Assessing Voice User Interfaces: The vAssist System Prototype

Conference Paper · November 2014

DOI: 10.1109/CogInfoCom.2014.7020425

CITATIONS
8

READS
230

9 authors, including:

Gennaro Cordasco
Università degli studi della Campania "Luigi Vanvitelli", Caserta, Italy
105 PUBLICATIONS **637** CITATIONS
SEE PROFILE

Maria Teresa Riviello
Istituto Internazionale per gli Alti Studi Scientifici
32 PUBLICATIONS **140** CITATIONS
SEE PROFILE

Anna Esposito
Università degli Studi della Campania "Luigi Vanvitelli"
222 PUBLICATIONS **1,719** CITATIONS
SEE PROFILE

Gerard Chollet
French National Centre for Scientific Research
311 PUBLICATIONS **2,576** CITATIONS
SEE PROFILE

Assessing Voice User Interfaces: The vAssist System Prototype

Gennaro Cordasco*†, Marilena Esposito¶, Francesco Masucci*, Maria Teresa Riviello*†, Anna Esposito*†, Gérard Chollet‡, Stephan Schlögl§, Pierrick Milhorat‡, and Gianni Pelosi¶

*Department of Psychology, Second University of Naples, Italy.
†International Institute for Advanced Scientific Studies (IIASS), Italy.
Email: {gennaro.cordasco, anna.esposito}@unina2.it

‡Institut Mines–Télécom Télécom ParisTech Paris, France.

§ Department of Management, Communication and IT, MCI Management Center Innsbruck, Austria.

¶ Integrazioni & Sistemi, Rome, Italy.

***Abstract*—This paper reports on the results of the first lab trials evaluating the vAssist (Voice Controlled Assistive Care and Communication Services for the Home) system prototype with Italian users.**

vAssist is an European Project aiming to provide specific voice controlled home care and communication services for elderly. An important vAssist objective is a multilingual Voice User Interface (VUI) in three different languages: Italian, French and German. Lab trials were foreseen in these three different countries to assess the vAssist VUI prototype on realistic user expectations and requirements. The assessment was made letting 43 Italian elderly interact with the VUI prototype in 4-5 defined scenarios, exploiting a Wizard-of-Oz (WoZ) paradigm and administering to them three questionnaires aimed to measure their perception of the system's usability, learnability and intuitivity. Qualitative and quantitative scores suggested that VUIs are very powerful communication interfaces and were greatly appreciated because of the simplification they provide in the elder everyday use of technological products, such as mobile phones, tablets, and computers.

I. Introduction

In recent decades two emerging areas, information technology and telecommunications, converged into Information and communication technology (ICT). Then came the need to integrate the results from cognitive sciences through the use of ICT devices. Cognitive infocommunications (CogInfoCom) was created in response to this necessity to investigate the link between the areas of ICT and cognitive sciences, as well as the devices and application developed to integrate these sciences [3], [8], [1]

A key aspect of CogInfoCom is related to the study of tools which allow the communication and interaction between different users as well as between users and machines, using various channels. This paper investigates this particular aspect. We report the outcome of a first evaluation of a user-system voice interface developed within the vAssist project.

vAssist [13] is a European Project aiming to provide specific voice controlled home care and communication services for two target groups of older persons: seniors suffering from chronic diseases and persons suffering from (fine) motor skills impairments.

vAssist aims to enhance the perceived quality of healthcare services and to reduce the costs related to their production and delivery by achieving what is called "channel independence" i.e, allowing the users to request its services through the existing home devices such as PCs, TVs, mobile phones and/or tablets, exploiting either speech and/or graphical and text interfaces. In particular for speech, it is well known that the existing VUIs (which include advanced algorithms for automatic speech synthesis – Text to Speech (TTS) – and recognition – Automatic Speech Recognition (ASR)) are context and speaker dependent and not able to handle free forms of conversations. In the vAssist project this setting is further complicated by the poor speech of the end-users (generally dysarthric speech) because of their motor skill deficiencies and other disabilities (such as hearing losses and reduced attention) due to chronic diseases.

