Supporting Video Authoring for Communication of Research Results

Katharina Wünsche
katharia.wuensche@univie.ac.at
Faculty of Computer Science,
University of Vienna, and
Austrian Academy of Sciences
Vienna, Austria

Laura Koesten
laura.koesten@univie.ac.at
Visualization and Data Analysis,
Faculty of Computer Science,
University of Vienna
Vienna, Austria

Jian Chen
chen.8028@osu.edu
Computer Science and Engineering, and
Translational Data Analytics Institute
The Ohio State University
Ohio, USA

Torsten Möller
torsten.moeller@univie.ac.at
Visualization and Data Analysis,
Faculty of Computer Science,
and Research Network Data Science
University of Vienna
Vienna, Austria

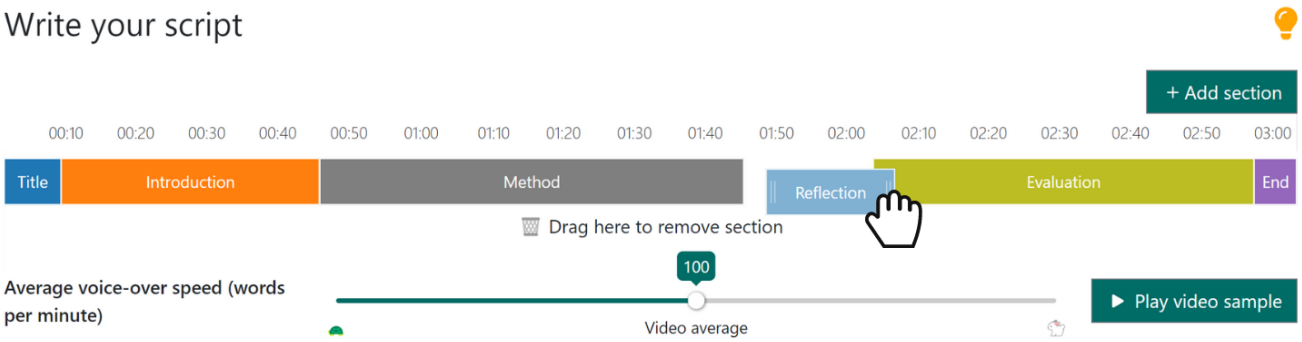

Figure 1: Interactive outline refinement in Pub2Vid. Screenshot of the script writing page.

ABSTRACT

Video summaries of scientific publications have gained more and more popularity over the last years, requiring many researchers to familiarize themselves with the tools and techniques of video production which can be an overwhelming task. This paper introduces a video structuring framework embedded into the authoring tool Pub2Vid. The tool supports users with the creation of their video outline and script, providing real video examples and recommendations based on the analysis of 40 publication summarization videos which were rated in a user study with 68 participants. Following a four-tier evaluation methodology, the application's usability is assessed and improved via amateur and expert interviews, two rounds of usability tests and two case studies. It is shown that the tool and its recommendations are particularly useful for beginners due to the simple design and intuitive components as well as suggestions based on real video examples.

CCS CONCEPTS

• **Human-centered computing** → *Interactive systems and tools.*

KEYWORDS

publication summarization video, video abstract, video analysis, video recommendations, science communication, script authoring

IMX '23, June 12–15, 2023, Nantes, France

ACM ISBN 979-8-4007-0028-6/23/06.
https://doi.org/10.1145/3573381.3596157

ACM Reference Format:
Katharina Wünsche, Laura Koesten, Jian Chen, and Torsten Möller. 2023. Supporting Video Authoring for Communication of Research Results. In *ACM International Conference on Interactive Media Experiences (IMX '23), June 12–15, 2023, Nantes, France.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3573381.3596157

1 INTRODUCTION

Videos can play an essential role in the dissemination of scientific findings, facilitating exchange with other scientists as well as with the general public. Research has shown that they can facilitate the transport of scientific messages and help audiences to process and recall the presented content [10, 24]. In addition, more and more publication authors are invited to produce video summaries of their publications for venues such as CHI [16], SIGGRAPH [3] or the IEEE CG&A journal [28], which can be a challenging and time-consuming task as many beginners do not know how to start with the creation of such a video [32]. The aim of this work is to support authors with this task by identifying recipes for successful videos and embed them into a web application that can be used for the video authoring process. As a result, we present a set of video recommendations together with a tool which allows users to interactively define their video outline and formulate a video script.

While automated video generation is already possible in theory [13, 34, 44], generating satisfying video summaries based on a scientific publication is challenging. Publication summarization approaches usually focus on text output [18, 33, 39, 61] or slide generation [22] (as described in more detail in Section 2). The diverse methods proposed for automatic video generation (e.g., [13, 47, 62]) focus on the automatic algorithmic process and have not considered factors important to engaging video production. Moreover, to the best of our knowledge, only a few tools are known that support users with the structuring and script writing process [40], and none of them are specialized for scientific videos. Our approach is a first attempt to close this gap.

To support researchers with the production of their publication summarization videos, we conducted both quantitative and qualitative data collection with researchers of diverse experiences with producing video summaries of their work and identified requirements for a video authoring system. We present a solution that provides users with a guided workflow for the first steps of the production process, the planning and structuring part, and allows users to flexibly use their own resources and techniques for the final video creation. We developed Pub2Vid, a web application that provides users with recommendations for their videos and supports them in defining an outline and writing a script for their voice-over, which can then be exported as a presentation template or a text file. The recommendations and examples in the tool are based on existing scientific videos in the field of visual computing which were analyzed with respect to features such as their duration, structure, and the usage of visual and audio elements. The same videos were rated during a user study, leading to a set of baseline video ratings that were used to derive recommendations for summarization videos, incorporating both qualitative and quantitative feedback. The application was evaluated in multiple rounds of usability tests, case studies, and interviews with both, experts and nonprofessionals from different domains. The results show that the application was considered helpful and easy to use by users of different backgrounds and levels of experience with video creation.

This work presents four main contributions:

- A video structuring framework embedded into a web application that allows users to plan their videos and receive suggestions together with real video examples, accessible under the domain https://pub2vid.web.app/.
- A set of video recommendations based on ratings of 68 participants and a detailed structure analysis of 40 paper summarization videos in selected fields of the computer science domain.
- A four-tier evaluation method, consisting of two rounds of usability tests, amateur and expert interviews, and two case studies.
- A dataset consisting of the feature and structure analysis of 40 videos together with ratings and comments by 68 participants.

2 RELATED WORK

We present three main topical areas that have influenced our recommendations and the design of our Pub2Vid video authoring system: publication summarization, video guidelines and video authoring tools.

2.1 Summarizing scientific publications

Content summarization is the process of using machine learning techniques or human readers to generate a concise and meaningful summary of a large piece of text, such as news or videos, [25, 37], literature survey [61], or presentation slides [22]. Different approaches in automatic summarization have been followed over the last decades. Most of them cover the summarization of text or documents, but recently also other media types have gained attention. Even though the automatic generation of video summaries has been explored already for other domains, such as web content [34, 44] and longer videos [38], systems that focus on scientific publications are uncommon as most publication summarization systems are restricted to text output [18, 33, 39, 61]. *DOC2PPT* [22] is an exception here, being an automatic presentation generator that creates slides from scientific documents by extracting sentence and figure embeddings from an input publication. While it would be easy to transform these slides into a video, the generator does not take storytelling techniques or other human input into consideration which would still have to be incorporated manually.

2.2 Video guidelines and analysis

Guidelines and recommendations for videos can be retrieved from various sources, ranging from conference and journal websites [2, 15, 27] over blogs [20, 59] to scientific literature. While the majority of publications about science video guidelines is based on literature reviews in the field of educational science and psychology [11, 50, 52, 53], only a few related papers could be found that analyse existing videos [46] and combine these results with audience reactions such as likes or comments [21, 54]. Other approaches for retrieving rules for educational videos run user studies [35] and surveys [9] with viewers.

Built on top of these successes, we focus on key video features that introduce creativity, informativity, and enthusiasm in scholarly video summaries, considering similar features to other approaches (e.g. sound and visual elements [9, 54]) and new ones such as specific video sections.

2.3 Video Authoring Tools

Traditional video editing software like *Adobe Premiere Pro* [4], *Final Cut Pro* [8] and *iMovie* [7] provide users with tools for assembling video and audio recordings, images and text overlays. Most of them provide a timeline-centered interface and offer different transition styles and video effects that can be applied to the clips. We also use a timeline component in Pub2Vid but integrated it to support the structural composition of video segments instead of single assets such as sound and images.

