Making web video accessible – interaction design patterns for assistive video learning environments

NIELS SEIDEL, Technische Universität Dresden

This paper discusses four interaction design patterns that constitute assistive solutions to enhance the accessibility of audible and visual information inside learning environments. These patterns result from a content analysis of 118 video environments that were designated for learning. The analysis yields a pattern language of currently 40 patterns that describes common solutions of recurring problems in the design and development of video learning environments. Beside the complete representation of the accessibility patterns CLOSED CAPTIONS, TRANSCRIPT, ZOOM, and SHORTCUT COMMANDS, this paper draws attention to possible ways of structuring the pattern language.

Categories and Subject Descriptors: H.5.2 [**Information Interfaces and Presentation**] User Interfaces

Additional Key Words and Phrases: Interaction Design Patterns, Hypervideo, Interactive Video, Accessibility

Introduction

This article is about four interaction design patterns for assistive technologies to make videos more accessible for impaired people. The patterns demonstrate how interaction designers and software developers can help impaired people to overcome the impairment when learning with videos. To achieve that, video learning environments need to be improved and extended in order to become more accessible.

The introduced video accessibility patterns are not isolated in their domain. All patterns are part of a pattern language of currently 40 individual patterns for the design and development of video learning environments. A video learning environment means a software application that enables and supports learning by means of instructional videos as the primary learning resource. Before the patterns gain the reader's attention, we will have a closer look at the surrounding pattern language and its structure. Before that, accessible media in terms of video will be explained.

Author's address: Niels Seidel, Technische Universität Dresden, International Institute Zittau, Markt 23, 02763 Zittau, Germany; email: niels.seidel@tu-dresden.de.

EuroPLoP '15, July 08 - 12, 2015, Kaufbeuren, Germany

ACM 978-1-4503-3847-9/15/07. . . $15.00
http://dx.doi.org/10.1145/2855321.2855339

Accessibility of videos

Access to audiovisual media is limited to audiences that are able to process both audible and visual information. In case a user is only capable to perceive and process either audible or visual information the missing channel needs to be translated in order to make the video accessible. For users that are deaf or hearing impaired, the information delivered through the audio channel needs to be transferred to a visual representation (e.g. as text equivalent or to enable lip reading). Blind audiences or those who have low vision would benefit from an audio representation that includes the visual content.

Some of these special needs can be addressed during the production of the video. Possibilities include, among others, [Brookes 2000] to (1) move the background narrator into the picture; (2) let them face the camera when speaking; (3) have only one person speaking at a time; and (4) show the lips while speaking. Unfortunately we are facing many situations where these recommendations can not be fulfilled. For instance, when a lecturer gets video captured automatically or when we are using authentic footage that has not been designed for any narration.

Since video files are not delivered as stand-alone learning resources, users with a motoric or visual impairment should be able to make use of a video player environment. Some guidance on how to achieve accessible web video applications can be derived from two recommendations of the *Web Accessibility Initiative* (WAI) of the W3C: *Web Content Accessibility Guidelines 2.0* (WCAG) [Caldwell et al. 2008] and *WAI-Accessible Rich Internet Applications Suite* [W3C 2011]. Regarding web videos, WCAG aims to make audiovisual media more accessible. Therefore, alternatives to time-based media such as captions for audio content and audio descriptions or extended audio descriptions are suggested. If the relevant visual information is not already covered by the audio information, audio descriptions should provide "information about actions, characters, scene changes, on-screen text, and other visual content" [Caldwell et al. 2008]. In addition to that, WAI-ARIA focusses on the dynamic web applications that should become accessible for impaired users, whether they are using screen readers, Braille displays, or other assistive technologies.

It could be stated that accessibility is not only a question of providing suitable media representation in terms of timely unrelated media, but also of how these representations can be aligned with the user interface of a video learning environment. Both recommendations provide guidance on a quite abstract level, considering either the media or application development. Good practice examples on how to incorporate these recommendations into a user interface were not within the scope of the *Web Accessibility Initiative*. Hence, this article aims to demonstrate the essence of solutions on how WCAG2 and WAI-ARIA can be put into practice. Addressing these issues with design patterns aims to increase the number of accessible video (learning) environments.

Target audience

Prima facie, the reader of this paper could think that the presented user interface solutions may appear obvious, but unfortunately they have been rarely put into practice so far. This is surprising insofar as non-impaired users could benefit from the assistance as well. Furthermore, the proposed scope of learning activities could be broadened to other video-based activities such as web-based TV or interactive television (e.g. [Kunert 2009]).

Hence, the pattern language is dedicated to four stakeholder groups that can be involved in the development of video-based learning environments. First of all, the patterns are proposed to sensitize interface designers about temporal aspects of layout and navigation in videos. Experienced designers will find examples on how others have solved certain problems. The patterns presented in this paper highlight design decisions that could enhance the accessibility and usability of audiovisual media application.

Closely related to that stakeholder group are software developers who want to design and implement the user interface of a video-based learning environment. For them, the patterns express the design space of possible interface components and solutions. The patterns can also be used as a decision support tool for choosing the

right video player or software framework. However, code examples or hints on how to implement certain solutions have been abandoned in favour of more timeless interface designs that neither depend on specific frameworks, nor on web specifications or browser implementations.

Content management providers as a third audience might be motivated to overcome non-interactive web videos by enabling time-dependent content and time-related interactions – especially for impaired users.

Finally, the patterns can be used by instructors who want to evaluate or draft learning environments regarding the needs of the dedicated audience. The presented pattern language can be used to assess and compare certain design features of commercial or free learning environments. Cross references between single patterns highlight dependencies and concurring design solutions. Thus, instructors should become aware of good design within the design space of video learning environments in order to support and aid learning with respect to their purposed didactical task. Furthermore, instructors are also learning content designers, who make use of extensive tools to enrich the modality of video for their heterogeneous audience of learners.

