Taylor & Francis
Taylor & Francis Group

International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hihc20

HierVid: Lowering the Barriers to Entry of Interactive Video Making with a Hierarchical Authoring System

Weitao You, Zhuoyi Cheng, Zirui Ma, Guang Yang, Zhibin Zhou & Lingyun Sun

To cite this article: Weitao You, Zhuoyi Cheng, Zirui Ma, Guang Yang, Zhibin Zhou & Lingyun Sun (03 Nov 2023): HierVid: Lowering the Barriers to Entry of Interactive Video Making with a Hierarchical Authoring System, International Journal of Human–Computer Interaction, DOI: 10.1080/10447318.2023.2267859

To link to this article: https://doi.org/10.1080/10447318.2023.2267859

Published online: 03 Nov 2023.

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2023.2267859

Taylor & Francis
Taylor & Francis Group

HierVid: Lowering the Barriers to Entry of Interactive Video Making with a Hierarchical Authoring System

Weitao You[a], Zhuoyi Cheng[b], Zirui Ma[a], Guang Yang[c], Zhibin Zhou[d], and Lingyun Sun[a]

[a]College of Computer Science and Technology, Zhejiang University, Hangzhou, PR China; [b]School of Software Technology, Zhejiang University, Hangzhou, PR China; [c]Alibaba Group, Hangzhou, PR China; [d]School of Design, The Hong Kong Polytechnic University, Hong Kong, ROC

ABSTRACT

Interactive videos have been applied to various areas due to their engagement potential and efficiency improvement of information communication. However, creating interactive videos can be challenging because of a lack of novice-oriented guidance in current platforms, and the logic-building process when authoring interactive videos. To address these challenges, we obtained insights from four creativity support tool designers, proposed a series of hierarchical interactive video structures based on existing narrative structures, and presented the HierVid system. The system is designed as a Template-Module-Unit Mode-based hierarchical authoring platform grounded on three design requirements, and we conducted two user studies to evaluate HierVid. The results showed that novice users could get started to use and understand the functions easily, and the system allowed users to use and explore freely, with an enhanced efficiency compared to the bilibili platform. In conclusion, our research and design of HierVid offer guidance and support for novice users, making interactive video authoring quicker and more accessible.

KEYWORDS

Interactive video; hierarchical system; authoring system; narrative structure

1. Introduction

Video is an essential medium for conveying information and communicating. Among various visual communication channels such as text (Moura et al., 2016) and image (Occa & Suggs, 2016), video has been widely utilized mainly due to its efficiency and effectiveness in information conveyance (Goldberg et al., 2019; W. Li, 2023). However, the lack of interactivity in traditional videos does not allow reciprocal communication, and fails to let users feel in control during the experience, in that the above two aspects are considered two of the dimensions of interactivity (Gao et al., 2009). For example, users can hardly choose different ingredients or adjust the quantity when following a finely made cake-making tutorial video.

Compared to videos, media like websites and interactive fiction (IF) (Farias & Martinho, 2021) have engaged users in the experiencing process by allowing them to participate and control. For example. it is natural to click on hyperlinks and jump elsewhere when browsing a website, or be led to different endings if the users make different choices in IFs. Such processes can enhance user experience (Sutcliffe & Hart, 2017) and the effectiveness of contents (Fidan & Debbag, 2023), while interacting with videos remains a relatively new concept for many people.

In recent years, more focus has been cast on the interaction between users and all sorts of media that are originally considered as one-way information conveyors, including videos. Using interactive videos to replace traditional linear videos is becoming increasingly common in various processes such as online teaching and learning (Cattaneo et al., 2019; Sauli et al., 2018; Su & Chiu, 2021), advertising (Belanche et al., 2020; Gu et al., 2022), and even filming (Green et al., 2017).

Despite that interactive videos are applied to various fields, the number of user-generated interactive videos is much lower than traditional videos. While many interactive video authoring platforms (IVAP), such as dot.vu[1] and hihaho[2], support crafting and playing interactive videos, most platforms are geared towards business users, resulting in low quantities of interactive videos created by personal users. Furthermore, creating an attractive and logically fluent interactive video can be challenging for most non-professional users.

The challenges of crafting interactive videos can be attributed to two main factors. First, interactive videos usually have more intricate structures than traditional videos, making them less intuitive for non-professional users to understand. For example, in a traditional house tour video, scenes are displayed in a definite order, while in an interactive video, the displaying order is determined by users, indicating that branching and loop structures are exerted. Second, those who are unfamiliar with interactive videos have no idea where the proper place is to insert interactions or interactive structures, and how to plan an appropriate storyline when making such videos. Consequently, the quantity of

CONTACT Lingyun Sun sunly@zju.edu.cn College of Computer Science and Technology, Zhejiang University, Hangzhou, PR China

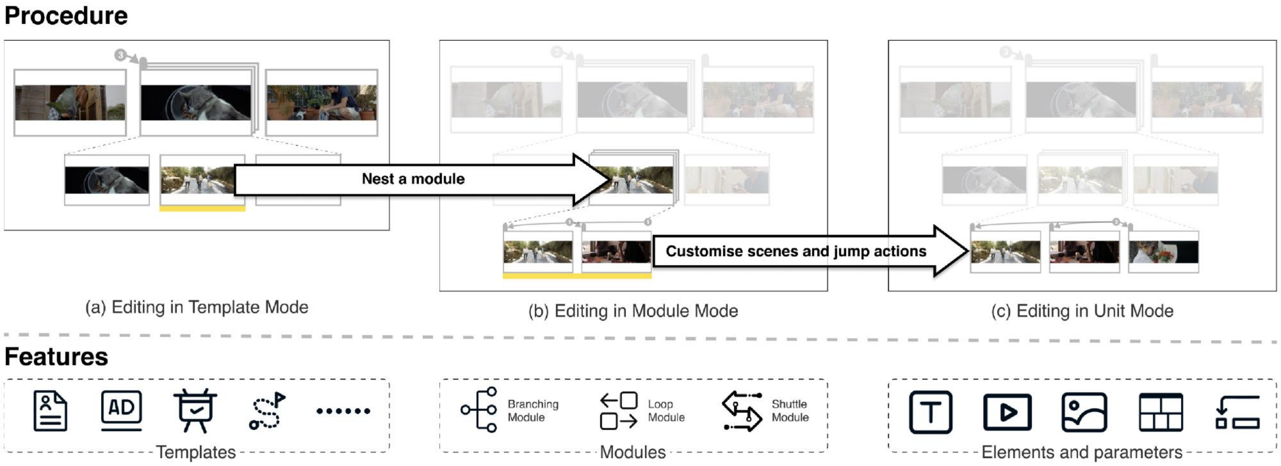


Figure 1. Example of the authoring process of building the interactive video structure hierarchically, and the main features in each mode. HierVid provides a Template-Module-Unit Mode based interface with different functions. It can guide novice users while retaining advanced functions for the users to explore and enhance their skills.

user-generated content (UGC) of interactive videos remains low, and their exposure to the public is insufficient, thus are relatively rarely used in real-life environments.

To tackle these challenges, we investigated existing IVAPs, gained insights and concluded design requirements from creativity support tool designers, and then borrowed lessons from previous work about text structures (Saunders-Smith, 2009) and narrative structures (Escalas et al., 2004; Partarakis et al., 2022). We proposed a series of interactive video structures that can hierarchically decompose and construct interactive videos, and developed the HierVid system, a hierarchical IVAP designed based on the design requirements and the proposed interactive video structures. Novice users can easily get started using our system, and feel free to use and explore the system. The hierarchical design with different features aims to offer users an easy-to-use yet powerful authoring tool with different features and functions provided in different modes (Figure 1). We evaluated HierVid regarding the ease of use of the system, flexibility, and authoring efficiency.

Our contributions are as follows: (1) a series of interactive video structures that can be used to decompose and construct interactive videos. (2) The design and implementation of HierVid, a hierarchical IVAP that applied the proposed interactive video structures. (3) Verification of the system accessibility, flexibility, and efficiency, both quantitative and qualitative results are included.

2. Related work

Relevant prior work includes research on tools for authoring interactive videos and video structure.

2.1. Interactive video authoring tool

Current IVAPs in the market are either for general goals or designed specifically for authoring online education videos. IVAP for general goals are platforms that are suitable for making interactive videos for various goals, including advertising, education, or entertainment (Aubert et al., 2012; Mendes et al., 2020). Online education or tutorial is an area that needs various unique functions because interactivity brings more convenience and experience to such a way of learning. Therefore, authoring platforms for creating interactive e-learning videos have also thrived (Bao et al., 2019; Layona et al., 2017; Ouh et al., 2022).

While the areas the platforms oriented might be different, we concluded three common methods to impart interactivity to videos in the authoring process according to the cases provided by Meixner (Meixner, 2018): add clickable elements, build multiple paths by creating links between clips, and add navigation to clips in a full video. Clickable elements include hotspots, annotation, hyperlinks, etc. (Cattelan et al., 2008; Gaeta et al., 2014; Meixner et al., 2016). These elements will not affect the video structure unless "jump action" is added to these elements. A common practice to build multiple paths between clips is to link them in a visual storyboard (Dellagiacoma et al., 2020; Meixner et al., 2014; Shipman et al., 2005). Such an interface provides users with a canvas to distribute and create links between videos. In addition to the above two methods, adding navigation to a video (Aubert & Prié, 2005; Chu et al., 2017; Truong et al., 2021) is the simplest way to enable users to interact with videos. The navigation panel provides entrances for different parts of a video.

Our system is designed for authoring interactive videos for general use, so we abstracted several general structures of interactive videos. Also, although the visual storyboard interface allows more freedom for video authors, we cast more focus on enabling novice users to create interactive videos. The system we developed consisted of three modes with corresponding functions and features, and it mainly focused on adding click-to-jump elements to build multiple paths.

2.2. Narrative structuring methods and applications

Narrative is an effective tool to convey information (Cunningham et al., 2014) and has been widely studied in the fields of video games, advertising videos, texts, etc. (Crovato et al., 2016; Ryan, 2015; Saunders-Smith, 2009). An early study made by Barbara Stern proposed the two important elements of narrative structure: chronology and causality (Stern, 1998), and Jerome Bruner proposed its previous theory from the perspective of psychology (Bruner & Bruner, 1990). Later, Jennifer Escalas et al. consolidated the two theories and applied them to advertising (Escalas et al., 2004).