Web-based video generation services usually provide users with templates and resources to reduce the preparation time and lower the barrier for users with less experience in video creation. Platforms like *Animoto* [6], *Canva* [14] and *Wibbitz* [60] offer storyboard editors, various visual effects, and basic editing utilities, often combined with functionality for searching stock images or music and predefined animations. However, their domains of use are marketing or social media rather than scholarly work. Thus their design guidelines may not always be applicable to scholarly article summarization videos.

Intelligent authoring systems allow users to create videos tailored to a user-defined input, e.g., a website or a data file. *DataClips* [5] is an authoring tool for data-driven videos which allows users to import data in form of a table and visualize it using predefined clips. *URL2Video* [17] generates videos from web pages by automatically selecting and arranging elements of the page and rendering the content into a video which can then be refined manually. *Katika* [31] allows users to author explainer motion graphics videos by writing a script which is then turned into single shots that can be adjusted and composed into a final video. *Katika* is suitable for users who already know how they want their final video to look but does not provide much support with structuring a video nor with recommendations or examples. The application should therefore be seen as complementary to our approach, tackling a different step in the video production process. Other authoring systems rely on one or more existing videos (e.g. [36, 51, 58]) which are then re-assembled according to user input such as text. Although they are tackling a different use case than ours, they can be of interest for future extensions of our system.

3 METHODOLOGY

The methodology followed during this research can be split into three parts: baseline video ratings and analysis leading to a set of video recommendations, the system design process and the 4-tier evaluation of the application.

3.1 Baseline video ratings

The recommendations and examples provided by Pub2Vid are based on a detailed analysis and baseline video ratings gained through a user study.

3.1.1 Video selection and analysis. A total number of 40 scientific videos was selected using the YouTube search by retrieving the videos with the highest number of views and a duration between 1 and 10 minutes of the following venues: CHI 2019 [15], SIGGRAPH 2019 [1], VIS 2019 [29] and EuroVIS 2019 [30], the IEEE CG&A journal [28] as well as two videos of the OpenAI YouTube channel [43]. Since the analysis was conducted in 2020 and 2021, venues and videos from 2019 were chosen in order to find a balance between recency and elapsed time as more recent videos might not have been online long enough to reach noteworthy numbers of views. All of the videos cover computer science topics, mostly in visual computing disciplines, i.e. human-computer interaction, visualization or computer graphics, which match the research fields of the four members of the analysis team. The selected fields are of major interest for such an analysis due to the fact that they strongly rely on visuals for scientific communication. The videos were randomly distributed between the members, with each member analysing 20 videos and each video being analysed by two people.

The data gathered in the analysis consists of four parts: (1) general information about the video, (2) visual and audio elements that were used, (3) timestamps for the different video sections and (4) a subjective rating. The general information contains the video URL, its title and duration as well as the number of cuts. The identification of visual elements was done using a predefined list of possible elements, including (instructional) animations, charts, diagrams, highlighting in the visualizations, infographics, photos, prominent transitions between scenes (e.g., fading in/out), section headings, slides (e.g., a recorded PowerPoint presentation), subtitles, user interface, video-over (e.g., a small video of the speaker in the bottom right corner), and live-action recordings. The same was done for the audio elements, which could include music, sound effects, and voice-over. For both categories, there was also the option for free text entry. For the coding of video sections, a set of common publication sections was used and refined over the course of the analysis. The final section were defined as follows:

- Title: usually contains paper/video title, author and/or institution names, conference name etc.
- Video Outline: a short description of what is to be presented
- Introduction: gives basic information about the topic/work, can include preview of final results, contributions (reflection) or related work
- Related Work: overview of existing work about the same or a similar problem
- Method: a particular procedure for accomplishing or approaching something, especially a systematic or established one [49]
- System Architecture: describes structure and behavior of the system, including its components and subsystems
- Results: outcome(s) of the development or design process, can include guidelines and recommendations
- Demo: a demonstration of the research outcome or technique, can include use cases or possible applications
- Evaluation: judging or calculating the quality, importance, amount, or value of the presented tool/technique [57]
- Reflection: summary of the presented work, synthesis of the key points; similar to conclusion, discussion and contributions
- Forward Looking: outlook of possible further developments of the presented tool/technique; similar to impact, implications or future work sections
- Ending Slide: final slide/image, can show the paper title, authors etc., or something like "Thank you"

The subjective rating could range from 1 (worst) to 5 (best) in each of the following three categories: creativity, enthusiasm, and informativity. Their choice is based on previous work ([19, 23, 26, 48]) and was refined after discussion with the analysis team and interview partners. After the first round of coding, discrepancies between the results of the two coders were analyzed and discussed, resulting in a unified section encoding for each of the videos.

3.1.2 Study procedure. To directly receive feedback from a potential target audience on existing scholarly article summaries, a user study was conducted with a total of 68 participants who were asked to watch and rate a subset of the 40 videos. The participants were recruited via mailing lists and/or university classes and had a background in computer science (75%) or data analytics (20%). The group consisted of 24 women and 44 men with an average age of 23.7 years (ranging from 20 to 46 years, standard deviation of 5.3). The most common highest level of education was a high school diploma (43%), followed by a bachelor's degree (34%), a master's degree (10%), and a Ph.D. (7%).

Participants were provided a web interface in which they played the videos, rated them, and optionally wrote down notes or comments. The system was designed to evenly distribute the videos in a round-robin fashion. However, due to an initial misconfiguration in the distribution mechanism, three of the videos were rated more often (38 to 41 ratings) while the least rated video only got 13 ratings. The median amount of ratings per video was 18.

In the user interface shown to the study participants, we provided an explanation for each of the three categories at the top of the page:

- Creativity refers to how creative you find the video itself; this does not necessarily have to match the creativity of the research paper's topic.
- Enthusiasm should reflect how enthusiastic the creators of the video appear to you, which does not necessarily have to match your personal enthusiasm about the topic.
- Informativity is supposed to reflect how much information you think is communicated through the video.

The combined results of the video analysis and the user study were used for the formulation of video recommendations. Their statistical significance was tested using a two-sided T-test for each of the rating categories with $\alpha = 0.05$. The equality of the variances of the two groups, e.g., "introduction section present" or "not present", was tested using the Levene test.

3.2 Design process

The tool development process followed an iterative, user-centered approach, consisting of an analysis phase in which users and their needs were identified, a conception phase with the aim to define user interactions and possible workflows within a system to support the planning of videos, a design phase that resulted in multiple prototypes which were refined in several iterations and an evaluation phase in the form of interviews and usability tests.

We conducted semi-structured one-on-one interviews with four experienced researchers from our research group who were confronted with the task of summarizing publications or research projects in form of a short video. They were asked about their intentions for the final result, prior experience with video editing software, challenges with the creation of their video and wishes for a tool to support them with future videos. One of the participants (I_1) had a background in digital humanities and design and was experienced with creating non-scientific videos. Two other interviewees (I_2 and I_3) had a background in visual data analysis and I_3 had some experience with the creation of scientific videos for university courses. I_4 had a physics background and no experience with video creation at all. At least two interviews were conducted per person, one in the beginning and one at the end of the video production. Based on these results, low- and high-fidelity prototypes were developed, covering different potential workflows. These were also discussed with the interview partners in separate meetings.

3.3 Evaluation

The evaluation follows a four-tier-evaluation methodology consisting of (1) amateur and (2) expert interviews, (3) two rounds of usability testing and (4) two case studies conducted with students from different academic backgrounds.

3.3.1 Amateur and expert interviews. After the prototyping phase and before the first round of usability tests was conducted, I_1 to I_4 were also asked for feedback on the final application. The study was exploratory and participants were asked about their general opinion as well as useful and missing features. Additional interviews with two experts for scientific videos were conducted during later stages of the development, i.e. between the two rounds of usability tests, in order to get ideas for final improvements and additional features which could be added in the future. They were selected because of their years of expertise with video creation both as publication authors and journal editors asking researchers for videos about their papers. Similar to the other interviews, they were introduced to the web application and asked for their general opinions and possible improvements.

3.3.2 Usability testing. Two rounds of usability testing were conducted: the first one (UT_1) with 5 participants, mainly with a computer science background, and the second one (UT_2) with 10 participants from different research fields such as social sciences, political science and life sciences. All participants were recruited from the extended work environment of the project team and were not involved in the project otherwise. The tests were conducted online via Zoom, using Google Forms for the pre- and post-questionnaires that included prior experience with video creation, confidence about creating scientific videos, an adapted version of the System Usability Scale [12] and open questions regarding positive and negative aspects of the application. The tasks that had to be completed during the test remained the same for both rounds and covered the full workflow of the application by making the users plan a hypothetical video about a scientific paper or thesis they had recently written. A document containing the questionnaires as well as the task descriptions can be found in the supplemental material.