Pattern language of video learning environments

The pattern language of video learning environments describes common solutions of recurring problems in the design and development of video-based learning environments. In that context, a video-based learning environment is a type of learning environment where learning resources mainly consist in audiovisual media. They provide tools and functions that aid learning by the means of accessing, organizing, and authoring the video content. Furthermore, they facilitate computer- or video-mediated communication among peers and instructors.

A complete list including a short description of each pattern is part of another pattern paper focussing on video annotations. It has been published following the same procedure as with the present paper (see [Seidel 2015]). A brief overview of the pattern mining and writing process can be found in a previous paper [Seidel 2014].

The pattern language has two layers. The first layer is concerned with interactions and manipulations inside a single video. The second layer covers user interactions with one or more videos as part of a greater whole. According to Schwan [2005], the first layer is about *micro interactivity* in a video player and the second about *macro interactivity* in a whole collection of videos (e.g. video portal, archive of lecture recordings, xMOOC plattform).

Below these basic layers, multiple patterns can be categorized by looking at higher-level objectives of design solution. The following six categories represent these objectives:

(1) **Access to time-based information:** includes solutions making audio-visual information that is structured in a temporal manner accessible or more easily accessible.

(2) **Contribution of contents:** learners are enabled to contribute, edit, or review additional contents.

(3) **Structure of content:** navigational representation of media arrangement as a whole including the relations between the videos.

(4) **Support of self-regulated learning:** design solutions that support learners to manage, regulate, and organize their learning process.

(5) **Layout:** spatial arrangement of objects within the available design space.

(6) **Monitoring:** trace and monitor ongoing or past user activities.

The pattern map in Figure 1 offers an alternative perspective on the relations among the patterns. Each layer is visualized by an island whose biggest cities are named after three relatively abstract solutions: BASIC CONTROLS, ANNOTATIONS, and ADD VIDEO. All of them subsume several other patterns or are supplemented by other patterns that could not exist without them. A ANNOTATED TIMELINE for instance, depends on BASIC CONTROLS. By the same token, we consider CLOSED CAPTIONS to be a special type of ANNOTATIONS. Moreover, these patterns could be confirmed by a high number of example applications[1]. BASIC CONTROLS-related patterns concern the delivery

[1] See the list of the 118 examined video learning environments at http://designingvideointerfaces.nise81.com/ (last accessed 2015/11/30).

and control of audio and video information, for instance zooming the moving images, manipulating the playback or changing the display mode. The patterns attached to ANNOTATIONS cover problems that arise when the video gets enriched by other media content (e.g. SYNCHRONIZED MAP) or if it is being semantically annotated (e.g. TEMPORAL TAGS). ADD VIDEO indicates the need for handling more than one video, e.g. as part of a collection or in a video management system to support learning, as mentioned above.

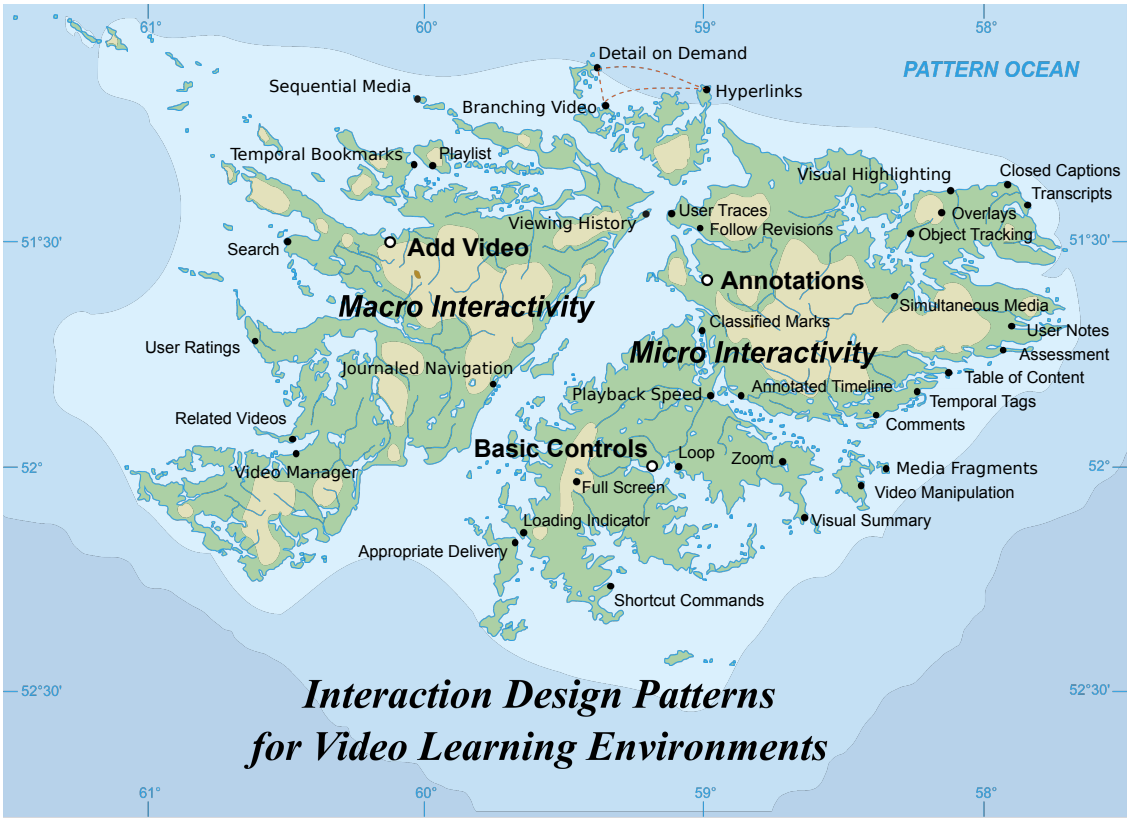

Fig. 1. Map of Video Interaction Design Patterns.

