



Audio rendering smart TV apps through mobile devices

Daniel Costa¹ · Carlos Duarte¹

Published online: 1 March 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

Abstract

TV service providers now offer a variety of features, including broadcast-related ones like electronic programme guides, catch-up and recording features, but also Internet access and a variety of TV applications. These new features turn TV devices into more versatile and interesting platforms, but also clog the screen with more content than ever. Since this content is mainly visual, this means that TVs have even more inaccessible content for visually impaired people. This paper presents the design of a solution that audio renders the TV application's user interface through a mobile device. Resorting to a mix of accessibility experts and user studies, we compared multiple feedback versions containing different contextual information. Participants reported that the use of repetitive sentences should be avoided; concise feedback feels smoother and quicker but, for some, lacks information while extended feedback can be annoying and take too much time though is appropriate for learning phases. For menus, most participants suggested to include the position of an element and the number of elements. The results helped to identify the critical information to convey to the user and to tailor two modes differing in the amount of contextual information provided, suitable for differently skilled users. Additionally, we condensed the findings into a list of design guidelines which can be generalized to other auditory interfaces meant to be operated by a VI user.

1 Introduction

Smart TV, also known as Connected TV, refers to any TV or set-top box that can be connected to the Internet and access content beyond the broadcast content that public digital services or private cable providers usually offer. Most Connected TVs provide that content through TV applications, similarly to what happens with mobile devices. Examples of popular TV applications include Netflix, Facebook or YouTube. The user interface of these applications tends to be simpler than desktop oriented applications. Nevertheless, it is highly reliant on visual content which can pose serious accessibility barriers to users with visual impairments. The variety of information displayed has been growing in content and complexity from the old analog TVs, through digital receivers and current Smart TVs. The lack of feedback about what is rendered on the screen is one of the main causes for these users not fully enjoying the capabilities of their televisions or stopping them from upgrading to newer sets [15].

Driven by a universal access vision, it is important to find a reliable solution with improved accessibility and accepted by visually impaired (VI) users. This is a complex task due to (1) the multiple existing Smart TV and TV applications that do not share common interface metaphors and elements, and (2) the characteristics of the target users and the requirement to convey clearly what is happening in the TV screen. This feedback must contain accurate information and at the same time not burden the user with too much detail, which would risk annoying users.

TV accessibility has been the focus of research efforts [12, 19, 21] but these have mainly concentrated on elderly people. There is a lack of literature detailing auditory feedback designed for VI users and TV applications. This work pursues this opportunity to contribute to a more accessible TV experience.

Smart TVs are characterized by an integration trend with mobile devices. In this paper we explore how this capability can be used to ameliorate the accessibility barriers of Smart TV, by allowing users to interact with their televisions through a device that already offers better accessibility solutions. Toward that end, we developed a software infrastructure that interprets the interface of applications rendered on TV and connects it with a mobile application. A VI user can then control the TV application through the mobile

✉ Daniel Costa
thewisher@lasige.di.fc.ul.pt

¹ LASIGE, Departamento de Informática Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisbon, Portugal

application. This is similar to existing TV remote control applications for mobile devices, but with enhanced features designed for the target population. The system is able to analyze and interpret the user interface of TV applications and build a document containing its structure and element properties. Then, the mobile application audio renders this information to the user.

Determining which information should be presented to fit the users' needs is the goal of the studies presented. We experimented with the rendering of labels of focused items, orientation of menus, listing of elements, elements around the focused one, segmentation of the interface into blocks and indexation. Different versions were implemented that present the contextual information with more or less detail.

The design process followed was iterative and involved representatives of different stakeholders. The initial version of the TV interface interpretation software and mobile application was grounded on a literature analysis and experience from past projects. This version was evaluated by accessibility experts, which resulted in an improved version. The second version was then evaluated in user trials with VI users, leading to suggestions for further improvements to be realized in a third version. This version was tested again in a new user study.

The findings helped to identify and tailor two different feedback modes: (1) *Concise*, which conveys short but critical information to the user; and (2) *Verbose*, which conveys additional contextual information that assists less experienced users. Additionally, the data gathered with the conducted studies allowed us to propose a set of guidelines that can be applied to other auditory interfaces.

In the remainder of this paper, we start by presenting background information about the problems VI users face operating their television sets. This is followed by related work on existing solutions supporting VI TV consumers, and on the integration of mobile and TV devices. Afterward, we present the methodology followed, the proposed technical solution, and describe in detail both the expert and the user evaluations. We conclude with a summary of the lessons learned so far and planned future developments.

2 Background and related work

This section presents related work in the area of TV accessibility. First, it identifies accessibility barriers and describes available solutions for generic TV accessibility, such as Audio Description (AD). Then, it focuses on TV applications and the integration of mobile and TV platforms.

2.1 The use of TV by visually impaired people

Comcast and the American Foundation for the Blind (AFB) report the results of a survey [14] of people with visual disabilities (626 visually impaired participants) showing that a majority spend four or more hours watching TV per day, almost as much as sighted users. Sixty-five percent of those surveyed encountered problems with looking up what's on TV and 53% experienced difficulty in following along with key visual elements. Less than half are aware of assistive technologies like audio description and talking TV guides. Those who are aware report that assistive technologies like audio description, text-to-speech and voice control are helpful as they watch television.

In [36], Oliveira et al. report the results of a study that had the goal of identifying VI users' problems and needs concerning the consumption of television. Ten participants were involved in the study of which five were partially blind. The authors reported that the participants watch TV on average 2 to 3 hours a day. The genres enjoyed the most were movies and series (42%), talk-shows and game shows (32%) and information (26%). Regarding interaction limitations, two participants stated they need help to adjust the television volume and other basic functions. The majority of the participants also feel that Digital TV introduced barriers to the way they consume TV. This was mostly caused by interactive services such as the Electronic Program Guide. Some participants reported that they had issues with the lack of feedback using the menus and got lost. All participants reported using the remote control and learned by themselves the location of the keys. On the bright side, participants find the Audio Description service very useful, more so if they could customize some of its properties such as the narrator's voice, speed and volume. They would also like to see features such as audio feedback of the different options available in the interface, possibility to switch language and access to a list of available audio description supporting channels.

2.2 Available solutions for VI people

2.2.1 Assistive technologies

“Assistive device” or “assistive technology” refers to any device that helps a person with any type of impairment or disorder to communicate with other people or interact with a device [33]. Assistive technologies often refers to tools, software or devices that help a person to clearly understand what is being said or shown and to easily express thoughts and actions.

One of the most used types of assistive technologies for VI users is the screen reader. Screen readers are able to

convey to the user the information presented on the screen through audio. One example of such technology is the Job Access With Speech (JAWS) [20], capable of controlling the operating system as well as navigating through Web pages in a computer. Another example, also widely used, is the NVDA screen reader [35]. Screen readers are essential for people with visual impairments and are used on other types of devices as well. People with visual impairments interact with their mobile devices using software like Google Talkback [23] for Android devices, or VoiceOver [5] for iOS devices.

Bigham et al. proposed WebAnywhere [9], a web-based and self-voicing web browser which can be accessed from any web browser without the need for any software installation. The content is sent to a server-side component that converts text to speech. Similarly, De Rosa and Justice [17] describe WebReader, a screen reader based on JavaScript for the web which does not require additional software and is free to use. WebReader can read all the headers or links; read all the headers of a given type (e.g., H1); stop the prompt and read again the current or the previous header or link, or move to the next; read the main content; identify and focus the main content. Ashok et al. [7] combine the screen reader with speech recognition in Capti-Speak and get better accuracy and usability results when comparing with traditional keyboard-controlled screen readers.