3.3.3 Case Studies. After the usability tests and resulting improvements were completed, the application was subjected to a more extensive evaluation in the context of two university courses: one for theater and film students and another interdisciplinary course with a main focus on law and education.

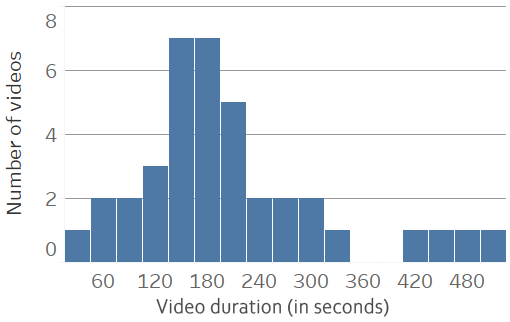

Figure 2: Histogram of video durations with bins of 30 seconds.

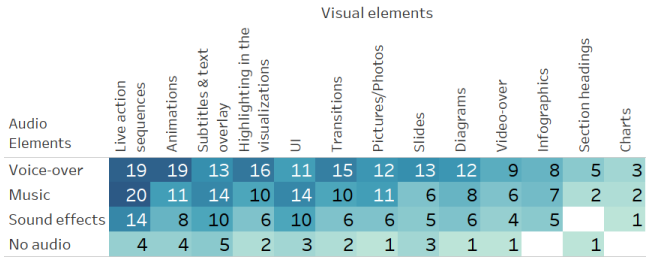

Figure 3: Heat map of visual and audio elements that appear in more than two videos. Each cell indicates how many videos share the given combination of visual and audio elements, e.g., there are 19 videos that share a voice-over and live action sequences.

Testing the generalizability of the tool's interface and recommendations to other domains or related use cases, the students involved in the law course were confronted with the task of creating educational videos about privacy and Internet security for children. 11 groups of two to four people had to produce two videos using the tools Toonly [56] or Doodly [55]: one for children from six to ten years and another one for children from ten to 14 years, both covering the same topic, e.g., "Social Media". After being provided with a handbook for Pub2Vid which describes the project in general and the main functionality together with a tool walk-through using screenshots as well as a tutorial video in form of a screen recording, the students could choose whether they wanted to use the tool, which 4 out of the 11 groups did. After the course, they were asked to fill out a questionnaire about their experiences with Pub2Vid and why they chose to use the tool or not.

As the students in the theater and film course were not required to plan or produce a full video using the tool but had to give written feedback on their first impressions as well as their considerations regarding the tool's usefulness, they were not provided with a tutorial video nor an extensive introduction, but had access to the same handbook as in the law course. A total of 17 multi-page feedback documents were submitted which are summarized in the Evaluation section of this paper.

4 VIDEO ANALYSIS

The goal of the video analysis was to identify patterns in the video structures and link them to video popularity by means of rating scores. The results of the video analysis consist of two parts: general video information as well as user ratings for each of the videos. The full data can be found in the supplemental material of this paper.

4.1 General video information

The majority of the selected videos has a duration of two to four minutes with an overall average of 222 seconds and a standard deviation of 92. A histogram depicting the distribution of video durations can be found in figure 2.

Figure 3 shows a heat map of visual and audio elements appearing in the same video. 50 percent of the analyzed videos contain a combination of music and live-action recordings, almost the same amount of videos use such recordings together with a voice-over or combine a voice-over with animations. Infographics, section headings, and charts are the least frequently used visual elements. The majority of videos use sound, i.e. a voice-over, music, and/or sound effects.

The distribution of the different video sections over the relative progress of a video can be seen in figure 4. It is clearly visible that the majority of videos start with a title slide, often followed by an introduction section, and finish with an ending slide. While methodology and system architecture sections are slightly more likely to appear in the first half of a video, evaluation, and result sections usually appear in the second half. Demo sections can appear almost anywhere within the video, similar to reflection sections which are not used that frequently but show occurrences both in the beginning and at the end of some videos.

4.2 Study results

The user study extended the existing video analysis results with ratings from 1 to 5 in the three categories "informativity", "creativity" and "enthusiasm", providing an average number of 20 ratings per video. Aggregated over all participant ratings, the average score was 3.5 for creativity, 3.3 for enthusiasm, and 3.6 for informativity with a standard deviation of 1.2 in all three categories. The overall results show that visual and audio elements as well as sections and their duration can have an impact on a video's rating. Figure 5 visualizes the average video ratings based on the presence or absence of different audio elements, namely music, sound effects, a voice-over or any of the three elements. It can be seen that videos containing a voice-over have significantly higher informativity scores while those with music and sound effects have higher ratings in creativity and enthusiasm. Videos that contain no sound at all were rated lower in all three categories which also matches the comments written by the participants, stating that these videos were usually perceived as disengaging and hard to follow. Moreover, the comments negatively highlighted computer-generated voice-overs as boring and irritating.

For visual elements, the difference between the two groups of videos was often not as crucial. However, for live-action sequences, significant growth in creativity and enthusiasm scores could be observed, leading to an average of 2.9 vs 3.7 for creativity and 2.8 vs. 3.5 for enthusiasm. Single users also wrote positive comments about the "movie-like" experience when watching the videos.

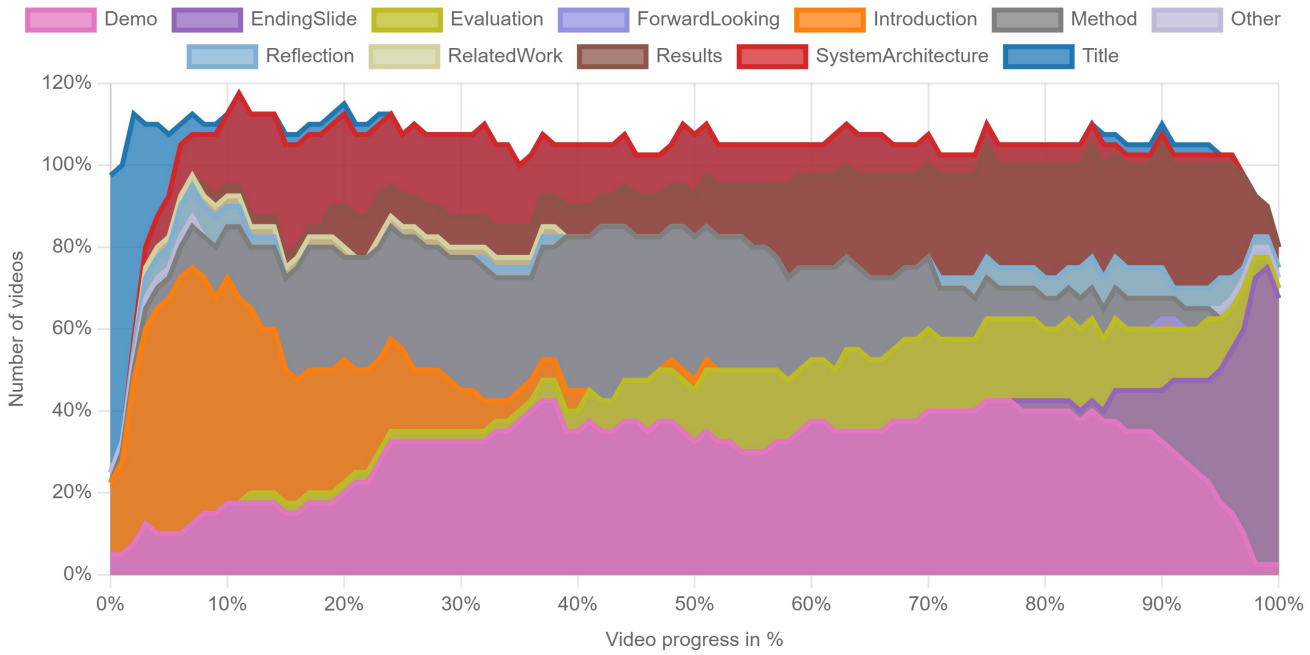

Figure 4: Distribution of video sections over temporal video progress. Note that overlapping sections can lead to a total percentage of more than 100.

As it can be seen in figure 6, the presence and duration of single sections also showed an impact on the videos' ratings. While videos containing a methodology section showed higher average informativity scores and those including an introduction had higher ratings in both, informativity and enthusiasm, the presence of a reflection section even led to higher scores in all three categories. Moreover, the total number of sections as well as the absolute video duration in seconds shows a strong correlation with the average informativity score.

