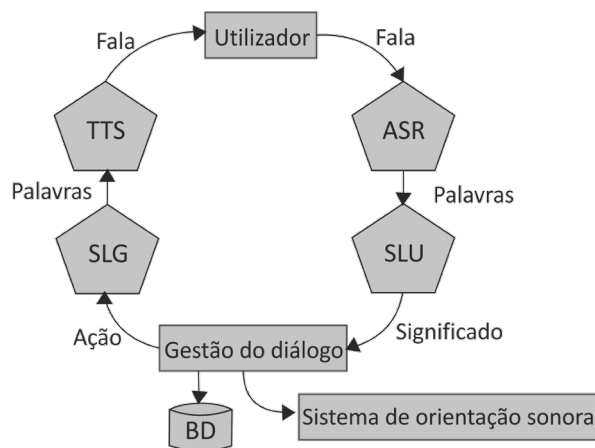


Figure 3: Main components of the flow of dialogue (personal file, 2012)

Finally the user response is converted into text representation, and then into acoustic speech synthesizer for text speech.

The SDSs should be subject to usability evaluations in order to ensure the quality of the dialogue between the user and the system. When the system is already finished or under implementation, the usability evaluation serves to refine some parameters to make it work better depending on the target audience public (Stavropoulou, 2009).

Aware of this point of view, telecom companies are increasingly investing in the development of usability methods evaluation in order to improve their services, assuming a development process much more focused on the user, and a character of high innovation ahead of its competitors.

ITU-T Rec P. 851 (2003) the services based on SDSs recommendation defines usability as "the degree of adequacy of service to meet the user requirements, including effectiveness, and efficiency of the system and resulting in user satisfaction." In accordance with this recommendation the quality of SDSs can be decomposed into different aspects:

- *Speech Input and output quality*: it refers to aspects of intelligibility, naturalness, and the effort required to listen to the system messages;
- *Dialogue asymmetry*: refers to the effect of the dialogue initiative and ability to control the interaction;
- *Dialogue cooperativity*: information, evidence, relevance, and elements of meta-communication, such as feedback, error recovery, and help messages system;
- *Communication efficient*: refers to the speed of interaction, size and smoothness of the dialogue;
- *Task efficiency*: refers to the success and ease of performing the task;
- *Comfort*: refers to the personality of the system including friendliness, education, natural system, and the user effort required during the interaction, such as ease of communication, agitation, etc.

3. DATA COLLECTION FOR USABILITY EVALUATION

The methods used in this article are based on the observation of the user performance interaction with the system in the real environment. The evaluation aimed to capture, analysis, segmentation, annotation and interpretation two types of data: the objective parameters and the subjective parameters of user interaction with the system. This article presents this topic describing how the two types of data were obtained.

3.1. Sample Selection

10 blind volunteers were recruited, including 7 men and 3 women (aged 33 to 48 years old) through two Institutes: "Associação de Cegos e Amblíopes de Portugal" – ACAPO and "Centro de Reabilitação de Areosa" – CRA. Three of the participants had used the Navmetro® system previously, while the other three had never used it. The absence of hearing impaired and the perfect English understanding were also taken into account in the selection of volunteers.

3.2. Materials

About the materials were used: two digital cameras, Canon Power Shot ELPH, a Nokia C300 mobile phone, and a Nokia mobile phone E65.

3.3. Procedures

The usability evaluation was performed at the Trindade's metro station, in the Porto city - Portugal. To perform the evaluation, a permission document signed by the secretary of the

Masters Course in Industrial Design at the Faculty of Engineering of University of Porto was given to the Metro do Porto S. A.

The evaluation was conducted in three phases: the first one was the evaluator's familiarity with the system, in this one week situation, the evaluator could see details about the system in order to understand it, in the second one there was a pilot evaluation to enable the procedures including the quality of the audio and video recording and the quality of the questionnaires, in addition to measuring the average time of the routes to be evaluated.

After clarification of the aims and procedures, the participants signed an informed consent form and answered the first survey containing 11 questions regarding their previous experience with Navmetro® system and others SDSs.

10 participants performed six routes within the Trindade's metro station. The routes included simulated trip to specific destinations, like the Vilar do Pinheiro's, Campainha's and Câmara Gaia's metro stations, as well as the exploration of existing resources in the station, including: bathroom, parapharmacy, and the tickets store.

