Examining Audience Retention in Educational Videos - Potential and Method

Stefan Oppl
Department of Continuing Education Research and Educational Technologies
Danube-University Krems
Krems, Austria
stefan.oppl@donau-uni.ac.at

Anja Gutmann, Ivan Lazic, Florian Mühlburger
Department of Business Informatics - Communications Engineering
Johannes Kepler University Linz
Linz, Austria
anja.gutmann@jku.at, ivan.lazic@jku.at,
florian.muehlburger@jku.at

***Abstract*—In-depth analysis of viewing behavior is an important factor informing educational video design. Research has largely addressed the support of such analyses for individual learners or has relied on aggregating user interactions with video players to derive potential points of interest. Only in recent years, with rising use of commercial video hosting platform, the data necessary to analyze viewing behavior on a larger scale and independently of active user interactions with learning content has become widely available. This article proposes to examine audience retention data to obtain in-depth information about the usage of educational videos and calculate metrics that can serve as a starting point for qualitatively examining the effects of didactic video design decisions. We propose a set of metrics and present a toolchain for extracting the necessary data from YouTube Analytics and calculating these metrics. The usefulness of these metrics is showcased on three different scenarios of use that are backed with real-world analytics data from flipped classroom courses. Our results show that the detailed audience retention data provided by video hosting platforms can provide insights that go beyond the current state-of-the-art in in-video learning analytics.**

***Index Terms*—Video Analytics, In-Video Viewing Behavior, Formative Metrics**

I. Introduction

Instructional videos have been recognized as essential conveyors of learning content in online and blended learning contexts since several decades [1]. Until a few years ago, technical constraints on data transmission bandwidth and platform performance, however, have hampered their widespread deployment. These constraints have largely vanished today with a nearly ubiquitous availability of high-bandwidth mobile data connections and the consolidation of the diverse field of video encoding and container standards [2], [3]. In the educational sector, these developments have led to a renaissance of learning videos, which are deployed in a variety of educational settings and different formats [4], [5].

As the importance of video as a component in digital learning support designs rises, research is inevitably confronted with the questions of how to design effective learning videos. While plenty of heuristically developed sets of guidelines have been developed in this field, there is little empirical evidence on the effects of videos in educational designs [4]. Even the metrics to assess these effects are unclear [6]. Research in the field of technology-enhanced learning has extensively studied the effects of learning materials on students learning success [7]–[11]. This outcome-oriented view is useful for summative evaluation of learning artifacts (such as educational videos) but is of limited use for formative evaluation, i.e., evaluation activities that aim at identifying the potential for improvement in these artifacts.

From a design science perspective [12], examining the process of artifact usage and matching the findings with the (deliberately designed or unintentionally incorporated) properties of the artifact provides the foundation for further design cycles. These cycles eventually lead to the development of general design theories [13] that can inform the design decisions for future artifacts. This article considers education videos as design artifacts and aims at contributing to the set of instruments available to assess their usage in learning processes. Current research in this field focuses on determining the role of videos as elements in educational designs [10] or examines navigational activities of learners for an in-depth view on video usage to improve video design [14] or to support learners during video consumption [15]. Continuously collected timeline-based data has been recognized as an additional source of potentially relevant information for examining learner's behavior when watching educational videos [16]. Which metrics can be derived from such data and how they can be useful for assessing video design decisions, however, has hardly been addressed in educational design research so far. Other disciplines, such as online marketing [17] or web engineering [18], have developed and used metrics based on such data and generally refer to them as audience retention metrics. The lack of use of these metrics in educational video design can potentially be attributed to challenges in obtaining the necessary data learning analytics systems usually do not allow for such an in-depth view on learning videos [19].

This paper sets out to explore the potential of using audience retention data in the process of assessing educational videos for effective design. Its contributions are twofold: First, it examines the information derivable from audience retention data on a conceptual level and identifies the expected added value for instructional video design. Second, it introduces a set of technical tools that can be used to obtain fine-grain audience

978-1-7281-2464-3/19/$31.00 ©2019 IEEE

retention data from videos hosted on the YouTube-platform, which collects the required data and process them in a way that allows for in-depth analysis.

The remainder of this paper is structured as follows: we first revisit the state-of-the-art in the analytics-based evaluation of educational videos and provide an overview of the historical and current foci of research. We then explore the potential of using audience retention data to determine detailed metrics about video consumption behavior that can eventually be used to inform the design of educational videos. In section 4, we present the set of tools developed to extract audience retention data from YouTube Analytics. Section 5 focuses on the algorithms used for automatically calculating the proposed metrics. Section 6 shows example applications of these metrics that demonstrate the potential of using audience retention data for formative analytics. We conclude with some remarks on the limitations of our current research and the potential for future development.

II. Related Work

In this section, we examine the state-of-the-art of analyzing educational videos. This literature review is used to understand and identify the different approaches in analyzing educational videos and which metrics are already used. The purpose of educational videos examined in the different approaches was insignificant for literature selection - we did not distinguish whether the analyzed videos are used in MOOCs (massive open online course), blended learning environments or in flipped classrooms.

Kleftodimos and Evangelidis [14] propose a framework for analyzing learner behavior in educational videos. The aim of this framework is capturing user events in educational videos, such as a user starting a video, entering or exiting a slide, pausing or resuming a video, or working with an interactive item, such as a quiz question. This data is captured based on user sessions. A session is considered the period from a login action of a user until leaving the web-page hosting the examined videos. Data analysis is based on the occurring user events during the videos viewed within a session, their video sections (e.g. slides with the same label) and the attempted interactive elements and quiz questions. The authors rely on the availability of slide-based videos, since events can be more easily identified than in camera-stream-based videos. In further research, the same authors [20] proposed some metrics based on their framework, to get deeper insight into learner behavior. They describe two different types of metrics: the learner engagement metrics and the video popularity metrics. The learner engagement metrics allow to focus on variables of a single learner and contain figures like the number of videos started by a learner or the number of days on which views occurred (e.g., before an exam). In contrast, the video popularity metrics allow to focus on data within a group of learners, like the number of times a video is started or abandoned. In a next step, those metrics are analyzed with the help of clustering algorithms to identify sets of videos that share common characteristics. By implementing the framework in their interactive video based learning environment, Kleftodimos and Evangelidis [19] refine the original data scheme and mention that sections play an important role in video analytics. That is the reason why they capture data when a section is entered, either while viewing the video sequentially or after a jump.