2.2.2 Audio description

Audio Description (AD) is the only TV specific solution for VI people mentioned in [8]. AD, by rendering the visual only information of broadcast content through speech, is known to make the visual content of films, soap operas, documentaries and other kinds of broadcast TV programmes more enjoyable, interesting and informative for VI spectators. Major TV content producers (like Netflix¹) and film studios (like Disney²) are increasingly including AD in their offerings.

The result is an increase in confidence and self-esteem as this group of users is able to discuss TV programmes without the fear of having misinterpreted the narrative or without the need for a family member or friend to describe the situation presented in the scene [40].

Despite being a very useful tool for the VI, AD is not capable of dealing with other important issues of modern TV sets. If a VI person cannot interact and navigate through the services and features provided by the TV, no amount of AD will be of use to them.

2.2.3 Audio rendering TV's screens

A European Blind Union report [42] states that, in order for a blind person to manage their digital receivers without any assistance, certain adaptations are needed. These include features like audio feedback of on-screen menus and for channel identification, font customization, increased audio description and an easy-to-handle remote control.

In [26], the authors present an Electronic Programme Guide (EPG) application which can be controlled using a mobile application or a remote control both endowed with speech recognition. The application's user interface is conveyed to the user through text-to-speech.

In the industry, companies such as Apple, Samsung and Google seem motivated to address the accessibility problems of their products. Apple TV offers access to VoiceOver, the built-in screen reader from Apple, commonly used by VI people who are iPhone owners. VoiceOver tells the users exactly what is on the TV screen through text-to-speech and incorporates gestures, which users are already familiar with, in their remote control [4]. Similarly, Samsung TV's Voice Guide [38] enables the television to read the text presented on the screen (for every menu and the EPG) as well as other important information such as volume, current channel and programme information. Android TV follows the same steps and allows the use of TalkBack, the assistive technology available on Android smartphones [22]. Comcast's X1 [13] offers a "talking guide" featuring a female voice that reads aloud selections like program titles, network names and time slots as well as DVR and on demand settings.

2.2.4 Haptic feedback

When browsing the internet, haptic solutions have been implemented to inform the presence of HTML elements on a user interface by using different techniques such as force feedback [30] or tactile pin representations [29]. Kuber et al. [28] proposed a structured approach to design assistive haptic feedback for use when exploring the Web. The use of Braille was also applied in a web browser for smartphones in [25]. In spite of being a crucial medium for blind and partially sighted people, the number of people able to read braille is small³ and the braille display devices are expensive.

As mentioned before, solutions already exist that provide visually impaired people with the semantic content of video (e.g., audio description) for TV. However, SensiTV [1] aims to convey emotions present in movies. Affi et al. focus on emotion recognition in multimedia content and

¹ <https://www.netflix.com/browse/audio-description>

² https://www.youtube.com/watch?v=O7j4_aP8dWA

³ European Blind Union, Making Information Accessible for All. URL: <http://www.euroblind.org/publications-and-resources/making-information-accessible-all>

the exploration of different modalities that can be used to translate those emotions. SensiTV makes use of lights, vibrations, emoticons, background mood music and the properties of subtitles to express the emotion to the user. Regarding VI users, vibrations and background music would be the most important modalities for this goal. Jieun Kim et al. [27] address the problem similarly, using different emotion recognition techniques, by providing a handheld device that applies different vibration patterns (amplitude, duration, delay and number of repetitions) accordingly to the emotion they wish to show. Additionally, Ariyasu et al. [6] synchronize broadcast programmes with haptic feedback through sonic transducers to enhance the viewer experience.

Haptic feedback can prove to be useful also in the context of TV applications with the use of a mobile device or vibrating remote control, especially for transmitting cues of existing events.

2.3 Connected TV applications

Current TV platforms offer access to popular applications familiar to users on other devices. These applications can either be installed using the popular app stores, such as the ones from Apple or Google if the TV runs their operative systems or accessed as Web-based applications (i.e., based on HTML5 and JavaScript). Support and access to these applications should also be provided for the VI.

In [15], the authors, after analyzing several commercially available TV platforms, report that most of them use a Web-based runtime environment. Consequently, it can be expected that these TV platforms bring to their users similar accessibility barriers they already face on Desktop platforms. Solutions designed to improve TV accessibility should resort to the knowledge that is already available for the design of Web applications for personal computers. This analysis groups TV applications found on current TV platforms into two classes: (1) TV applications, usually simpler and well designed, taking into account the characteristics of TV sets though with less functionality and content than their Web counterparts; (2) Web applications that can be accessed through the browser, if provided by the TV platform, which are more complex and suited to a desktop view.

Projects such as GUIDE [12] tried to solve some of the accessibility issues of TV applications by providing automatic adaptation of the user interface (e.g., contrast, font size, etc.). Although they were designed for elderly people, some are suitable for low vision users. Adaptations changing the layout of the application, however, were not well received by content producers. Therefore, a solution tailored for this group of users is of great interest and value for both viewers and content creators.

2.4 Integrating TV and mobile devices

Modern TV platforms are already integrated with mobile devices, which can be a key factor to support multimodal, accessible interaction. Currently, these devices provide, through a variety of applications [16], features such as programme guide information, related content on the web, or synchronization between shows and content [32, 41].

By taking advantage of the integration of second screens (i.e., mobile devices with their full capabilities) with TV platforms, these could be used as an alternative input and output device. This creates an opportunity for those users that are already familiar with the assistive features of their mobile devices. Additionally, it offers a solution that does not occlude the main television screen, which is particularly important when a VI spectator shares the living room with family or friends.

3 Methodology

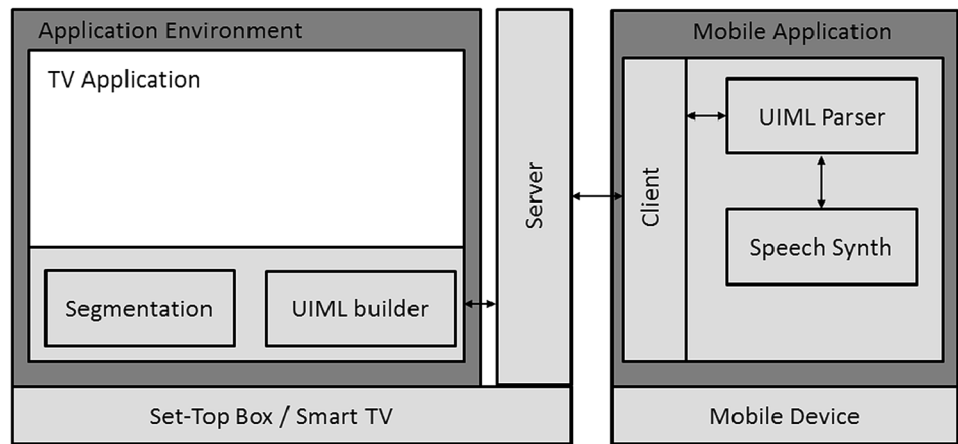
For the development of our proposed solution we followed an iterative design process, according to user-centered design (UCD) [34] principles, ensuring an active participation and involvement of target users. In our process, we also introduced expert design cycles in an attempt to streamline the initial phases of the development process. The first step in UCD is to define the context of use. For our solution, the users are VI people, the product is an assistive technology for TV platforms and the environment is the living-room (or other rooms where TV is consumed). After the initial definition, the development process followed iteratively, with four phases already completed: Early Requirements phase; Accessibility Expert Evaluation phase; and two user study phases. Between each phase, there is an implementation period where the requirements elicited in the previous phase are implemented in increasingly high-fidelity prototypes.

