Personalised and Accessible TV Interaction for People with Visual Impairments

Daniel Costa
LASIGE, Departamento de Informática
Faculdade de Ciências, Universidade de Lisboa
1749-016 Lisboa, Portugal
thewisher@lasige.di.fc.ul.pt

Carlos Duarte
LASIGE, Departamento de Informática
Faculdade de Ciências, Universidade de Lisboa
1749-016 Lisboa, Portugal
caduarte@fc.ul.pt

ABSTRACT

Connected TV refers to any TV or set-top box that can be connected to the Internet providing access to additional content such as TV applications, Web browsing and communication with other devices in the network. These new features turn TV devices into more versatile and interesting platforms, but also clog the screen with more content, making TVs more inaccessible for people with visual impairments. This paper presents the design of a solution that takes advantage of multiple modalities, personalisation and allows users to control TV applications from their mobile devices, taking advantage of the accessibility features of the latter.

CCS Concepts

•**Human-centered computing** → **Accessibility technologies;** *Accessibility systems and tools; Touch screens; Gestural input; Auditory feedback;*

Keywords

Connected TV; Human computer interaction; Personalisation; Adaptive systems; Multimodal systems; User centered design; Accessibility; Blind users

1. INTRODUCTION

TV service providers now offer a variety of features, including electronic programme guides, catch-up and recording features, recommendation services, but also Internet access and a variety of TV applications that are unrelated to broadcast content. TV applications, similarly to what happens with mobile devices can be accessed through an app store usually available in most TV platforms. Examples of popular TV applications include NetFlix, Facebook or Youtube.

Several accessibility issues can be found in these entertainment devices and their applications. TV applications are highly reliant on visual content which can pose serious accessibility barriers to visually impaired (VI) users. The lack of feedback about what is rendered on the screen is one of the main reported causes for these users not fully enjoying the capabilities of their televisions [10, 16, 4].

Additionally, the remote control is the universal input device for TV. It is commonly agreed that the remote control design tends to exclude many VI users, raising issues such as text size, labeling and tactile feedback [3, 15].

This paper summarizes the design and implementation of a prototype which is able to audio render the information about the TV application's user interface and send navigation commands by activating buttons using the assistive technology (e.g. TalkBack) of the users' smartphone and other alternative modalities. To support personalisation a user interface event logger to detect certain patterns that could indicate user difficulties was also implemented. Adaptations are suggested to the user accordingly with the patterns that are found with the goal to ameliorate the accessibility of the TV applications and overall interaction.

WOODSTOCK '97 El Paso, Texas USA

DOI: 10.475/123_4

2. RELATED WORK

This section summarizes existing research regarding the main topics of this work. First the problems VI users face when using their TV and then the advantages of having multiple modalities and adaptive features for building accessible interactions.

2.1 Accessibility Barriers of TV

Usually, people with some kind of disability need assistance when interacting with most consumer electronic devices. This is even more pronounced when considering visually impaired people and devices strongly based on visual interfaces such as the TV.

Oliveira et al. [10] present the results of a questionnaire that aimed to identify visually impaired users' problems and needs concerning the consumption of television. Participants reported they need help to adjust the television volume and other basic options. Most of the participants also feel that Digital TV brings difficulties in the way they watch TV because it comprises interactive services such as the Electronic Program Guide (some users reported that they had feedback problems using the menus and got lost). On the other hand, participants find the Audio Description service very useful if they could change some properties such as the narrator's voice, speed and volume. Moreover, they would like to see features such as voice feedback of the different options, switch language, and access to a list of favorite channels with audio description. Additionally, in the first stage of the authors' work in [9] it was reported by the participants that the implementation of programs with audio description; subtitles amplification; possibility to modify the contrast of the TV image and subtitles; possibility to change the speed of the audio information transmitted; and dubbing are the most important requirements for an accessible TV.

In [4] similar results were found. The most reported difficulties were the access to the programme guide and recording a show. Participants were asked if they used or had knowledge of any type of assistive technology for TV. Only 2 out of 26 have reported using radio provided Audio Description, a service offered by the public broadcast service of their country. In this study it was patent the lack of accessibility of the participants' TV sets, as well as some of the requirements of the target users for a better solution. Overall, the participants of these studies showed dissatisfaction with the level of accessibility of their TV services and they pointed out the main cause for this to be the lack of feedback TV users have to deal with.