While the structuring methods above remain at an abstract and general level, other research (Crawford, 2013; Horton, 1990; Meadows, 2002; Ryan, 2006, 2015) studied a detailed part, namely, the structure of the plot, which is also called the narrative structure. Illustrations for different narrative structures are found in these researches, and they can be categorized into Node Based Structure and Linear Story Structure (Jackson & Latham, 2022). Node Based Structures are considered to be an apt approach to add interactivity, while according to Ryan (Ryan, 2015), Linear Story Structures are also able to support interactions in a structure that "*[…] giving no choice but to move forward but becomes interactive through optional side branches that lead to "roadside attractions"*".

Given so many different narrative structures are proposed, Gu et al. grouped them into five categories with distinct representative graphs (Gu et al., 2022). Among the five categories, the third category representing branching structures is commonly seen in many domains, like games (Moser & Fang, 2015) and AR stories (C. Li et al., 2022; W. Li et al., 2023). There are also design examples that either included existing narrative structures in their design or analytical processes (Baumer et al., 2020; Jackson & Latham, 2022), or proposed new narrative structures based on previous research (Partarakis et al., 2022; Reyes, 2017). Another classification method also proposed five categories and drew graphic organizers for each (Saunders-Smith, 2009). Despite this method originally being a structuring method for informational text (or, non-fictional text) based on the relationship or function of the context, some overlaps can be found between the above two classifications.

The core structure of our system is inspired by previous theoretical research. However, instead of directly applying current structures, we extract a set of common structures for a natural implementation of interactions and better user experience in our system.

2.3. Multimedia authoring tools attributes and modeling

An interactive video authoring system is a multimedia authoring system. As there are many existing works on multimedia authoring attributes and modeling, Wijaya et al. made a brief review of the current attributes and models (Wijaya et al., 2021, 2022) that described the details of the structures and construction of multimedia authoring systems. In an interactive video authoring system, descriptive files are often applied to describe the structure of interactive video in detail. For example, Hjelsvold et al. and Meixner et al. used SMIL format to construct an interactive video authoring system (Hjelsvold et al., 2001; Meixner et al., 2014), and Monserrat et al. utilized JSON as the descriptive language of system attributes (Monserrat et al., 2014).

As attributes describe how the interactive video structure is represented as a file, where the structural descriptions focus on the spatial and temporal logic of interactive videos (dos Santos & Muchaluat-Saade, 2012), the modeling methods delve into the organization of temporal logic, together realizing the logical conversion of interactive videos. The modeling methods used to describe the parallel structure in multimedia authoring tools include three main categories: Petri Nets, Hoare Logic, and LOTOS (Wijaya et al., 2022). We find examples like, the interactive video-based learning platform in the work of Magdin et al. adopted Petri Nets model (Magdin et al., 2011), and some adopted LOTOS model to realize the description of narrative structure, which generated RT-LOTOS modeling for SMIL 2.0 structural document (Sampaio & Courtiat, 2004).

Our system applied JSON-based structure descriptive file and LOTOS modeling method, and provided some preset template structure at the same time. Doing so can maintain the expressiveness of the system while avoiding being demanding to users.

3. Design study

To understand the design logic of common IVAP, and how novice users use these platforms, we conducted a design study to obtain design requirements that can inform our system design.

3.1. Methodology

We invited four creativity support tool designers (one male D1 and three females D2–D4) to try some of the existing IVAPs and get feedback on their advantages and drawbacks. Although the designers had no experience in interactive video making, they excel in using and designing other types of creativity support tools, which enables them to quickly catch the important points of an unfamiliar creativity support tool. Their work experience ranges from 3 to 4 years, covering design areas like visualization tools (D1 and D4), VR-based interior design tools (D2), and video generation systems (D3). Before the study, we first categorized existing IVAPs' editor patterns by analyzing 17 authoring platforms both in the market and in academic papers (Table A1). These platforms matched the following criteria: (1) The platform is designed for general using scenarios instead of specific or professional using scenarios like education, (2) the platform provides free access, or introduction to its mechanism is available, (3) the platform is operated on PC, in the form of software or web-app, and (4) branching or jumps between different time points or videos is possible. We studied each tool's authoring workflow and the design of the editor interface, and concluded five individual and one hybrid paradigm of editor patterns.

We finally selected six platforms with different combinations of editor patterns for the novice users to test, including: bilibili[3], dot.vu, eko[4], hihaho, Mindstamp,[5] and spott.[6] We asked them to try to use each platform for at most 30 min using several videos we provided. In the process, they were required to focus on the design logic and function of each platform, and the logic of interactive video they authored during the process could be ignored. This process lasted about 3 h in total, and then, we interviewed them about their insights on each platform's pros and cons, what optimizations can be made to each platform, and how should we design a system that is suitable for novice users.

3.2. Insights and discussion

To reach the goal that novice users can easily get started with the system, all four designers suggested that our system should provide templates and novice guidance on either the workflow or the functions for novice users. For template, D1 reckoned it necessary because "interactive video authoring is of certain difficulty in getting started," and he further added that, without templates, novice users cannot exploit the values brought by interactivity in interactive videos. In terms of novice guidance, all were satisfied with the platforms that provided certain user guidance (bilibili, dot.vu and eko), and met difficulties to get started when using the platforms that lacked user guidance (hihaho, Mindstamp and spott). However, not all novice guidance was properly designed. D1 and D4 both mentioned that the novice guidance of bilibili platform is clear, but it is introductory guidance rather than operational guidance, among which the latter type would do greater help to a novice user. D2 and D3 said that, although dot.vu platform did provide a seemingly clear novice guidance, it was presented by text and cost abundant time to read." They should consider simplifying the instructions and add some actual guidance that can lead the users to go through the process" (D2).

Simplicity is yet another key concern of D2 and D4. The bilibili and Mindstamp platforms presented examples of what a clear and simple system looks like. Compared to eko and hihaho platforms, which have a special operating logic, Mindstamp allows users to operate within the video directly in a straightforward way (D2). The dot.vu and spott platforms, on the contrary, looked too complicated, with too much information presented at once (dot.vu) or too many function panels integrated and presented within a single interface (spott), said D4. D2 added that dot. vu gave her a bad impression because it looked very difficult to use at first sight, and the information organization was not explicit. The simplicity of the interface, workflow and functions can lower the learning cost, making it easier for novice users to get started. Also, simplicity has a direct impact on authoring efficiency. Either the lack of simplicity or over-simplicity can lead to inefficiency (D4). For example, the preview function is a frequently used function during interactive video-making processes, but the bilibili platform barred users from previewing before completing a set of text information, which is an anti-intuitive design, and choked the workflow (D1). D2 thought that the tree structure enabled users to quickly build a simple interactive video, while D3 pointed out that, making modifications took a lot of time, because the text information and relationships between videos need to be modified one by one. And for the Mindstamp platform, although praised for its simple interface and straightforward operation, its simple and limited editing function heavily increased the time consumption for even completing a very simple interactive video (D2).

Different platforms also hold different considerations on whether to present all functions at once or separate some of them into different modes. The bilibili, eko, and hihaho platforms by default will hide some advanced functions or settings, which can decrease novice users' learning pressure (D4). D2 also recommended that platforms such as spott and dot.vu should consider segmenting authoring modes so that novice users will not get lost in an information sea. Also, in the experience process, D3 mentioned that, once you have grasped the basic operation of platforms like bilibili, and eko, the transition from using basic functions to using advanced functions is quite fluent and easy. She explained that, in bilibili platform, this was brought by a reasonable segmentation. The basic functions were introduced in the guidance, the intermediate functions remained unintroduced but can be explored by users easily, and the advanced functions were hidden. In eko platform, the fluent transition is a result of the elaborately designed template. They can serve as examples of the functions' usage, and also a way for users to explore different functions and settings.

3.3. Design requirements

Based on the insights from the four designers, we concluded three design requirements, denoted as R1–R3, for designing our interactive video authoring system, HierVid.

R1: Get users started using easily. Our primary goal is to lower the barriers to entry for novice users. Compared to traditional linear videos, interactive video-making is more demanding because of the required logic-building process. Thus, to address this problem, we should design a simple way of authoring, and provide adequate and clear guidance for novice users (D2). First, providing templates or patterns to users is strongly suggested by all designers, because they can be the starting points or references for novice users, and we can see such practices in some animation authoring tools (W. Li, 2022; Ma et al., 2022). Then, D2 suggested that the workflow should be designed in a way that the users are probably familiar with, thus lowering the learning cost. Compared to the guidance that is isolated from the system, such as a video tutorial, we should integrate the guidance into the system (D1, D4). Aside from common tutorials that showed the basic functions of the system, we should also inform novice users of the workflow (D2, D4). Furthermore, designing functions in a" What You See Is What You Get" way can also serve as implicit guidance throughout the whole authoring process.

R2: Offer users with enough flexibility. In general, our system should provide flexibility to some extent, which can be achieved by segmenting functions into different modes properly, supported by D2, D3, and D4. By doing so, users can decide to what extent they would like to use the system, adapt to the interactive video authoring process progressively, and being able to explore more complex functions spontaneously (D3). The theory proposed by Shneiderman also supports this. They stated that we should "Design with low thresholds, high ceilings, and wide walls" (Shneiderman, 2007) when designing creativity support tools. We also adopted the *Curation* → *Customization* → *Creation* spectrum proposed by Ma et al. (Ma et al., 2022) for the balance between ease of use and expressiveness, which also supports the rationality of designing a flexible system. To achieve this, we can set template as part of the system, which is supported by all four participants. Templates mentioned in R1 can not only lower the barriers to entry, but can also help the system with flexibility building in that they can be divided into smaller parts. By designing functions based on them, we can build a progressive workflow.

R3: Enhance the efficiency and avoid interruption. Our system should make the authoring and modifying process efficient. This can be achieved by avoiding the lack of simplicity and being over-simple (D4). We should not include unnecessary functions that choke the workflow (D1), such as a required information completion, and also avoid operations that are so limited that even a simple project takes up abundant time. The workflow and functions should be clear and simple (D2, D3), so that time used on getting familiar with the system decreases.