In the first phase, requirements were collected by gathering information from the literature, a survey with VI participants, and a usability study of off-the-shelf TV applications with VI users. From this phase, it resulted the design of the first prototype. The description of its architecture and components can be found in the following section.

The Accessibility Expert Evaluation phase makes use of the perspective and knowledge of experts to collect new information to improve the initial prototype. This evaluation screens the prototype for possible barriers and bugs that can frustrate target users. Improvements were then implemented to start preparations for the next phase. A detailed description of the procedure and results from this phase can be found in Sect. 5.

In the first User Study phase, we again collect information directly from the end users. Being the first User Study

Fig. 1 System’s architecture



in the development process, we were not looking for a summative assessment that would compare accessibility advantages and improvements of our proposed solution with other solutions. Therefore, a formative assessment is preferable to understand the users’ expectations about the system and their overall experience while using it. The results from this phase can be found in Sect. 6.

After analyzing the results of the first user study, further improvements to the interactive prototype were proposed and implemented. Section 7 presents these improvements and reports results from a follow-up User Study.

The studies involving VI users from a supporting institution were authorized and approved by the psychology department of the institution, with the head of the department serving as a gatekeeper between the researchers and the volunteers. In all studies, a written consent form was handed out to each participant informing the goals of the study and how the procedure would be conducted. For VI participants, the consent was read out loud and for those who could not sign it, an audio confirmation was recorded. All data were anonymously collected.

4 Early requirements and first prototype

A literature analysis was conducted to understand the problems VI users face when using their TV sets at home. We searched the ACM Digital Library for articles using keyword search. The keywords used in the search included “TV,” “interaction,” “accessibility,” “usability” or “user study.” We also included in the analysis other articles that were referred by the relevant articles we identified in the original search.

The findings from the analysis guided a survey with 26 visually impaired people in order to characterize their TV experience. The more frequently reported difficulties were accessing the programme guide and recording a show with their TV sets at home. Participants were asked if they used or had knowledge of any type of assistive technology for TV.

Only two have reported using radio provided Audio Description, a service offered by the public broadcast service of their country. In this study, the lack of accessibility of the participants’ TV sets was evident, 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 the lack of feedback was singled out as the main cause for this.

Additionally, a usability study with 5 VI participants was conducted where it was concluded that TV applications, from the former Opera TV, now called VewdOS⁴ platform, were not accessible as no participant could complete the requested tasks. Although the participants were using a known assistive technology for web content (JAWS), it was clearly difficult for the users to access the different sections of TV applications. The source of the problems originated from TV applications being designed in a different way from regular web pages. While JAWS navigation is mostly based on headers and links, TV applications expect keyboard commands to navigate and retrieve information, thus causing the problems found.

Taking all this into account, we proceeded to the design and implementation of a prototype that is expected to cope with the barriers found in these studies.

4.1 Architecture

Figure 1 depicts the architecture for the assistive technology solution. The components are distributed between two devices: the Set-Top Box (or Smart TV) and the mobile device. In the Set-Top Box, reside the components related with the extraction of the content presented on the screen (e.g., menus, applications, etc.) and in the mobile device are the components which decode and convey the content to the

⁴ VEWD OS. Website: <https://www.vewd.com/products-services/vewd-os/>

VI user through speech synthesis. The mobile application also serves as an alternative method of input for controlling the TV, sending commands that will be converted to key events (as if sent by a remote control) in the Set-Top Box.

4.2 Set-top box application

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 [18], the User Interface Markup Language (UIML) was chosen. UIML is an XML-compliant language that supports a declarative description of a user interface in a device-independent manner. The two most important nodes in a UIML document are the structure and style, where the content of the application is described. The module responsible for building this description is the UIML builder. Working together with the UIML module to assist in the identification of the UI structure, the Segmentation module is responsible to group elements into *blocks*. The two modules are explained in the following sections.

The proposed approach has two major advantages. First, it is universal, simply requiring that a converter from the language used for the TV application and UIML be developed (which we have done to convert from HTML and CSS to UIML). Second, it is independent of the application. It will work for new applications and it will continue working when existing applications are updated.

4.2.1 UIML builder

As mentioned before, most TV applications are Web-based. For this reason, our initial implementation of the UIML builder targeted Web technology. In the future, implementations targeting other application languages can be developed.

The process starts when the TV application is loaded and the script is injected by the browser running the Web application. We implemented a browser (based on Electron⁵) to perform the script injection, but Smart TV or Set-Top Box manufacturers could easily do the same with their devices' browsers. The UIML builder goes through the DOM tree of the loaded application in search for the visible elements.

From these elements, it starts building the structure of the application in the UIML document. This section is composed by <part> tags with *id* and *class* properties. The *id* is reused from the HTML element *id* if it has one; otherwise, the UIML builder generates an *id* for the element. The *class* property matches the tag name of the element. Because not all elements are eligible in a UIML document, the application's structure in UIML is made to be as close as possible to the Web application. Simultaneously, the style section is also built by taking information from CSS and HTML nodes, including text from *label* and *alt* properties.

4.2.2 Segmentation

Before sending the UIML document to the mobile device, additional information regarding the visual presentation of the TV application is collected. The segmentation module identifies visual groups of elements, which are cataloged as *blocks* and their orientation. The purpose of this information is to help the screen reader on the mobile device convey to the user possible menus, or other meaningful groups of items, as well as their screen orientation (horizontal or vertical), offering the user a better perception of which command to send to navigate within the application.

In order to segment the application, we use a page segmentation tool. As described by [3], Web pages are typically designed for visual interaction and include a number of visual segments. Usually, they are used to segment a Web page into a number of logical sections such as headers, footers or menus. Page segmentation tools try to identify these segments automatically. Block-o-Matic [39] was integrated into the Segmentation module for this purpose. Block-o-Matic takes advantage of three different sources of information: content, geometric and logical structure of the Web page. The content distinguishes the HTML elements used to hold and organize content. The geometric structure allows representing the page's objects based on their visual presentation while the logical structure describes the connections between blocks. The outcome of this process is a segmented Web page.

By analyzing the blocks created by Block-o-Matic, a new property is added to the style section of the UIML document which includes the block *id* to which an element belongs to and its orientation (calculated based on the geometric properties of the block). When this process is completed, the UIML document is sent to the server, which will forward it to the connected mobile device. Every time, the browser receives a navigation command (i.e., key event), and a new UIML is generated to update the UI status on the mobile device also.

⁵ Electron Website: <https://electronjs.org/>

4.3 Server

This system uses a regular client–server architecture with socket based connections. The server component is responsible for routing messages between the TV application environment and the mobile application environment. Messages from the Set-Top Box are mainly UIML documents while the mobile device sends mostly key codes, representing the commands issued by the user.

4.4 Mobile application

The mobile application is the main interface through which VI users interact with the TV applications and menus in the Set-Top Box. The application parses the information received from the Set-Top Box and conveys it to the user through a speech synthesizer. The current version was implemented for Android devices and is compatible with Talk-Back. Its interface simulates a remote control with adapted features for the blind. The initial version of the interface makes available to users Navigational (i.e., Left, Right, Up and Down), Confirmation, Read all, Repeat and Stop commands. We retained a visual interface for users with residual vision. Its design (size, contrast and spacing) was later improved with the feedback from the different studies.