2.2 Alternative Interaction for TV

The advantages of alternative modalities to ameliorate the accessibility for desktop and mobile devices have already been researched in the past. Azenkot and Lee [1] conducted a survey with 169 people (65 with low-vision or blind). The authors found that VI users used speech for input more frequently than sighted people. VI users were also more satisfied with this modality than sighted people. On screen gestures are already used by VI users when interacting with their mobile devices, however the use of on-body interaction and mid-air gestures are also being studied as another alternative [8]. Duarte et al. [6] presented a multimodal approach to control a mobile device, using a combination of mid-air gestures and speech. In their use case, a SMS application, VI users performed better using this approach than the traditional methods. A focus on input alternatives for VI users is yet to be researched in the TV environment.

Recently, TV platforms, media and entertainment centers have been trying to implement gestures and speech recognition. Samsung, Android, Apple and Video Game Systems such as Playstation or XBox already provide some of these interaction features. Taking advantage of this can help to improve the accessibility of such platforms for VI users. Additionally, modern TV platforms are already integrated with mobile devices, which can be a key factor to support multimodal, accessible interaction that could replace the traditional remote control.

2.3 Personalising based on Interaction Patterns

The advantages of multimodal interfaces can be better explored by introducing adaptive capabilities [13]. By monitoring the user's interaction and the context of use, the system can adapt automatically, improving its ability to interact with the user, building a user model based on partial experience with that user. These adaptive capabilities are important when dealing with users with different physical and cognitive characteristics, preferences and knowledge.

According to Raufi et al. [12], there is a myriad of adaptation techniques that support sighted users through information space. However, serious presentation and navigation issues arise when adaptation is targeted for VI people. It is suggested that specific adaptation techniques should be applied in order to achieve a successful adaptation for these users. For instance, information can be colored, zoomed, altered or rendered in alternative modalities such as auditory, based on the type of content, environment context and user's characteristics.

For an adaptive user interface to perform adaptations, information about the users must be gathered. This user information is either prompted directly from the user or captured automatically by the system. A possible data source for capturing relevant information in an implicit way is the sequence of interaction events, also described as user interface events. Raheel [11] states that creating this "user interaction profile" and making the system able to learn and interpret each user's actions, leads to proper and better adaptations.

Usually, desktop and mobile devices generate user interface events as natural products of their normal operation. Hilbert and Redmiles' survey [7] examines computer-aided techniques used by HCI researchers to extract usability-related information from such events. The analyzed events indicate user behavior with respect to the components that make up an application's user interface (e.g., key presses with respect to application input fields or mouse clicks with respect to application buttons, menus, lists, etc.). Because these events can be captured automatically and because they indicate user behavior with respect to an application's user interface, they have long been regarded as a fruitful source of information regarding application usage and usability [7].

From Balbo et al.'s work [2] four different patterns could be identified: 1) Direction Shift, usually happens when a user stops progressing along a branch of a task tree; 2) Action Cancellation, occurs when a user backtracks immediately after taking an action; 3) Irrelevant Actions, when the user performs irrelevant actions during a task; 4) Action Re-occurrence, when the user performs an elementary action repeatedly. Additionally, Shah [14] proposed the following patterns: 1) Vertical/Horizontal Mouse Movement, this pattern happens when a user is unable to proceed with the execution of a task and starts to visually explore the interface for other options (usually this is reflected in the motion of the mouse pointer); 2) a Steady Scroll Down pattern means that the user is engaged in reading the page while Quick Up/Down Scroll pattern means the user is searching and skimming for information; 3) Page Hoping happens when a user spends short periods of time in a page as he/she is backtracking for a familiar page; 4) Hub and Spoke navigation pattern occurs when the user keeps returning to a familiar page after accessing an unfamiliar one, not accessing more than two pages away.

Although the presented techniques and patterns were designed and identified to evaluate user interaction and the usability of applications in an offline setting, these should be studied to inspire a real-time analysis of the user interface events prompted and use it for adaptation and personalisation purposes.

3. PERSONALISED ACCESSIBLE INTERACTION WITH TV APPS

Taking into account what is reported in the literature and the results of previous studies we proceeded to the design of a prototype capable to interact with TV applications, convey information and explore personalisation features.

3.1 Architecture

The components are distributed between two devices: the Set-Top Box (or Connected TV) and the mobile device. In the Set-Top Box (STB) reside the components related with the extraction of the content presented on the screen (e.g., menus, applications, etc.) and the execution of the commands received from the mobile device. On the mobile device are the components which decode and convey the content to the VI user through speech synthesis, and alternative methods of input for controlling the TV, which are converted to key events (as if sent by a remote control) in the Set-Top Box. The user events are logged and analyzed and adaptations are suggested to the user.