Several approaches propose to use data that can be derived from users' interactions with the video player software. Chen et al. [16] argues that video analysis can be focused on learner click-stream behavior and peak analysis, since these data allow to derive, how students consume the different sections of a video. Kim et al. [21] propose to analyze the in-video dropout, which occurs when students start watching a video, but leave before finishing it, and the interaction peaks (i.e. cumulated occurrences of play, pause or skip video events). Giannakos et al. [22] focus on collecting data such as video navigation of each student, the relationship between students views and learning performance (with the help of a video-based assignment) and the global peak of each video. By doing so, they manually derive an understanding on how students learn and interact with videos. Ullrich et al. [23] propose to use machine learning techniques for the analysis of students viewing patterns. They focus on collecting data from viewing patterns such as rewinding, skipping forward and watching sections repeatedly, and then put them into relation with variables like student performance, the course category or the rating of the course teacher.

Some authors also try to use timeline-based viewing data about videos for analysis instead of event-based interaction data. Garnett and Button [24] take a hybrid approach and analyze, learning platform logs and YouTube Analytics with respect to the extent educational videos are being accessed by students. In addition, they also examine in-video audience retention and mention, that taking a closer look at timestamps, where the video was stopped significantly often, might provide hints at problematic sections in videos. Lau et al. [25] used learner analytics in video-based lectures (VBL) and collected data like the total view count, the total percentage of videos viewed and the audience retention along four different levels (VBL series at global, series, individual video and feedback levels). Within this study, the focus was on the analysis of the audience retention, which is provided by YouTube Analytics. They propose to exclude viewers, who stopped a video within the first thirty seconds, to only get the target audience. For analyzing the drop of audience retention, a regression analysis was used and the resulting curves have also been examined by extracting time points at which the audience retention rose up by more than 10 percent per minute. Further, they analyzed the relation between the audience retention and the video length.

Most of the reviewed approaches concentrate on the analysis of educational videos with a focus on a single learner view. The differences appear in the kind of metrics which are collected and examined and if those metrics appear from a single video or a whole session of videos. Especially the framework developed by Kleftodimos and Evangelidis [14] analyses data from a whole session. Chen et al. [16] and Kim et al. [21] concentrate on the video peak metric with a strong focus on click-streams and Ullrich et al. [23] observe the relation between viewing patterns and functions like the course teacher. All those approaches focus on analyzing educational videos based on user-generated events. In contrast to this, Garnett and Button [24] and Lau et al. [25] examine the properties of a video via analyzing timeline-based audience retention data. While Garnett and Button [24] just mentioned that this metric could be interesting, Lau et al. [25] examined it in more detail. They, however, did not look into the potential of automatic feature extraction from audience retention data to inform in-depth analysis of video sets, but rather focus on discussing observed properties of audience retention for their particular use case. Event-based approaches, such as those proposed by Kim et al. [21] or Chen et al. [16], hint at the usefulness of detailed in-video analytics to derive potential improvement measures. In this article, we aim at examining the potential of using audience retention data for detailed analysis of in-video viewing characteristics.

III. Audience Retention as a Data Source to Guide Educational Video Design

As could be seen in the review of related work, in-depth analysis of viewing characteristics of educational videos largely has only been addressed on an individual level (i.e., observing the viewing process of single learners) or has relied on aggregating user interactions with video players to derive potential points of interest, not analyzing viewing characteristics per se, but active user interventions in the viewing process). Some studies [24], [25] have argued that an analysis of aggregate viewing behavior along the time-line of a video could inform future video design decisions. Audience retention has been suggested as a data source to be used in this context (ibid.). The potential value of this approach is also backed by evidence from industry, where this metric is used to optimize viewer engagement in and, eventually, the revenue of a video[1]. Audience retention is a time-line-based data source that represents the share of users watching a particular point in a video over its duration. Existing work, however, uses graphical representations of these data for visual analysis only [24] or aggregate it to single-value metrics [25], hardly taking advantage of the richness of the available data.

We propose to perform quantitative analyses of the comprehensive time-line-data to inform further qualitative analysis of the impact of design decisions in videos. From the study of related work and observations in real-world data from around 100 educational videos with different educational designs used by over 400 students in four bachelor-level courses over a duration of two years, we have identified quantitative metrics that characterize the relevant properties of an audience retention graph and provide a foundation for qualitative analysis. These metrics are described in detail in the following subsections.

Before that, we give a brief account on the technical background on obtaining audience retention data to provide a foundation to assess the reliability of the available data: Audience retention data is obtained by streaming video platforms by counting the number of times a particular portion of a video has been played in the user's video player. The granularity of the available data is dependent of the technical implementation of the video streaming platform. Usually, video data is downloaded incrementally in portions of several seconds. These incremental portions are be counted and used to calculate audience retention for the timeframe covered by this portion. More fine grain data would be available, if streaming data protocols like RTSP would be used. For reasons of technical compatibility, however, such approaches are rarely used on platforms with larger market share with a focus on asynchronous delivery of prerecorded video content. Audience retention is usually represented as a fraction of the overall number of views of a video, where a view is usually specified as the number of times a video has been started (independently of how long it has been watched). Consequently, it can rise above 100 percent, if a part of a video frequently is played repeatedly within a single viewing session. The resulting data can be presented as a graph visualizing audience retention over the duration of a video. Prior research (e.g., [25]) and visual analysis of existing graphs obtained in our own research shows that this graph has several features that can provide insights into how viewers watch the respective video. These features consequently can be used to reflect on video design decisions. In the following, we discuss these features and elaborate on the metrics that describe them. We focus on the general characteristics of the features here and do not yet specify how they can be determined by automatic graph analysis. The latter topic is covered in section V. Figure 1 gives an overview about the features discussed in the following.