Although it can be argued that introducing an additional application to control the TV application leads to an increase in the cognitive load of the user, there are two mitigating factors: (1) the mobile application is replacing the remote control; therefore, the number of devices stays the same; (2) mobile devices are the most accessible platforms at the moment.

4.4.1 Navigation

TV-based applications are usually navigated using the four direction keys available in a remote control. The mobile application sends these commands by converting them into key codes and a Set-Top Box script converts them into key events which are interpreted by the TV application.

4.4.2 Screen reader

Audio is the primary source of entertainment from the TV but also the source of information about interfaces and content for the visually impaired. VI users are used to overlapping audio sources with the use of assistive technologies in mobile and desktop environments. Mind that the audio sources are maintained in two different devices (entertainment audio comes from the TV, description of the UI comes from the mobile application). The content

from the TV application is parsed from the UIML document and composed together in sentences by the Screen Reader module. 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 navigation such as other options from the menu or possible navigation directions;
- **Content feedback.** When the user explicitly asks the system to read the interface by pressing the Read button, the application conveys to the user all the content present in the TV application’s UI.

Three navigation feedback options were implemented, containing more or less information regarding the options available to the user.

- **F1—Focus.** This version only informs the user about the current focused element by saying “*The focus is on <element>*”. It provides minimal information but gives the user quick feedback;
- **F2—Siblings.** This version also includes information about the siblings of the focused element, informing the user about possible selections in the current menu. The sentence is formed in the following way: “*The focus is on <element> and there are other <n> elements: <element₁ ... <element_n>*”;
- **F3—Navigation Map.** This version tells the user which element will be selected for each navigable direction. The sentence is built according to the following template: “*The focus is on <element>. If you go <direction_{xx} ...*”.

Regarding content feedback, two different versions were implemented:

- **L1—Linear**, where every element is read sequentially;
- **L2—Blocks**, where the information is grouped in segments or blocks. With this option the sentence is built as follows: “*Begin block. <element₁ ... <element_n>. End block.*” for each block identified.

5 Accessibility expert evaluation

This section presents the results of the second phase, including the experts’ evaluation and the resulting improvements to the prototype.

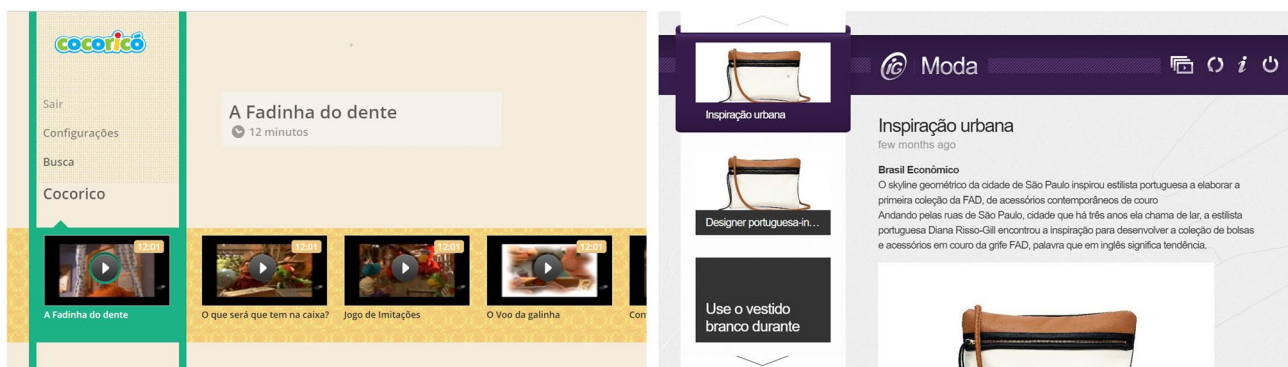


Fig. 2 Vewd store’s Cocorico (left) and IG Moda (right) applications

5.1 Methodology

Five accessibility experts were recruited via snowball sampling and ranged from researchers with more than 10 years of experience in this field to PhD students (with at least 3 years of experience) that are actively researching in the area of accessibility in fields such as mobile and web accessibility, or automated accessibility evaluations, focusing on visually impaired users. The participants evaluated themselves in a 5-point Likert scale in terms of accessibility expertise level ($M = 4.6, SD = 0.55$), familiarity with assistive technologies ($M = 3.6, SD = 1.67$) and TV applications ($M = 2.8, SD = 1.3$).

The Set-Top Box was simulated in a desktop computer. Participants did not have access to the monitor, and an Android phone was provided with the application. This study was made in a laboratory environment and each individual session lasted around 1 hour. We asked the experts to “think aloud” while experimenting with it. We registered their thoughts, comments and suggestions. While tasks were asked in no specific order (see Table 1), a list of the applications’ features was provided to the experts to make sure the participants tried and assessed all of the functionalities. Two TV applications from the Vewd store, using different templates, were employed in the evaluation (Fig. 2). The three different conditions for focus feedback (**Focus**, **Siblings** and **Map**) and the two regarding screen reading (**Linear** and **Blocks**) were randomly assigned during the session and all tested.

5.2 Results

This section presents the results of the analysis of the registered thoughts, comments and suggestions of the accessibility experts.

5.2.1 Navigation feedback

Overall, the differences between **Focus**, **Siblings** and **Map** were easily perceived by the experts.

Focus presented less contextual information to the user and participants felt somewhat lost. One participant thought there was only one option in a menu while another stated “The only thing I know is that the focus is on Pause but I do not know anything else.” Another participant found this to be the more pleasant version to interact with.

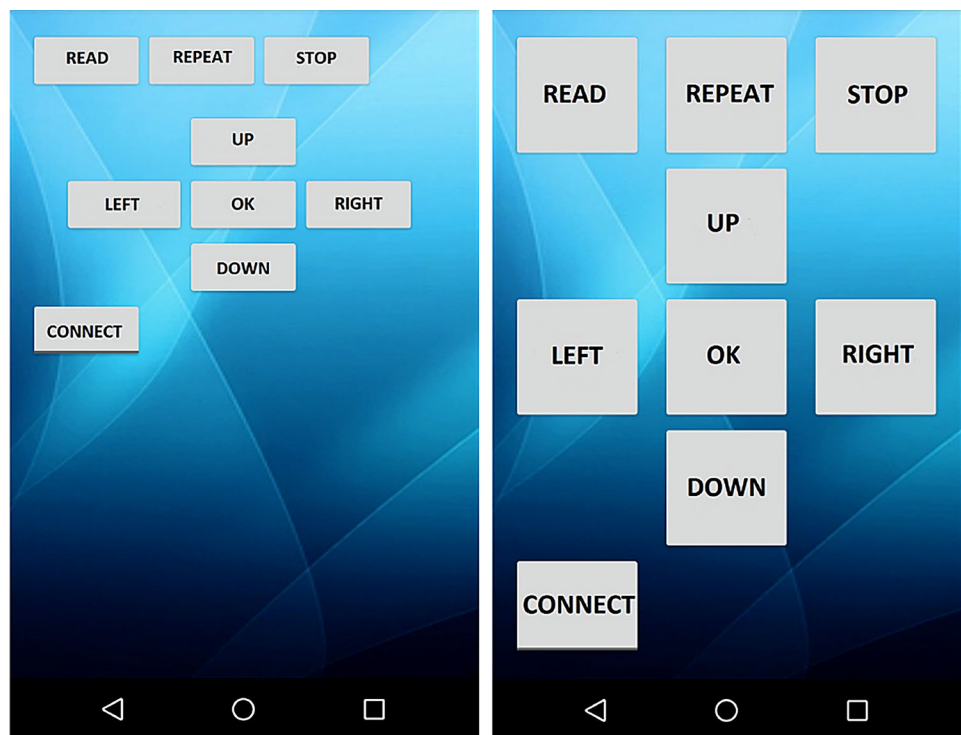
Siblings gives additional information about the selectable elements in a menu. Most participants felt that in spite of having this “general information” of what they can select, “only by experimenting” they can know the direction of the element they want to focus on.