4. Interactive video structure

To meet the first and the second design requirements, and inspired by the Template-based pattern in the research, we decided to build a system that starts from the Template level, and then deconstructs the template into smaller parts, to form a progressive using experience. By achieving this, we should first understand the common structures of interactive video. As no previous research on decomposing the structure of interactive video was found, we first researched the structure of other media. Efforts have been made to decompose the structure of videos according to their narrative structure (Escalas et al., 2004), and various text structures were proposed to clarify interactions (Saunders-Smith, 2009), which were later used to illustrate video structure (Gu et al., 2022).

Inspired by the definition of narrative structure, and the classification of text structures, we extended and modified the previous work by (1) defining the basic components that build up an interactive video, (2) proposing a hierarchical structure for logically decomposing and constructing interactive videos. In our model, the structure of an interactive video is a three-level hierarchical structure, namely, Interactive Video Unit (IVU), Interactive Video Module (IVM), and Interactive Video Template (IVT). Take the Cooking Tutorial from hihaho.com[7] as an example, we truncate this video and apply the hierarchical model to the footage, as depicted in Figure 2.

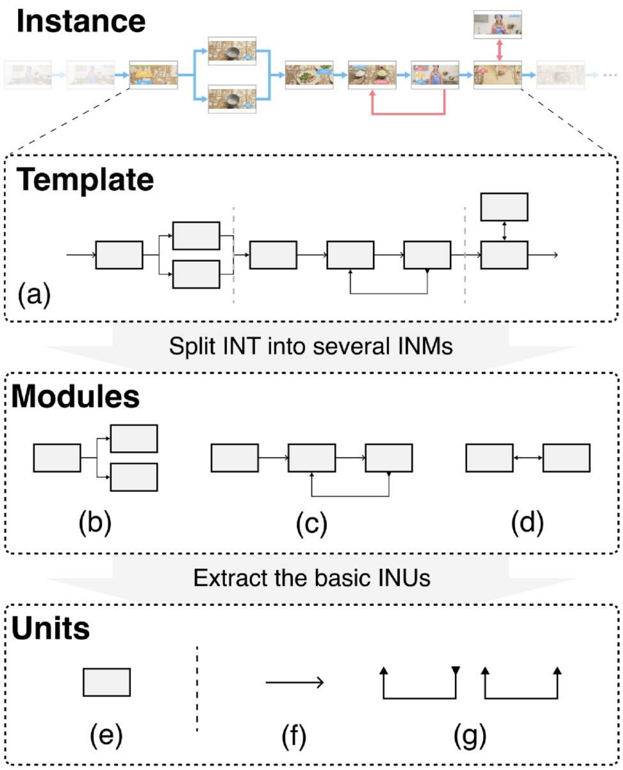


Figure 2. We illustrated part of the structure of an interactive video from hihaho. The general structure of the interactive video was considered an IVT (a). It consisted of three types of IVMs (b-d), from which we could extract smaller parts, the three different IVUs: e) event; f) linear action, illustrated by line arrow; g) non-linear action, illustrated by triangle arrow and reversed triangle (representing the end and the start of the arrow respectively).

4.1. Interactive Video Unit

IVU is the collective name of Events and Actions. Events or Actions alone are meaningless for crafting an interactive video, and they should be combined to build up an interactive video. The two types of IVU respectively symbolize the visual and narrative compositions of interactive videos, and the direction of the plot.

4.1.1. Event

Events are scenes with inner attributes, such as the number of split screens, the layout of split screens, and the presence of interactive elements. These inner attributes do not affect the structure of the whole interactive video but are of great importance to providing diversity in the practical authoring environment. Abstract visualization of Events is a rectangle with a black border and grey background (Figure 2(d)).

4.1.2. Action

Actions describe the direction of the plot, in other words, how events are ordered from the perspective of narrative. They are visualized by directed arrows, indicating how one event moves to another, which is either triggered by clicking interactive elements (e.g., button) or denotes a natural transition to the subsequent event when finishing playing.

Events in a traditional video equal to clips that are ordered chronologically without backtracking or skipping.

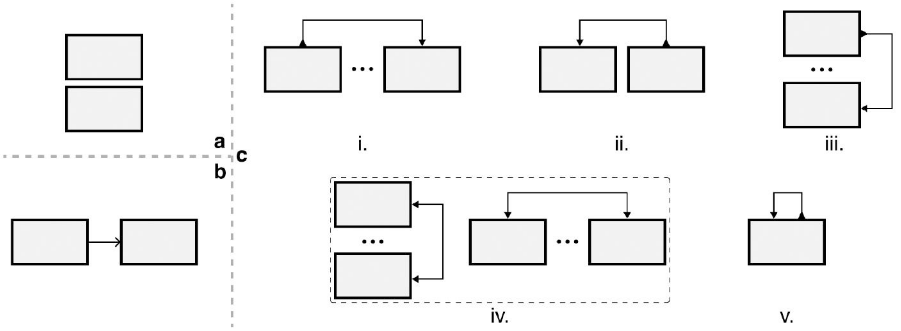


Figure 3. We identify three basic combination modes of IVUs. Mode (a) illustrates the combination mode between events without actions; mode (b) illustrates the combination mode of events linked only by linear actions; mode (c) includes 5 sub-modes, focusing on events linked by non-linear action.

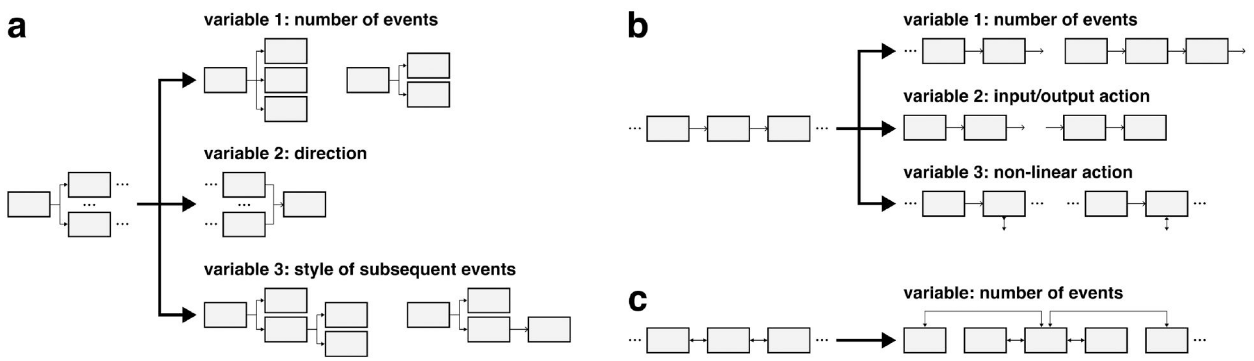

a
variable 1: number of events
variable 2: direction
variable 3: style of subsequent events
b
variable 1: number of events
variable 2: input/output action
variable 3: non-linear action
c
variable: number of events

Figure 4. We defined three types of IVMs with presented variants: a) the Tree Structure, where the number of branches, the direction of the structure, and the subsequent style of each branch are variable. b) The Sequential Structure also possessed three variables, including the number of events in the sequence, the presence of input and output action, and the presence and position of non-linear actions. c) For the Shuttle structure, the only variable is the number of branches. Notice that the difference between branch structure and shuttle structure is whether the action is one-way or two-way.

Such videos have a linear timeline, whereas, in an interactive video, the timeline is non-linear. We identified two types of Actions: Linear Action and Non-linear Action, and visualized them using arrows with different styles (Figure 2(e,f)). The basic attribute that separates the two types of action is direction. Linear Action represents the traditional chronological logic that one Event points to its subsequent one, while non-linear action embodies all the other unorthodox directions, concerning five sub-types of combination modes of events and actions (Figure 3(c)).

4.1.3. Basic combination modes of IVUs

As events or actions cannot function alone to form an interactive video, we define modes that combine at least two IVUs to get them collaborated.

Event and event without linking Action are either mutually irrelevant or parallel, among which the former relationship is meaningless, so we will discuss the latter one. A parallel relationship is that one Event is unable to directly move to another. The two events are visually adjacent but not temporally sequential, instead, they are counterparts on the timeline, as depicted in Figure 3(a). Stated simply, they are two branches of the same branching root.

Event and linear action together can form only one combination (Figure 3(b)), because of the restraints brought by the definition of Linear Action. We call such a relationship a Serial Relationship, indicating that the visually preceding event moves to the subsequent and adjacent one.

Event and non-linear action together form five possible combinations: (1) along-the-time direction, pointing to visually subsequent but non-adjacent Event (Figure 3(c.i)), (2) oppose-the-time direction, pointing to visually preceding event (Figure 3(c.ii)); (3) pointing to temporally parallel event (Figure 3(c.iii)); (4) jump back and forth between two events, we can regard this as the combining two non-linear actions (Figure 3(c.iv)); (5) pointing to self, which can form a circulation (Figure 3(c.v)).

4.2. Interactive Video Module

An IVM is the assembly of IVUs abides by combination modes' rules. We identify three types of IVMs based on the classification of non-fiction text structure (Saunders-Smith, 2009). There are originally five types of text structures: (1) Compare and Contrast, (2) Cause and Effect, (3) Sequence or Procedural, (4) Question and Answer, and (5) Exemplification. We combined text structure types with similar graphic organizers (1 and 5, 2 and 4), and the newly defined three types are Tree Structure, Sequential Structure, and Shuttle Structure. All IVMs are imparted with modifiable attributes to meet various needs while retaining their key features, as illustrated in Figure 4.