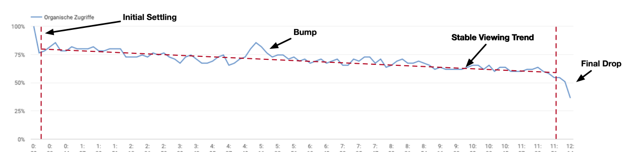

Fig. 1. Audience Retention Graph Features

A. Initial Settling

Videos usually show a drop in audience retention in the first few seconds, after which the graph settles on a rather stable or slowly declining level. This drop is characterized by the *initial settling* feature. There are three metrics that describe this feature:

Settling time is the duration until the graph starts to progress on a more stable level. The shape of the graph during the settling time is characterized by a steady decline, i.e., each data point shows a lower audience retention than the one before. Following this characterization, we identify the end of settling time at the data point, where audience retention remains stable or increases in comparison to the former data point. The metric allows to identify and subsequently examine the portion of the video that usually causes the highest viewer losses. It is

[1]e.g., https://wistia.com/learn/marketing/understanding-audience-retention or https://www.vidyard.com/blog/4-video-marketing-metrics/

interesting for analyzing single videos as well as for comparing sets of videos. If sets of videos show similar settling times, they should be examined for common characteristics (such as intro screens, etc.) that might be causing the drops.

Initial retention drop is the percentage of audience retention loss until the end of the settling time. By definition, the initial audience retention at the beginning of the video is 100 percent. The metric is thus calculated by subtracting the audience retention value at the end of the settling time from the initial value of 100. The metric provides hints at the amount of viewer losses during settling time and is mainly interesting for comparing videos to each other. Videos with different initial drop values should be examined for differences in the features or information provided during settling time that might influence audience retention. For low amounts of overall video views, rather small absolute losses can lead to significant initial drops. In such cases, it thus makes sense to look at the *absolute initial retention drop* value, which is calculated by multiplying the initial retention drop value with the overall amount of views.

Drop slope is a metric for the steepness of the initial drop. It is calculated by dividing the initial retention drop by the settling time. It's value is negative by definition. Different drop slopes indicated differences in the rate of audience loss during the initial drop and are interesting mainly for analyzing single videos. Lower negative values indicate that the losses happen more slowly (i.e., more distributed over time) and might not be attributable to a single feature of the video introduction. Higher negative values indicate quick audience losses (i.e., more densely located around the very beginning of the video) and might hint at issues with the first few seconds of a video.

B. Final Drop

Similarly to the initial settling, videos usually also show a conspicuous decline of audience retention in their final stage. This decline is characterized by the *final drop* feature. The metrics and their usage are also similar to those of the initial settling feature:

Duration of final drop is the duration from where the graph starts its final decline until the end of the video. Again, within this duration, each data point shows a lower audience retention than the one before. The start of the final drop consequently is the data point, after which audience retention steadily drops until the end of the video. The metric allows to identify when the viewer losses in the final stage of the video start to show. It is interesting for analyzing single videos as well as for comparing sets of videos. If sets of videos show similar settling times, they should be examined for common characteristics (such as outros holding little to no information of interest, etc.) that might be causing the drops.

Final retention drop is the percentage of audience retention loss over the duration of the final loss. The metric is thus calculated by subtracting the audience retention value at the end of the video from that at the start of the final drop. The metric provides hints at the amount of viewer losses during the final drop and is mainly interesting for comparing videos to each other. Videos with different initial drop values should be examined for differences in the features or information provided during settling time that might influence audience retention. For low amounts of overall video views, rather small absolute losses can lead to significant final drops. In such cases, it thus makes sense to look at the *absolute final retention drop* value, which is calculated by multiplying the final retention drop value with the overall amount of views.

Drop slope is a metric for the steepness of the final drop. It is calculated by dividing the final retention drop by its duration. It's value is negative by definition. Different drop slopes indicated differences in the rate of audience loss during the final drop and are interesting mainly for analyzing single videos. Lower negative values indicate that the losses happen more slowly (i.e., more distributed over time) and might not be attributable to a single feature of the video outro. Higher negative values indicate quick audience losses (i.e., more densely located around the very end of the video) and might hint at issues with the last few seconds of a video.

C. Stable Retention Trend

After the initial settling time, the development of audience retention usually follows a relatively stable linear trend, either remaining stable on a certain level or showing a slow decline over time. This stable trend usually can be observed until the very end of a video, where another drop in audience retention occurs. This portion of the graph between the two drops is characterized by the *stable retention trend* feature. There are three metrics that describe this feature:

Average retention level gives an idea of how high audience retention is during the main part of the video. It is calculated as the arithmetic average over all data points after the initial settling and the final drop. The metric is mainly of indicative value to identify videos with unusually high or low audience retention. It is of limited use for detailed analysis of videos, as it does not consider the linearity and slope of the trend, which are discussed in the following.

Slope of trend is a metric describing the development of the trend over time. It is determined by calculating a simple linear regression over the data points after the initial settling and the final drop. The slope of the simple linear regression function (usually denoted as β) is used as the relevant metric here. Values close to zero indicate a stable audience retention with no losses of viewers over the main part of the video. Usually, the slope will be slightly negative, where related work [25] and our own data indicates that different types of videos are characterized by different values in this metric. As an example, captures of in-class lectures seem to exhibit higher negative values for this metric, whereas explanatory videos produced in a studio setting often have value close to zero, indicating a stable plateau in audience retention.

Linearity of trend describes how well the shape of the graph after initial settling and final drop can be explained by a linear function as calculated for determining the slope of the trend. It is calculated as the coefficient of determination for the simple linear regression already used above (usually denoted as r^2).

Values close to 1 represent a good fit of the linear function and indicate a linear trend line with few data points that do not fit the model. Lower values indicate that the linear function is a bad predictor for the actual shape of the graph. In the context of our timeline-based data, this usually indicates graphs with high variances over time, which can be caused by viewers skipping or replaying parts of the video. As such parts are usually interesting for in-depth analysis, low values for this metric usually justify a closer look at the bump feature metrics described in the following and could also be used as a filter when working with larger sets of videos.