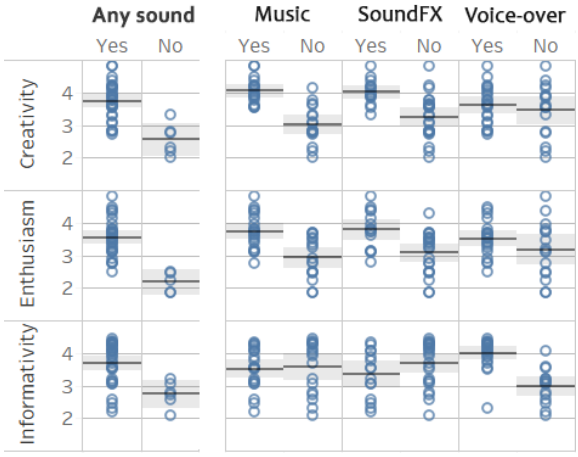

Figure 5: Average video ratings based on the presence of single audio elements. Each circle in the plot represents a video, the black line in each cell indicates the average rating over all videos of a specific group and the grey area visualizes a 95% confidence interval.

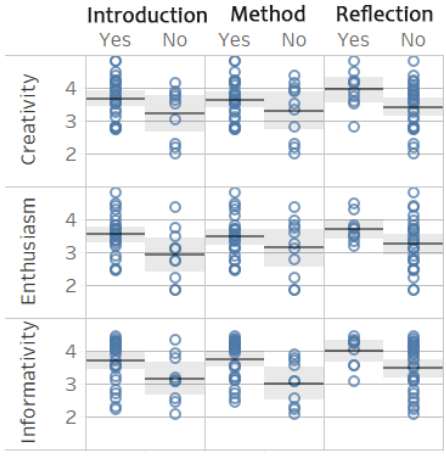

Figure 6: Average video ratings based on the presence of single sections (b). Each circle in the plot represents a video, the black line in each cell indicates the average rating over all videos of a specific group and the grey area visualizes a 95% confidence interval.

5 DESIGN PROCESS

The requirements for a tool supporting researchers with the authoring of their video summaries were analysed based on interviews and incorporated into multiple iterations of prototype development. Their condensed version is implemented as a web application which, together with its design decisions, is discussed in the second half of this section.

5.1 Requirement analysis and prototyping

Participants I_1, I_2, and I_4 used similar workflows for video production that started with planning and outlining, then gathering figures and footage, writing a script and recording a voice-over using various software. I_3 had a different approach, starting with presentation slides, writing a script in the speaker notes, and recording the voice-over with the presentation. As a reason for this differing workflow, he stated that the reorganization of scenes, i.e. slides, can be done on the fly or with very little effort and that he is used to this workflow from teaching university classes. The participants identified a need for real-world examples and an estimated script reading duration to simplify the process.

The requirements identified during the interviews are also reflected in related literature ([31, 32, 41, 42]), resulting in the following goals for the application's design: (1) startup-support, especially for inexperienced users, by providing a low-level entry and real-world examples, (2) adaptability to individual needs and specifications, e.g., maximum video duration, (3) integration or compatibility with existing workflows and tools, (4) flexibility for assembling and reorganizing the video structure, and (5) script writing functionality and an estimated voice-over duration based on the video script.

Multiple prototypes were developed that suggest different workflows and functionalities. While one of them was based on classic slideshow presentation software such as Microsoft PowerPoint and another one rather resembled video editing software like Adobe Premiere, the third prototype focused more on the storytelling aspect, starting with the video script and only then filling the scenes with assets such as images or videos. Since the latter provided space for integrating the most useful features and received the best feedback from the interview partners, it was decided to use this approach for the further development and integrate single functionalities from the other prototypes that were positively highlighted.

5.2 Walkthrough and design decisions

The final application is written in Vue.js with a Python backend and can be accessed at https://github.com/katharinawuensche/Pub2Vid. A video describing the interface as well as the workflow which was itself structured using Pub2Vid can be found in the supplemental material of this paper. The following section covers the most important interface components and design decisions.

5.2.1 Settings. The video settings available to the user concern the desired style (creative, informative, or enthusiastic), set via a triangle picker, and the video duration, set via a slider (fig. 7a and b). By clicking on the "Show matching videos" button (fig. 7c), real videos matching these settings can be played along with their overall structure which can also be reused for the user's project. Based on the configuration, suggestions for the video structure are computed and displayed as suggested outlines (fig. 7d). By clicking on one of the templates, the user can select it for further adjustment in the next step.

Design decisions. The aim of the triangle picker is to encourage users to reflect on the trade-off between different settings for the video style. Unlike independent sliders for each variable, it makes them dependent to prevent users from maximizing all values at once. This design decision is based on the strong correlation of the video ratings in the three categories. The timeline containing suggested outlines was redesigned after the first round of usability tests, initially showing only one suggestion which was automatically used as a template for further editing. Users attempted to modify the outline by dragging or editing its sections. To avoid interfering with the conceptual separation of settings and detailed structure refinement, users were given a choice of three different outline suggestions, while detailed refinement was kept on the script page.

5.2.2 Script. The script section contains tools for refining and writing the video script. The video structure can be adjusted by adding, moving, or deleting sections (fig. 8a&b). Moreover, the average speed of the voice-over can be set (fig. 8c) and compared with real video examples (fig. 8d). This value is used to estimate the reading time of each section and provides users with a warning if their script exceeds the planned section duration.

The lower half of the interface contains the area for writing the video script (fig. 8e) where the text for each section can be entered and compared with the planned duration ("Est. reading time"). Parts of the text can also be formatted and provided with a background color to exclude them from the calculation, e.g. because they are instructions (pauses, sound effects, etc.).

Design decisions. Video duration is one of the few specifications authors get from a venue and the key feature limiting the amount of text that can fit in the script. As the analysis showed that a video's structure and its style can influence each other, it was decided to use these two factors for the computation of video outlines and recommendations. Since the refinement of the overall video outline and the writing of the actual script are closely coupled tasks, it was decided to display the controls for both of them on the same page. If users realize that the script for one of their sections exceeds its desired duration, they can adjust it without having to navigate to another page. The suggested sections mainly relate to computer science publications and videos which is why they might not fit perfectly into other academic fields such as social sciences. There is, hence, the option to add a section called "Other" as well as the possibility for users to rename their sections according to their needs.

5.2.3 Recommendations and Export. Based on the settings made in the beginning, the tool recommends visual and audio elements for the video.[1] By clicking on the corresponding entry, another window with example scenes from real videos using the element is opened. Additionally, sources for free images, icons, and background music are listed that redirect to external websites. In the last step, the

[1] See https://pub2vid.web.app/recommendations

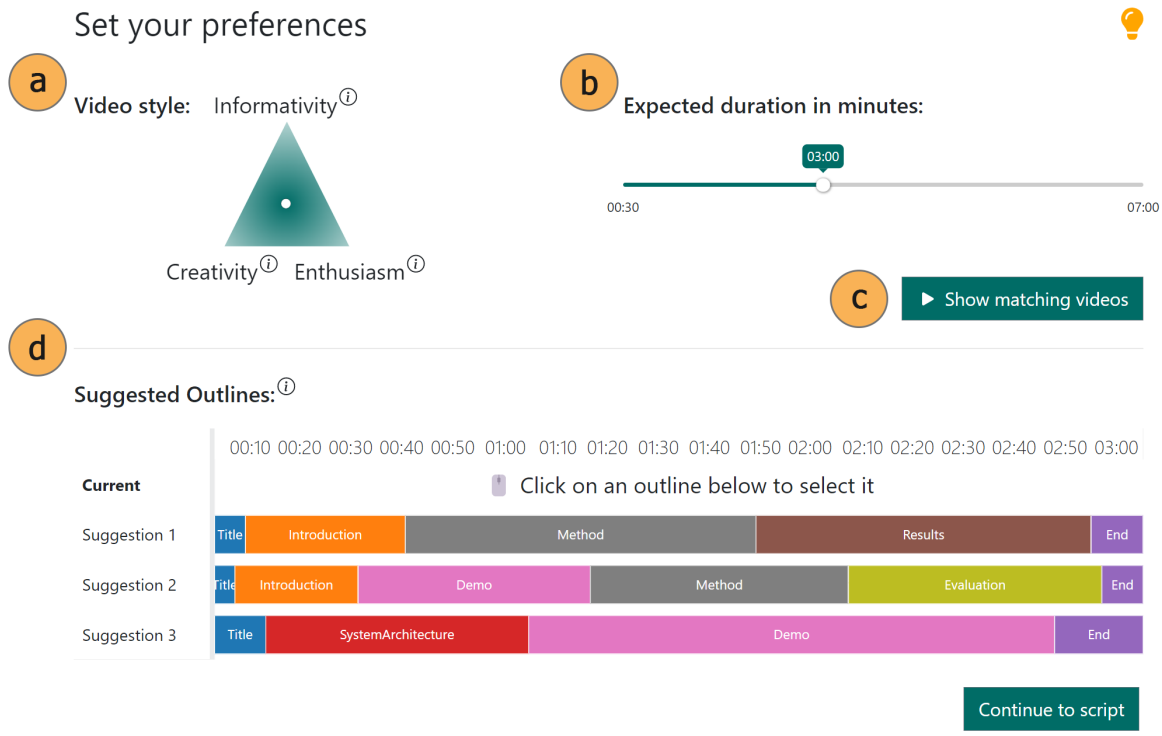

Figure 7: Screenshot of the settings page, including components for setting the video style (a) and duration (b), a button for playing real video examples in an overlay window (c), and a timeline component for the user to choose between suggested outlines (d).