4.2.1. Tree structure

The graphical organizers of type 2 and 4 both illustrate a hierarchical branching structure (a branching root with at

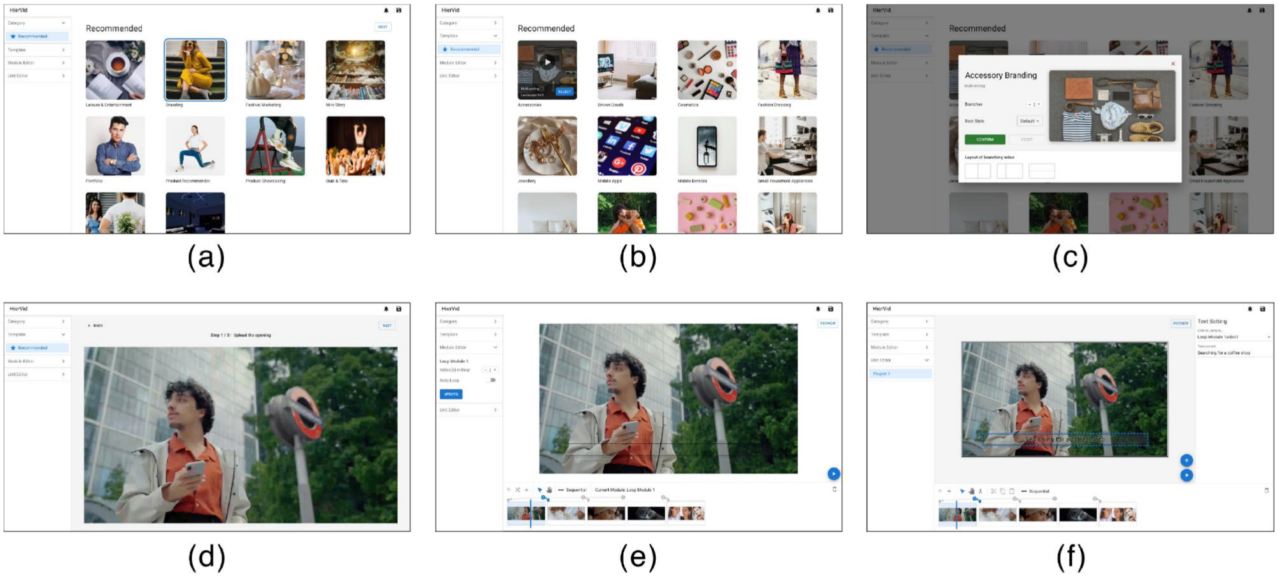


Figure 5. An Overview of the HierVid platform interface. a) Category Selecting Page, b) Template Selecting Page, c) Parameter Modification Pop-up Window, d) an example of Template Filling Page, e) Module Mode Page, and f) Unit Mode Page. Template Mode consists of interface (a)–(c).

least 2 branches). The original text structures convey the information that: (1) questions/causes must precede answers/effects and (2) placing branches before or after the branching root both makes sense. The two features are considered to be the core features of Tree Structure (Figure 4(a)), thus its variables focus on the position of the branching root and styles of branches. We can modify the structure by defining the number of branches, the sequence of branching root and branches, and the subsequent style of each branch.

4.2.2. Sequential structure

The core feature of sequential structure (Figure 4(b)) is the main storyline consisting of a series of events linked by linear actions. The number of events in the main storyline, the presence of input/output linear action, and the presence and position of non-linear action are three variables for sequential structure, making it possible to change the length of the structure, and form loops and shortcuts.

4.2.3. Shuttle structure

The graphical organizers of type 1 and 5 are represented by shuttle structure (Figure 4(c)). This structure is similar to the tree structure because of the branching structures. The distinction between the two structures is that Tree Structure generally proceeds along the time, while Shuttle Structure allows the branching root and branches to jump to each other. We use "shuttle" to describe such a relationship, and the number of branches is the only variable of this structure.

4.3. Interactive Video Template

An IVT can be considered a combination of IVMs and Events. A simple IVT may consist of only one IVM whereas an advanced IVT typically contains multiple IVMs. For example, the footage of Cooking Tutorial from hihaho.com is composed of three IVMs, as depicted in Figure 2.

5. System overview

Guided by the three design requirements in Section 3, and the interactive video structures proposed in Section 4, we designed and implemented HierVid, an IVAP that is a three-level hierarchical system. The workflow of HierVid is a step-by-step process (Figure 1), involving Template Mode, Module Mode, and Unit Mode, and detailed operations represented by the changes in video structure (i.e., Timeline in HierVid) are depicted in Figure 7.

5.1. Template mode

Template mode consisted of Category Selection Page, Template Selection Page, Parameter Modification Pop-up Window, and Template Filling Page (Figure 5(a–d)). We preset some categories as well as templates based on commonly-seen video content and storyline. Users can preview the final effects of each template on the Template Selection page, and the preview video content will change as the parameters change.

The Parameter Modification Pop-up Window will appear when users select a template. Take the "Accessory Branding" template as an example (Figure 5(c)), parameters users can modify include: the number of branches, the style of the branching root, and the layout of the branching root. The Template Filling Page would show up once the users confirmed the template and parameters, which was a series of guiding pages that varied between templates (Figure 7(c)). On these pages, users could upload clips according to the caption and adjust text buttons. The preview page is the last page of the guiding pages, where users can operate the interactive video made with templates. Users without the need for detailed editions can stop here and save the video.

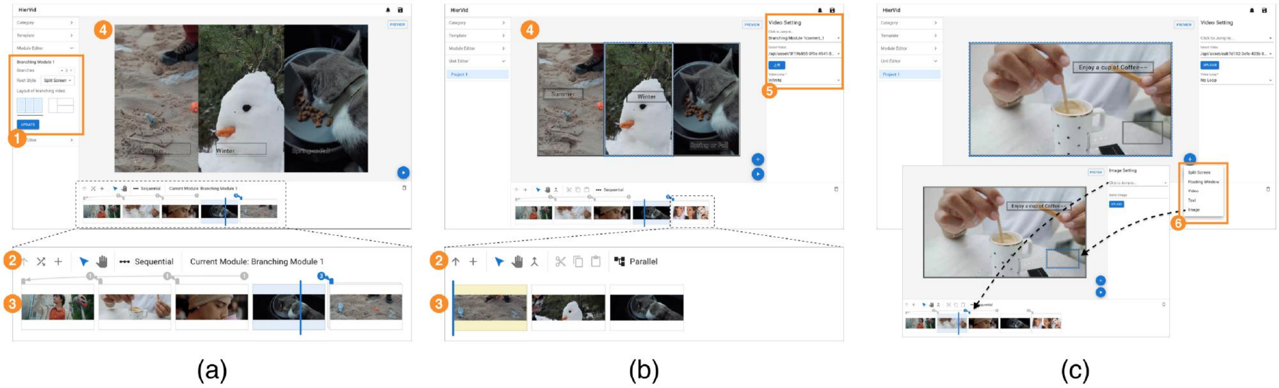
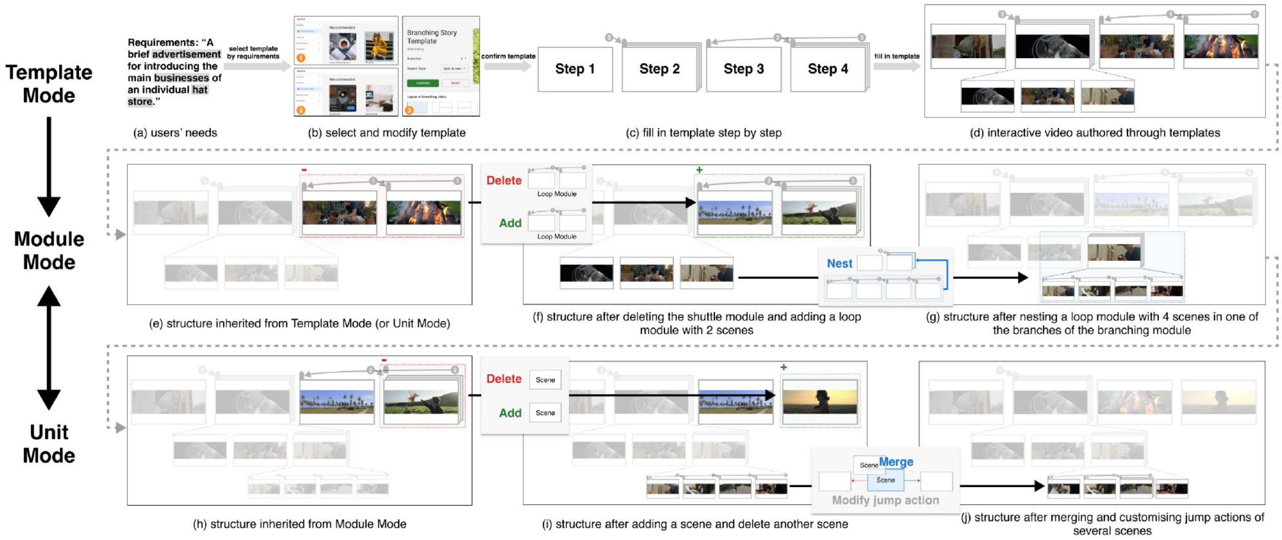


Figure 6. Overview of the HierVid's Module Mode (a) and Unit Mode Interfaces (b), (c). both present a Main View (a-4, b-4), a Parameter Setting Panel (a-1, b-5), and a Timeline Editor (a-2, 3, and b-2, 3), and other functional buttons. Users can upload and examine the current effects in the Main View, while editing the video structure mainly in the Timeline Editor, which consists of Toolbar (a-2, b-2) and Timeline (a-3, b-3). the blue and yellow background of Timeline indicates the relationship between current Events or piles, sequential and parallel respectively. Once the indicator is placed on a certain Event or pile, related arrows and symbols will be highlighted with blue. (c) is an example of operating elements in Unit Mode.



(a) users' needs
(b) select and modify template
(c) fill in template step by step
(d) interactive video authored through templates
(e) structure inherited from Template Mode (or Unit Mode)
(f) structure after deleting the shuttle module and adding a loop module with 2 scenes
(g) structure after nesting a loop module with 4 scenes in one of the branches of the branching module
(h) structure inherited from Module Mode
(i) structure after adding a scene and delete another scene
(j) structure after merging and customising jump actions of several scenes

Figure 7. The workflow of creating an interactive video using all three levels in HierVid platform. We presented the procedure by timeline to amplify the detailed changes. Users can select an appropriate template (b) according to their needs (a), and fill in the templates (c) to obtain a simple but complete interactive video (d). They can also choose to advance to Module Mode (e) to make further editions using Modules as presented in (e)-(g). for professional users, we also provide Unit Mode enabling full control of the structure as presented in (h)-(j). They can switch between Module Mode and Unit Mode whenever they need.

Figure 7(a–d) presents the main steps of authoring an interactive video using template mode.