For the objective parameters collection were made movies and audio recording through conference calls between two mobile phones.

For the subjective parameters collection each of the 10 participants responded to 18 specific questions after completion of each course (questionnaire 2), a total of 36 responses for each question, and 37 questions regarding the overall impression of the system (questionnaire 3). The questionnaires were adapted from ITU-T Rec P. 851 (2003). It was important in this study to perform the specific questionnaires at the end of each of the six paths in order to preserve the information perceived by the user, avoiding omission of important elements in each path.

The audio and movie analysis were done using the Praat software and Windows Live Movie Maker, respectively. The subsequent static analysis was performed using SPSS software.

4. PRELIMINARY RESULTS

This topic presents the preliminary results of the system usability evaluation that are still in-progress. The methods used in this initial evaluation, adapted from Möller (2004) aims to predict which parameters benefit most aspects of usability, such as those exposed in topic 3.2, for, in safely way, find out which improvements should be made in the system

4.1. Principal Components Analysis

The analysis of subjective parameters was performed by extracting the principal components of the set of questions answered by the participants in both questionnaires 2 and 3. This analysis helped to reduce the amount of questions drawn from the questionnaires, grouping them consistently (Möller Krebber, Smeele & 2005). This analysis revealed six major components that could be explained covering 78% of the variance of cumulative factors.

Table 1 shows the valuations corresponding to each question judged by the user. According to the ITU-T Rec P. 851. (2003) recommendation, the components with values ± 6 were grouped and interpreted and are in bold:

Questions	Components					
	C1	C2	C3	C4	C5	C6
1 Overall impression	-,78	-,07	-,52	,16	,07	,07
2 Reliability	,77	-,23	,05	-,07	-,02	,10
3 Pattern interaction	,19	,61	,54	,22	,34	,20
4 Inconstancy	,40	-,40	-,46	-,05	,23	,55
5 Unpredictable interaction	-,39	-,07	,11	,43	-,63	,08
6 Fast to the goal	,06	-,01	,59	,12	,11	-,37
7 A lot of mistakes	,69	-,53	-,02	,23	-,15	-,10
8 Reacted politely	,60	,56	-,11	,19	-,26	-,04
9 Easy error recovery	,02	,36	-,61	-,27	-,28	,04
11 I prefer not to use the system	-,54	-,27	-,11	-,61	,41	,09
12 Boring interaction	,09	,57	-,56	-,26	,34	,26
13 Satisfaction	,78	,07	,52	-,16	,22	-,07
14 Difficult to use	-,42	-,55	,38	-,00	,02	-,20
15 Interaction control	,16	,22	,74	-,14	-,05	,17
16 I will use the system	,63	-,08	,45	,18	,13	,07
17 I felt relaxed	,57	,35	,12	,16	-,14	-,13
18 Repetitive interaction	-,50	,61	,21	,52	-,22	,28
19 Annoying interaction	-,45	,04	,45	,54	,02	,07
20 Frustrating interaction	-,52	,35	,25	,42	,53	-,08
21 Inflexible system	-,87	-,01	,20	-,21	,14	,25
22 Easy to learn how to use	-,30	,18	,36	-,73	,27	,17
23 Slow interaction	-,39	-,22	,08	-,49	-,15	,74
24 Introduction is too long	,19	-,16	,30	,52	,24	,70
25 I like to use the system	,63	-,25	,59	-,08	-,31	,09
26 Conclusion is too long	,64	-,14	-,18	,35	-,43	,26
27 Voice sound is clear	,40	,60	,11	-,58	,00	-,03
28 Dialogue is natural	,64	,34	-,01	,16	-,56	,20
29 Location - floor	,55	,09	,48	-,15	,78	,11
30 Guidance - stairs	,53	-,26	,50	-,33	-,21	,15
31 It's easy to get lost in the orientation	-,07	-,48	,21	,08	-,78	,23
32 Concentration level	,34	-,40	,40	,56	,35	-,10
33 Did not do what I wanted	,003	-,87	,10	-,16	,13	,017
34 Reagiu de forma rápida	,68	-,04	-,46	-,07	,16	,40
35 Information is clear	,49	,67	-,31	,28	,10	-,15
36 System is friendly	,61	,42	-,11	,31	-,46	,21
37 Dialogue is clear	,81	-,57	-,01	-,02	-,01	-,06
38 I know the forward way	,24	,12	-,25	,31	,63	-,06
39 Complete information	,40	,69	-,32	,21	,24	-,34