D. Bumps

Bumps are portions of the audience retention graph after the initial settling and before the final decline, during which audience retention rises significantly over its prior level (where ”significantly” here is not to be understood as a statistical term, but refers to a portion of the graph that conspicuously rises over its surrounding area - the exact algorithms to automatically identify bumps is described in section V. Figure 2 shows a bump and its relevant features. Bumps can occur several times in a graph and are caused by viewers skipping over earlier parts of a video (thus reducing the relative audience retention value there) or replaying the part covered by the bump (thus increasing relative audience retention there). As such, identified bumps are indicators for parts of a video that draw particular interest of viewers and thus justify closer examination.

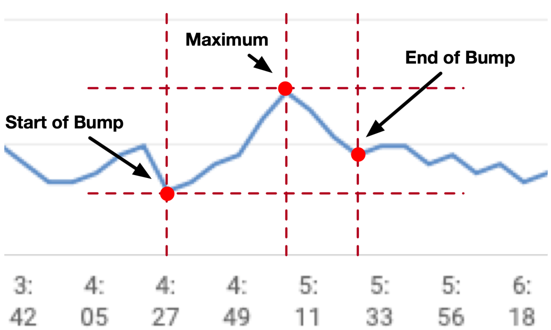

Fig. 2. Features of Bumps

Each bump is characterized by four metrics:

Start of bump is the time-stamp at which the rising slope of the bump starts. For in-depth analysis of bumps, it is usually useful to start reviewing the video from the start of the bump rather than from the time-stamp of its maximum value.

Length of bump is the duration from the start of the bump until the end of the bump. The end of the bump is identified as the time-stamp at which the declining slope of the bump ends (i.e. does not decline further) or drops below the level of the start of bump. The latter criterion has been determined heuristically are being useful to reliably determine the end of a bump in graphs that show higher negative values for the slope of retention trend, i.e., also decline on a global level even outside bumps. The length of a bump indicates the portion of the video that might be interesting for in-depth analysis. Furthermore, our own data indicates that different lengths of bumps are connected to different types of content in the video that draw interest. As an example, explanations of figures or examples in a video tend to produce longer bumps (spanning across the whole duration of the explanation), whereas visually identifiable changes of topics (e.g., slide changes in lecture recordings or switching speakers) tend to lead to shorter bumps (indicating that viewers briefly skim into the respective part of the video but then partially skip over it after some seconds).

Overall retention rise is the percentage of audience retention gain from the start of a bump to its maximum value. It is calculated by subtracting the retention value at the start of the bump from the maximum retention value inside the bump. The use of this metric is twofold. First, it serves as a filter criterion, as not all rises in the graph can be characterized as a bump. If the overall rise does not exceed a heuristically determined threshold, a bump candidate is not considered further. Second, the value of audience retention rise serves as an indicator of the perceived importance of the video content covered by the bump. Higher values thus again might point at parts of a video that more important for in-depth analysis.

Slope of retention rise is a metric describing how quickly audience retention rises within a bump. It is calculated as the slope of the linear function passing through the data points identified as the start of bump and maximum value of the bump. It is positive by definition. Higher values indicate that quick audience gains after the start of bump. Such quicker gains can be observed, if viewers jump to similar starting positions when skipping over or replaying parts of the videos. Such high values are often observed, if navigation links are provided for the video (e.g., chapter marks). Lower values, i.e., slower audience gains, are often caused by seeking behavior of viewers, i.e., viewers who quickly skim through the video until they find an interesting part which they then continue to watch. In such cases, viewers usually do not start at exactly the same video positions, causing the audience retention rise to be spread over a longer duration of time.

In addition to these metrics, which are mainly relevant for in-video analysis, the *overall number of bumps* can be an interesting global metric for a video that, similarly to the linearity of trend metric, can be used as a filter criterion to identify videos exhibiting particularly interesting features.

IV. Tool Support for Data Extraction

Fine-grain data on audience retention can hardly be obtained from video players usually embedded in learning platforms. As the review of related work has shown, most analytical approaches rely on event-based data, such as click streams or slide changes, that can be captured from available meta-data or user interactions. Audience retention data can only be obtained when capturing is decoupled from user interactions and solely based on the rendered portions of the video under examination.

Such data for example is captured by video hosting providers such as YouTube or Vimeo and can also be obtained in self-hosted video platforms via third-party analytics solutions such as Matomo[2]. We here make use of the data generated by YouTube, as it can be accessed for any video without any immediate monetary costs. Aside this, we have chosen to deliver our learning videos via a widely used hosting provider due to the high compatibility for video playback on a plethora of devices and available bandwidths, which we could not obtain in the self-hosted solutions we deployed earlier.

The audience retention data collected by YouTube can be obtained via its analytics interface[3]. Unfortunately, the data that can be exported for further processing there does not include detailed audience retention information. It thus needs to be extracted directly from the user interface. The relevant information is dynamically loaded as a graph in SVG-format and can thus be scraped in way suitable for further processing. Furthermore, learning analytics usually requires data from larger sets of videos. Manual extraction of the raw data from the analytics web page would be cumbersome in such cases and require large effort. For this reason, we have developed a tool chain to extract raw data from the YouTube Analytics website for an arbitrary number of videos. The components of the tool chain are shown in Figure 3 and are described in the following. The raw data is subsequently post-processed to calculate the metrics described above. This post-processing is subject of the next section.

A. Preparation of Meta-data for Analysis

In order to automatically extract the data from YouTube Analytics, it is necessary to manually specify some meta-data for each video to be analyzed. The mandatory meta-data includes the unique video ID and the timeframe for which the data should be retrieved. YouTube Analytics allows to specify the start and the end date of the time period for which the audience retention data should be displayed, thus allowing to filter for particular timeframes of interest in analysis (e.g., the duration of a particular course) or even comparing data from different timeframes (e.g. for comparing video usage during the term with an exam preparation phase).

The videos and the corresponding meta-data are entered in a spreadsheet, which provides the foundation for the actual data scraping process. The current spreadsheet template allows to specify up to four timeframes per video, which are linked by a common stem in the automatically generated unique ID for each of the resulting datasets.