5.2. Module mode

Users can enter module mode page through the "Edit" button on the preview page to customize the structure of the interactive video through modules. We currently provide three types of Modules that adopted the three structures defined in Section 4 (Figure 4): branching module (tree structure), loop module (sequential structure), and comparing module (shuttle structure). Module mode page is composed of three sections: the main view (Figure 6(a-4)), Parameter Setting Panel (Figure 6(a-1)), and Timeline Editor (Figure 6(a-2,a-3)). Users can preview the current clip and check the settings of interactive elements in the main view. The Parameter Setting Panel is for users to modify existing Modules in the timeline. Timeline Editor is the core part of the module mode page, consisting of a Toolbar and Timeline, where the order of modules and relationships between events are displayed and edited. The detailed introduction to timeline editor is as follows:

Toolbar. From left to right (Figure 6(a-2)): (1) Previous Level. If users have expanded one or more piles to enter another Timeline level (Figures 7 and 6(b)), they can click here to return to the previous level. (2) Nest Module. This tool can replace an Event in a Module with a new Module, thus users can nest the two modules (Figure 7(f,g)). (3) Add Module. Add module tool enables users to add one of the three modules at the rear of the current timeline level (Figure 7(e,f)). (4) Select Mode. In this mode, we can drag or click to place the indicator, and double-click to expand the piles. (5) Move Mode. Users can drag Modules to reorder them, note that this mode is disabled at the parallel level (yellow background). 6) Delete Module. This tool is put on the right side of the Toolbar, users can delete the current Module with this tool (Figure 7(e,f)). In addition to these tools, Toolbar also displays information about the current timeline level and module.

Timeline. Timeline in HierVid is a video-based editing mode instead of the keyframe-based mode in traditional video editors (e.g., iMovie), which is a common practice in IVAPs that support multi-video editing. The thin blue bar is an indicator to inform which Event and Module they are currently at (also displayed in the Toolbar and Parameter Setting Panel).

Timeline holds 2 features that differ from existing IVAPs: (1) applying colors and arrows to signify relationships between events, and (2) being collapsible in a "deep" dimension. Timeline with a blue background indicates that current modules will be played sequentially (Figure 6(a-3)), and a yellow background indicates that current events or pile of events (hereinafter referred to as "Pile") are branches of the same branching root (Figure 6(b-3)). Sequential or parallel relationships are displayed in the toolbar. The arrows represent actions triggered by clickable elements. However, linear actions and non-linear actions do not differ in style here. The collapsible feature works as a directory tree does, we can double-click a Pile to enter another level of timeline. Branches collapse in the blue background and sequential events collapse into one branch in the yellow background.

5.3. Unit mode

If users become familiar with module mode, and require full control of the structure, they can work in unit mode, which extends beyond the structures we proposed in Section 4, to author interactive videos with creativity and freedom. In the main view, in addition to functions provided in module mode, users can select and modify elements in each event. Parameter Setting Panel is also a place to modify the settings of elements, such as Actions and loop setting (Figure 6.5). Timeline is the same as that in module mode, while the Toolbar differs, and we also provide elements in unit mode for detailed editions in each event.

Tools. We designed tools for users to freely edit the structure (Figure 6(b-3)). From left to right: (1) Previous Level. (2) Add Event. This tool adds an empty event at the rear of the present Timeline level, users can upload videos and add elements to it (Figure 6(h,i). (3) Select Mode. (4) Move Mode. Allows users to drag and re-order events or piles. (5) Merge Mod (e). Merge Mode is used to drag and drop the selected event or pile to a target event or pile, and this operation forms a new pile (Figure 6(i,j)). (6) Delete Event. Delete current event or pile (Figure 6(h,i)). Also, the information of the timeline level is displayed in Toolbar.

Elements. In Unit Editor, HierVid provides some elements for customizing and designing the style of each scene (Figure 6.6). (1) Split screen. This element allows users to partition a scene or an existing segment of a scene into two parts. The "branch root style" in the parameter setting panel of templates and modules shows the effect of adding split screens. (2) Floating window. This element floats on the top of the background video(s), we can upload videos and images to it. (3) Video. Video elements allow users to add a video floating on top of background video(s), it can also be added to an empty Split screen or Floating window. (4) Image. Similar to the video element (Figure 6(c)). (5) Text. Equals to what we call the text button before. The settings users can change include size, position, text information, and action after a click.

6. Evaluation

To examine if we had fulfilled the three design requirements, we conducted two user studies. Study I was a between-subject experiment across two platforms, the bilibili IVAP which is the most commonly-used IVAP in China, and HierVid developed by our team. Participants were required to complete a well-defined mission of replicating the structure of a given interactive video. Study II was an open-ended mission of conditional interactive video creation that only used the HierVid system. Specifically, we hoped to find answers to the following questions:

Q1. Will participants find the workflow and functions of the system easy to understand, and can they get started to operate our system easily?

Q2. What are the practical demonstrations of HierVid's flexibility during the using process?

Q3. Can HierVid help participants reach higher efficiency than those who use the bilibili platform?

6.1. Preparations

To answer the previous questions, we decided to choose a platform from the previously researched platforms (Table A1) as the control condition of our experiments. We filtered recent IVAPs according to the following criteria: (1) support editing multiple clips in one project, (2) relationships between video clips are displayed in a certain form, (3) not equipped with Template-based Mode, and (4) not designed as a hierarchical system. The first three factors heavily affect final results and user experience, so we should ensure a consistent experience on these points. On the other hand, the latter two factors are what we are studying, and are also aspects of our system that differ from most existing platforms. We finally choose bilibili for comparison in the following studies.

6.2. Study I: Well-defined video replication mission

6.2.1. Participants

We sent a study invitation to two social platforms and received 30 sign-ups through the screening survey. We finally recruited 22 participants and categorized them into two groups: (1) true novices consisted of 16 participants (seven female and nine male), who had neither video-making experience nor knowledge of interactive video, denoted as RT1 to RT16, and (2) six advanced novices (four female and two male) who reported themselves as experienced traditional video makers and had some knowledge on interactive media, denoted as RA1–RA6. None of them had experience in interactive video making. The 22 participants were undergraduate or graduate students between 18 and 30 years old, from different majors. Participants were evenly divided into two conditions considering gender and experience, to use different platforms in the replication mission.

6.2.2. Conditions

There were two conditions in Study I: the bilibili IVAP and HierVid system designed by our team. The advanced function in the bilibili platform, which enables users to impart weights to options, was banned to keep the quality of produced videos across the two platforms consistent. The HierVid system provided the participants with the functionality described in Section 5, except for the Unit Mode that we chose not to evaluate to control the experiment's duration.

6.2.3. Materials

We exploited interactive videos from wirewax[8] (currently vimeo[9]) and hihaho[10] to explain the features and structures of the interactive video we were studying to the participants. We also recorded a video of operating an interactive video[11] for participants to replicate in the formal experiment. The latter video was selected because its structure consisted of several basic structures we identified in Section 4, and the structure could be reproduced in more than one way on both platforms. Material videos used in the mission were free clips from coverr.co.[12] The contents of the material videos were mainly daily activities and scenes, to avoid the divisions caused by experience differences.

6.2.4. Procedure

Two of the authors hosted the two conditions separately with the same process. We grouped the participants according to their experiment time period, and each group consisted of 2-3 participants. Before the replication mission started, the experimenter held a 20-min online group training on the platform they would use in the mission. Then, each participant entered a separate online meeting room to do the replication mission. We first sent the material videos to the participants and asked them to browse and rename these videos within 5 minutes. After this, the participants started the 20-min replication mission to replicate the structure of the interactive video we provided, using the material videos. They were also required to rationalize the storyline, preventing them from filling the videos in the structure randomly. Finally, the participants were required to receive an interview to answer some questions related to their performances and collect feedback on user experience. We recorded the complete experiment processes for further analysis with the informed consent of the participants.

6.3. Study II: Open-ended video creation mission

6.3.1. Participants

We sent another study invitation to the same platforms mentioned in Study I, and received 8 sign-ups. We recruited seven participants and divided them into the same two groups in Study I: (1) three true novices (two female and one male) with neither traditional video-making experience, nor knowledge of interactive video, denoted as CT1–CT3, and (2) four advanced novices (two female and two male) with traditional video making experience and some knowledge on interactive video, denoted as CA1–CA4. None of them had experience in interactive video making. The 7 participants were undergraduate or graduate students between 18 and 30 years old, from different majors.

6.3.2. Materials

Some participants from Study I said in the interview that the provided materials lacked variety, thus building the storyline would take up lots of time. Therefore, in Study II, we replaced the material videos presenting similar content with new ones from coverr.co, thus providing more diverse scenes to allow more possibilities for combination.

6.3.3. Procedure

The experimenter hosting Study II was the same one in Study I that hosted the experiment using the HierVid system. We split the participants into three groups, each consisting of 2–3 participants, to give a 20-min group training on the HierVid system. Participants were required to follow the steps during the training. Then, each participant entered a separate online meeting room to do the creation mission. First, we sent the material videos to the participants and asked them to browse and rename these videos within 5 min. After this, the participants started the 20-min conditional creation mission that required them to create at least one, at most five interactive videos using no less than five material videos in each interactive video. Furthermore, if the participants intended to author more than one interactive video, they were required to use the functions in Module Mode in at least one interactive video. Finally, the participants were interviewed to answer some questions related to their performances and collect feedback on user experience. We recorded the complete experiment processes for further analysis with the informed consent of the participants.

7. Results

The results included subjective opinions extracted from interviews, experimenters' observations of participants' behaviors, and quantitative results. In the interview sessions, we solicited subjective perspectives of their user experience, including questions like, whether it was easy to get started and if the system provided them with enough flexibility. Generally, the interview results, together with the observations demonstrated that HierVid was easy to understand and get started for novice users (R1), and they considered

Table 1. The gist of interview and observation results concerning R1.

Codes	Part of the quotes/observations
Get users started using easily	″Any functions can be seen at a glance″ (RT16); ″It's very simple to operate the system, even novices can quickly get started″ (RT7); ″The icon of branching layout is a What You See Is What You Get visualization (RT15, RT16)″.
Helpful novice guidance	″The instructions of each step (in the Template Filling Page) help me know what to do″ (RT1, RT2, and RT14); Change the parameters and check the preview of each template (RA4).

Table 2. The gist of interview and observation results concerning R2.