In addition, when interacting through voice, users are not aware of the VUI constrained capabilities, and tend to use conversational speech, expecting to be understood as it happens in their ordinary verbal interaction with other humans. Being misunderstood, and forced to reproduce several times the request, will contribute to a user's feeling of frustration and disappointment and may eventually lead to the refusal of such a voice interface. Therefore, it is necessary to develop VUIs that are able to manage the dialog in a way that overcomes technology constraints and conforms to a user's expectations. Given that the current technology is not able to provide algorithms that process and understand free-forms of conversations, the appropriate design of the dialog management is crucial to ensure the user satisfaction. In order to do so, the present paper focuses on identifying dialog complexities and user needs through user–VUI interactions with primary Italian vAssist users who evaluated two tele-medicine services (called "PillBox" and "DailyCare") before implementing them in the final VUI prototype. Additional service scenarios were also evaluated. The Wizard-of-Oz (WoZ) paradigm[1] was applied to collect natural speech data that will be used to improve the speech recognition system.

[1]The Wizard of Oz (WoZ) simulation method, where some (or all) application functionality is simulated by a wizard behind-the-scenes, is a very common and practical approach for testing the effectiveness and efficacy of a system prototype according to the users' requirements and expectations.

A. Italian Lab Trials

The Italian lab trials focused on the speech-based interaction between the elderly (primary users) and the vAssist system prototype within real life scenarios. This interaction was implemented in order to evaluate the effectiveness of the integrated voice recognition and synthesis system, and the WoZ paradigm was exploited.

The lab trial's outcomes consist of:

- natural speech data (real interaction voice recordings and related scripted texts).
- data on the users' evaluation of the system focusing on quality of experience, quality of service provided, and usability.

The lab trial's evaluation results will help improve the development process of the speech-based interaction, providing insights into aspects of real user-system interaction aspects and information about its acceptability and its impact on elderly users' lives.

B. Paper Organization

The rest of the paper is organized as follows: Section II describes the lab policies adopted to start the evaluation process. Section III provides a description of the assessment procedures conducted with the recruited elderly Italian users, whereas the system evaluation results are illustrated in Sections IV.

II. Lab policies

Some preliminary lab procedures were necessary to set up the experiments.

A. System Configuration

In order to run the vAssist system prototype, it was essential to install and configure a virtual machine, which implements such a system, on a dedicated laptop. Then, a wireless router, connected to the modem available in the location hosting the lab trial, was set to provide Internet access. This procedure allowed the laptop hosting the vAssist System to exploit the online Google ASR service [4]. In addition, a second laptop was connected to the same wireless router so that a Wizard of Oz web application [10] deployed on the vAssist system was available on this second machine. Finally, tests aiming to check that the configured apparatus were working correctly – recording wave files and decoding texts as output of the speech recognition and synthesis systems – were performed.

B. Scenarios

The scenarios represent the base–dialog underlying the speech-based interaction between the vAssist system prototype and the users. The base–dialogs of the used scenarios (called "Pill-Box" and "Daily Care") describe the assistive care services vAssist is aimed to provide – such as recording and reporting users' medical data and prescriptions, and monitoring users' health status and activities. These dialogs are constrained to the required service and were developed inside the project, although only for English. Thus, it was crucial to translate the preexisting "Pill Box"and "Daily Care" scenarios into Italian, and adapt them to the Italian social context and behavior. Indeed, some expressions, especially those related to health care, had to be adjusted to fit the Italian ordinary way of verbalizing and allow a natural and easy user–system interaction. In addition to "Pill Box" and "Daily Care" scenarios, two new scenarios, "Communication Services" and "Personal Assistant Services" were designed. These scenarios describe new vAssist voice services aimed at facilitating users to be in touch with family and friends, and organizing their (free) time. The "Communication Services" scenario offers services such as sending emails and SMS text messages, sharing information, documents and/or photos with family and friends. The "Personal Assistant Services" scenario allows users to make a call, get information about the weather, set an alarm, arrange appointments, search and book a restaurant, ask for the cinema schedule, and for searching information on the web. Once the scenarios were defined, pilot tests were performed in order to be sure that the dialog was fluid and worked appropriately.

C. Adaptation of Pre- and Post-Interviews and Questionnaires

The methodology to evaluate the lab trial, and thus the feasibility and the success of the vAssist system prototype, is based on user face to face interviews and questionnaires, measuring qualitative and quantitative data throughout all phases of the system evaluation process, i.e. before, during and after the users–system interaction. Qualitative and quantitative data were collected through the pre-interview and post interview and through questionnaires. The pre-interview, administrated before the user–system interaction, collected users' demographic data and information on their experience level with modern technology. The post-interview – administrated soon after each user–system interaction process – focused on users' subjective judgment about the quality of their interaction with the vAssist system, underlining positive and/or negative aspects and benefits they were expecting when requiring the services offered by the system. During the interaction, the "Simple Ease Question" (SEQ) questionnaire [11] measured for each scenario how easy/difficult the user–system interaction was (according to the users' perception).