Map solved this problem for some participants: “Now I can understand the order of the menu as well as the submenu” and it “gives more clues about my location.” However, one expert stated that the feedback “should be faster. It is more clear but complex and tiring, it should be simpler.”

Table 1 Tasks performed in both TV applications

Application	Task Description
Cocorico	Locate yourself
	Navigate to a specific menu item
	Choose a video and try to pause it
	Return to the list of videos
IG Moda	Describe the structure of the application
	Locate yourself
	Navigate to a specific menu item
	Choose a news article
	Retrieve a specific piece of information

Fig. 3 Mobile app UI: Old (left) and new (right)



5.2.2 Content feedback

When comparing **Linear** with **Blocks**, the experts preferred the latter. One expert mentioned the blocks information helps to understand the structure of the page but “there should be more context,” such as position and orientation of the blocks. **Block** makes “sense in an exhaustive reading” scenario but “it should read the menus first, then the content.” Overall, **Linear** confused the participants as most could not differentiate informative content from menu items.

5.2.3 Interaction

Regarding interaction, some of the issues found are related with the design options implemented in the TV applications. One had circular menus while the other did not and some menus were difficult to reach. One expert commented that “it makes no sense to go to the upper menu from the article. I think people with visual impairments would have difficulties to navigate in this application.” Additionally, most participants had problems when navigating in the video player. This is mostly a result of the video player implemented in the TV applications, which have a navigation bar that disappears after a few seconds (and requires a key press to enable it again). This behavior of the video player was not reported by the system. One of the participants tried to pause the video using the Stop button in the mobile application.

The navigation in the TV application had some issues specific to the navigation modes. **Siblings** does not inform

users about the orientation of the menu causing some participants to navigate left or right when they should go up or down or vice versa. One expert reported that he would prefer **Map** while learning the TV application but then he would go with **Focus** (with less feedback). It was also suggested that the four directions should be reduced to two, similar to the interaction modes provided by TalkBack and VoiceOver.

5.2.4 Suggestions and comments

Regarding feedback, the experts suggest that, instead of repeating the same content, the system should tell the user that there are no more options in the direction the user is navigating, when that is the case. Feedback should also be more fluid and the sentence “The focus is on ...” should be removed as “it becomes annoying”. Additionally, more context should be added to the content feedback in **Blocks** regarding position and orientation. The Stop button should be used to skip options, instead of completely stopping the feedback.

Regarding the mobile application’s interface, the experts suggested to scale up the buttons and maximize the space used, as this will reduce problems for VI users when exploring the interface with their fingers using TalkBack.

New feature suggestions include different modes for more experienced users (with less feedback and more functionalities) and a button to localize the user in the application (similar to the navigation feedback).

5.2.5 Implications for the next iteration

Overall, this study has produced important feedback to improve the assistive technology and no major issues were found that would prevent a study with VI participants. Taking into account, the suggestions from the accessibility experts an updated prototype was implemented. The interface was updated with larger buttons (see Fig. 3). The Stop button now skips portions of the sentence making it possible to go through the different options quickly. **Blocks** and **Siblings** have been upgraded to include information about the orientation of the content. In **Blocks**, feedback now informs if the current menu is vertically or horizontally oriented. In **Siblings** feedback also provides the orientation of each block. The additional feedback aims to help the user locate herself or himself in the page, the content and which directional keys to use to navigate within the menus.

Other issues found during the study were addressed. When users navigated too fast, the system would lag behind, raising synchronization issues between the feedback being provided by the mobile application and the content being rendered in the television. This was solved by optimizing the Set-Top Box script and interrupting the speech in the mobile application when required. Not all suggestions regarding feedback were implemented in the updated prototype as we were expecting to hear the perspective of our target users as well. Therefore, after completing the additions and a stable version of the system was ready, the study with VI users was prepared.

6 User study

This section presents the results from the third phase of this design process, a user study of the improved prototype.

6.1 Methodology and participants

Similar to the previous study, the focus of this phase is a formative assessment of the prototype. We followed the same methodology that was described in the previous study, but using the improved prototype.

We recruited 12 visually impaired participants from an institution supporting visually impaired people. The participants ranged from 25 to 62 years of age ($M = 44.9, SD = 12.23$), all male with some kind of visual impairment (Seven participants reported to have some residual sight, the remaining were blind). On average, participants lost their vision since their early twenties ($M = 20.8, SD = 15.94$). From observation of their interaction with desktop and mobile devices (all based on screen

readers), it was discerned that participants do not rely on sight when using interactive devices even though they report to be partially sighted.

Participants evaluated themselves in a 5-point Likert scale in terms of familiarity with assistive technologies ($M = 3.67, SD = 0.89$) and TV applications ($M = 1.92, SD = 0.90$). Note that most of the participants were students in the institution and still learning some of the assistive technologies available for the different devices, mainly screen readers. For this study, only users of smartphones, the main interaction device in our system, were recruited. When using computers, most participants report using NVDA (8) or JAWS (4), while when using mobile devices they use VoiceOver (5) or TalkBack (7).

This study follows the same setup as the previous evaluation and each session lasted around 90 minutes. We explained the UI and functionalities before starting the session and let participants experiment until they felt comfortable. Participants used a smartphone provided by us and the study was conducted in the institution facilities. Participants were asked to “think aloud” while experimenting the prototype and performing the tasks (see Table 1), with an observer registering their comments and suggestions.

6.2 Results

This section presents the results of the analysis of the observations and transcriptions collected in the user study.

6.2.1 Navigation feedback

Most participants felt they need more information when using **Focus**. “This way a person feels lost” or “It lacks information” were some of the comments. However, three of the participants reported that less information is better because it is less confusing and more useful after knowing the application. Only one participant said explicitly that this was the preferred feedback mode.

There was no clear preference between **Siblings** and **Map**. Each was chosen by 5 participants. It was mentioned several times that **Siblings** presents “too much information.” **Map** was not considered an annoyance, except when the entire article was read when giving the directions to the user.

Most of the participants issued a common request for the three versions: notification of the current option (numerical) position in the menu and the total number of items in the menu (similar to what is done in TalkBack or VoiceOver).

6.2.2 Content feedback

There was no clear preference between **Linear** and **Blocks**. The former was selected by 4 participants while 5 chose the

latter, with the remaining showing no preference. **Blocks** were considered useful by most but “boring.”

6.2.3 Interaction

Usually, iPhone users employed sequential navigation while Android users mainly used exploration. Some participants used a mix of the two methods. Participants that used sequential navigation, swiping left and right, were slower and interrupted the speech more often.