Pattern format

A pattern starts with the name. To indicate the subjective confidence about the solution, [Alexander 1979] one to three asterisks behind the pattern name express the dissemination of the solution. An asterisk corresponds to a tercile of the patterns ordered by the number of known examples. Considering the "Rule of Three", the absolute frequency per pattern ranges from 3 to 42 known examples of the total number of 118 examined video learning environments. Patterns without an asterisks are still considered as a proposal.

The first part of the pattern is a short description of the context, followed by three diamonds (⌘⌘⌘). In the second part, the problem (in bold), the background, and the forces are described, followed by three diamonds. The third part offers the solution statement (in bold) as well as positive and negative consequences of the pattern application including a discussion of possible implementations. In the second to last part of the pattern, known application examples are illustrated by screen shots. The final part consists of a list of related patterns of the underlying pattern language. The list is complemented by patterns of related pattern languages such as the *iTV patterns* of Kunert [2009] or the more general *Interaction Design Patterns* presented by Tidwell [2011].

Subsequently, there will be four sections regarding the patterns CLOSED CAPTIONS, TRANSCRIPT, ZOOM, and SHORTCUT COMMANDS. All these patterns represent assistive solutions that support the perception and control of audiovisual learning resources by the means of interactive video players on the level of micro-interactivity. CLOSED CAPTIONS and TRANSCRIPT can be considered as competitive solutions to augment audible information as well as descriptions of visual content. Therefore, they are considered as SIMULTANEOUS MEDIA. A promising pattern, ZOOM assists users with low vision as well as those who want to comprehend details of the visual information. SHORTCUT COMMANDS approaches the universal need to access graphical user interfaces with a few finger movements and draws attention to common shortcuts for both keyboard and multi-touch devices.

Closed Captions ∗ ∗ ∗

An instructional video is usually accompanied by spoken word in a certain language.

⌘ ⌘ ⌘

Audio information in a video can not be understood by audiences that are deaf, hard of hearing, or that have insufficient language competencies. These audiences are hardly able to participate in learning.

For individuals that are deaf or hard of hearing, the audible cognitive channel of a video is not or only partly available. Besides the visual information, meta data is considered as a source of information for these target groups. TABLE OF CONTENT, TEMPORAL TAGS or an abstract in the VIDEO MANAGER as well as other textual ANNOTATIONS can help to get an overview of the learning content. However, meta data lacks detailed information.

If the volume is too low, it can be increased to some extent by using the volume controls (BASIC CONTROLS). In noisy environments or in situations where listening to sound is unfavorable (e.g. in a open space office), audio cannot be played loud enough for sufficient comprehension without using headphones. Often, there is also no alternative to videos with low quality audio.

Difficulties in understanding that are caused by a high speaking rate can be approached by reducing the PLAYBACK SPEED.

If you can only listen to the audio, it can be hard to extract the spelling of words, names or technical terms. To lower these language barriers, phrases can be displayed simultaneously with the video, e.g on slides or pictures. But these (SIMULTANEOUS MEDIA) elements cover only a subset of the overall sum of spoken words. Especially context information as well as grammatical structures get lost, which is awkward for language learners. Furthermore, it is not always affordable, technically, to present information besides the video (e.g. on small display devices).

⌘ ⌘ ⌘

Support comprehension by providing closed captions of transcripted speech and other relevant audiovisual information. Let the user decide whether to display these subtitles in the bottom of the video frame or not.

The availability of closed caption opens three different ways to perceive the content: the video without captions, the video with the captions, and the captions itself as an alternative representation of the audible contents. Providing diverse entry points can help to improve the video accessibility for different target groups. For impaired users who are deaf or hard of hearing, captions are essential in order to be able to comprehend and understand. But also non-impaired people can benefit from the written representation, e.g. to capture difficult terms, spellings or a foreign-language speech. This practice correlates with the modality principle of Mayer [2009]. The principle states that words should be presented as speech rather than on-screen text, if learners have sufficient competencies and abilities to understand the spoken words. Because videos are often dedicated to a diverse audience with different language competencies and abilities, it is advisable to provide captions that can be turned on and off. Since closed captions are a special type of OVERLAYS, the user can hide or close them. That's why we call them "closed captions" in contrast to "open captions" or subtitles that are visible all the time.

Captions consist of a synchronized text as an equivalent of spoken words and dialogues. Thus, the contained information is more detailed and precise, compared to other descriptions or meta data. Technically, captions are separate from the video file, which makes them flexible and exchangeable. Hence, it is possible to declare multiple caption files to cover different languages. Furthermore, they can be accessed by screen readers. Various data formats are commonly used[2] while none of them gained acceptance as a standard for video players.

Captions should always be presented as in sync with the corresponding auditory information as possible. In terms of the amount of words and the frequency of changing subtitles, reading it all can become a challenging task

[2] For example: *W3C Timed Text*, *QTtext/QuickTime* text, *SubViewer* (*.sub), and *SubRip* (*.srt).

for the user. Since space is limited, Brookes [2000] recommends to show no more than 15 words at a time. For that reason common words can be abbreviated. In case of coincidental complex visual information and/or difficult subtitle contents, users should be allowed to reduce the PLAYBACK SPEED. Alternatively, a TRANSCRIPT can be read while the video is paused. To afford an appropriate readability against the background image, the font size, color/contrast, and position of the subtitles should be customizable. Adding a ticker border and shadow to the font brings further clarity. A simple font should be preferred to support readability. Sometimes different text colors are used to identify who is speaking.

In case a video is completed by a simultaneous text-based media representation (SIMULTANEOUS MEDIA) like slides, it is unfavorable to add another visual information source for risk of a cognitive overload. The split-attention effect of concurring visual information sources can be seen as the main cognitive disadvantage of captions.