export page[2] allows users to review their script and download it as a Microsoft Word file or a PowerPoint template which contains the script in the speaker notes and has predefined settings for the presentation slides matching the section lengths.

Design decisions. To address user irritation caused by the recommendations page appearing after the export step, the two pages were rearranged, causing the workflow to be finished with the script export. Suggestions for image and audio resources were integrated on the page to address difficulties in finding royalty-free options reported by the interview partners. Two different download functionalities were integrated based on observed user preferences and a limitation in the PowerPoint format. The "sript only" option was made available to preserve the formatting in the script which would get lost in the PowerPoint speaker notes such as font size and background color.

6 EVALUATION

The following section addresses the results of the evaluation of the application and its recommendations. These include interviews, usability tests, and case studies.

6.1 Amateur and expert Interviews

6.1.1 Nonprofessionals. Since I_1 to I_4 were already involved in the development and prototyping process before, the majority of features was not new to them, but their feedback was valuable nevertheless. I_1, I_2, and I_4 all stated that the tool would have been helpful to them during the creation of their videos and that the suggested workflow matches their own. I_2 mentioned the timeline as a particularly useful feature, while I_1 and I_4 both positively highlighted the estimation for the voice-over duration. I_4 also mentioned the video and resource suggestions as well as the questions related to the sections as helpful. Regarding possible improvements, I_1 suggested adding a justification for the single recommendations to the interface, e.g., "32% of the matching videos use this structure". He further questioned whether the PowerPoint export is useful as most universities or conferences provide their own presentation templates. Since I_3 used a different workflow than the one suggested by the tool, he was unsure whether he would use the application for future videos. However, he highlighted single features that could still be useful to him, namely the interactive timeline and the estimated voice-over duration. Moreover, he suggested adding an upload functionality for slides which could, in combination with extraction of the video outline and script from the existing presentation, make the tool also compatible with his workflow.

[2]https://pub2vid.web.app/export

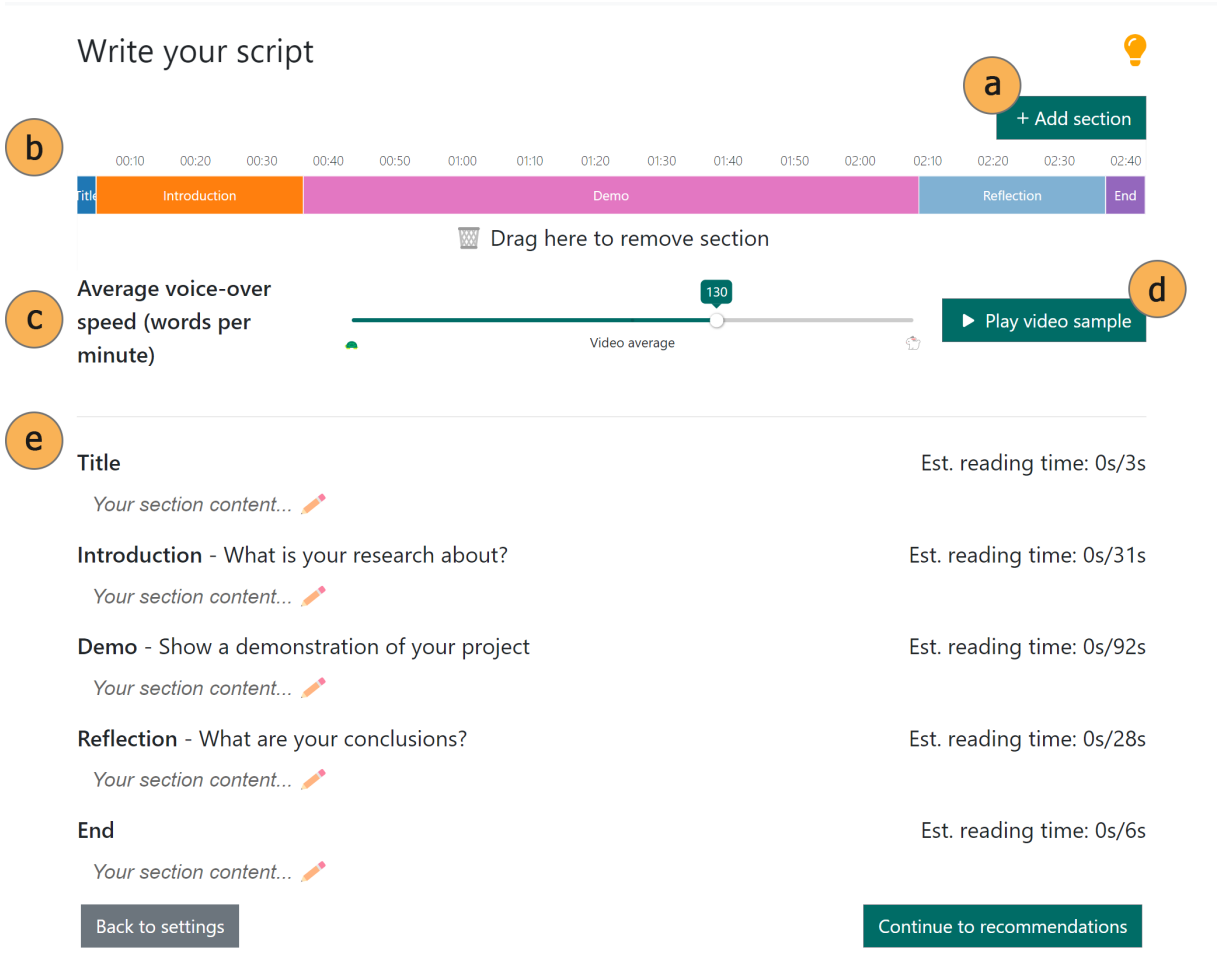

Figure 8: Screenshot of the script page, including components for adding new sections (a) and adjusting the video outline (b), setting the desired voice-over speed (c) and playing real video examples (d) as well as text inputs for writing the video script (e).

6.1.2 Experts. Being shown an initial version of the interface, the two science video experts suggested showing multiple outline suggestions instead of a single one and adding a justification for these suggestions. Moreover, they put a special focus on the video examples and recommended adding metadata for them, e.g., keywords or YouTube likes and dislikes. Adding an option to save one to three videos as a template and making them available on all pages, not just the settings screen should provide users with further guidance during the video structuring. Finally, they proposed adding an MP4 export functionality that generates a color-coded timeline that can be easily imported into video editing programs.

6.2 Usability testing

6.2.1 Qualitative results. The tool was generally well-received by the fifteen usability testers, who found it easy to use and appreciated its simple design. They found the feature of considering the video style in advance to be new but useful and felt that the tool could be particularly helpful for planning and structuring videos. The estimated voice-over duration and recommended image and audio resources were also highlighted positively.

Some participants suggested adding hints and more information on tool usage, which were added in the second round of usability tests (UT_2) through a short description at the beginning of each step in the web interface. In response to participants of the first round (UT_1) trying to interact with the suggested outline on the settings page, the number of suggestions was raised to three predefined templates from which users could choose in UT_2 (as discussed in section 5.2.1). Letting users define custom sections was another requested feature that was added after UT_1.