Table 1: Principal components analysis (rotation Varimax)

C1 group: general impression, the system is reliable, the system has made many mistakes, the system responded in polite company, I liked using the system, I felt satisfied with the system, I will use the system in the future, are related to overall user satisfaction with the system;

C2 group: interaction pattern, the sound of the voice system is clear, the system did not always what I wanted, the information is related to the complete satisfaction of the interaction dialogue;

C3 group: was easy to retrieve errors and I felt in control of the interaction with the system are related to the asymmetry of the interaction dialogue interaction with the system;

C4 group: is easy to learn to use the system is related to the ease in learning to use the system;

C5 group: I always knew the path I should follow, it's easy to get lost on the guidance provided by the system, it is easy to get lost with the guidance provided by the system are related to meeting the guidance provided by the system;

C6 group: the introduction of the system is very long, the interaction is related to the slow speed of communication and consequently with efficiency in communicating with the system.

The reliability of each question ± 6 was analyzed for each of the components using the indicator Cronbach Alpha (α). In this study it was found components with $\alpha \geq 8$ as being appropriate. The results according to table 2, show that C1, C2 and C3 satisfy this criteria.

Component	Alpha (α)
C1	,96
C2	,84
C3	,80
C4	,72
C5	,63
C6	,50

Tabela 2: Cronbach value for each component

Some parameters have been adapted goals and ranked according to ITU-T Series P. Supplement 24 (2005) and are described below:

Dialogue parameters: dialogue total duration (DD), duration of speech user (DFU), duration of the speech system (DFS), delayed response user (ARU), delayed response system (ARS), returns system (RS), returns user (RU), questions the system (QS), conceptual efficiency (CE);

Meta-communication parameters: help messages requested by the user (MAU), help messages provided by the system (MAS), interruptions (barge-in), returns Correction System (RCS), returns the correct user (RCU), maximum waiting time for user response (time-out), rejections system (RJS);

Voice input parameters: user responses that are correctly recognized (RCR), incorrectly recognized (RIR), or partially recognized (RPR);

Cooperativity parameters: system responses that are appropriate (RAS), inappropriate (RIS), or that can't be observed (RNO) in accordance with the principles of cooperativity (Bernsen & Dybkjaer 1996).

4.2. Correlation between Objective and Subjective Interaction Parameters

In order to correlate the objective and subjective interaction parameters it was used the linear regression model, taking into account subjective parameters as dependent variables and objective parameters as independent variables. This model predicts, for simple mathematical calculations, which are variations of the dependent variables as a function of variations of the independent variables, which one wants to test. Through the product obtained by the variation coefficient of the independent variable was possible to make this comparison.

	Significant predictors	R ²
C1	- 0.349 ARS, - 261 RS, + 0.103 EC	0.517
C2	- 0.470 QS, + 0.206 EC, + 0.187 <i>time-out</i> , + 0.201 RAS	0.401
C3	+ 0.397 RAS, - 0.113 RCU, + 0.067 MAS	0.214

Table 2: Linear regression model with the most significant components

According to the table 3, one can appreciate the following conclusions regarding each of the three most significant components:

Concerning the predictor: $- 0.349 \text{ SRD} - 0.261 \text{ RS} + 0.103 \text{ EC}$, any of the 4 variables: system response delay (- 0.349 SRD), system return (-0.261 RS), and efficiency concepts (+ 0.103 EC) can affect strongly C1 group, but it is clear that the variable -0349 SRD affects more negatively the predictive analytics. Thus, to increase C1, including the overall satisfaction of the system should reduce the delay in the system response. This parameter is calculated from the moment the user

stops talking until the moment the system starts talking and is, according to the analysis, 60% higher in the early stages of the dialogue.

Concerning the predictor: $- 0.470 SQ$, $+ 0.206 EC$ $+ 0.187$ time-out, $+ 0.201 ASR$, any of the 4 variables: system questions ($-0.470 SQ$), efficiency concepts ($+ 0.206 EC$), time-out ($+ 0.187$ time-out), and appropriate system responses ($+ 0.201 ASR$) may strongly affect C2 group but it can also be seen that the variable $- 0.470 SQ$ affects the negative the predictive analysis. Therefore, to increase the C2, related to the satisfaction of the interaction dialogue one should try to reduce the number of the system questions.