Closed captions hardly include all auditory information. Text equivalents of acoustic events like noise, music or intonations are often excluded to express the spoken words. In the following example, the speaker is set at the beginning of the line, while the acoustic event follows within square brackets:

> JACK: I got the machine ready. [engine starting]

Providing comprehensive context information may support the understanding, but requires the user to read a lot more. Instead, a TRANSCRIPT is a more suitable representation for details like this.

Transcribing a video precisely requires much effort and time. Although automatic speech-recognition provided some advances, in most cases a video needs to be transcribed manually. A good authoring tool similar to USER NOTES can help save time. Alternatively, the task can be outsourced to volunteers or companies who are offering such services.

Examples.

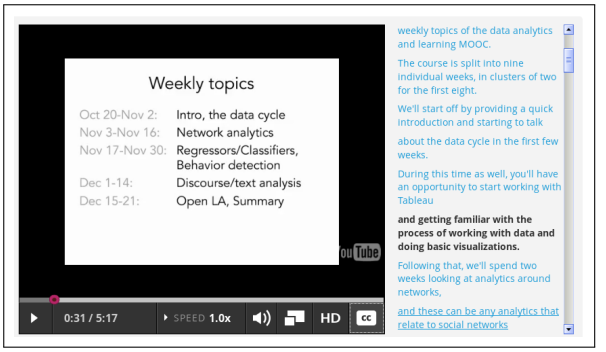

Fig. 2. *edX* provide closed captions for the videos included in their *Massive Open Online Courses*. The representation could also be considered as a TRANSCRIPT because the whole subtitle text is readable at a time while only the corresponding paragraph is highlighted. See https://www.edx.org/ (last accessed 2015/11/30).

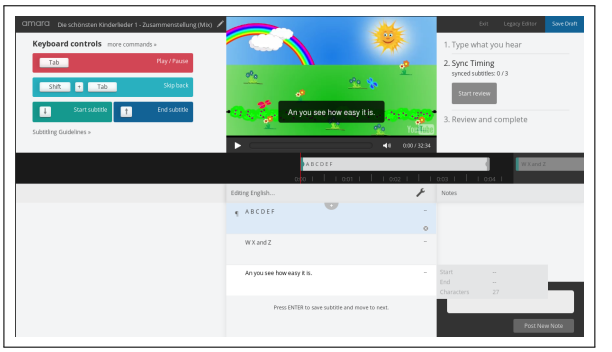

Fig. 3. *Amara* is an advanced tool to annotate video subtitles in different languages. The creation process consists of three steps: 1) select a video URL as well as a source and a target language, 2) type what you hear by the help of keyboard shortcuts, 3) synchronize the lines to their exact positions, 4) review the captions and publish them, including a translation for the title and abstract. See https://amara.org/ (last accessed 2015/11/30).

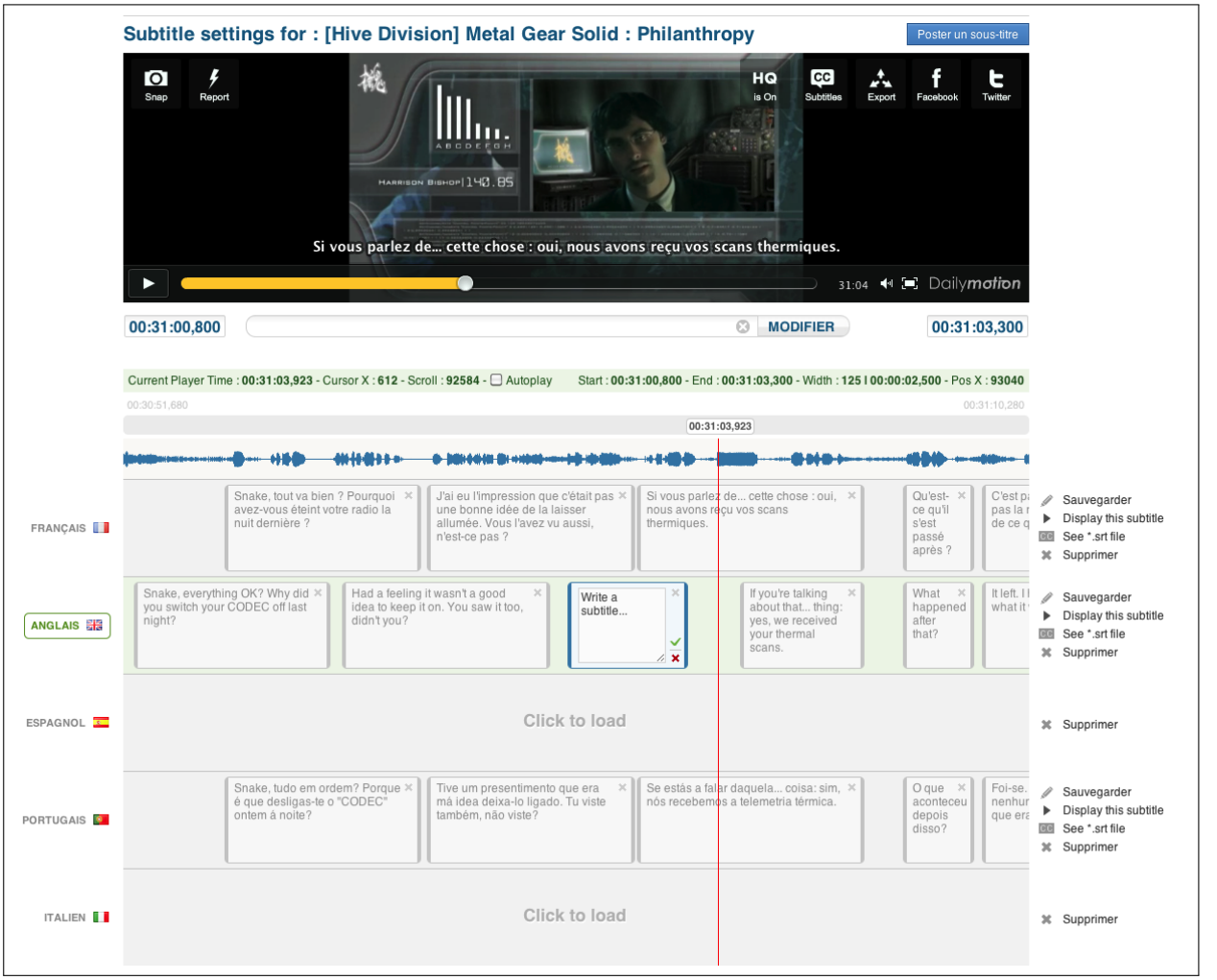

Fig. 4. *Dailymotion* provides a multi-track interface to annotate subtitles/captions in several languages. It is also possible to import and export srt-files containing the captions. See http://www.dailymotion.com/ (last accessed 2015/11/30).