B. Data Scraping

The actual audience retention data needs to be extracted from the respective YouTube Analytics page. As already mentioned above, this page is dynamically generated and the respective data is only retrieved via Javascript during runtime and thus cannot be captured by creating dumps of the static webpage.

[2]https://matomo.org/docs/media-analytics/

[3]http://youtube.com/analytics

The specified meta-data is used to construct the URL requesting audience retention data for a specific video in a specific timeframe. The respective URL parameters include the unique YouTube-ID of the video and the start- and end-date of the timeframe of interest in Unix time format. The generated links are passed to a script, which dumps the dynamically loaded content (cf. lower part of Figure 3). The script iterates over all requested datasets and opens a browser for each of them. A snippet of Javascript injected in the web-page loaded in the browser window then allows to dump the HTML version of the currently displayed web-page (including all dynamically loaded elements). The dumped content is then saved to an HTML file for further processing.

The data of interest is contained in an SVG-graph embedded in the HTML-file. We use regular expressions to extract the SVG graph and the meta-data required to calculate the actual audience retention values from the data contained in this graph. The resulting data are then passed on to the data transformation component, which performs the calculation of audience retention values.

C. Data Transformation

SVG graphs contain information tailored to be drawn on 2D-canvases using absolute coordinates. This data needs to be transformed to represent the actual audience retention information. Transformation needs to account for offsets and scaling factors that are used to draw the graph at a particular position of the canvas at the right size. The extracted SVG contains data on the structure of the diagram enclosing the graph line, which can be used to calculate the offsets and scaling factors.

Data transformation is performed in a Java-program, which transforms the absolute coordinates into relative audience retention values and saves them to a CSV-file for each dataset. These files are used for metric calculation, which is described in the next section.

V. Automatic Determination of Metrics

The audience retention metrics as described above are calculated from the extracted raw data using the R software toolset [26]. Figure 4 shows a graphical representation of the metric calculation result. The graph is cut-off after the settling time and before the final drop.

In the following, we outline how we determine the different metrics. For identifying the end of the initial settling and the beginning of the final drop, we determine the local minima and maxima in the data representing the overall graph. The *settling time* is the timestamp where the first local minimum was found (after around 79 seconds in the graph depicted in Figure 4). The *start of the final drop* is the timestamp at which the last local maximum was found (around 38 seconds before the end of the graph depicted in Figure 4) . Based on this information, we can determine the metrics for *initial settling* and *final drop*.

For further processing, we trim the dataset to cut off the initial settling region and the final drop. The *Stable Retention*

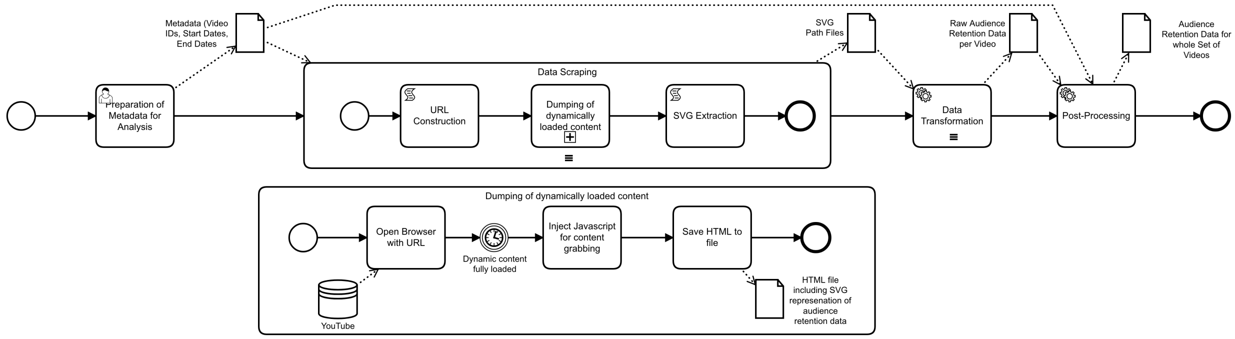

Fig. 3. Toolchain

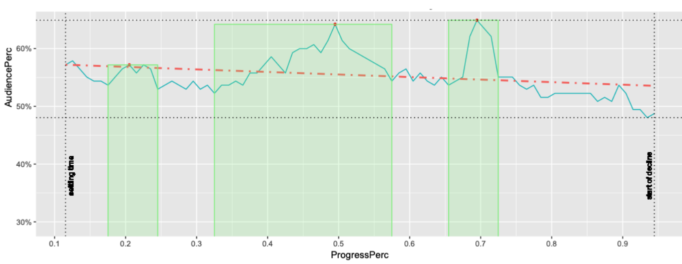

Fig. 4. Example of identified Audience Retention Metrics

Trend metrics are then calculated using R-functions on the remaining dataset. The average retention level for the graph in Figure 4 is 55.4%. The red dash-dotted line in Figure 4 shows the result of the simple linear simple linear regression. The estimate for β is -0.044, the value for r^2 is 0.0963. The β-value hints at an only slowly declining retention trend in the stable area, which matches the graphical interpretation of Figure 4. The rather low r^2-value points at relatively high variances in audience retention in the stable area, indicating a graph with features worth further analysis.

The areas marked in green in Figure 4 indicate the identified *bumps* that should be examined further. Bumps are found by iterating of the trimmed dataset and looking for locally increasing audience retention. As soon as an initial increase is found, we follow the increasing values, where we also accept intermediate declines as long as they do not exceed a certain threshold (e.g. as can be found three times in the rising slope of the second bump identified in Figure 4). The threshold is adapted dynamically in proportion to the overall increase found until the decline occurs. Declines that occur after high increases thus do not as easily lead to finishing the rising slope than those occurring after smaller increases. If a decline exceeds this threshold, we assume that we have reached the highest point of the bump (marked with red dots in Figure 4) and then follow its decline. The declining slope is followed until an increase is found in the graph again or if retention value drops below that of the start of the bump (the three bumps in Figure 4 are examples for the former criterion). Finally, the *overall retention rise* for each identified bump candidate needs to exceed a heuristically determined threshold. In this way, small rises (like at the very end of the graph in Figure 4) are dropped and are not included in the list of bumps. Having determined the time-stamps for start, end and maximum of each bump the allows to determine all other metrics specified above.