Regarding interaction with TV applications, most difficulties were related with the disappearing video player menu and scrolling news articles. The disappearance of the video player menu was not perceived by the VI participants as there is no feedback other than the visual. This was explained to the study participants and they overcame the issue by issuing two consecutive commands. The second issue happens when a scrollable article is focused. In this situation, the Down (and Up) button switches its behavior from navigation to scroll until the end of the article is reached. Once again, there was no feedback other than the visual, which meant that the participants did not perceive this change.

Importantly, some participants had difficulty in understanding the concept of the mobile device serving as a remote control and realizing that there were two different applications running. These participants expected the TV application to run in the smartphone, instead of using the smartphone as a controller of the TV application. This resulted in different behaviors that are representative of this lack of understanding. When trying to pause a video, several participants tried to use the Stop button of the mobile interface. In other situations, participants used the Read button to activate the Info button in the TV application. Additionally, VI users are used to linear navigation in their applications, not to the horizontal and vertical navigation of TV applications. However, after a short time interacting with the application these difficulties receded.

Some participants swiped through all the options in the mobile interface expecting new options and content from the TV application.

6.2.4 Suggestions and comments

All participants reported that using the mobile device to control the television was an adequate solution. However, one said “it would be more inclusive if the remote control had audio feedback,” while another would prefer an interaction similar to the smartphone.

Most participants stated that the feedback sentences should be generally shorter, removing the “The focus is on ...” from the beginning and the “If you go <direction> ...” in **Map**. Several participants suggested to add the index

of menu items. The Repeat button should only repeat the navigation feedback as the content can be triggered with the Read button, making Repeat redundant.

The majority of the participants suggested to have two modes of feedback: Verbose mode where more context and feedback is described to the user (**Siblings** or **Map**); and a Concise mode where less information is conveyed to the user, useful when she or he is already familiar with interaction and applications (**Focus**).

6.2.5 Implications for the next iteration

Following the participants’ suggestions, two modes of feedback were implemented: the **Verbose** mode which offers more contextual information about the interface of the TV applications to the user (this mode includes the possibility to choose between **Siblings** or **Map** and always makes use of **Blocks**); and the **Concise** mode where minimal information is reported to a more experienced user (**Focus** and **Linear**).

The “Repeat” button is now named “Locate” and prompts the navigation feedback as the previous functionality was deemed redundant. Additionally, information considered useless by the participants was removed from the rendered sentences such as “The focus is on...”. When **Map** finds a huge piece of information (i.e., greater than 100 characters), it is no longer described, instead “Block of text” is rendered. All navigation feedback includes the index and total number of elements in the menu.

By scanning HTML elements for tags or properties that hint for the presence of a video player in the application, two commands are now sent (one to activate the video player control and one to perform the actual user action) when the focus is on the video player. This adaptive technique prevents the VI user from having to deal with the disappearing menu issues and having to perform two actions instead of one. Additionally, an earcon [10, 43] was implemented to warn the user if a command did not produce any change in the TV application. This feature helps the user to understand if a menu is circular or not; if a certain command is useless in that context; and avoids the repetition of feedback if nothing has changed.

7 User study with updated prototype

This section presents the results from a new user study which aims to evaluate the improvements originating from the previous phase. It is also the last validation study before a new iteration cycle that will implement adaptive and multimodal features which are expected to further improve the accessibility of TV applications for VI users.

7.1 Methodology and participants

The methodology for this study is similar to the ones conducted before, with all the procedures done in the same way as the previously presented studies.

Five participants were recruited from the same institution as the previous study, all with experience using smartphones and assistive technologies from different devices. Participants ranged from 33 to 57 years of age ($M = 47.6, SD = 9.15$), three males and two females with a visual impairment (only one participant reported to have some residual sight, the remaining were blind). Participants evaluated themselves in a 5-point Likert scale in terms of familiarity with assistive technologies ($M = 4.20, SD = 0.45$) and TV applications ($M = 1.80, SD = 0.84$).

7.2 Results

This section presents the results of the analysis of the transcriptions and observations resulting from this user study.

7.2.1 Interaction

While interacting with the smartphone, only one of the participants used the swipe technique to cycle between the buttons in the mobile application. Similar to the previous phase, this user presented difficulties situating himself in the TV application as this participant kept swiping and interrupting the speech synthesizer. Specifically, when asked to select a predetermined video or option, the participant ignored the feedback given by the mobile application and swiped the screen making TalkBack speak over the information needed to perform that task. The remainder of the participants used finger exploration to find the desired buttons. Those who opted for this technique had smoother and quicker interactions, expressing no difficulties interacting with the mobile application and consequently with the TV applications.

7.2.2 Feedback modes

In general, the participants could understand the differences between the two versions of the **Verbose** mode (**Siblings** vs **Map**) as well as the differences between the **Concise** and **Verbose** modes. No major difficulties were found when interacting with the different modes.

All participants agreed that the **Verbose** mode “describes everything” while the **Concise** mode presents less information and “it’s more direct.” When comparing **Siblings** and **Map** after the session, participants report that the former presents the information in a more extensive way which can be “too much” and “takes too much time” (specially on menus with several options). However, it gives the user a notion of what “is there.” The latter is simpler, more useful

and “gives indications of where can I go”. In the end, four out of five preferred **Map** while the other participant prefers **Siblings** “because I like to understand all the available options”.

All participants agreed that the **Concise** mode has the information they need to navigate and interact with the TV applications although they would prefer to start with the **Verbose** mode to learn it first.

8 Design guidelines

After carefully analyzing the results from the three studies presented, we condensed the findings into a list of design guidelines which can be generalized to auditory interfaces running in other devices and environments meant to be operated by a VI user. The guidelines are listed below.

Beginner and expert users of auditory interfaces need different levels of descriptions. For some participants, it became evident that less information, allowing for quicker interaction and rapid understanding of what is happening on the screen, is more important than detailed but exhaustive descriptions. For others, understanding the layout of the TV application is more important, since it allows for smoother and safer interaction. Alonso et al. [2] also proposed that the presentation should make provision for several detail levels. On the one hand, novice VI users will want to receive as much information as possible of each interface element as they learn to use the application. On the other hand, expert users will only want to receive the information that they need to do the job. This highlights the importance of providing different options or modes that fit the preferences of the user (which can change with time and experience). Taking this into account, it is recommended that an auditory interface has at least the two modes presented in this paper.

Concise descriptions for expert users of auditory interfaces should include at least the label of the focused element, its index and total elements in the container. Alonso et al. [2] points out that the element’s name has to appear first in speech to enhance navigation speed and Rajapakse et al. [37] concluded that there should be sufficient amount of depth cues to identify the current position, specially in a two dimensional audio interface. In our studies some participants mentioned that the inclusion of an index of the focused element and the total number of elements in the menu should be included in all feedback versions. This is an indication that VI users rely mainly on this information in their navigation, thus an auditory interface that offers concise feedback should at least include this important information.

When the auditory interface is not linearized the orientation of the container should be conveyed. In [37], the authors reported that VI users perform better with linearized

solutions due to the frequent usage of linear interaction models in their current day-to-day activities. However, the authors suggest that for nonlinearized navigation there should be an easily understood navigational layout. Experts from the first study identified issues concerning navigation within the TV application. These were related with the lack of feedback regarding the menus' orientation, causing participants to navigate in wrong directions. An auditory interface that does not implement linear navigation should include the orientation of the container in its feedback.