The video examples and outlines displayed underneath the timeline on the settings page were found to be confusing or ignored by some participants of UT_1, leading to the implementation of a button to display them on demand inside a popup. In UT_2, users complained about the 30-second step limit for setting the duration and the lack of exact section length in the timeline. One participant asked for a video output instead of just a text or presentation file and another one suggested letting users change the colors of

sections as well as marking them as fixed during the structuring and outline refinement process. Regarding the predefined sections, participants of UT_2 mentioned that some of them were not applicable to their domain, e.g., "System Architecture" and that they were missing titles such as "Theory" which would be subject to a future extension of the underlying corpus and analysis.

6.2.2 Quantitative results. Both rounds of usability tests show average SUS scores of 93, despite the fact that the participants of UT_2 had different backgrounds and less experience with computer science. The detailed results and exact values can be found in the supplemental material of this paper. Even though more participants of UT_2 thought that they would need the support of an experienced user to be able to use the system and slightly more users thought that they needed to learn a lot before getting used to working with it, none of them gave a rating of more than 1 (completely disagree) for the statements regarding inconsistencies and cumbersomeness. Interestingly, the two users giving a neutral response regarding their confidence while using the system were the only ones who had not created a video before which might have led to lower confidence in video planning in general. Overall, the results support the observations made during the qualitative evaluation that the system is user-friendly even for inexperienced users.

6.3 Case studies

6.3.1 Law course. Out of the seven groups that handed in the feedback questionnaire for the Pub2Vid application, four decided to use the tool for planning their videos, arguing that the tool provides a "good way of structuring a video", "getting an overview of the outline" and "estimating the duration of the script". The groups that decided not to use the tool argued that it would "limit the creative freedom for structuring the video", that "telling a story for children differs too much from a scientific presentation" and that they "worked spontaneously and driven by [their] ideas".

Table 1 shows the quantitative results of the questionnaire. It can be seen that groups that decided to use the application needed less time for both, setting up a story and recording the voiceover based on their script. While there seems to be no difference in confidence when it comes to video production between the groups that used the tool and those that did not, the confidence value for the structuring process seems to be lower for students who decided to use the tool. However, the causal relationship is still unclear: students might have decided to use the tool because they felt insecure about the structuring process or they felt less confident after having used the tool.

Out of the four groups that used the tool, two stated that they found the suggested outlines helpful and three reported that the video examples were useful - even though they came from a different domain. Furthermore, it was suggested to let users set custom colors for their sections and to export the video outline as a timeline as well as make the renaming of sections more intuitive.

6.3.2 Theater and film course. The positively-highlighted features of the application match those mentioned in the other rounds of evaluation. Single participants also mentioned the questions next to the section titles, e.g., "How did you evaluate your solution or findings?", and the option to select a real video's structure for one's own outline as helpful. The participants criticized the responsiveness of the timeline and the 7-minute video duration limit as well as the unclear meaning of "enthusiasm" in the video style categories. They also suggested not separating creativity and enthusiasm and criticized the inability to edit video outlines on the settings page and the potential inaccuracy of estimated reading time.

Some of the reviews also included suggestions or requests for new features such as integrating a basic video creation functionality, allowing users to select their scientific field in the beginning of the workflow, and enhancing the corpus of underlying videos. Minor improvements include providing more links for video and audio resources, adding examples for recording and video editing software, and reworking some of the interface labels. Moreover, it was recommended to mention the PowerPoint export functionality earlier and use the tool for presentation generation instead of restricting it to videos.

7 DISCUSSION

Our paper focuses on both, the analysis of existing publication summarization videos and the development of a prototype that supports users with the authoring of their videos as well as insights gained from its detailed evaluation. We discuss our findings in more detail in the following section.

7.1 Video recommendations

Based on the results of the general video analysis and the evaluation studies, we can derive the following recommendations for publication summarization videos within the given scope, i.e. ranging from two to seven minutes and with a focus on visual computing.

- (R_1) **Use sound.** Videos that contain neither music nor sound effects nor a voice-over were rated significantly lower in all three categories.
- (R_2) **Choose the right type of sound.** For an informative video, provide a voice-over. For a creative or enthusiastic one, use music and/or sound effects.
- (R_3) **Avoid using a computer-generated voice for your voice-over.** Videos that made use of a synthetic voice for the voice-over received more negative feedback and were perceived as disengaging and hard to follow.
- (R_4) **Reflect on your work.** Videos that contain a reflection section show a higher average rating in all three categories.
- (R_5) **Give an introduction.** Videos that contain an introduction section show a higher average rating in enthusiasm and informativity.
- (R_6) **Use live action sequences.** Videos that contain live-action sequences show a significantly higher creativity and enthusiasm score.
- (R_7) **Be thorough.** Longer videos and videos with a higher amount of sections tend to have a higher informativity score.
- (R_8) **Explain your methodology.** Videos that contain a methodology section show a significantly higher informativity score.

Despite the different methodological approaches, some of these recommendations are also supported by other related work, especially R_1, R_2 ([45, 50, 52, 53]) and R_4 ([50, 52]). Since many existing guidelines refer to instructional videos rather than to paper summaries or video abstracts, they do not mention concrete publication sections

Used the tool	Number of group members	Confidence with structuring/producing the video (1-5, per group member)	Time needed for setting up a story (in h)	Time needed for writing the script (in h)	Time needed for recording the voice over (in h)	Time needed for final video production (in h)	Points awarded for the video (0-55)
no	3	4/4; 4/2; 4/3	4.5	1.5	6	16	51.5
no	3	4/3; 4/4; 4/3	2.5	4	10	35	49
no	2	5/5; 5/5	4.5	6	3	40	55
yes	3	4/3; 3/4; 5/3	-	-	-	-	55
yes	2	4/3; 3/3	2	12	1	30	48.5
yes	3	3/2; 3/4; 4/4	2	3.5	2	20	55
yes	2	4/4; 3/4	0.75	1	0.33	16	49.5

Table 1: Quantitative results of the law course feedback, including the number of group members, the confidence with structuring and producing the video per group member, the estimated time needed for the individual video creation steps (setting up a story, writing the script, recording the voice-over and producing the video) as well as the points awarded for the final video. Since one group only gave very imprecise information about their estimated time needed, e.g., "several hours" instead of a numeric value, their results are not shown here.

such as a methodology or reflection section. However, Van der Meij and Hopfner [52] propose to include both a preview and a review section of the task covered in an instructional video which can be considered analogous to R_5 and R_4. What is remarkable about other video recommendations is that the majority ([11, 45, 50, 52, 53]) recommends keeping the videos or sections as short as possible. While this may appear contradictory to R_7, it has to be kept in mind that the recommendations proposed here are based on a limited subset of videos and should hence only be applied to videos within a similar scope regarding their duration and domain. It is therefore not stated in R_7 to create an infinitely long video in order to maximize its informativity score but to stay within the upper end of the two to seven minutes range. Finally, it should be highlighted that the ratings in the three categories of creativity, enthusiasm, and informativity show high correlations and that it is discouraged to only focus on one of these aspects. Instead, a balance between them should be maintained and can be beneficial for the overall viewer experience.

7.2 Video script authoring

Based on the existing approaches in the field of document summarization ([18, 22, 33, 39, 61]) and automated video creation ([34, 38, 44]), it can be said that summarizing scientific publications as videos is still hard to automate and requires human-in-the-loop solutions instead. Data-driven recommendations can support researchers with the planning and structuring of their own videos as it is done by Pub2Vid as a proof of concept. Our approach solves challenges in getting started with video creation, e.g., lacking confidence when not knowing what to begin with and spending too much time or having to re-start the video production phase due to a skipped planning and structuring step [32].

The evaluation of the application shows that it strikes a balance between flexibility and guidance. Real video examples are appreciated even by users of other domains, but their exact structures are rarely used for new projects. It is therefore important to provide users with an intuitive solution for modifying and rearranging the sections of their outline. Allowing users to define custom sections ensures generalizability to non-computer science videos as stated by participants of the usability tests and case studies. The visual and audio elements in the tool can have an impact on a video's rating, but this depends on the content and suitability for presentation. The estimated voice-over duration and interactive timeline were often highlighted as users' favorite features, and suggestions for image and audio resources were also well-received as they directly help with saving time for script re-recording and resource searching.

7.3 Limitations and future work

The scope of the underlying data is still limited. As the videos are not evenly distributed over the range of durations nor in the dimensions of the three rating categories, the reliability of the recommendations provided by the tool strongly depends on the input values and is less significant for edge cases. Our selection of venues is even more limited and may have an impact on the analysis results, even though the user ratings did not show any venue-specific comments. Extending this data with additional videos and ratings would hence allow for more versatile and reliable recommendations.