This parameter is calculated by the number of inquiries to the user, and is 92% higher in the early stages of the dialogue. It appears that increased the use of the system causes the automatic registration of certain information, such as those related to the intended destination user. However, other related information, e.g., acquisition and validation of the ticket is always applied and become annoying to most users.

Concerning the predictor: $+ 0.397 ASR$, $- 0.113 UCR$, $+ 0.067 RSM$, at the same way, any of the 3 variables: appropriate system responses ($+ 0.397 ASR$), user correct returns ($- 0.113 UCR$) and help system messages ($+ 0.067 RSM$) may strongly affect C3 group, and the $+ 0397 ASR$ affects more positively the predictive analysis. In the final analysis, it appears that to enhance C3, related to the asymmetry of the dialogue interaction, one must increase the number of appropriate system responses.

The system responses are appropriate when they not violate the principles of dialogue cooperativity (Bernsen, 1996). The principles described by the author, the most commonly found in the notes held relate the amount of information provided by the system and confirmation. Thus, the system should not provide more information than necessary, must provide feedback to each user response also should provide instructions sufficiently clear to the user including the vocabulary used. Note that users are reluctant in relation to the term "resources" used to designate the bathroom, the chemist shop and purchase tickets.

5. FUTURE WORKS

The Navmetro® evaluation process is in progress, and the next steps include analyzing more consistently variations of subjective parameters as a function of changes in the types of system errors. These parameters were not taken into account in this analysis since they deserve greater attention by the evaluator.

From this study it was found that the main objective of this type of evaluation is to provide SDSs guidelines which can be evaluated automatically using linear regression models. The automation function of this type of evaluation is important so that the development team does not need to resort to more subjective information, which often becomes expensive and costly. A simple manipulation of the significant predictors obtained in this study can provide strong evidence for the improvement of the system.

6. REFERENCES

- [Bernsen 1996] N.O. Bernsen. Cooperativity in human-machine and human-human spoken dialogue. *Discourse Process*, 21, pp. 213-236
- [CRPG 2007] Elementos de caracterização das pessoas com deficiência e incapacidades em Portugal, Vila Nova de Gaia: CRPG
- [Freitas 2008] D. Freitas. Sistema de orientação sonora. 10^o Encontro de engenharia de áudio da AES Portugal, pp. 1-6
- [Freitas 2008] D. Freitas. Sistema complementar de navegação pessoal na rede do Metro do Porto para pessoas com deficiência visual. Relatório final: Faculdade de Engenharia da Universidade do Porto. Projeto POSC 978/4.2/c/REG.
- [Hoover 2011] S. Hoover. *Designing mobile interfaces*, CA: O'Reilly Media
- [INE 2001] Recenseamento geral da população. Disponível em:
<http://www.ine.pt/xportal/xmain?xpgid=ine_mainexpid=INE> [Acessado em 03 março de 2012]
- [ITU-T Rec. P. 851 2003] Subjective quality evaluation of telephone services based on spoken dialogues systems: Geneva, International Telecommunication Union
- [ITU-T Series P. Supplement 24 2005] Parameters describing the interaction with spoken dialogue systems: Geneva, International Telecommunication Union
- [Möller 2005] S. Möller. Evaluating the speech output component of a smart-home system. *Speech communication*, 48, pp. 1-27
- [Möller 2004], S. Möller. *Quality of telephone-based spoken dialogue systems*. Germany: Springer, pp. 94-461
- [Presidência do Conselho de Ministros 2007]. Resolução do Conselho de Ministros nº 9/2007. Diário da República, I Série, nº 12 de 17 de Janeiro. Disponível em:
<<http://dre.pt/pdf1s/2007/01/01200/03660377.pdf>> [Acessado em 17 março de 2012]
- [Spiliotopoulos 2009] D. Spiliotopoulos. Spoken dialogue interfaces: integrating usability. *USAB*, p. 484-499
- [Weinschenk 2000] S. Weinschenk. *Designing Effective Speech Interfaces*. NY: Wiley, pp. 20-98
- [WHO 2009] World health statistics. Geneva. Disponível em:
<http://www.who.int/gho/publications/world_health_statistics/en/index.html> [Acessado em 15 abril de 2012]