Earcons should be used in auditory interfaces to signal non-circular menus or actions without effect. Participants could identify circular menus correctly, whereas participants navigating non-circular menus experienced several problems such as insisting in the same direction although the feedback returned was always the same. To mitigate this issue, it is recommended that earcons are implemented warning the user that their action did not produce effects. Brewster and Crease [11] could overcome menu selection issues and menu slips or mis-selections with the use of earcons.

It should be conveyed to the user when user interface elements disappear or hide after a period of inactivity. Sometimes user interfaces contain elements that disappear after a period of inactivity. This should be avoided. If not, then it should be conveyed to the user (e.g., through an earcon) or overcome by adaptive features making this issue unnoticeable. For instance, in Mercator [31] whistling sounds are used to notify the appearance or disappearance of pop-up windows. In our case study, the problem of the video player was resolved with an additional command sent to activate the video player control menu before sending the intended one. Additionally, this solution avoided the need to perform two commands in a short period of time to maintain the menu visible.

Auditory interfaces should be designed in a way that do not clash with audio-based assistive technology. It was identified in the studies that sometimes the auditory feedback was interrupted by the native assistive technology of the mobile device mainly provoked by using the swipe technique. Although the system was compatible with it from an input perspective, the feedback produced by the assistive technology clashed with the feedback produced by the mobile application. If the auditory interface runs in parallel with other assistive technologies, an effort should be made to avoid this problem. One solution could be the use of concurrent speech renders with different pitch properties [24].

9 Conclusions and future work

The development of assistive technology is a challenging process. The participation of end users in the development process of interactive solutions has been recognized as very

important. We could argue it is paramount when the interactive solution targets impaired populations. For that reason, employing UCD methods is mandatory. In the design process reported in this article, we aimed to combine expert based evaluations with user studies. The former aim to assess and identify accessibility problems in the system's concept, design, interaction and user interface. Specially in the earlier stages of development, this informal method of evaluation can be very beneficial for the development process. When conducted prior to user testing it can reduce the number and level of accessibility barriers that participants would face. Although one can argue results can be influenced by the bias of the experts, gathering a diversified set of experts in terms of experience and fields inside human–computer interaction (e.g., web accessibility, mobile accessibility) mitigated this effect and provided a reliable and sound source for feedback.

User studies are one of the most important tools of the UCD approach as it involves the target users in the development process. Although one can expect that users would not give feedback as technically detailed as experts, the amount of knowledge users have about the tools they are used to in their every day lives can be surprising. While many of the conclusions reached by the experts were also found by the VI participants, other identified requirements were never mentioned by the experts. While this could be an argument for not conducting an expert evaluation, it has to be pointed out that the resources (human and time) required for the expert evaluation are much lower, thus justifying it.

The input received from the users helped in the design of two feedback modes that were not previously planned but emerged naturally from the observed interactions. Following the participants' feedback, the grouping of the different contextual feedback options led to the **Concise** and **Verbose** modes. The amount of information the **Verbose** mode offers about the content and navigation of the TV application was indicated by the participants as suitable for the initial learning phase. However, participants did not enjoy such informative content for long periods of time, reporting that it can be “too much” and “takes too much time.” The **Concise** mode offers less information but gives a quicker and responsive interaction which is the ultimate goal of the majority of the users after some time spent in the **Verbose** mode.

Comparing with existing solutions, our proposal has the potential to increase the accessibility of TV applications by providing users of assistive technology with more information about the contents of TV applications and how the content is structured, and by offering users the possibility to switch between feedback modes, with two modes suited to different contexts: one for when the user needs more information about the TV application, maybe due to not being familiar with it yet, and another for when the user does not need all that information and a better user experience can be achieved with shorter, more focused feedback. Additionally,

the proposed solution was tested with existing TV applications and did not require any changes to the applications, which represents an advantage for vendors of TV applications that will be able to reach a larger audience without the need to adapt their products.

Future work comprises the integration of multimodal interaction, including speech recognition for text input or for giving commands, haptic feedback and adaptive features, the need for which was identified in the reported studies: automatically adapting the interaction with TV applications to the way the user interacts with her or his smartphone (especially in trying to address the differences between swipe and exploration methods of interaction with mobile assistive technologies) and adapt the control application features to the level of expertise of its user.

Acknowledgements This work was supported by FCT through the LASIGE Research Unit, ref. UIDB/00408/2020 and ref. UIDP/00408/2020 and by the Reitoria da Universidade de Lisboa through the PhD Grant BD2015. We would like also to thank Fundaçao Raquel e Martin Sain for providing the volunteers for the studies.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Affi, D., Dumoulin, J., Bertini, M., Mugellini, E., Abou Khaled, O., Del Bimbo, A.: Sensitiv: Smart emotional system for impaired people's tv. In: Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video, ACM, New York, NY, USA, TVX '15, pp. 125–130, (2015) <https://doi.org/10.1145/2745197.2755512>
- Alonso, F., Fuertes, J.L., González, Á.L., Martínez, L.: User-interface modelling for blind users. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) *Computers Helping People with Special Needs*, pp. 789–796. Springer, Berlin (2008)
- Andrès Sanoja, S.G.: Block-o-matic: a web page segmentation tool and its evaluation. In: 29^{ème} journées “Base de données avancées”, BDA'13 (2013)
- Apple Accessibility: Tv that everyone can control. <http://www.apple.com/accessibility/tv/vision/>, accessed 18 September (2018)
- Apple Inc Vision accessibility—iphone—apple. <https://www.apple.com/lae/accessibility/iphone/vision/>, accessed 07 July (2019)
- Ariyasu, K., Kawakita, H., Handa, T., Kaneko, H.: Tactile sensibility presentation service for smart tv. In: 2014 IEEE 3rd Global Conference on Consumer Electronics (GCCE), pp. 236–237, (2014) <https://doi.org/10.1109/GCCE.2014.7031287>
- Ashok, V., Borodin, Y., Puzis, Y., Ramakrishnan, I.V.: Captispeak: a speech-enabled web screen reader. In: Proceedings of the 12th Web for All Conference, ACM, New York, NY, USA, W4A '15, pp. 22:1–22:10, (2015) <https://doi.org/10.1145/27455.2746660>
- Bhowmick, A., Hazarika, S.M.: An insight into assistive technology for the visually impaired and blind people: state-of-the-art and future trends. *Journal on Multimodal User Interfaces* pp. 1–24, (2017). <https://doi.org/10.1007/s12193-016-0235-6>
- Bigham, J.P., Prince, C.M., Ladner, R.E.: Webanywhere: a screen reader on-the-go. In: Proceedings of the 2008 International Cross-disciplinary Conference on Web Accessibility (W4A), ACM, New York, NY, USA, W4A '08, pp. 73–82 (2008). <https://doi.org/10.1145/1368044.1368060>
- Brewster, S.A.: Using nonspeech sounds to provide navigation cues. *ACM Trans. Comput.-Hum. Interact.* **5**(3), 224–259 (1998). <https://doi.org/10.1145/292834.292839>
- Brewster, S.A., Crease, M.G.: Correcting menu usability problems with sound. *Behav. Inf. Technol.* **18**(3), 165–177 (1999). <https://doi.org/10.1080/014492999119066>
- Coelho, J., Duarte, C., Biswas, P., Langdon, P.: Developing accessible tv applications. In: The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility, ACM, New York, NY, USA, ASSETS '11, pp. 131–138, (2011). <https://doi.org/10.1145/2049536.2049561>
- Comcast: Comcast launches new accessible tv features ahead of 2018 paralympics. (2018). <https://corporate.comcast.com/stories/comcast-launches-new-accessible-tv-features-ahead-of-2018-paralympics>. Accessed 26 April 2019
- Comcast Study: Visually impaired adults tune-in to television almost as much as general public. <https://corporate.comcast.com/press/releases/study-visually-impaired-adults-tune-in-to-television-almost-as-much-as-general-public>. Accessed 26 April 2019 (2018)
- Costa, D., Duarte, C.: Visually impaired people and the emerging connected tv: a comparative study of tv and web applications' accessibility. *Univ. Access Inf. Soc.* **16**(1), 197–214 (2017). <https://doi.org/10.1007/s10209-016-0451-6>
- Courtois, C., D'heer, E.: Second screen applications and tablet users: constellation, awareness, experience, and interest. In: Proceedings of the 10th European Conference on Interactive Tv and Video, ACM, New York, NY, USA, EuroITV '12, pp. 153–156, (2012). <https://doi.org/10.1145/2325616.2325646>
- De Rosa, A., Justice, D.: Webreader: A screen reader for everyone, everywhere. In: Proceedings of the 13th Web for All Conference, ACM, New York, NY, USA, W4A '16, pp. 10:1–10:2, (2016). <https://doi.org/10.1145/2899475.2899510>
- Duarte, C., Costa, D., Feiteira, P., Costa, D.: Building an Adaptive Multimodal Framework for Resource Constrained Systems, Springer London, London, pp. 173–191. (2015). https://doi.org/10.1007/978-1-4471-6708-2_9
- Epelde, G., Valencia, X., Carrasco, E., Posada, J., Abascal, J., Diaz-Orueta, U., Zinnikus, I., Husodo-Schulz, C.: Providing universally accessible interactive services through tv sets: implementation and validation with elderly users. *Multimed. Tools Appl.* **67**(2), 497–528 (2013). <https://doi.org/10.1007/s11042-011-0949-0>
- Freedom Scientific Blindness solution: Jaws®. <http://www.freedomscientific.com/Products/Blindness/JAWS>. Accessed 18 September (2018)
- Godard, N., Pecci, I., Isokoski, P., Weslide: Gestural text entry for elderly users of interactive television. In: Proceedings of the 11th European Conference on Interactive TV and Video, ACM, New York, NY, USA, EuroITV '13, pp. 55–58, (2013). <https://doi.org/10.1145/2465958.2465963>
- Google Android tv help: android tv accessibility features. <https://support.google.com/androidtv/answer/6123320?hl=en>. Accessed 18 September (2018)
- Google Get started on android with talkback. <https://support.google.com/accessibility/android/answer/6283677?hl=en>. Accessed 07 July (2019)
- Guerreiro, J., Gonçalves, D.: Faster text-to-speeches: enhancing blind people's information scanning with faster concurrent