After the whole user–vAssist system interaction process, questionnaires investigating a user's judgment about the perceived quality of the lived experience and the provided services were administrated. In particular, post-questionnaires were:

1) the "AttrakDiff's attractiveness scale" [5], [6], for assessing the system usability effectiveness, efficiency, enjoyment, engagement and appeal of using;
2) the "Software Usability Scale (SUS)" [2], for evaluating the system learnability (the easiness with which users are able to effectively interact with the system);
3) the "INTUI questionnaire" [12], for evaluating the intuitivity of the system (the novice user's ability to deal with the system without any effort).

The Italian and English detailed versions of the material exploited for the evaluation can be requested by email to `giannipelosi@isspa.it`.

D. Participants Recruitment

The primary users of the vAssist system are male and female elderly who suffer from chronic diseases (e.g. diabetes, rheumatism, arthritis, high blood pressure, cardiovascular diseases, etc.) and/or fine-motor skill impairments. In order to contact and involve the persons fulfilling the abovementioned criteria, the recruitment took place among the guests of two retirement homes, and a senior citizens center.

The first step was to contact the directors of the abovementioned associations and plan a meeting with each of them separately. The main goal was to provide information about the vAssist project, and apply for both their and the associations' guests' permissions and consensus in order to run the system prototype. In addition, a request was formulated to reserve two dedicated rooms and provide Internet access.

The directors provided a list of guests fulfilling the system requirements, which were consequently contacted and asked to participate in the experimentation. An agenda of appointments was set in accordance with the participants' availability and the time required to perform each trial (approximately 60 minutes). Pilot trials were separately performed in the three locations in order to check that the system worked appropriately.

III. Evaluation Procedure

A total of 42 participants evaluated the vAssist system prototype. Among them, 33 (group A) tested the vAssist prototype interacting with the system within the "Pill Box", "Daily Care", and "Communication Services" scenarios, whereas 9 (group B) tested the system through the interaction based on the "Personal Assistant Services" scenario.

A. Setting

Before inviting each user to start the evaluation, the experimenters prepared and placed the laptops to be used for the evaluation so that:

- the first laptop hosting the vAssist system prototype was set on a desk, situated in a dedicated room, in front of which the test user would have been invited to sit;
- the second laptop, supporting the Wizard of Oz software, was placed on a desk in a second room.

Two experimenters were simultaneously involved in each lab trial: one, sitting next to the user, guiding and supporting her/him in the interaction with the vAssist system; the other, staying in a separate room, controlling the voice-based dialogue between the vAssist system and the users through the Wizard of Oz software opportunely developed for the system services (Figure 1).

B. Procedure

The structure of the lab evaluation procedure was almost the same for the two user groups (A and B), and consisted of the 9 steps listed below. However, as already claimed in section III, there were some differences (highlighted in items 4 and 6) in the evaluation procedure experimented by the two groups.