VI. Demonstration of Use

We demonstrate the usefulness of the proposed audience retention metrics by presenting three different showcases for it. These showcases demonstrate differences in audience retention for different timeframes on the same video, the effects of different video types on audience retention, and usage of audience retention data to identify audience-drawing portions of a video for further in-depth analysis. Detailed descriptions of the showcases are given in the following subsections.

A. Comparing different timeframes

Audience retention in our experience has been shown to provide useful metrics to examine different video consumption patterns of students in different phases of a course, which in turn might hint at different learning behaviors. A large share of our data has been obtained in flipped-classroom-style courses [27], where videos are available continuously over the whole duration of the course. The potential relevancy of a video, however, changes over time. Each video is assigned to a particular in-class session, for which it should be watched as a preparation. After the in-class session, the video might be useful for further reference when applying the presented concepts in case-studies that go beyond the in-class exercises. Finally, the videos can be used for exam preparation towards the end of the course. These three usage scenarios can be attached to different timeframes, allowing to extract different audience retention datasets. Figure 5 shows the audience retention graphs of a single video in two different time frames. The upper graph shows the data obtained during preparation for the in-class session, whereas the lower graph shows the data obtained during preparation for the exam.

Figure 5 shows that the initial settling time and the start of the final drop are nearly identical. Viewers' behaviors in the main part of the video, however, differ significantly.

The stable viewing trend metrics for both timeframes are given in Table I. Just from these metrics, one can recognize,

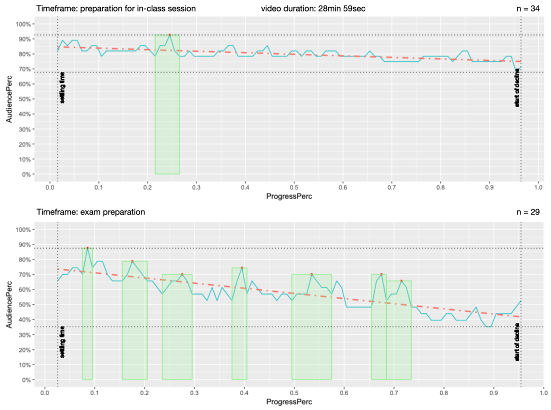

Fig. 5. Comparison of audience retention over different timeframes

that the audience retention during in-class preparation appears to have been higher than for exam preparation (as indicated by the *average retention level*). Furthermore, the audience retention seems to have remained more stable over time in the former case (as shown by the *slope of trend* value), indicating that there has been only little loss of viewers over time. Finally, the values for *linearity of trend* indicate some variance in the graph, hinting at the potential presence of bumps that could be examined further.

TABLE I
COMPARISON OF METRICS OVER DIFFERENT TIMEFRAMES

	ARL[a]	**SoT**[b]	**LoT**[c]
In-class preparation	79.9%	−0.10	0.50
Exam preparation	57.8%	−0.34	0.67

[a]Avg. retention level
[b]Slope of trend (β)
[c]Linearity of trend (r^2)

Visual analysis of the bumps marked in Figure 5 shows that viewers appear to have viewed the video much more selectively during exam preparation. The relatively steep *slopes of retention rise* for the different bumps hint at rather targeted seeking behavior, that seems to be supported by the video (e.g., by visual hints on the current topic that allow to easily identify areas of interest).

B. Comparing different video types

Audience retention also seems to be impacted by different video types. The set of videos we build our hypotheses on contains four different categories of videos that all show distinct features in audience retention that are consistently present across all videos of a given category. The first and largest category covers videos presenting conceptual input for flipped classroom lectures produced in a studio-setting, the second category covers introductory overview videos, the third category includes recording of in-class Q&A-sessions, and the fourth category contains screen-captures of software tutorials and worked examples. Figure 6 shows an example for category 4 at the top and an example for category 3 at the bottom. All other examples so far (e.g., those shown in Figure 5) are representatives of category 1.

Fig. 6. Comparison of audience retention in different video types

The graphs shown in Figure 6 and the metrics given in Table II indicate fundamentally different audience retention characteristics than for videos of category 1. First, audience retention is much lower (i.e., the drop during initial settling is higher) and generally declines more strongly over time. This rather strong decline appears to be common for in-class recordings, as it could also be observed in the study of [25]. This is a significant finding, as videos of category 1 usually show a different, much more stable audience retention in the main part of the video, in particular when used for preparation for in-class sessions (while leading to similar average retention rate as in the study of [25], where the videos were deployed in a similar setting).

TABLE II
COMPARISON OF METRICS FOR DIFFERENT VIDEO TYPES

	ARL	**SoT**	**LoT**
Screen capture (category 4)	32.8%	−0.31	0.90
In-class recording (category 3)	33.5%	−0.24	0.43

see Table 1 for abbreviations

The differences between videos of category 3 and 4 become obvious when visually examining the graphs. While the explanation of the worked example at the top of Figure 6 shows no bumps at all, but a steadily declining audience retention (also indicated by the high value of the *linearity of trend* metric), the recording of the Q&A-session shown a much more varying graph. Content-wise analysis of the bumps shows that they occur when particular questions are discussed and illustrated by examples on the blackboard. The *slopes of retention rise* in this case are not as steep as in the examples discussed above (cf. Figure 5), indicating more a cumbersome seeking process (which is not surprising given the non-optimal camera perspective in the examined video). If questions as answered without any visual hints (e.g., drawings on the blackboard), no bumps can be observed. This hints at the importance of visual cues for seeking behavior in videos.