Codes	Part of the quotes/observations
Offer users with various choices	″I like the templates that I can select and modify freely″ (RT1); ″Without Templates, I may not be able to author a (complete) interactive video″ (RT2); ″I can create complicated structures using Modules and Module-based functions″ (RT8).
Facilitate users exploring functions	Tried to make detailed modifications to the structure (RT15, RT16, and CT1); ″The Template-Module-Unit structure is reasonable″ (CA2).

that the hierarchical designed system allowed them enough freedom to use and explore the system (R2). For the statistical results, we found that the time consumption of completing the same mission using HierVid system is less than using the bilibili platform (R3) using a Mann–Whitney U test.

7.1. Novice-friendly system

In general, we grouped the interview and observation results that support R1 into two codes (Table 1): Get users started using easily and Helpful novice guidance.

Five participants (RA1, RA4, RT7, RT15, and RT16) found our platform user-friendly. They commented that “The interface and structure are simple and intuitive” (RA1 and RT16) and “any functions can be seen at a glance” (RT16), “It's very simple to operate the system, even novices can quickly get started” (RT7). RT15 and RT16 sketched in the statement by adding that the “branching layout” is a solid function because it is designed to be a “What You See Is What You Get visualization” and “can support authoring various effects.” RT16 also commented that “the simplicity of system and function design is to lower its barriers to entry, to enable everyone a friendly and intuitive platform for authoring their own interactive videos.”

Besides, the participants generally felt guided and instructed by the information provided in the Template Filling Page (RT1, RT2, RT4, RT7, RT14, RT15, and RA1) and the Template-Module-Unit process presented in the sidebar (RT8 and RA4), which helped them to have a better understanding of the feature of interactive videos and the workflow of authoring an interactive video (RT2). In conclusion, the system fulfilled the first design requirement (R1).

7.2. Flexible hierarchical structure

The results supporting R2 were also grouped into two codes, say, Offer users with various choices and Facilitate users exploring functions. The gist of the results was presented by Table 2. Eight out of eleven participants (RT1, RT2, RT4, RT7, RT8, RT14, RT15, RA1, and RA4) using the HierVid system in Study I found the Templates or Modules to be a preferable or important function: “I like the templates that I can select and modify freely” (RT1). RT2, RT7, and RA1 believed that template was the core part of the system, and RT2 added that “…I think functions subsequent to Template Mode are not as that important, because I consider obtaining a complete interactive video to be the most important thing… without Templates, I may not be able to author a (complete) interactive video.” Three participants (RT2, RT14, and RT15) mentioned that they found the Template Mode convenient because “I can directly apply these templates” and “I can use the Templates by simply uploading materials and modifying the buttons… I don't feel restrained by using templates”.

For modules, RA4 liked the preset actions in module, compared to other platforms that “I may need to set the jumping relations myself later (after setting videos)”. RT8 thought it was the core feature because function like Nest Module was module-based, he could create complicated structures using modules and module-based functions.

Moreover, although we did not open Unit Mode for participants, 3 novice participants (RT15, RT16, and CT1) tried to make detailed modifications to the structure, inferring that our platform can help users without video-making experience and little knowledge about the interactive video quickly familiarize the authoring process. The feedback indicated that HierVid could facilitate users exploring advanced functions, and the design of the system had an osmosis effect on enhancing their skills. Therefore, we considered the second design requirement (R2) to be achieved.

7.3. Enhanced efficiency

We calculated the time the 22 participants expended to complete at least two branching structures, and data from 18 participants (nine for each platform) were accepted (Figure 8). Data from the other 4 participants were excluded because they either did not consider the video's storyline (RT9) or their structures diverged greatly from the target video (RT12, RT13, and T14). On average, the participants took 10.98 min (sd = 0.70) using the HierVid system and 13.64 min (sd = 2.98) using the bilibili platform. Median time assumption in HierVid group and bilibili group was 11.12 and 13.58 min, the distributions in the two groups differed significantly (Mann–Whitney $U = 17$, $n_1 = n_2 = 9$, $P < 0.05$, Cohen $d = 1.23$). This indicated that our system was more efficient and faster to use, thus R3 had been achieved.

In the bilibili platform condition, 7 out of 11 participants were interrupted by the incompletion of material filling, titles, or option descriptions when intended to preview the video or close the parameter setting panel. RT3 commented that "It should allow me to preview first rather than forcing me to complete the information." RT6 kept browsing the material video because the uploaded videos lack thumbnails, and they cannot play material videos in the editor. RA3 and RT3 mixed up the story module and the jump module. These struggles accounted for a longer completion time. With our platform, RA5 mixed up the move mode function and nest module function, and RT8 forgot how to enter and check the folded videos, which caused some confusion, but they were able to correctly use these functions after reminding. CT3 and CA1 did not switch back to select mode after using move mode but were aware of this independently after several attempts.

8. Discussion

As demonstrated above, the three design requirements were supported by the results, but the participants also proposed valuable suggestions either interested us, or helped us think about the future direction (see Table 3). We concluded from the results that an IVAP needs heuristic functions and HierVid do provide such functions, and that there are some gaps between the traditional video authoring mode and the interactive video authoring mode, which may disappoint traditional video makers and hinder the transition from the former mode to the latter mode.

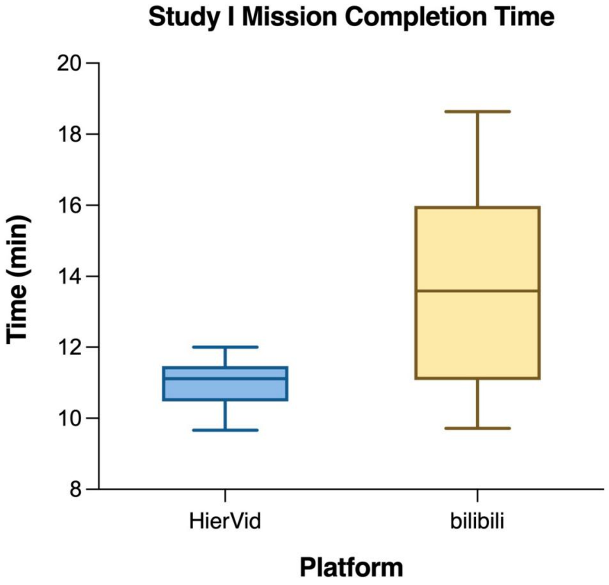


Figure 8. The time assumption of 18 participants (9 valid data per platform) replicating the interactive video structure in Study I.

8.1. Heuristic functions

As the video materials we provided did not have apparent relationships with each other, building a reasonable and logical storyline depends on their associative ability. All users from both study I and II using the HierVid system stated that it is harder to work out a storyline than build the structure, as RT8 said "It is simple to replicate the structure, while requiring a reasonable storyline makes the mission become difficult." In contrast, all participants with the bilibili platform thought building structure was more difficult than relating the materials. However, four participants (RT2, CT1, CA3, and CA4) using HierVid found some functions could spark inspiration when building the storyline. Three out of the four participants (RT2, CT1, and CA4) said "the Modules and Templates with preset branches may prompt me to category materials into different classes … -thus I can create different storylines" (RT2), compared to what another 3 participates (RT5, RA2, and RA6) using bilibili platform said: "I need to name the materials, which forced me to think relations between videos" (RT5 and RA2), inferring that the bilibili platform only forces the users to rethink the relations without providing any substantive help that can trigger new thoughts. These opinions noted that novice users, especially those who did not have any video-making experience, need the help of inspiration-triggering mechanisms during their first few attempts.

8.2. The gap between two authoring modes

Our system was designed with minimalized functions to lower the barriers to entry for novice users, and most of the current functions in the system cater to interactivity, thus may disappoint experienced traditional video makers. Three participants with traditional video-making experience (RA4, RA5, and CA2) felt uncomfortable with the lack of a cutting function. While the cutting function *is* a basic function in

Table 3. The gist of Discussion Section presenting findings from interview and observation results apart from results concerning the design requirements.

Findings	Codes	Part of the quotes/observations
Heuristic functions	Hard to work out a storyline	"It is simple to replicate the structure, while requiring a reasonable storyline makes the mission become difficult" (RT8-HierVid); "I think it's easy to create a story, while it's difficult to build the structure I need." (RT6, RA3-bilibili).
	Spark inspirations	"The Modules and Templates with preset branches may prompt me to category materials into different classes … thus I can create different storylines" (RT2-HierVid); "I need to name the materials, which forced me to think about relations between videos" (RT5, RA2-bilibili).
The gap between two authoring modes	Disappoint experienced traditional video makers	"I wanted to cut the material video but there's no such function" (RA4, RA5, CA2); "I think you the system should allow me to add audio or animations" (CA2).

traditional video authoring platforms, it is uncommon in an IVAP. And CA2 also hoped to add audio and transition animations to the interactive video in the authoring process. How to fill the gap between the traditional video-making mode and interactive video-making mode? How to make a fluent transition between the two modes? How to appeal to such groups to use our system without overshadowing the basic functions of authoring interactive videos? These are all questions that are worth further exploration.

9. Limitations

Although the statistics and participants' feedback presented overall positive results of the HierVid system, there are also limitations in our work and the participants provided constructive advice for future improvements.

First, the functions of our system were limited. We provided only three templates and three kinds of modules, and button is the only interactive element. HierVid could be extended to support more using scenarios and fulfill more needs by expanding the category of templates and modules, and implements interactive elements such as sliders, switches, etc.

Second, the experiment design can be optimized. The participant number is limited to a sample size of 7 in Study II, which can lead to biased results. Additionally, to avoid the homogenization of produced videos, the material videos used in the experiment were deliberately made weak in correlations, resulting in the problem that many participants took more time struggling to build a storyline than we expected. Moreover, a between-subject design cannot ensure separations between groups, especially when they have different backgrounds with diverse thinking modes. For example, we observed that RT8 and RT15 excelled in organizing logic, thus they spent relatively less time completing the mission although they were true novices.

Lastly, although we required the participants to rationalize the storyline, it is still unclear what difficulties would users face when using true and meaningful material videos in the real world. Future work could conduct studies to understand how would users use HierVid when they face real needs.