The selection of the categories used for setting the video style was criticised by single users and should be reflected upon. Since the average video ratings showed strong correlations for these categories, it could be argued that their number should be reduced, e.g., by combining creativity and enthusiasm which show the highest correlation. While some users felt limited in their creativity when having to set a video style, others stated that reflecting upon the style in advance was helpful for them. One possible solution for this discrepancy could be to make the style selection optional and include all videos as a base for the suggestions and recommendations when no value is selected. Even though the users' satisfaction with the presented suggestions has been partially assessed during the evaluation, the system would benefit from an additional round focusing on improving these aspects.

The tool's name "Pub2Vid" might raise false expectations in users as it suggests a video output instead of a script or a presentation template. Since a video creation functionality was often requested as an additional feature, adding a simple video editor should be subject to future work on the application. Moreover, an extended export functionality that supports file formats of common video editing tools could also improve the symbiosis of this video structuring tool and the later video creation.

While university students and staff are part of the main target group of our application and the scientific videos, an expansion of the user study to include other groups of people would make our results even more generalizable. In addition, it would be interesting to compare the analyzed videos with the associated publications to assess their completeness.

8 CONCLUSION

This paper has four main contributions: (1) a video structuring framework that supports users with the authoring of their videos, (2) a set of video recommendations based on the analysis of 40 publication summarization videos, (3) a 4-tier evaluation methodology to assess the usefulness and applicability of the recommendations and a tool implementing them, and (4) the underlying data in form of analysis results of the 40 videos together with ratings and comments of 68 user study participants.

We introduce the tool Pub2Vid to support researchers in planning and structuring videos summarizing their publications. It is designed based on multiple stages of prototyping, addressing the needs of real users involved in the development process. The tool provides recommendations for a video outline, audio and video elements, and a script writing interface that estimates the duration of the voice-over. The final video script can be downloaded as a text file or presentation template. An evaluation of the tool was conducted with users from various domains and showed positive results in two rounds of usability tests. Moreover, two case studies were conducted with students of law/educational sciences and theater/film courses, resulting in positive feedback regarding the tool's workflow, usability, and features. Despite the limitation of the underlying video corpus to visual computing fields, the results indicate the tool's generalizability to other domains.

Future work on this topic and the project itself should address the collection of more data, i.e. more baseline videos, also from other domains, as well as the implementation of a simple video creation functionality such that, after having finished the video structuring and script writing, researchers can start creating their scenes and fill them with assets such as recordings, images and sound. Simplifying the full video creation process can lower the barrier for researchers to summarize their publications in form of a video and make their work accessible to an even broader public.

REFERENCES

[1] ACM SIGGRAPH. 2019. SIGGRAPH 2019. https://s2019.siggraph.org/. Accessed: 2022-09-12.

[2] ACM SIGGRAPH. 2021. Video Recording Recommendations. https://s2021.siggraph.org/wp-content/uploads/2021/04/SIGGRAPH-2021-Video-Recording-Recommendations-04_28_21.pdf. Accessed: 2022-09-12.

[3] ACM SIGGRAPH. 2022. SIGGRAPH 2022. https://s2022.siggraph.org/. Accessed: 2022-09-12.

[4] Adobe 2003. *Premiere Pro.* Adobe. https://www.adobe.com/products/premiere.html

[5] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Andres Monroy-Hernandez, and Pourang Irani. 2016. Authoring data-driven videos with dataclips. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2016), 501–510. https://doi.org/10.1109/TVCG.2016.2598647

[6] Animoto Inc 2022. *Animoto.* Animoto Inc. https://animoto.com/

[7] Apple 1999. *iMovie.* Apple. https://www.apple.com/imovie/

[8] Apple 2011. *Final Cut Pro.* Apple. https://www.adobe.com/at/products/premiere.html

[9] Jacob Beautemps and André Bresges. 2021. What comprises a successful educational science YouTube video? A five-thousand user survey on viewing behaviors and self-perceived importance of various variables controlled by content creators. *Frontiers in Communication* 5 (2021), 600595. https://doi.org/10.3389/fcomm.2020.600595

[10] Gordon H Bower and Michal C Clark. 1969. Narrative stories as mediators for serial learning. *Psychonomic Science* 14, 4 (1969), 181–182. https://doi.org/10.3758/BF03332778

[11] Cynthia J. Brame. 2016. Effective Educational Videos: Principles and Guidelines for Maximizing Student Learning from Video Content. *CBE—Life Sciences Education* 15, 4 (2016), es6. https://doi.org/10.1187/cbe.16-03-0125 arXiv:https://doi.org/10.1187/cbe.16-03-0125 PMID: 27789532.

[12] John Brooke. 1996. SUS—A Quick and Dirty Usability Scale. *Usability evaluation in industry* 189, 3 (1996).

[13] Rui Cai, Lei Zhang, Feng Jing, Wei Lai, and Wei-Ying Ma. 2007. Automated music video generation using web image resource. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, Vol. 2. IEEE, IEEE, Honolulu, HI, USA, II–737. https://doi.org/10.1109/ICASSP.2007.366341

[14] Canva 2013. *Canva.* Canva. https://canva.com/

[15] CHI. 2019. Welcome to chi 2019. https://chi2019.acm.org/. Accessed: 2022-09-12.

[16] CHI. 2022. Welcome to chi 2023. https://chi2023.acm.org/. Accessed: 2022-09-12.

[17] Peggy Chi, Zheng Sun, Katrina Panovich, and Irfan Essa. 2020. Automatic Video Creation From a Web Page. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. Association for Computing Machinery, New York, NY, USA, 279–292. https://doi.org/10.1145/3379337.3415814

[18] Shai Erera, Michal Shmueli-Scheuer, Guy Feigenblat, Ora Peled Nakash, Odellia Boni, Haggai Roitman, Doron Cohen, Bar Weiner, Yosi Mass, Or Rivlin, Guy Lev, Achiya Jerbi, Jonathan Herzig, Yufang Hou, Charles Jochim, Martin Gleize, Francesca Bonin, Francesca Bonin, and David Konopnicki. 2019. A Summarization System for Scientific Documents. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations*. Association for Computational Linguistics, Hong Kong, China, 211–216. https://doi.org/10.18653/v1/D19-3036

[19] Rosemary Feasey. 2012. *Creative science: Achieving the WOW factor with 5-11 year olds.* Routledge, New York, NY.

[20] Nate Fedrizzi. 2022. What Makes an Effective Science Video? https://ssec.si.edu/stemvisions-blog/what-makes-effective-science-video. Accessed: 2022-09-12.

[21] Flavio Figueiredo, Jussara M. Almeida, Fabrício Benevenuto, and Krishna P. Gummadi. 2014. Does Content Determine Information Popularity in Social Media? A Case Study of Youtube Videos' Content and Their Popularity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 979–982. https://doi.org/10.1145/2556288.2557285

[22] Tsu-Jui Fu, William Yang Wang, Daniel McDuff, and Yale Song. 2022. DOC2PPT: Automatic Presentation Slides Generation from Scientific Documents. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. Association for the Advancement of Artificial Intelligence, Online, 634–642. https://doi.org/10.48550/arXiv.2101.11796

[23] Kendall Goodrich, Shu Z Schiller, and Dennis Galletta. 2015. Consumer reactions to intrusiveness of online-video advertisements: do length, informativeness, and humor help (or hinder) marketing outcomes? *Journal of advertising research* 55, 1 (2015), 37–50. https://doi.org/10.2501/JAR-55-1-037-050

[24] A. Graesser, Brent A. Olde, and Bianca Klettke. 2002. How does the mind construct and represent stories. In *Narrative impact: Social and cognitive foundations*. Psychology Press, Hove, East Sussex, United Kingdom, 38–39.

[25] Richang Hong, Jinhui Tang, Hung-Khoon Tan, Chong-Wah Ngo, Shuicheng Yan, and Tat-Seng Chua. 2011. Beyond search: event-driven summarization for web videos. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 7, 4 (2011), 1–18. https://doi.org/10.1145/2043612.2043613

[26] Qian Huangfu, Hong Li, Sanshan Tang, Jianrong Wang, Qian Liu, and Guojun Chen. 2022. How teacher enthusiasm affects students' learning of chemistry declarative knowledge in video lectures. *Chemistry Education Research and Practice* 23, 4 (2022), 898–912. https://doi.org/10.1039/D2RP00095D

[27] IEEE CG&A. 2022. CG&A Video Guidelines. https://computer.org/csdl/magazine/cg/write-for-us/102681. Accessed: 2022-09-12.

[28] IEEE CG&A. 2022. IEEE CG&A. https://www.computer.org/csdl/magazine/cg. Accessed: 2022-09-12.