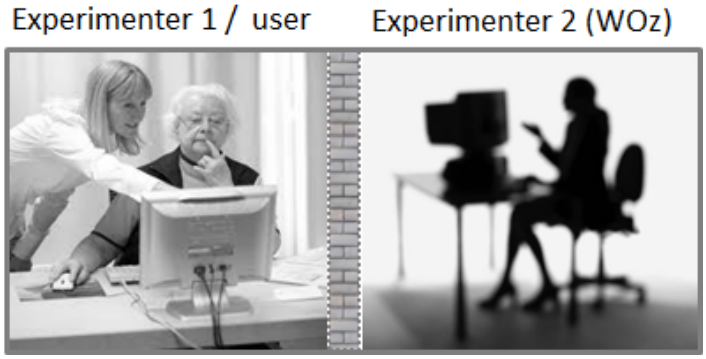
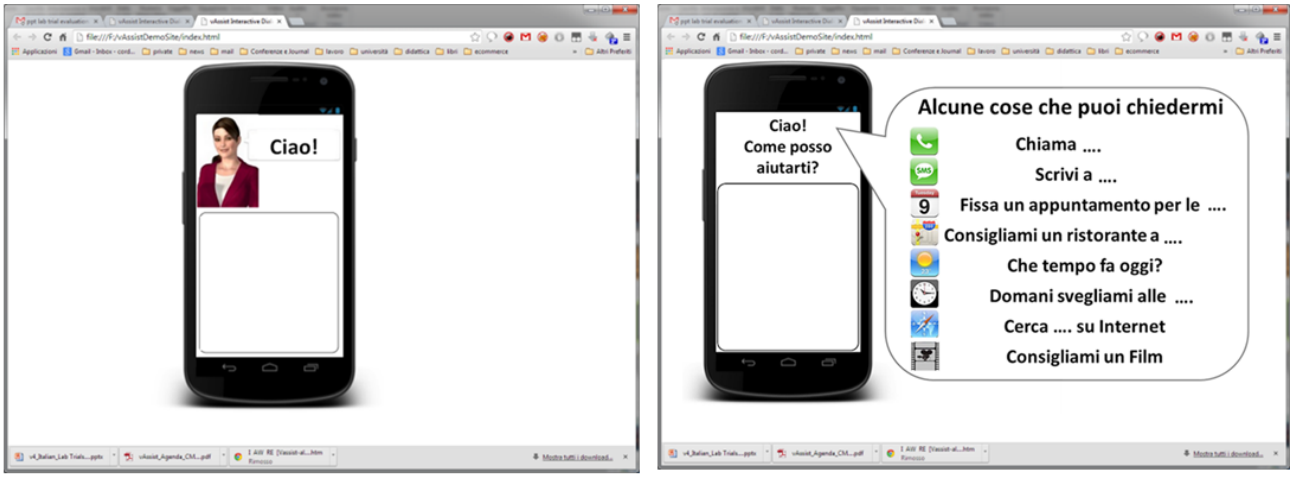


Fig. 1. System setting.

1) **Welcome (all users).** Before starting the lab evaluation, each user was informed about the vAssist project, its goals and what she/he was invited to do (reported in the following list of items):
   - to test the vAssist system prototype by vocally interacting with it on the base of the abovementioned scenarios;
   - to express her/his opinion about quality of the interaction and services provided through the interviews and questionnaires.
2) **Informed consent (all users).** All users were asked to sign an informed consensus (i.e. a consent form) for allowing the experimenters to record her/his voice and collect some of their personal data. The signature of the informed consensus was a prerequisite for their participation in the evaluation.
3) **Pre-interview (all users).** The evaluation procedure started administering the pre-interview, in order to collect information about a user's age, gender, job position, experience with touch devices, physical impairments, frequency of appointments with the doctor, etc.
4) **Presentation of the vAssist system interface (different for the two user groups).** Two slightly different system interfaces (consisting of a virtual smartphone shown on the laptop screen) were employed for each user group. Users belonging to group A interacted with the interface reported in Fig. 2 (a). It represents a smartphone with the picture of a female agent who is supposed to vocally interact with the user. No defined indications are presented, except a welcoming greeting (Hello!); the users belonging to group B saw the interface reported in Fig. 2 (b). It represents a smartphone displaying an invitation to use the system's services such as calling or texting someone, arranging appointments, searching for a restaurant, checking the weather, setting an alarm, searching the web or suggesting a movie to watch.
   It is worth to underline that the small differences in the two virtual interfaces produced a different user–system interaction approach. Group A users–system interactions were guided by the experimenter, who introduced the user through each scenario. Group B users selected the services offered by the system according to their preferences and therefore the dialog management was less constrained.
5) **Instruction on how to interact with the vAsisst system**



(a) (b)

Fig. 2. The vAssist system interface: (a) interface seen by users of group A; (b) interface seen by users of group B.

(all users). Each user was instructed to click the "press to speak" button, each time she/he started to speak to the system. In the case she/he was not able to click the button, she/he was helped by the experimenter who clicked it on her/his behalf.

6) **Interaction within scenarios (different for the two users' groups).**
   - **Gr. A:** Users from group A tested the vAssist prototype interacting with the system within the "Pill-Box", "Daily-Care" and "Communication Services" scenarios. The interaction unfolded on the scenarios' base-dialogs, however, the real dialog management changed from one user to another according to the answers she/he gave to the system's requests and answers.
   - **Gr. B:** Users from group B were invited to explore at least two services among those on the system's interface (cf. Fig. 2 (b)) in order to develop their interaction with the system.