C. Identifying audience-drawing video portions

As already briefly shown in the last section, the identified bumps and their metrics can provide valuable input for further

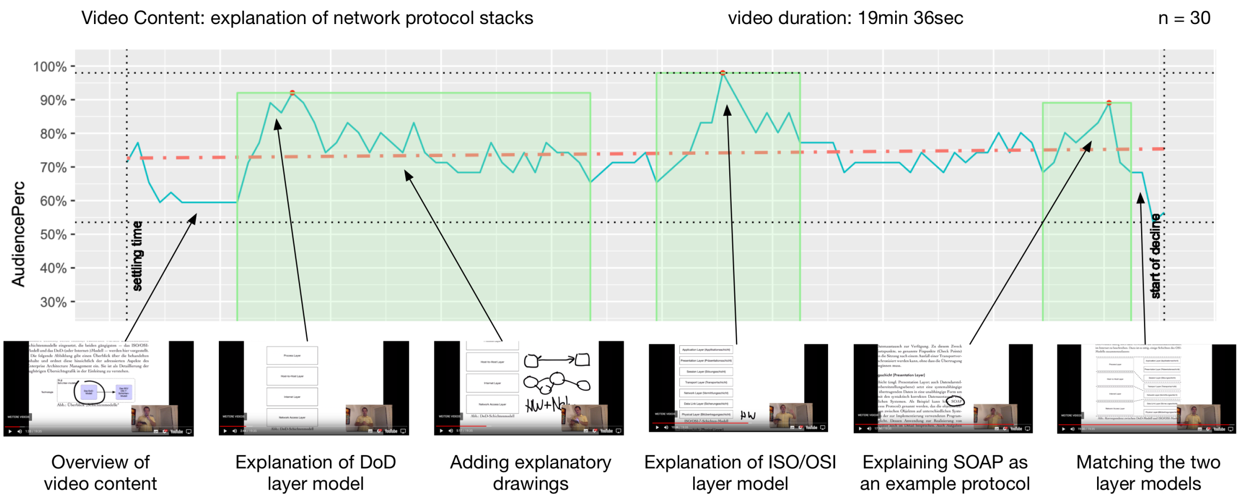

Fig. 7. Identification of audience-drawing video portions

in-depth analysis of a video. In particular, the specific features of the content and its presentation in the area of a bump might provide hints at educational video design elements that lead to higher audience retention. Our observations indicate that bumps are caused by different didactic elements. They, however, all have in common that they provide a visual cue for learners that a potentially interesting section has started. In the example shown in Figure 7, these visual cues are provided by a recording of the lecture notes. The teacher manipulates these lectures notes, scrolling to different areas and occasionally also adding explanatory drawings.

Bumps appear to be mainly connected to figures in the lecture notes appearing in the video. This is the case for the two highest bumps shown in Figure 7 (represented by the second and fourth example screenshot at the bottom of the figure. Audience retention remains relatively high after the first bump (indicated by the long duration of the bump), which appears to be caused by an additional explanatory drawing created to further illustrate the discussed figure. The final bump is not caused by a figure, but also features a visual cue. In this case, the teacher marks the acronym of an Internet protocol used as an example by circling it with a digital pen (represented in the last but one example screenshot). As the protocol presented there also plays a prominent role later on in the course, learners might have been prompted to revisit this particular section during exam preparation (this assumption is backed when reviewing the data for different timeframes of this view, where the bump is only present in the timeframe covering exam preparation).

Not only the bumps, however, might be interesting to review - also the sections leading to audience retention drops could be of interest to identify content or presentation styles that should be avoided in future recordings. One example for such content elements can be identified at the beginning of the video (represented by the first example screenshot). The teacher gives an overview about the content before starting the actual presentation. The drop in audience retention recognizable in the graph exactly covers that particular portion of the video. This might be an indicator that such introductions are perceived to be of limited value (this assumption is backed by the low average retention rates for videos of category 2 as described in the former section, which include introductory overview videos). An interesting pattern can also be observed in the very final portion of the video (represented by the last screenshot in Figure 7), where the teacher provides concluding remarks and compares the two models that were discussed in the video. Although repeatedly mentioned to by important in the video and accompanying in-class sessions, this comparison leads to a steep decline in audience retention, even before the final drop occurs. While the actual reason remains unclear, one might attributed the drop to the redundancy of the visual cue. The teacher uses a figure, which contains similar information as the two former figures. This could lead to the assumption that hardly any new information would be presented and consequently lead to skipping over it. This, however, could only be confirmed, if additional event-based data on viewer behavior (e.g., as proposed by Kim et al. [21]) was available.

VII. Conclusion

In this paper, we have examined the potential of audience retention data as a source for metrics on educational videos. We have explored potential of these metrics to point at certain video characteristics and their suitability to identify regions within a video that reflect critical positive or negative didactic design decisions. We could show that detailed audience retention data available in video hosting platforms can provide insights that go beyond the current state-of-the-art in in-video learning analytics, which usually rely on user-generated event-streams. We have introduced a toolchain supporting data extraction for large sets of videos and have presented algorithmic support for automatically determining the metrics and potential regions of interest for further in-depth analysis.

The research presented here suffers from several limitations. Mainly, the algorithms for determining bumps (regions of interest) require further validation with other sets of audience retention data to validate to heuristically determined parameters encoded there. The current values consistently deliver appropriate results for our own set of educational videos, but have not been used on videos with other types of content presentation or audience retention patterns. Furthermore, the potential of combining the information retrievable from audience retention data with those of user-generated click streams. The additional semantics introduced by this data source would allow to significantly improve the deduced metrics used for high-level interpretation (e.g., it would enable to distinguish bumps caused by replayed sections from those caused by skipping over other sections). Finally, the toolchain developed for data extraction and metric calculation needs further technical stabilization and improvements in terms of usability. The current version of the source code of the developed scripts and applications is provided for download on the Zenodo long-term repository[4].

Our future research address these issues and will initially focus on validating the algorithms used for metric calculation. In parallel, we will explore the potential learnings on video design decisions that can be drawn by calculating audience-retention-data-based metrics on real world data obtained from larger sets of educational videos.

REFERENCES

[1] J. Foertsch, G. Moses, J. Strikwerda, and M. Litzkow, "Reversing the Lecture/Homework Paradigm Using eTEACH® Web-based Streaming Video Software," *Journal of Engineering Education*, vol. 91, no. 3, pp. 267–274, 2013.