3.2 Set-Top Box

In order for an assistive technology such as this to convey to the VI user what is happening on the TV screen, it needs to have knowledge of the TV application's user interface. In order to be able to build a solution as generic as possible, one of the elicited requirements for this technology is to be language independent and compatible with any type of TV application (e.g., JAVA, HTML, etc.). For this reason we opted to use a standard User Interface Description Language (UIDL) to represent the TV application's interface so that it could be manipulated in the mobile application. Because it has been successfully used in a TV platform before [5], the User Interface Markup Language (UIML) was chosen.

When this information is extracted and the UIML document is built, it is sent to the connected mobile device. Every time a navigation command (i.e., key event) is received, a new UIML is generated to update the UI status on the mobile device also.

3.3 Mobile Application

The mobile application is the main interface through which VI users interact with the applications and menus of the STB. The application parses the information received from the STB and conveys it to the user through a speech synthesizer. The current version was implemented for Android devices and is compatible with TalkBack. Its interface simulates a remote control with adapted features for the blind. The mobile application's user interface makes available to users navigational buttons (i.e., Left, Right, Up and Down), confirmation (OK) button, Localize button, Read Screen button and Enter Speech Command button.

3.3.1 Alternative Modalities

TV based applications are usually navigated using four direction keys. The mobile application sends these commands by converting them into key codes and in the STB script they are converted into key events which are interpreted by the TV application. Independently from which input source it comes, the command goes through the component that interprets and sends the key code of the intended command. The goal is to integrate several alternatives such as speech and gestures to be made available to the user.

In terms of output, the content from the TV application is parsed from the UIML document and composed together in sentences. This feedback can be triggered by two different actions:

Navigation feedback. After a navigation command is sent and a new UIML is received containing information about the newly rendered interface, the application informs the user about the focused element and any additional information relevant for the navigation (e.g. index, orientation, etc). This feedback is also triggered by the Localize button.

Content feedback. When the user explicitly asks the system to read the interface by pressing the Read Screen button, the application conveys to the user all the content present in the TV application's UI. There are two feedback modes: **Concise** and **Verbose**. The first one gives less but sufficient information to the users while the latter adds contextual information (e.g. orientation of the menu). Verbose mode has 2 variations regarding Navigation feedback: **Siblings** and **Map**. The former lists the available options in the same menu of the focused element while the latter informs the user about the options around it.

3.3.2 Personalising the System for the User

The system gathers information from the user in two ways: explicitly by using a quick survey in the first interaction (e.g. experience with assistive technologies, severity of the visual impairment, etc.) and implicitly by storing the interaction events generated by the user.

Every interaction event logged contains the following information: an *id* identifying uniquely each event, *description* of the event, *type* and *orientation* of the current menu, *index* of the focused element, *modality* used in this event, *feedback* mode and a *timestamp*. Based on the literature presented in section 2.3, detection of the following patterns was included:

Irrelevant Actions. Actions that result in the same index will be considered irrelevant. This pattern can indicate the user is confused or does not know what to do next.

Action Re-occurrence, when the user performs an elementary action repeatedly. This can mean the user is not perceiving the feedback correctly so he or she persists in repeating actions.

Quick Scroll pattern means the user is searching and skimming for information. This can mean the user is an expert user of the TV application and is scrolling to quickly get to the required content, or there is a problem to locate it.

Lost Awareness. This pattern was implemented to understand if we can detect if an user is lost in the TV application. If the user is always using the localize feature one can assume there is a problem with the consistency of the TV application's UI.

The analysis of these patterns occurs after the total of user actions performed reaches a certain amount (i.e. a more frequent user of the system will trigger this analysis sooner than other user). The selected threshold represents several days of use, which prevents adaptations being triggered too often and annoying the user and affords a good volume of data to be analyzed. Additional patterns will be implemented as the iterative development process continues.

This information will then be the main source for the component responsible for adaptations to analyze patterns and decide which adaptation makes sense to suggest or apply. Several aspects can be adapted such as the amount of contextual information in the content and navigation feedback (Concise vs Verbose), speed and pitch of the speech synthesizer, font size and contrast values of the TV application and so on. This adaptations can be triggered in two ways: 1) after the user pattern analysis; and 2) immediate detection of a problem. In the former, adaptations are permanent, they are suggested by the system and only applied if confirmed by the user. In the latter, "quick fix" adaptations are applied but only momentarily. Additionally, the users have access to these variables through the settings giving total freedom for a more personalised experience.

The triggered adaptations depend on the detected patterns and current values of these adaptable variables. If Irrelevant Actions are detected and the feedback mode is set to Concise, then the system would suggest to change to Verbose as it offers contextual information that would help the user avoid this pattern. However, if Verbose is already selected, then a change in the navigation feedback is suggested (Siblings vs Map). A decrease of the audio render speed is suggested if Re-Occurences of the Read Screen action are detected as it can aid in memorizing better the information conveyed. On the other hand, Quick Scroll patterns can indicate an expert user and Concise mode is suggested. An example of a "quick fix" adaptation is the decrease of the volume of the STB when speech recognition issues are detected. These adaptations will be validated by user studies.