   In both cases the real dialogs evolved differently depending on a user's request and/or answers and error-recovery questions.
7) **Post-scenario questionnaires (all users).** Soon after each user completed the request of a system's service (embedded in a specific dialog scenario), she/he was invited to fulfill the "Simple Ease Question" (SEQ) questionnaire, indicating how easy or difficult that task was, according to her/his opinion.
8) **Post-interviews and questionnaires (all users).** After the whole user–system interaction process, each user was post-interviewed and guided to complete the "AttrakDiff's attractiveness scale", the "Software Usability Scale" (SUS); and the "INTUI" questionnaire (see Section II-C for details).
9) **End of the lab trial.** After each user completed the evaluation procedure, light pastries and fruit juices were arranged and handed out to each participant, to make her/him comfortable and express the experimenters' gratitude for the offered availability.

Each user/vAssist interaction data (wave recordings and the texts) were stored in an appropriate folder and were considered as a project outcome.

IV. Quantitative Results

Table I gives an overview of the collected data (groups A and B). It provides an insight into the users' demographic and information about their physical impairments, frequency of appointments with the doctor, need to reduce doctor visits and experience level with modern technology. It is worth to note that users from group B show proportionally higher experience with touch screens than those from group A.

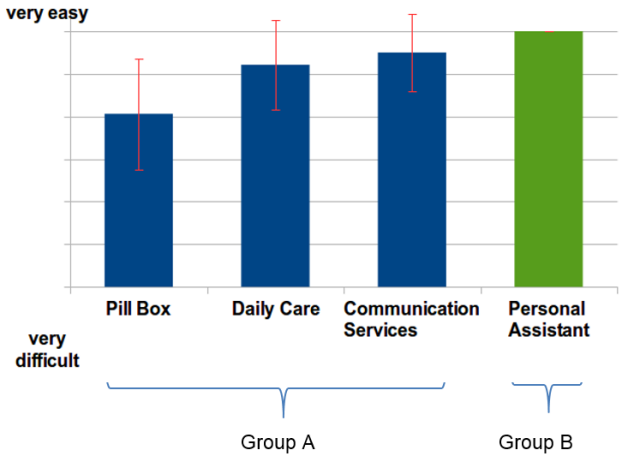


Fig. 3. Results of SEQ.

A. Task easiness: Single Ease Question's Results

The SEQ questionnaire was administrated after the users completed each service's scenario and aimed to assess, separated by scenario, how easy/difficult it was for the users to interact with the services. Results (cf. Fig. 3) show that, for Group A users the "Daily Care" and "Communication Services" scenarios were perceived as easier than the "Pill

Data		Italian Users Group A		Italian Users GroupB	
Participants		33		9	
Pathology		16 Chronic Diseases	17 Fine Motor Skill Impairments	7 Chronic Diseases	2 Fine Motor Skill Impairments
Age (avg / med / range)		69.94 /68/ 65-92		67/ 65/ 65-74	
Gender		23 M	10 F	5 M	4 F
Retired		26/33		6/9	
Experience with touch-screen devices	none	16/33		4/9	
	I tried it once	11/33		1/9	
	I use it sometimes	5/33		1/9	
	I own a touchscreen device	1/33		3/9	
Frequency of visits to the doctor	more than once a week	4/33		0/9	
	once a week	5/33		0/9	
	two times a month	11/33		0/9	
	at least once in a month	13/33		4/9	
	fewer than once a month	0/33		5/9	
Level of perceived stress for visits to the doctor	not stressful at all	11/33		8/9	
	little stressful	11/33		0/9	
	rather stressful	7/33		1/9	
	very stressful	4/33		0/9	
Reduction of doctor visits requested		20 Y	13 N	1 Y	8 N

TABLE I. DESCRIPTION OF GROUP A AND B SAMPLES AND PRE-INTERVIEW RESULTS.

Box" scenario. Indeed, many users felt a little uncomfortable to answer its specific questions. It is worth to underline that users from group A completed the "Pill Box" and "Daily Care" scenario, whereas only $12/43$ went through the "Communication Services" scenario because they were worried about their privacy. All the users from group B rated the interaction with the "Personal Assistant Services" as easy. The results of the SEQ questionnaire were elaborated according to [9].