Related Patterns. TRANSCRIPT, OVERLAYS, ANNOTATION, TELEPOINTER [Schümmer and Lukosch 2007, S. 359]

Transcript ∗∗

An instructional video is usually accompanied by spoken language.

⌘ ⌘ ⌘

Auditory information of a video is volatile and can be neither accessed or skimmed at the users own pace.

CLOSED CAPTIONS only present a fraction of the overall audio content that corresponds to the scene the user is currently watching. The spoken word of previous or upcoming scenes is not visible at the time. While a faster playback rate (PLAYBACK SPEED) improves the situation, the user is still not able to read or skim the subtitles at his own pace. This becomes problematic if there is a very low or very high density of subtitles within a certain period of time. A TABLE OF CONTENT or a cloud of TEMPORAL TAGS can help to get an overview of the content, while details are not included.

Thus, the auditory information is part of the self-contained video file: It is neither searchable by the user, nor by the system.

⌘ ⌘ ⌘

Provide a transcript of all auditory information as a text equivalent beside the video. Make this transcript navigable and highlight paragraphs that correspond to the current playback position.

"Transcripts allow deaf/blind users to process content through the use of refreshable Braille and other devices. Screen reader users may also prefer the transcript over listening to the audio of the web multimedia." [Brookes 2000] In contrast to CLOSED CAPTIONS, other audible cues such as descriptions of music, noise or intonations are typically not included. Therefore, the written narration appears in a flow of paragraphs instead of time-dependent snipped dialogs. This enables the user to skim the text much quicker than listening to the narration at a higher rate. In the same way as non-impaired users would increase the PLAYBACK SPEED, some screen reader users prefer to listen to the reader at a much higher rate. Thus, they can access a transcript much faster.

The transcript text should be displayed stationarily rather than inside a scrolling panel that is tiring for the reader [Brookes 2000]. Brookes [2000] recommends at least 15 seconds for 50 words so that the user has time to look away from the display. Single words, sentences or paragraphs can be selected in order to navigate to the corresponding playback time. During the playback, the corresponding section of the transcript should be highlighted.

Unlike a TABLE OF CONTENT or an abstract in the VIDEO MANAGER, written transcripts provide not just an overview but also an accurate representation of the spoken words. Thus, a transcript is an additional representation of the provided video learning resource.

In the same way, it facilitates searching and browsing for the user. Besides, it is a remarkable source to enable SEARCH inside and across multiple videos.

The presentation of a transcript requires sufficient space beside the video frame. This can be a problem on small display devices or in applications that include other time-related or time-dependent components, e.g. SIMULTANEOUS MEDIA in terms of presentation slides. Compared to CLOSED CAPTIONS, the effort to create transcripts containing all auditory information and their relation to the video is much higher. On the other hand, blind and deaf people derive a greater benefit from a self-containing text equivalent as opposed to in sync sentences, i. e. subtitles. Transcripts are usually not considered as user-generated content ANNOTATIONS whereas USER NOTES describe a way on how to achieve personal transcripts.

Although users without special needs benefit from a transcript (e.g. to search inside the video) this type of SIMULTANEOUS MEDIA does not have to be visible all the time. In order to avoid a split attention effect, transcripts should be made available on demand.

In case a video is completed by a simultaneous text-based media representation (SIMULTANEOUS MEDIA) it is unfavorable to add another visual information source for risk of a cognitive overload. The split-attention effect of concurring visual information sources can be seen as the main cognitive disadvantage of captions.

Examples.

Fig. 5. *Zeugen der Shoah* is a collection of over 900 video interviews of holocaust survivors. Each interview has been transcribed to be provided below the video player. The current playback position of the 30 minutes long videos gets highlighted. Each word can be clicked to call the corresponding scene inside the video. See http://www.zeugendershoah.de (last accessed 2014/10/14).

Fig. 6. *VideoReadr* allows a simultaneous presentation of audio transcripts next to a video. The corresponding paragraph is highlighted while the text scrolls up and down automatically. In favor of a better accessibility, the font size can be increased. See http://www.videoreadr.com/ (last accessed 2015/11/30).