10. Conclusion

We have presented HierVid, a hierarchical authoring platform for novice users to author interactive videos. The design of HierVid was guided by our understanding of previous IVAPs, the insights from four designers with creativity support tools designing experience, and the study of interactive video structures. The system consisted of 3 editor patterns: Template-based Mode, optimized Directory Tree View, and Parameter Setting Panel, which was easy to get novices to start using and understanding the system, provided adequate flexibility to users, and enhanced efficiency. We conducted a user evaluation with two studies, and sample outcomes supported the system's prospects, with potential future work in further lowering the barrier to entry for novice users, and enhancing the authoring experience.

Notes

1. https://dot.vu
2. https://hihaho.com
3. https://member.bilibili.com/platform/upload/video/interactive
4. https://eko.com
5. https://mindstamp.com
6. https://spott.ai
7. https://hihaho.com/showcase/cooking-tutorial/
8. https://www.wirewax.com/showcase/gallery/#8199813/
9. https://vimeo.com
10. https://hihaho.com/showcase/cooking-tutorial/
11. https://www.wirewax.com/showcase/gallery/#8071116/
12. https://coverr.co

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work is supported by the National Key R&D Program of China (2022YFB3303301), and the National Natural Science Foundation of China (No. 62107035 and No. 62006208).

References

Aubert, O., & Prié, Y. (2005). *Advene: Active reading through hypervideo* [Paper presentation]. Proceedings of the Sixteenth ACM Conference on Hypertext and Hypermedia (p. 235–244), Salzburg, Austria. https://doi.org/10.1145/1083356.1083405

Aubert, O., Prié, Y., & Schmitt, D. (2012). *Advene as a tailorable hypervideo authoring tool: A case study* [Paper presentation]. Proceedings of the 2012 ACM Symposium on Document Engineering (p. 79–82), Paris, France. https://doi.org/10.1145/2361354.2361370

Bao, L., Xing, Z., Xia, X., & Lo, D. (2019). Vt-revolution: Interactive programming video tutorial authoring and watching system. *IEEE Transactions on Software Engineering*, *45*(8), 823–838. https://doi.org/10.1109/TSE.2018.2802916

Baumer, E. P. S., Blythe, M., & Tanenbaum, T. J. (2020). *Evaluating design fiction: The right tool for the job* [Paper presentation]. Proceedings of the 2020 ACM Designing Interactive Systems Conference (p. 1901–1913), Eindhoven, Netherlands. https://doi.org/10.1145/3357236.3395464

Belanche, D., Flavián, C., & Pérez-Rueda, A. (2020). Consumer empowerment in interactive advertising and EWOM consequences: The PITRE model. *Journal of Marketing Communications*, *26*(1), 1–20. https://doi.org/10.1080/13527266.2019.1610028

Bruner, J., & Bruner, J. S. (1990). *Acts of meaning: Four lectures on mind and culture* (Vol. 3). Harvard University Press.

Bulterman, D. C., Hardman, L., Jansen, J., Mullender, K., & Rutledge, L. (1998). Grins: A graphical interface for creating and playing SMIL documents. *Computer Networks and ISDN Systems*, *30*(1–7), 519–529. https://doi.org/10.1016/S0169-7552(98)00128-7

Cattaneo, A. A., van der Meij, H., Aprea, C., Sauli, F., & Zahn, C. (2019). A model for designing hypervideo-based instructional scenarios. *Interactive Learning Environments*, *27*(4), 508–529. https://doi.org/10.1080/10494820.2018.1486860

Cattelan, R. G., Teixeira, C., Goularte, R., & Pimentel, M. D. G. C. (2008). Watch-and-comment as a paradigm toward ubiquitous interactive video editing. *ACM Transactions on Multimedia Computing,* *Communications, and Applications*, *4*(4), 1–24. https://doi.org/10.1145/1412196.1412201

Chang, H. B., Huang Hsu, H., Liao, Y. C., Shih, T., & Tang, C. T. (2004). *An object-based hypervideo authoring system* [Paper presentation]. 2004 IEEE International Conference on Multimedia and Expo (ICME) (IEEE Cat. no.04th8763) (Vol. 3, p. 2219–2222), Taipei, Taiwan. https://doi.org/10.1109/ICME.2004.1394711

Chu, J., Bryan, C., Shih, M., Ferrer, L., & Ma, K. L. (2017). *Navigable videos for presenting scientific data on affordable head-mounted displays* [Paper presentation]. Proceedings of the 8th ACM on Multimedia Systems Conference (p. 250–260), Taipei, Taiwan. https://doi.org/10.1145/3083187.3084015

Crawford, C. (2013). Interactive storytelling. *The video game theory reader* (pp. 259–273). Routledge.

Crovato, S., Pinto, A., Giardullo, P., Mascarello, G., Neresini, F., & Ravarotto, L. (2016). Food safety and young consumers: Testing a serious game as a risk communication tool. *Food Control*. *62*, 134–141. https://doi.org/10.1016/j.foodcont.2015.10.009

Cunningham, L. S., Reich, J. J., & Fichner-Rathus, L. (2014). *Culture and values: A survey of the western humanities*. Cengage Learning.

Dellagiacoma, D., Busetta, P., Gabbasov, A., Perini, A., & Susi, A. (2020). Authoring interactive videos for e-learning: The elevate tool suite. *International conference in methodologies and intelligent systems for techhnology enhanced learning* (pp. 127–136). Springer. https://doi.org/10.1007/978-3-030-52538-5_14

dos Santos, J. A. F., & Muchaluat-Saade, D. C. (2012). Xtemplate 3.0: Spatio-temporal semantics and structure reuse for hypermedia compositions. *Multimedia Tools and Applications*, *61*(3), 645–673. https://doi.org/10.1007/s11042-011-0732-2

Escalas, J. E., Moore, M. C., & Britton, J. E. (2004). Fishing for feelings? hooking viewers helps. ! *Journal of Consumer Psychology*, *14*(1), 105–114. https://www.sciencedirect.com/science/article/pii/S1057740804701370 https://doi.org/10.1207/s15327663jcp1401&2_12

Farias, M., & Martinho, C. (2021). An approach to multiplayer interactive fiction. *International conference on interactive digital storytelling* (pp. 48–60). Springer. https://doi.org/10.1007/978-3-030-92300-6_5

Fidan, M., & Debbag, M. (2023). Comparing the effectiveness of instructional video types: An in-depth analysis on pre-service teachers for online learning. *International Journal of Human–Computer Interaction*, *39*(3), 575–586. https://doi.org/10.1080/10447318.2022

Gaeta, M., Loia, V., Mangione, G. R., Orciuoli, F., Ritrovato, P., & Salerno, S. (2014). A methodology and an authoring tool for creating complex learning objects to support interactive storytelling. *Computers in Human Behavior*, *31*, 620–637. https://doi.org/10.1016/j.chb.2013.07.011

Gaggi, O., & Celentano, A. (2002). *A visual authoring environment for prototyping multimedia presentations* [Paper presentation]. 4th International Symposium on Multimedia Software Engineering, 2002. proceedings (pp. 206–213), Newport Beach, CA. https://doi.org/10.1109/MMSE.2002.1181614

Gao, Q., Rau, P. L P., & Salvendy, G. (2009). Perception of interactivity: Affects of four key variables in mobile advertising. *International Journal of Human-Computer Interaction*, *25*(6), 479–505. https://doi.org/10.1080/10447310902963936

Goldberg, M. H., van der Linden, S., Ballew, M. T., Rosenthal, S. A., Gustafson, A., & Leiserowitz, A. (2019). The experience of consensus: Video as an effective medium to communicate scientific agreement on climate change. *Science Communication*, *41*(5), 659–673. https://doi.org/10.1177/1075547019874361

Green, D. P., Bowen, S., Hook, J., & Wright, P. (2017). *Enabling polyvocality in interactive documentaries through" structural participation* [Paper presentation]. Proceedings of the 2017 Chi Conference on Human Factors in Computing Systems, (p. 6317–6329), Denver, CO. https://doi.org/10.1145/3025453.3025606

Gu, C., Lin, S., Sun, J., Yang, C., Chen, J., Jiang, Q., Miao, W., & Wei, W. (2022). What do users care about? research on user behavior of mobile interactive video advertising. *Heliyon*, *8*(10), e10910. https://doi.org/10.1016/j.heliyon.2022.e10910

Hjelsvold, R., Vdaygiri, S., & Léauté, Y. (2001). *Web-based personalization and management of interactive video* [Paper presentation]. Proceedings of the 10th International Conference on World Wide Web (p. 129–139). https://doi.org/10.1145/371920.371969

Horton, W. (1990). *Designing and writing online documentation: Help files to hypertext*. John Wiley & Sons, Inc. https://dl.acm.org/doi/book/10.5555/77488

Hsu, H. H., Shih, T. K., Chang, H. B., Liao, Y. C., & Tang, C. T. (2005). Hyper-interactive video browsing by a remote controller and hand gestures. *Embedded and ubiquitous computing – EUC 2005 workshops* (Vol. 3823, pp. 547–555). Springer Berlin Heidelberg. https://doi.org/10.1007/11596042_57

Jackson, D., & Latham, A. (2022). Talk to the ghost: The storybox methodology for faster development of storytelling chatbots. *Expert Systems with Applications*, *190*, 116223. https://doi.org/10.1016/j.eswa.2021.116223

Layona, R., Yulianto, B., & Tunardi, Y. (2017). Authoring tool for interactive video content for learning programming. *Procedia Computer Science*, *116*, 37–44. https://doi.org/10.1016/j.procs.2017.10.006

Li, C., Li, W., Huang, H., & Yu, L. F. (2022). Interactive augmented reality storytelling guided by scene semantics. *ACM Transactions on Graphics*, *41*(4), 1–15. https://doi.org/10.1145/3528223.3530061

Li, W. (2022). *Animaton: Scriptable finite automaton for animation design in unity3d game engine* [Paper presentation]. Proceedings of the 2022 International Conference on Pattern Recognition and Intelligent Systems (p. 97–102). Wuhan, China https://doi.org/10.1145/3549179.3549196

Li, W. (2023). *Compare the size: Automatic synthesis of size comparison animation in virtual reality* [Paper presentation]. 2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (Hora) (p. 1–4). Istanbul, Turkiye. https://doi.org/10.1109/HORA58378.2023.10156731

Li, W., Li, C., Kim, M., Huang, H., & Yu, L. F. (2023). *Location-aware adaptation of augmented reality narratives* [Paper presentation]. Proceedings of the 2023 Chi Conference on Human Factors in Computing Systems, Hamburg, Germany. https://doi.org/10.1145/3544548.3580978

Ma, J., Wei, L.-Y., & Kazi, R. H. (2022). *A layered authoring tool for stylized 3d animations* [Paper presentation]. Proceedings of the 2022 Chi Conference on Human Factors in Computing Systems (p. 14), New Orleans, LA. https://doi.org/10.1145/3491102.3501894

Magdin, M., Cápay, M., & Mesárošová, M. (2011). *Usage of interactive video in educational process to determine mental level and literacy of a learner* [Paper presentation].14th International Conference on Interactive Collaborative Learning (pp. 510–513), Piestany, Slovakia. https://doi.org/10.1109/ICL.2011.6059637

Meadows, M. S. (2002). *Pause & effect: The art of interactive narrative*. Pearson Education.