B. Usability: Software Usability Scale's Results

The Software Usability Scale (SUS) was administrated at the end of the whole users–system interaction. It evaluated the easiness of the user interaction with the system. The total SUS score was 89.23 (SD= 8.47) and is graded as A+ (=excellent), according to Sauro and Lewis' scale [7]. Thus, the system was perceived to be very easy to use. The evaluation results for the two groups are similar (cf Fig. 4), however group B users seem to appreciate system usability slightly more than group A.

C. Attractiveness: Results of AttrakDiff's Scale

The AttrakDiff's attractiveness scale evaluated users' perception of system efficiency, enjoyment, engagement and appeal of using. The scale investigated four possible system's dimensions:

- the Pragmatic Quality (PQ).

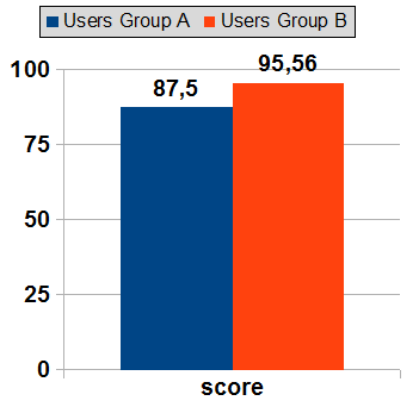


Fig. 4. SUS scores.

- the Hedonic Quality-Identity (HQ-I)
- the Hedonic Quality-Stimulating (HQ-S)
- the Attractiviness (ATT)

As Fig. 5 shows, the average scores shown on the y-axis (blue for group A and red for group B), All four dimensions lay in the highest region. Therefore, in addition to being satisfied and stimulated, the users were also very attracted by the system.

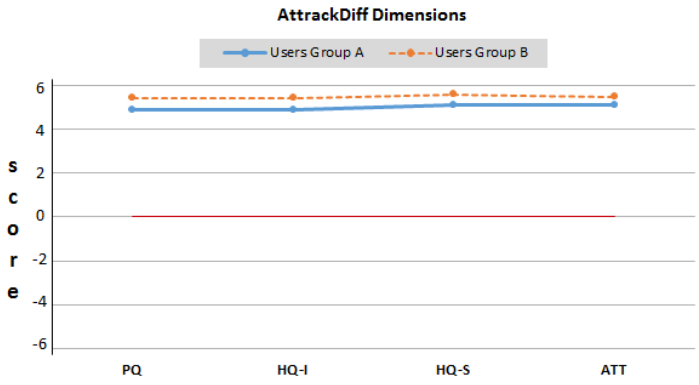


Fig. 5. Attrakdiff's dimensions scores.

Analyzing the users PQ and HQ dimensions scores (where HQ denotes the average between HQ-I and HQ-S) it was also possible to assess the "character" the users attributed to the system. Fig. 6 displays how, according to these scores, both groups of users perceived the system as "desired".

All the results of the AttrakDiff's "attractiveness" scale were elaborated according to [6].

D. Intuitivity: INTUI Questionnaire Results

The INTUI questionnaire investigated the users' ability to deal with the system without efforts, i.e. the intuitivity of the system. Only the "effortlessness", "gut feeling" and "magical experience" components of the intuitive interaction were considered. Indeed, the "verbalizability" component, which checks the users' ability to verbalize the different phases of the system interaction, was discarded given the speech-based nature of the user–system interaction.

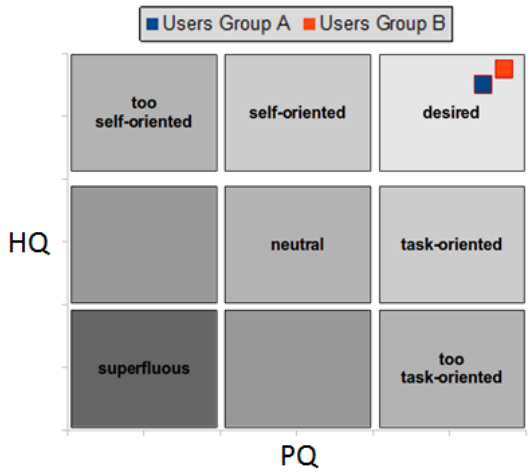


Fig. 6. Attrakdiff's System Character.

The questionnaire results, as reported in Figure 7, show that the users perceived the interaction with the vAssist system easy and enchanting, although it required some mental concentration.

The results of the INTUI questionnaire were elaborated according to [12].