4. CONCLUSION AND FUTURE WORK

The developed technology allows visually impaired TV users to control TV based applications, overcoming the accessibility barriers that new TVs raise by controlling them via a device they are already familiar with, and that possesses a higher accessibility level. With this solution it is possible to interact with TV applications from the Opera TV store and properly describe their content. By itself, this is already a positive contribution over the lack of existing solutions [4]. However, we are convinced that the availability of alternative modalities and the integration of adaptive features can improve the accessibility of these platforms.

Several user studies are already taking place or being planned to validate different components of the system. The first study relates with the feedback that is rendered via audio to the user. We want to understand which and how much information is needed for VI users to understand the structure of the TV applications. The second study focuses on the multiple modalities that are made available. We are planning to test and compare the time and errors to complete tasks using 4 different alternatives: TalkBack, screen touch gestures, mid-air gestures and speech. Additionally, we want to understand user preferences irrespective of the performance factor. Finally, we want to evaluate the adaptation process from the detection of the patterns to the effectiveness of the suggested adaptations.

5. REFERENCES

[1] S. Azenkot and N. B. Lee. Exploring the use of speech input by blind people on mobile devices. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '13, pages 11:1–11:8, New York, NY, USA, 2013. ACM.

[2] S. Balbo, S. Goschnick, D. Tong, and C. Paris. Leading web usability evaluations to wauter. 01 2005.

[3] A. S. Bhachu. Inclusive technologies for enhancing the accessibility of digital television. *SIGACCESS Access. Comput.*, (99):9–12, Jan. 2011.

[4] D. Costa and C. Duarte. Visually impaired people and the emerging connected tv: A comparative study of tv and web applications' accessibility. *Univers. Access Inf. Soc.*, 16(1):197–214, Mar. 2017.

[5] C. Duarte, D. Costa, P. Feiteira, and D. Costa. *Building an Adaptive Multimodal Framework for Resource Constrained Systems*, pages 173–191. Springer London, London, 2015.

[6] C. Duarte, S. Desart, D. Costa, and B. Dumas. Designing multimodal mobile interaction for a text messaging application for visually impaired users. *Frontiers in ICT*, 4:26, 2017.

[7] D. M. Hilbert and D. F. Redmiles. Extracting usability information from user interface events. *ACM Comput. Surv.*, 32(4):384–421, Dec. 2000.

[8] U. Oh, L. Stearns, A. Pradhan, J. E. Froehlich, and L. Findlater. Investigating microinteractions for people with visual impairments and the potential role of on-body interaction. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '17, pages 22–31, New York, NY, USA, 2017. ACM.

[9] R. Oliveira, J. F. de Abreu, and A. M. Almeida. Promoting interactive television (itv) accessibility: an adapted service for users with visual impairments. *Universal Access in the Information Society*, 16(3):533–544, Aug 2017.

[10] R. Oliveira, J. Ferraz de Abreu, and A. M. Almeida. An approach to identify requirements for an itv audio description service. In *Proceddings of the 9th International Interactive Conference on Interactive Television*, EuroITV '11, pages 227–230, New York, NY, USA, 2011. ACM.

[11] S. Raheel. Improving the user experience using an intelligent adaptive user interface in mobile applications. In *2016 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET)*, pages 64–68, Nov 2016.

[12] B. Raufi, M. Ferati, X. Zenuni, J. Ajdari, and F. Ismaili. Methods and techniques of adaptive web accessibility for the blind and visually impaired. *Procedia - Social and Behavioral Sciences*, 195(Supplement C):1999 – 2007, 2015. World Conference on Technology, Innovation and Entrepreneurship.

[13] N. Sebe. Multimodal interfaces: Challenges and perspectives. *J. Ambient Intell. Smart Environ.*, 1(1):23–30, Jan. 2009.

[14] I. Shah. Event patterns as indicators of usability problems. *Journal of King Saud University - Computer and Information Sciences*, 20:31 – 43, 2008.

[15] M. V. Springett and R. N. Griffiths. Accessibility of interactive television for users with low vision: Learning from the web. In *Proceedings of the 5th European Conference on Interactive TV: A Shared Experience*, EuroITV'07, pages 76–85, Berlin, Heidelberg, 2007. Springer-Verlag.

[16] E. B. Union. Digital tv accessibility: Report on the current status in european countries. Technical report, European Blind Union, 2008.