Meixner, B. (2018). Hypervideos and interactive multimedia presentations. *ACM Computing Surveys*, *50*(1), 1–34. https://doi.org/10.1145/3038925

Meixner, B., John, S., & Handschigl, C. (2016). Siva suite: An open-source framework for hypervideos. *ACM SIGMultimedia Records*, *8*(1), 10–14. https://doi.org/10.1145/2898367.2898371

Meixner, B., Matusik, K., Grill, C., & Kosch, H. (2014). Towards an easy to use authoring tool for interactive non-linear video. *Multimedia Tools and Applications*, *70*(2), 1251–1276. https://doi.org/10.1007/s11042-012-1218-6

Meixner, B., Siegel, B., Hölbling, G., Lehner, F., & Kosch, H. (2010). *Siva suite: Authoring system and player for interactive non-linear videos* [Paper presentation]. Proceedings of the 18th Acm International Conference on Multimedia (pp. 1563–1566), Firenze, Italy. https://doi.org/10.1145/1873951.1874287

Mendes, P. R. C., Guedes, L. V., Moraes, D. D S., Azevedo, R. G. A., & Colcher, S. (2020)., July). *An Authoring Model for Interactive 360 Videos* [Paper presentation]. 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW) (pp. 1–6), London. https://doi.org/10.1109/ICMEW46912.2020.9105958

Monserrat, T. J K. P., Li, Y., Zhao, S., & Cao, X. (2014). *L.ive: An integrated interactive video-based learning environment* [Paper presentation]. (p. 3399–3402). Toronto, Ontario, Canada. https://doi.org/10.1145/2556288.2557368

Moser, C., & Fang, X. (2015). Narrative structure and player experience in role-playing games. *International Journal of Human-Computer Interaction*, *31*(2), 146–156. https://doi.org/10.1080/10447318.2014.986639

Moura, M., Almeida, P., & Geerts, D. (2016). A video is worth a million words? comparing a documentary with a scientific paper to communicate design research. *Procedia Computer Science*, *100*, 747–754. https://doi.org/10.1016/j.procs.2016.09.220

Occa, A., & Suggs, L. S. (2016). Communicating breast cancer screening with young women: An experimental test of didactic and narrative messages using video and infographics. *Journal of Health Communication*, *21*(1), 1–11. https://doi.org/10.1080/10810730.2015.1018611

Ouh, E. L., Gan, B. K. S., & Lo, D. (2022). *ITSS: Interactive web-based authoring and playback integrated environment for programming tutorials* [Paper presentation]. Proceedings of the ACM/IEEE 44th International Conference on Software Engineering: Software Engineering Education and Training, (pp. 158–164). New York, NY. https://doi.org/10.1145/3510456.3514142

Partarakis, N. N. P., Doulgeraki, P. P. D., Karuzaki, E. E. K., Adami, I. I. A., Ntoa, S. S. N., Metilli, D. D. M., Bartalesi, V. V. B., Meghini, C. C. M., Marketakis, Y. Y. M., Kaplanidi, D. D. M., Theodoridou, M. M. T., & Zabulis, X. X. Z. (2022). Representation of socio-historical context to support the authoring and presentation of multimodal narratives: The mingei online platform. *Journal on Computing and Cultural Heritage*, *15*(1), 1–26. https://doi.org/10.1145/3465556

Reyes, M. C. (2017). Screenwriting framework for an interactive virtual reality film. *3rd immersive research network conference ILRN* (pp. 92–102). https://doi.org/10.3217/978-3-85125-530-0-15

Ryan, M. L. (2006). *Avatars of story* (Vol. 17). U of Minnesota Press.

Ryan, M. L. (2015). *Narrative as virtual reality 2: Revisiting immersion and interactivity in literature and electronic media*. JHU press.

Sampaio, P. N. M., & Courtiat, J. P. (2004). An approach for the automatic generation of RT-LOTOS specifications from SMIL 2.0 documents. *Journal of the Brazilian Computer Society*, *9*(3), 39–51. https://doi.org/10.1590/S0104-65002004000100004

Sauli, F., Cattaneo, A., & van der Meij, H. (2018). Hypervideo for educational purposes: A literature review on a multifaceted technological tool. *Technology, Pedagogy and Education*, *27*(1), 115–134. https://doi.org/10.1080/1475939X.2017.1407357

Saunders-Smith, G. (2009). *Non-fiction text structures for better comprehension and response*. Maupin House Publishing, Inc. https://books.google.com/books?id=-CQuwfUtEzcC

Shipman, F., Girgensohn, A., & Wilcox, L. (2005). *Hypervideo expression: Experiences with hyper-hitchcock* [Paper presentation]. Proceedings of the Sixteenth ACM Conference on Hypertext and Hypermedia, (p. 217–226), Salzburg, Austria.https://doi.org/10.1145/1083356

Shneiderman, B. (2007). Creativity support tools: Accelerating discovery and innovation. *Communications of the ACM*, *50*(12), 20–32. https://doi.org/10.1145/1323688.1323689

Stern, B. (Ed.). (1998). *Representing consumers: Voices, views and visions* (1st ed.). Routledge. https://doi.org/10.4324/9780203380260

Su, C. Y., & Chiu, C. H. (2021). Perceived enjoyment and attractiveness influence Taiwanese elementary school students' intention to use interactive video learning. *International Journal of Human–Computer Interaction*, *37*(6), 574–583. https://doi.org/10.1080/10447318.2020.1841423

Sutcliffe, A., & Hart, J. (2017). Analyzing the role of interactivity in user experience. *International Journal of Human–Computer Interaction*, *33*(3), 229–240. https://doi.org/10.1080/10447318.2016.1239797

Truong, A., Chi, P., Salesin, D., Essa, I., & Agrawala, M. (2021). *Automatic generation of two-level hierarchical tutorials from instructional makeup videos* [Paper presentation]. Proceedings of the 2021 Chi Conference on Human Factors in Computing Systems, Yokohama, Japan. https://doi.org/10.1145/3411764.3445721

Wijaya, M. C., Maksom, Z., & Abdullah, M. H. L. (2021). A brief of review: Multimedia authoring tool attributes. *Ingénierie Des Systèmes d Information*, *26*(1), 1–11. https://doi.org/10.18280/isi.260101

Wijaya, M. C., Maksom, Z., & Abdullah, M. H. L. (2022). A brief review: Multimedia authoring modeling. *Journal of Information Hiding and Multimedia Signal Processing*, *13*(1), 39–48. https://bit.nkust.edu.tw/ jihmsp/2022/vol13/N1/04.JIHMSP-1579.pdf

About the authors

Weitao You is a professor at the College of Computer Science and Technology, Zhejiang University. His research lies in design intelligence and computational aesthetics.

Zhuoyi Cheng is a master candidate at the International Design Institute of Zhejiang University. She is exploring creativity support tools and AI-assisted design areas.

Zirui Ma is a master candidate at the College of Computer Science and Technology, Zhejiang University. His research interests include human-computer interaction and large language model applications.

Guang Yang is a senior design director at Alibaba Group and a Ph.D. candidate attached to Polytechnic Institute of Zhejiang University. His research focuses on Human-computer interaction and user experience.

Zhibin Zhou is a Research Assistant Professor in the School of Design at The Hong Kong Polytechnic University. His prior research endeavours to capture the interaction between humans and AI in order to gain a greater understanding of AI as an emerging technology for empowering the user experience (UX).

Lingyun Sun is a professor at the College of Computer Science and Technology, Zhejiang University. His research interests include human-computer interaction and creative intelligence.

Appendix A. Seventeen researched platforms

Table A1. The business platforms and platforms designed in academic papers we researched.

Platforms	Editor pattern(s)	Multi-video edition	Features
bilibili	Storyboard View; Parameter Setting Panel	Yes	Customize button style; Impart weight to options; horizontally-developed storyboard
Blue Billywig	Multi-layer Timeline View; Parameter Setting Panel	No	Detailed action customization of each element; control the state of the video
Chang et al. (2004)	Storyboard View	Yes	Select and annotate meaningful video objects; select particular viewing path
dot.vu	Multi-layer Timeline View; Parameter Setting Panel; Template-based Mode	No	Goal-oriented template; unique guiding steps adapted to each template
eko	Storyboard View; Parameter Setting Panel; Template-based Mode	Yes	Goal-oriented template; partly preview; play selected branch; preset style and layout
Gaggi and Celentano (2002)	Storyboard View; Directory Tree View	Yes	Support multimedia type
GriNS (Bulterman et al., 1998)	Multi-layer Timeline View;	Yes	Presenting jump relationship in the timeline between elements
hihaho	Multi-layer Timeline View; Parameter Setting Panel	No	Various interactive elements; interact between projects
Hsu et al. (2005)	Storyboard View	Yes	Multi-modal interaction
Ivory Studio	Multi-layer Timeline View; Parameter Setting Panel	Yes	Combine functions for traditional video edition and interactive video edition
Luma1	Parameter Setting Panel	No	Detailed parameter setting
Mindstamp	Parameter Setting Panel	No	Add drawings; interact between projects
SIVA Suite (Meixner et al., 2010)	Storyboard View;	Yes	Automated shot detection; export to XML and flv files
spott	Multi-layer Timeline View; Parameter Setting Panel	No	Track object
Stornaway	Storyboard View; Parameter Setting Panel	Yes	Island with an entrance and multiple exits; drag to build between-video relationships
verse	Parameter Setting Panel	Yes	Edit elements in one video at a time