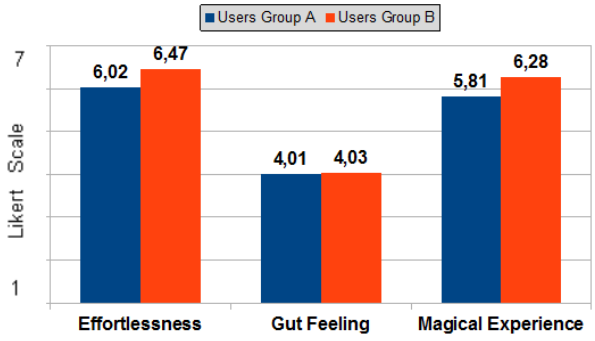


Fig. 7. INTUI Scores.

E. Discussion

In general the vAssist System Prototype was well accepted by all users and the users–system interaction was perceived as easy and pleasant. It is worth to underline that the interaction was speech-based, thus users were not requested to type or manually enter data. This could be the reason for the high scores attributed to the users' perception of system usability, learnability and intuitivity. Male users appear to be more enthusiastic than female users when interacting with the system. This can be due to the fact that technical devices attract more males than females. However, everybody thought that the system can be useful for their well-being. Users from group B seem to be slightly more comfortable in their interactions than users from group A. This can be attributed to their higher experience level with this kind of devices. However, also the services offered can play a role. Indeed, users from group B dealt with topics such as choosing a restaurant or a movie, checking the weather, etc. These services may have been perceived as more stimulating and funny. Nevertheless, the vAssist system prototype was considered very promising, since, on average, all users showed satisfaction in interacting with the system.

F. Problematic Aspects of the System

Difficulties related to user–system interaction concern the use of a specific button to be clicked in order to activate the voice recognition function. In general, users were not so confident in using the mouse to click the button and/or they forgot to press the button before starting to speak. They would have preferred to speak with the system without performing a manual command. Problems also occurred in the voice recognition system when a user's voice was very low or hoarse, and this should be taken into account considering that target users of the vAssist system are older persons often reporting voice quality distortions.

V. Conclusion

The results of this system evaluation are relevant as they show that Voice User Interfaces are well accepted by elderly users, are considered attractive, easy to use, produce a considerable user satisfaction and do not require technological expertise. All this comes from a static interface that does not exploit other communicative signals, such as facial expressions, gestures, and complex intonations. The simplicity and attractiveness of such interfaces is mostly based on the fluidity of the dialog management procedure, which in this case was implemented by a Wizard (the experimenters). The audio recordings collected provided a rich source of semantic frames for developing more naturalistic dialog management in context constrained VUI applications.

References

[1] P. Baranyi and A. Csapo. Definition and synergies of cognitive infocommunications. *Acta Polytechnica Hungarica*, 9:67–83, 2012.

[2] J. Brooke. SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189:194–200, 1996.

[3] A. Csapo and P. Baranyi. CogInfoCom channels and related definitions revisited. In *Intelligent Systems and Informatics (SISY), 2012 IEEE 10th Jubilee International Symposium on Subotica*, 2012.

[4] GoogleASR. https://dvcs.w3.org/hg/speech-api/raw-file/tip/speechapi.html, 2014.

[5] M. Hassenzahl. The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction*, 19:319–349, 2004.

[6] M. Hassenzahl. AttrakDiff(tm). Internet Resource http://www.attrakdiff.de., 2014.

[7] James R. Lewis and Jeff Sauro. The factor structure of the system usability scale. In *Human Centered Design LNCS Vol. 5619*, 2009.

[8] G. Sallai. Defining infocommunications and related terms. *Acta Polytechnica Hungarica*, 9(6), 2012.

[9] Jeff Sauro. Comparison of three one-question, post-task usability questionnaires. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1599–1608, 2009.

[10] Stephan Schlögl, Gavin Doherty, and Saturnino Luz. Wizard of Oz Experimentation for Language Technology Applications: Challenges and Tools. *To appear in Interacting with Computers doi: 10.1093/iwc/iwu016*, :, 2014.

[11] D.P. Tedesco and T.S. Tullis. A comparison of methods for eliciting post-task subjective ratings in usability testing. In *Usability Professionals Association Annual Conference*, 2006.

[12] Daniel Ullrich and Sarah Diefenbach. INTUI. Exploring the Facets of Intuitive Interaction. *Mensch & computer*, 10:251, 2010.

[13] vAssist web site. http://vassist.cure.at/home/, 2014.