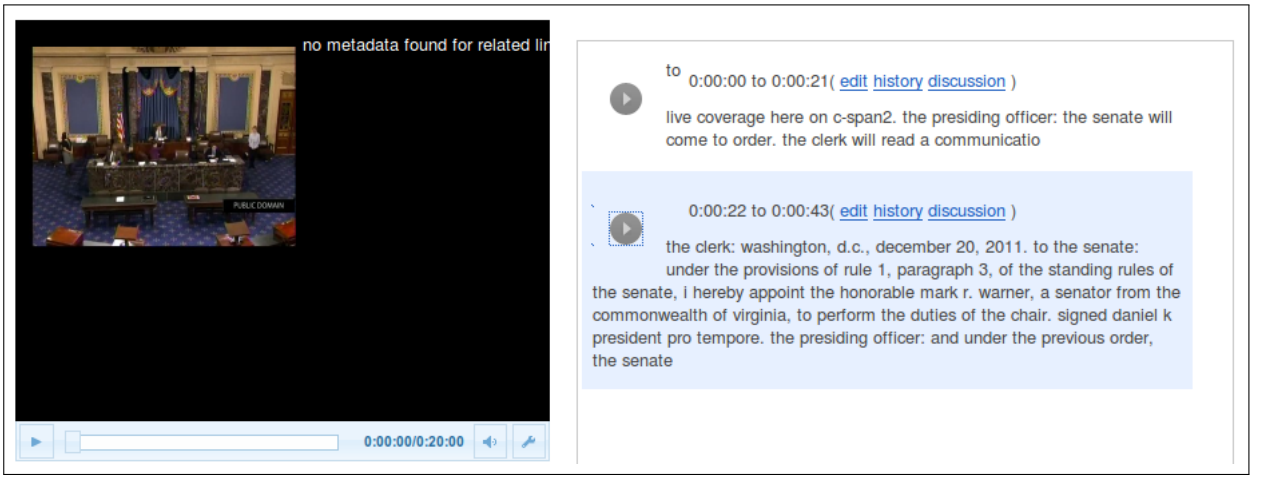

Fig. 7. *MetaVid.org* is a community driven open video archive that collects speeches of both houses of the U.S. congress [Dale et al. 2009]. Each video is split into segments that represent a navigable table of content including a transcript of the spoken words. See http://metavid.org/ (last accessed 2015/11/30).

Related Patterns. CLOSED CAPTIONS, ANNOTATION, TELEPOINTER [Schümmer and Lukosch 2007, S. 359]

Zoom ∗

A video or its synchronized contents contain visual details that are essential for comprehension.

⌘ ⌘ ⌘

Users miss essential visual details that are presented too small.

An in-depth observation can be helpful in order to recognize and understand certain dynamics. High Definition (HD) video is able to provide deep insights into the moving images. But these visual details get compressed inside the window of the video player. The barrier for observing the details lies in the frame size of the video within the possible display dimensions in FULL SCREEN MODE rather than in the image quality.

Using the browser's built-in zoom function could be a solution, but the necessary controls are hidden in the menu or the user needs to know the respective KEYBOARD COMMAND. Unfortunately, the built-in browser zoom is not available in FULL SCREEN mode.

Another way of dealing with visual details is the reduction of the PLAYBACK SPEED. Thus, dynamic visualizations can become observable in slow motion mode, but the depicted visual details stay as small as they were.

⌘ ⌘ ⌘

Therefore, it is advisable to provide a magnifier to let the user zoom in all time-dependent contents.

"A zooming feature can support the detailed processing of visual information in a video" [Krauskopf et al. 2012]. Video zoom can help to get a better understanding of visual details that are otherwise hard to observe or recognize. Compared to zoom functions for static images, it is important to know that the video continues to play while parts of it are shown in detail. If needed, the user can pause the video or navigate on the timeline (BASIC CONTROLS).

There are several ways of implementing a video zoom. For instance, buttons for zooming in and out can be provided to enlarge or shrink the video frame step by step. After zooming in, the user can pan the image in all directions. Visual content can also be magnified when hovering the frame. Alternatively, the magnification can be limited to a certain area.

VISUAL HIGHLIGHTING can be used to inform the user about segments that should be magnified in order to obtain a better understanding.

It should be possible to zoom into the relevant content in all display modes – be it in FULL SCREEN or normal presentation mode. In case there is any SIMULTANEOUS MEDIA presented aside the main video, it becomes necessary to extent the zoomable area to the relevant aspects.

Even if zoomed in, some dynamic visualizations may be too volatile to be observed. In that case the PLAYBACK SPEED can be reduced.

Related Patterns. FULL SCREEN, PLAYBACK SPEED, SIMULTANEOUS MEDIA.

Examples.
To date, only two desktop applications could be identified that provide a zoom function. However, zoom has become a standard feature for players on multi-touch devices. Only a few research prototypes utilize video zoom [Canessa et al. 2014; Zahn et al. 2012; Khiem et al. 2011] so far.

Fig. 8. *MediaSite* offers a function to zoom in and out while the video is scaled down to 50 % by default. For zooming autonomously, steps from 75 %, 100 %, 150 % up to 200 % are available. The example shows how small text becomes readable after zooming in. See example http://mediaweb.med.uni-magdeburg.de/ (last accessed 2015/11/30).

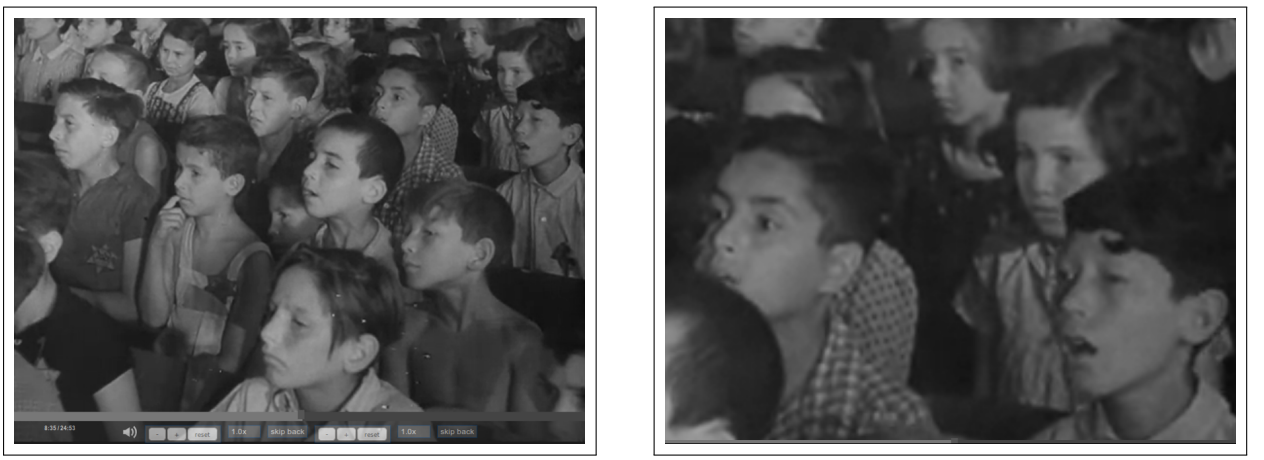

Fig. 9. *Theresienstadt explained* incorporates a video zoom by the help of *jQuery pannzoom*. Below the timeline are buttons to zoom in and out as well as to reset the size to the initial state. The example above demonstrates how it becomes easier for the learner to observe and identify people in the background. See http://theresienstadt-film.net/ (last accessed 2015/11/30).