[29] IEEE COMPUTER SOCIETY. 2019. VIS 2019. http://ieeevis.org/year/2019/welcome. Accessed: 2022-09-12.

[30] IEEE COMPUTER SOCIETY. 2019. WELCOME TO EUROVIS 2019. http://gpcg.pt/eurovis2019/. Accessed: 2022-09-12.

[31] Amir Jahanlou and Parmit K Chilana. 2022. Katika: An End-to-End System for Authoring Amateur Explainer Motion Graphics Videos. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 502, 14 pages. https://doi.org/10.1145/3491102.3517741

[32] Amir Hosein Jahanlou, William Odom, and Parmit Chilana. 2020. Challenges in Getting Started in Motion Graphic Design: Perspectives from Casual and Professional Motion Designers. In *Graphics Interface 2021*. Canadian Human-Computer Communications Society, Online. https://doi.org/10.20380/GI2021.06

[33] Jiaxin Ju, Ming Liu, Longxiang Gao, and Shirui Pan. 2020. SciSummPip: An Unsupervised Scientific Paper Summarization Pipeline. In *Proceedings of the First Workshop on Scholarly Document Processing*. Association for Computational Linguistics, Online, 318–327. https://doi.org/10.18653/v1/2020.sdp-1.37

[34] Murat Kalender, M Tolga Eren, Zonghuan Wu, Ozgun Cirakman, Sezer Kutluk, Gunay Gultekin, and Emin Erkan Korkmaz. 2018. Videolization: knowledge graph based automated video generation from web content. *Multimedia tools and applications* 77, 1 (2018), 567–595. https://doi.org/10.1007/s11042-016-4275-4

[35] René F. Kizilcec, Kathryn Papadopoulos, and Lalida Sritanyaratana. 2014. Showing Face in Video Instruction: Effects on Information Retention, Visual Attention, and Affect. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 2095–2102. https://doi.org/10.1145/2556288.2557207

[36] Mackenzie Leake, Abe Davis, Anh Truong, and Maneesh Agrawala. 2017. Computational video editing for dialogue-driven scenes. *ACM Trans. Graph.* 36, 4 (2017), 130–1. https://doi.org/10.1145/3072959.3073653

[37] Zechao Li, Jinhui Tang, Xueming Wang, Jing Liu, and Hanqing Lu. 2016. Multimedia News Summarization in Search. *ACM Trans. Intell. Syst. Technol.* 7, 3, Article 33 (Feb. 2016), 20 pages. https://doi.org/10.1145/2822907

[38] Rainer Lienhart, Silvia Pfeiffer, and Wolfgang Effelsberg. 1997. Video abstracting. *Commun. ACM* 40, 12 (1997), 54–62. https://doi.org/10.1145/265563.265572

[39] Elena Lloret, María Teresa Romá-Ferri, and Manuel Palomar. 2013. COMPENDIUM: A text summarization system for generating abstracts of research papers. *Data & Knowledge Engineering* 88 (2013), 164–175. https://doi.org/10.1007/978-3-642-22327-3_2

[40] Marcel Marti, Jodok Vieli, Wojciech Witoń, Rushit Sanghrajka, Daniel Inversini, Diana Wotruba, Isabel Simo, Sasha Schriber, Mubbasir Kapadia, and Markus Gross. 2018. CARDINAL: Computer Assisted Authoring of Movie Scripts. In *23rd International Conference on Intelligent User Interfaces (IUI '18)*. Association for Computing Machinery, New York, NY, USA, 509–519. https://doi.org/10.1145/3172944.3172972

[41] Naoko Nitta and Noboru Babaguchi. 2011. Example-based video remixing for home videos. In *2011 IEEE International Conference on Multimedia and Expo*. IEEE, Barcelona, 1–6. https://doi.org/10.1109/ICME.2011.6012080

[42] Naoko Nitta and Noboru Babaguchi. 2011. Example-Based Video Remixing Support System. In *Proceedings of the 19th ACM International Conference on Multimedia (MM '11)*. Association for Computing Machinery, New York, NY, USA, 563–572. https://doi.org/10.1145/2072298.2072371

[43] OpenAI. 2017. OpenAI - YouTube. https://www.youtube.com/c/OpenAI. Accessed: 2022-09-12.

[44] Mayu Otani, Yuta Nakashima, Tomokazu Sato, and Naokazu Yokoya. 2017. Video summarization using textual descriptions for authoring video blogs. *Multimedia Tools and Applications* 76, 9 (2017), 12097–12115. https://doi.org/10.1007/s11042-016-4061-3

[45] C. Plaisant and B. Shneiderman. 2005. Show Me! Guidelines for producing recorded demonstrations. In *2005 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC'05)*. IEEE, Washington DC, 171–178. https://doi.org/10.1109/VLHCC.2005.57

[46] Margret Plank and Jens Kösters. 2022. Improving guidelines for video abstracts: An analysis of the most popular videos abstracts in the TIB AV Portal. *Grey Journal (TGJ)* 18, 1 (2022).

[47] D. Shi, F. Sun, X. Xu, X. Lan, D. Gotz, and N. Cao. 2021. AutoClips: An Automatic Approach to Video Generation from Data Facts. *Computer Graphics Forum* 40, 3 (2021), 495–505. https://doi.org/10.1111/cgf.14324

[48] Thomas J Shuell. 1996. Teaching and learning in a classroom context. In *Handbook of educational psychology*. Prentice Hall International, New York, USA. https://doi.org/10.1016/B0-08-043076-7/02449-9

[49] John Simpson. 2008. The Oxford Pocket Dictionary of Current English.

[50] Maximilian Tandi, Lorena Niebuhr, and Eva-Maria Jakobs. 2020. How-To: Instructional Video. Recommendations for the Design of Software Video Trainings for Production Workers. In *ACHI 2020 : the Thirteenth International Conference on Advances in Computer-Human Interactions*. IARIA Press, Valencia, Spain, 378–383.

[51] Anh Truong, Floraine Berthouzoz, Wilmot Li, and Maneesh Agrawala. 2016. QuickCut: An Interactive Tool for Editing Narrated Video. In *UIST '16: Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. Association for Computing Machinery, New York, NY, USA, 497–507. https://doi.org/10.1145/2984511.2984569

[52] Hans van der Meij and Constanze Hopfner. 2022. Eleven Guidelines for the Design of Instructional Videos for Software Training. *Technical Communication* 69, 3 (2022), 5–23. https://doi.org/10.55177/tc786532

[53] Hans van der Meij and Jan van der Meij. 2013. Eight guidelines for the design of instructional videos for software training. *Technical communication* 60, 3 (2013), 205–228.

[54] Raphaela Martins Velho, Amanda Merian Freitas Mendes, and Caio Lucidius Naberezny Azevedo. 2020. Communicating Science With YouTube Videos: How Nine Factors Relate to and Affect Video Views. *Frontiers in Communication* 5 (2020). https://doi.org/10.3389/fcomm.2020.567606

[55] Voomly LLC 2022. *Doodly*. Voomly LLC. https://www.doodly.com/

[56] Voomly LLC 2022. *Toonly*. Voomly LLC. https://www.toonly.com/

[57] Elizabeth Walter. 2008. *Cambridge Advanced Learner's Dictionary with CD-ROM*. Cambridge university press, Online.

[58] Miao Wang, Guo-Wei Yang, Shi-Min Hu, Shing-Tung Yau, and Ariel Shamir. 2019. Write-a-Video: Computational Video Montage from Themed Text. *ACM Trans. Graph.* 38, 6, Article 177 (nov 2019), 13 pages. https://doi.org/10.1145/3355089.3356520

[59] Dustin Welbourne and Will J Grant. 2015. What makes a popular science video on YouTube. https://theconversation.com/what-makes-a-popular-science-video-on-youtube-36657. Accessed: 2022-09-12.

[60] Wibbitz 2011. *Wibbitz*. Wibbitz. https://wibbitz.com/

[61] Huiyan Xu, Zhijian Wang, and Xiaolan Weng. 2019. Scientific literature summarization using document structure and hierarchical attention model. *IEEE Access* 7 (2019), 185290–185300. https://doi.org/10.1109/ACCESS.2019.2960611

[62] Yipeng Yu, Zirui Tu, Longyu Lu, Xiao Chen, Hui Zhan, and Zixun Sun. 2021. Text2Video: automatic video generation based on text scripts. In *Proceedings of the 29th ACM International Conference on Multimedia*. Association for Computing Machinery, Online, 2753–2755. https://doi.org/10.1145/3474085.3478548