- speech. In: Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility, ACM, New York, NY, USA, ASSETS '15, pp. 3–11, (2015). <https://doi.org/10.1145/2700648.2809840>
25. Jassi, I.S., Ruchika, S., Pulakhandam, S., Mukherjee, S., Ashwin, T.S., Reddy, G.R.M.: Ember: a smartphone web browser interface for the blind. In: Proceedings of the 7th International Symposium on Visual Information Communication and Interaction, ACM, New York, NY, USA, VINCI '14, pp. 106:112, (2014). <https://doi.org/10.1145/2636240.2636847>
 26. Johnston, M., Stent, A.J.: Epg: speech access to program guides for people with disabilities. In: Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility, ACM, New York, NY, USA, ASSETS '10, pp. 257–258, (2010). <https://doi.org/10.1145/1878803.1878859>
 27. Kim K.J.R.H.J: 'heart-to-feel': A new audio description coding scheme for the visually impaired on affective cinematography and emotive vibration. In: Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video, TVX '14 (2014)
 28. Kuber, R., Yu, W., McAllister, G.: Towards developing assistive haptic feedback for visually impaired internet users. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA, CHI '07, pp 1525–1534, (2007). <https://doi.org/10.1145/1240624.1240854>
 29. Kuber, R., Yu, W., O'Modhrain, M.S.: Tactile web browsing for blind users. In: Nordahl, R., Serafin, S., Fontana, F., Brewster, S. (eds.) Haptic and Audio Interaction Design, pp. 75–84. Springer, Berlin (2010)
 30. Magnusson, C., Tan, C.C.S., Yu, W.: Haptic access to 3d objects on the web. In: Proceeding of the Eurohaptics 2006, p. 4 (2006)
 31. Mynatt, E.D.: Transforming graphical interfaces into auditory interfaces for blind users. *Hum.-Comput. Interact.* **12**(1), 7–45 (1997). https://doi.org/10.1207/s15327051hci1201&2_2
 32. Nandakumar, A., Murray, J.: Companion apps for long arc tv series: Supporting new viewers in complex storyworlds with tightly synchronized context-sensitive annotations. In: Proceedings of the 2014 ACM International Conference on Interactive Experiences for TV and Online Video, ACM, New York, NY, USA, TVX '14, pp. 3–10, (2014). <https://doi.org/10.1145/2602299.2602317>
 33. National Institute of Deafness and Other Communication Disorders (NIDCD) Assistive devices for people with hearing, voice, speech, or language disorders. <http://www.nidcd.nih.gov/health/hearing/pages/assistive-devices.aspx>, Accessed 26 February 2016, (2011)
 34. Norman, D.A., Draper, S.W.: User Centered System Design. New Perspectives on Human-Computer Interaction. L. Erlbaum Associates Inc., Hillsdale, NJ, USA (1986)
 35. NV Access (2019) Nvda. <https://www.nvaccess.org/about-nvda/>. Accessed 07 July (2019)
 36. Oliveira, R., Ferraz de Abreu, J., Almeida, A.M.: An approach to identify requirements for an itv audio description service. In: Proceedings of the 9th International Interactive Conference on Interactive Television, ACM, New York, NY, USA, EuroITV '11, pp. 227–230, (2011). <https://doi.org/10.1145/2000119.2000166>
 37. Rajapakse, R., Dias, M., Weerasekara, K., Dharmaratne, A., Wimalaratne, P.: Audio user interface for visually impaired computer users: in a two dimensional audio environment. In: Proceedings of World Academy of Science, Engineering and Technology (2012)
 38. Samsung Accessibility settings for your tv. <https://www.samsung.com/us/support/answer/ANS00062238/>. Accessed 18 September (2018)
 39. Sanoja, A., Gañçarski, S.: Block-o-Matic: a Web Page Segmentation Tool and its Evaluation. 29ème journées "Base de données avancées", BDA'13, <https://hal.archives-ouvertes.fr/hal-00881693>, poster (2013)
 40. Schmeidler, E., Kirchner, C.: Adding audio description. does it make a difference? *J. Vis. Impairment Blind.* 197–212 (2001)
 41. Silva, P., Amer, Y., Tsikerdanos, W., Shedd, J., Restrepo, I., Murray, J.: A game of thrones companion: Orienting viewers to complex storyworlds via synchronized visualizations. In: Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video, ACM, New York, NY, USA, TVX '15, pp. 167–172, (2015). <https://doi.org/10.1145/2745197.2755519>
 42. Union, E.B.: Digital TV Accessibility: Report on the Current Status in European Countries. Tech. Rep, European Blind Union (2008)
 43. Vargas, M.L.M.: Combining speech and earcons to assist menu navigation. In: Proceedings of the 9th International Conference on Auditory Display (ICAD2003) (2003)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.