Shortcut Commands ∗

Video players incorporate different designs and provide different user interfaces.

⌘ ⌘ ⌘

Graphical user interfaces are not accessible for the blind without a screen reader. These interfaces can also become frustrating for users who are not impaired in their vision, especially for frequently executed tasks.

According to Shneiderman and Plaisant [2004], it is crucial to consider the needs of diverse users from "novices to experts, age ranges, disabilities, and technological diversity" [Shneiderman and Plaisant 2004]. Users who are blind, who have low vision or hand tremors find it difficult or impossible to make use of a video player that only provides a graphical user interface [Caldwell et al. 2008]. One reason for that is the non-standardized user interface facilitating BASIC CONTROLS as well as additional controls (e.g. ZOOM or SKIP BACK).

⌘ ⌘ ⌘

Enbale keyboard shortcuts and touch gestures for the most frequent user interactions. Specifically, define those shortcuts that most of the video players have in common.

Many video players implement a set of core keyboard commands and touch gestures to provide BASIC CONTROLS. Especially blind people benefit from gestures since the graphical user interface is not accessible for them without a screen reader [Kane et al. 2011]. The shortcuts in Table I could be identified in many different video players (see example), including *YouTube* as the most common one.

Table I. Typical commands and the assigned keyboard shortcuts and touch gestures.

Command	Keyboard shortcut	Touch gesture
play/pause the video	space-bar	double tap
skip forward	→	drag to the right
skip backward	←	drag to the left
skip to 10% . . . 90% of the playback time	1 . . . 9	–
skip to beginning	0	–
volume up	↑	drag up
volume down	↓	drag down
toggle full screen	F	–
exit full screen mode or video	Esc	–

It is remarkable that the reading direction from left-to-right has become a universal standard for video players. Both the play button and the timeline are oriented to the right.

However, it is a challenging task to apply the recommendations for keyboard shortcuts and touch gestures on highly interactive video players. An ANNOTATED TIMELINE as well as an OVERLAYS or other types of ANNOTATIONS require complex interactions that are difficult to map on shortcuts.

In particular, it is challenging to make keyboard shortcuts platform-independent, e.g for different web browsers and operating systems. Further conflicts can arise through the use of single key press commands. Especially if the user is required to enter text (e.g. USER NOTES, COMMENTS or ASSESSMENT) or the same key is already mapped as a common user-agent keyboard command.

Other interactive elements should be made accessible by the use of Tab and Enter only. Coding recommendations on how to do that are provided by the W3C [W3C 2011; Caldwell et al. 2008] Although the mentioned shortcuts might be very common, they need to be communicated to the user, e.g. on a help page.

Examples.

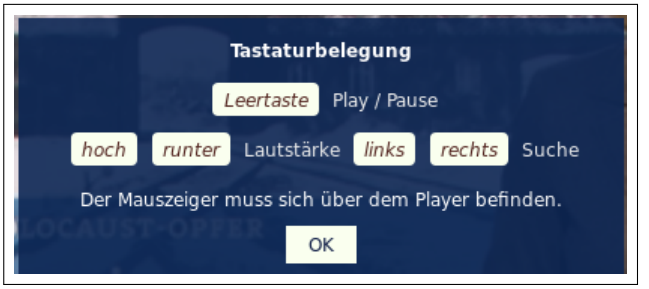

Fig. 10. The news cast *Tagesschau* is utilizing the*Projekktor* player, which supports common keyboard commands. See http://www.projekktor.com/ (last accessed 2015/11/30).

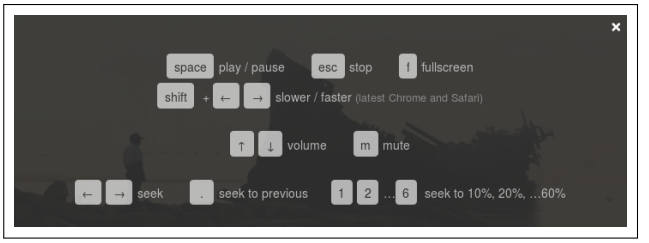

Fig. 11. The video player that is used for the interactive documentary *CloudsOverCuba* covers the most common shortcuts, including those to adjust playback speed. See http://cloudsovercuba.com/ (last accessed 2015/11/30).

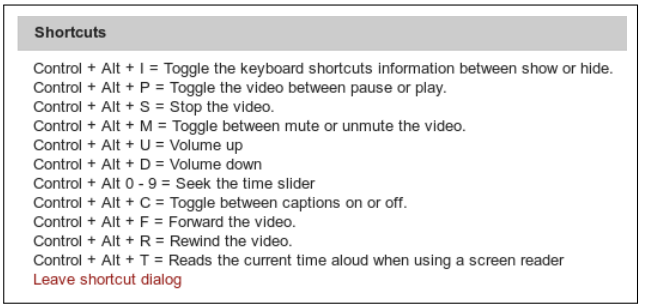

Fig. 12. *Lernfunk / Opencast Matterhorn* is a negative example because it requires the user to learn unusual keyboard short cuts. See http://www.opencast.org/matterhorn(last accessed 2015/11/30).