[2] R. H. Kay, "Exploring the use of video podcasts in education: A comprehensive review of the literature," *Computers in Human Behavior*, vol. 28, no. 3, pp. 820–831, May 2012.

[3] J. Copley, "Audio and video podcasts of lectures for campus-based students: production and evaluation of student use," *Innovations in Education and Teaching International*, vol. 44, no. 4, pp. 387–399, Nov. 2007.

[4] C.-M. Chen and C.-H. Wu, "Effects of Different Video Lecture Types on Sustained Attention, Emotion, Cognitive Load, and Learning Performance," *Computers & Education*, vol. 80, pp. 108–121, Jan. 2015.

[5] M. N. Giannakos, J. Krogstie, and T. Aalberg, "Video-based learning ecosystem to support active learning: application to an introductory computer science course," *Smart Learning Environments*, vol. 3, no. 1, p. 11, Jul. 2016. [Online]. Available: https://doi.org/10.1186/s40561-016-0036-0

[6] P. J. Muñoz-Merino, J. A. Ruipérez-Valiente, C. Alario-Hoyos, M. Pérez-Sanagustín, and C. Delgado Kloos, "Precise Effectiveness Strategy for analyzing the effectiveness of students with educational resources and activities in MOOCs," *Computers in Human Behavior*, vol. 47, pp. 108–118, Jun. 2015.

[7] E. D. Manley and T. M. Urness, "Video-based Instruction for Introductory Computer Programming," *J. Comput. Sci. Coll.*, vol. 29, no. 5, pp. 221–227, May 2014.

[8] A. E. Williams, N. M. Aguilar-Roca, and D. K. O'Dowd, "Lecture capture podcasts: differential student use and performance in a large introductory course," *Educational Technology Research and Development*, vol. 64, no. 1, pp. 1–12, Feb. 2016.

[9] T. Traphagan, J. V. Kucsera, and K. Kishi, "Impact of class lecture webcasting on attendance and learning," *Educational Technology Research and Development*, vol. 58, no. 1, pp. 19–37, Feb. 2010.

[10] F. V. O'Callaghan, D. L. Neumann, L. Jones, and P. A. Creed, "The use of lecture recordings in higher education: A review of institutional, student, and lecturer issues," *Education and Information Technologies*, vol. 22, no. 1, pp. 399–415, Jan. 2017.

[11] C. Orús, M. J. Barlés, D. Belanche, L. Casaló, E. Fraj, and R. Gurrea, "The effects of learner-generated videos for YouTube on learning outcomes and satisfaction," *Computers & Education*, vol. 95, pp. 254–269, 2016.

[12] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research." *Journal of management information systems*, vol. 24, no. 3, pp. 45–77, 2007.

[13] S. Gregor and A. R. Hevner, "Positioning and Presenting Design Science Research for Maximum Impact." *MIS quarterly*, vol. 37, no. 2, pp. 337–355, 2013.

[14] A. Kleftodimos and G. Evangelidis, "A Framework for Recording, Monitoring and Analyzing Learner Behavior while Watching and Interacting with Online Educational Videos," in *Proceedings of ICALT 2013*. IEEE, Jul. 2013, pp. 20–22.

[15] J. Kim, P. J. Guo, C. J. Cai, S.-W. D. Li, K. Z. Gajos, and R. C. Miller, "Data-driven Interaction Techniques for Improving Navigation of Educational Videos," in *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '14. New York, NY, USA: ACM, 2014, pp. 563–572. [Online]. Available: http://doi.acm.org/10.1145/2642918.2647389

[16] Q. Chen, Y. Chen, D. Liu, C. Shi, Y. Wu, and H. Qu, "PeakVizor: Visual Analytics of Peaks in Video Clickstreams from Massive Open Online Courses," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 10, pp. 2315–2330, Oct. 2016.

[17] D. Zigmond, S. Dorai-Raj, Y. Interian, and I. Naverniouk, "Measuring advertising quality on television: deriving meaningful metrics from audience retention data," *Journal of Advertising Research*, vol. 49, no. 4, pp. 419–428, 2009.

[18] Q. Yang, M. M. Amiri, and D. Gündüz, "Audience retention rate aware coded video caching," in *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2017, pp. 1189–1194.

[19] A. Kleftodimos and G. Evangelidis, "Using open source technologies and open internet resources for building an interactive video based learning environment that supports learning analytics," *Smart Learning Environments*, vol. 3, p. 9, Jun. 2016.

[20] ——, "Using metrics and cluster analysis for analyzing learner video viewing behaviours in educational videos," in *Proc. of AICCSA 2014*, Nov. 2014, pp. 280–287.

[21] J. Kim, P. J. Guo, D. T. Seaton, P. Mitros, K. Z. Gajos, and R. C. Miller, "Understanding in-video dropouts and interaction peaks inonline lecture videos," in *Proceedings of the first ACM conference on Learning at scale conference*. ACM, 2014, pp. 31–40.

[22] M. N. Giannakos, K. Chorianopoulos, and N. Chrisochoides, "Making sense of video analytics: Lessons learned from clickstream interactions, attitudes, and learning outcome in a video-assisted course," *The International Review of Research in Open and Distributed Learning*, vol. 16, no. 1, Jan. 2015.

[23] C. Ullrich, R. Shen, and W. Xie, "Analyzing Student Viewing Patterns in Lecture Videos," in *2013 IEEE 13th International Conference on Advanced Learning Technologies*, Jul. 2013, pp. 115–117.

[24] T. Garnett and D. Button, "A case study exploring video access by students: wrangling and visualising data for measuring digital behaviour," in *33rd International Conference of Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education*, 2016, p. 225.

[25] K. H. V. Lau, P. Farooque, G. Leydon, M. L. Schwartz, R. M. Sadler, and J. J. Moeller, "Using learning analytics to evaluate a video-based lecture series," *Medical Teacher*, vol. 40, no. 1, pp. 91–98, Jan. 2018.

[26] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2009.

[27] J. O'Flaherty and C. Phillips, "The use of flipped classrooms in higher education: A scoping review," *The Internet and Higher Education*, vol. 25, pp. 85–95, 2015.

[4] http://doi.org/10.5281/zenodo.1323648