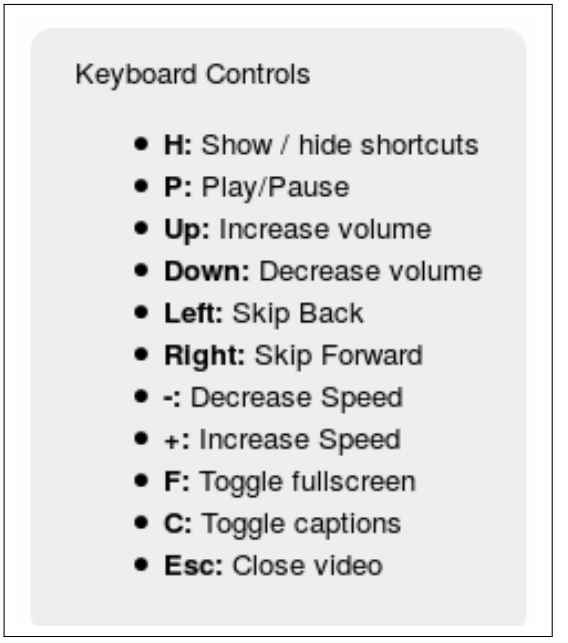

Fig. 13. *Coursera* additionally introduces short cuts to control the playback speed. See https://www.coursera.org/ (last accessed 2015/11/30).

Related Patterns. BASIC CONTROLS, see also KEYBOARD ONLY in Tidwell [2005].

ACKNOWLEDGEMENTS

The author expresses their deepest appreciation to their guardian Christian Kohls for the consistently thorough guidance while preparing this paper. Also, the feedback of all members of the writers' workshop E at the *EuroPLoP 2015* conference was highly welcome, therefore, I want to thank all of them.

REFERENCES

ALEXANDER, C. 1979. *The Timeless Way of Building*. Oxford University Press, New York.

BROOKES, C. 2000. Speech-to-text systems for deaf, deafened and hard-of-hearing people. In *Speech and Language Processing for Disabled and Elderly People (Ref. No. 2000/025), IEE Seminar on*. IET, 5/1–5/4.

CALDWELL, B., COOPER, M., REID, L. G., VANDERHEIDEN, G., CHISHOLM, W., SLATIN, J., AND WHITE, J. 2008. Web Content Accessibility Guidelines (WCAG) 2.0. Tech. rep., W3C.

CANESSA, E., FONDA, C., TENZE, L., AND ZENNARO, M. 2014. EyApp & AndrEyA - Free Apps for the Automated Recording of Lessons by Students. *International Journal of Emerging Technologies in Learning (iJET) 9,* 1, 31–34.

DALE, M., STERN, A., DECKERT, M., AND SACK, W. 2009. SYSTEM DEMONSTRATION: Metavid.org: A social website and open archive of congressional video. In *Proceedings of the 10th International Digital Government Research Conference*. Digital Government Society of North America, Puebla, Mexico, 309–310.

KANE, S., WOBBROCK, J., AND LADNER, R. 2011. Usable gestures for blind people: understanding preference and performance. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. CHI '11. ACM, New York, NY, USA, 413–422.

KHIEM, N. Q. M., RAVINDRA, G., AND OOI, W. T. 2011. Towards Understanding User Tolerance to Network Latency in Zoomable Video Streaming. In *Proceedings of the 19th ACM International Conference on Multimedia*. MM '11. ACM, New York, NY, USA, 977–980.

KRAUSKOPF, K., ZAHN, C., AND HESSE, F. W. 2012. Leveraging the affordances of Youtube: The role of pedagogical knowledge and mental models of technology functions for lesson planning with technology. *Computers & Education 58,* 4, 1194–1206.

KUNERT, T. 2009. *User-Centered Interaction Design Patterns for Interactive Digital Televison Applications*. Springer, Dordrecht / Heidelberg / London / New York.

MAYER, R. E. 2009. *Multimedia Learning - second edition*. Cambridge University Press, New York.

SCHÜMMER, T. AND LUKOSCH, S. 2007. *Patterns for computer-mediated interaction*. Wiley, Hoboken, NJ.

SCHWAN, S. 2005. Gestaltungsanforderungen für Video in Multimedia-Anwendungen. http://www.eteaching.org/didaktik/gestaltung /visualisierung/video/schwan.pdf.

SEIDEL, N. 2014. Interaction design patterns for design and development of video learning environments. In *Proceedings of the 19th European Conference on Pattern Languages of Programs*. ACM, New York, NY, USA, 20:1–20:12.

SEIDEL, N. 2015. Interaction design patterns for spatio-temporal annotations in video learning environments. In *Proceedings of the 20th European Conference on Pattern Languages of Programs*. ACM.

SHNEIDERMAN, B. AND PLAISANT, C. 2004. *Designing the User Interface: Strategies for Effective Human-Computer Interaction* 4. Edition Ed. Pearson Addison Wesley, Boston, MA, USA.

TIDWELL, J. 2005. *Designing Interfaces – Patterns for Effective Interaction Design* 1 Ed. O'Reilly Media, Sebastopol.

TIDWELL, J. 2011. *Designing Interfaces – Patterns for Effective Interaction Design* 2 Ed. O'Reilly Media, Sebastopol.

W3C. 2011. Accessible Rich Internet Applications (WAI-ARIA) 1.0 W3C Candidate Recommendation 18 January 2011.

ZAHN, C., KRAUSKOPF, K., HESSE, F., AND PEA, R. 2012. How to improve collaborative learning with video tools in the classroom? Social vs. cognitive guidance for student teams. *International Journal of Computer-Supported Collaborative Learning 7,* 2, 259–284.

Received February 2015; revised April 2015; accepted December 2015

EuroPLoP '15, July 08 - 12, 2015, Kaufbeuren, Germany

ACM 978-1-4503-3847-9/15/07…$15.00
http://dx.doi.org/10.1145/2855321.2855339