# Mimicking Real Users' Interactions on Web Videos through a Controlled Experiment

Antonia Spiridonidou, Ioannis Karydis, and Markos Avlonitis

Dept. of Informatics, Ionian University, Kerkyra 49100, Greece
{p10spyr,karydis,avlon}@ionio.gr

**Abstract.** The huge volume of available video content calls for methods that offer insight to the content without necessitating burdensome users' extra effort or being applicable to specific types or conditions. Preliminarily experimentation on collective users' interactions with the web video interface has been shown to offer such information to a great extend. This work, reports on the design, execution and results of a controlled user-experiment wherein participants are requested to view a video and identify their opinion on the importance of the scenes viewed in a realistic web-based video content viewing scenario, based on the interface of prominent video web-streaming provider. Initial results on the data collected show increased interaction on areas that were expected to attract attention whether related to the content or not.

**Keywords:** User-experiment, users' interactions, collective intelligence, stochastic patterns

## 1 Introduction

Nowadays, video content consumption and creation is easier than ever. On one side, widespread penetration of fast and highly interactive internet allows to an ever increasing number of users enjoying video content while on the other hand affordable storage as well as high-quality capturing devices have made creating such content an ubiquitous process with unprecedentedly high demand. The most popular web streaming video content service, YouTube [8], serves more than 1 billion unique users per month, while storing 72 hours of video every minute [9]. Accordingly, being able to make sense of the available content in a computerised manner, that is, being able to extract new and interesting information that is otherwise very difficult to be done due to the sheer volume of data, is of paramount importance.

Traditional content-based methodologies for the aforementioned data mining processes examine the actual content of each video in order to extract information. Nevertheless, their performance and capabilities fall short in certain occasions and thus research has recently focused on contextual or user-based semantics[1]. Such semantics rely on a broad spectrum of interactive behavior

[1] For a detailed discussion about the complementary character of the two approaches see [1]

and "social activities" users exhibit and perform in relation to video content consumption such as sharing with others, assigning comments/tags, producing replies by means of other videos or even just expressing their preference/rating on the content. Rich as these "social metadata" may be, they have also been critiqued [1] as offering extra burden in the usual consumption process, that mainly includes viewing and browsing, by necessitating extra user effort, leading to the long-tail effect as to their existence. Thus, "social metadata" aside, research [3, 5] has examined the interaction information during the core processes of video content consumption, i.e. during viewing and browsing.

The previously mentioned increased interactivity that Web 2.0 offered for the consumption of video content additionally assisted, through web-oriented architectures, in exposing content providers' functionality that other applications leverage and integrate in order to provide a set of much richer applications. In order to enhance the effectiveness of these applications, there is need for extensive studies of large users' interaction data. To this end, the design and implementation of controlled users' experiments has gained increasing attention. Indeed, controlled experiments provide sets of data of (almost) any desirable size under controlled conditions giving thus the possibility to study specific users' interactions properties for specific Web content. Accordingly, Gkonela and Chorianopoulos [3] utilising the SocialSkip platform [2, 6] collected a pioneering user-based interaction dataset by conducting a controlled experiment during video content consumption providing a clean set of data that was easier to analyse. The platform integrated custom interface videos from YouTube with querying form functionalities of the Google Docs API in order to create an environment that would allow for video content consumption as well as user querying in order to accumulate data.

In this way, the collective behavior of Web users watching the video content emerged by means of characteristic patterns in their activity leading to *collective intelligence* as to the importance of video content solely from users' interactions with the video player.

### 1.1 Motivation and Contribution

The dataset introduced at [3] was based on the following assumptions:

- all viewing interface buttons were made similar to a typical VCR device in order to take advantage of the existing cognitive model most users have related to controls of VCR devices,
- in order to fill-in a questionnaire, users browsed the available video content searching for answers within a specific time limit,
- the significance of scenes was predefined based on the questions users' were asked.

To begin with, the typical video content consumption interface supported by key players in the area such as YouTube and Vimeo [7] is highly differentiated to the interface found in VCRs. Web streaming interfaces usually begin

immediately without necessitating to press the play button, only include pause and play buttons while the main feature for browsing is the slider that allows arbitrary moving of the content's time as per the user's will backwards and forward. In addition, the querying process of the experimentation should be free of time limits in order to simulate the realistic usage scenario. Finally, in order to fully take advantage of the collective intelligence of users that may lead to previously unexpected results, experimentation should allow users to freely designate segments of the content thought of as important.

To address the requirements posed, the experiment proposed herein adheres to the following principles:

- the viewing interface includes the controls found in YouTube in order to simulate realistic user video content consumption,
- content viewing does not have any time limit, again in order to simulate realistic user video content consumption,
- the questionnaire requests free-text replies in order to ensure that users' declare the segments they thought of as most important without interference,
- the significance of segments is not predefined, allowing for true collective intelligence.

The rest of the paper is organised as follows: Section 2 presents the participants (Section 2.1), materials (Section 2.2) and the procedure (Section 2.3) of the methodology utilised in order to conduct the experiment. Next, Section 3 details the results received from the experiment conducted and the paper is concluded in Section 4.

## 2 Methodology

This Section details the experiment conducted under the principles discussed in Section 1.1.

### 2.1 Participants

Being delivered through web application, the experiment did not require the physical presence of the subjects and thus allowed users to undertake it at their own time and location of preference. The dissemination phase included informing the users of the experiment's URL link through emails as well as facebook messages.

Following the received link, users were requested to voluntarily participate in the experiment the duration of which would not exceed six and a half minutes to watch the video and then a few minutes answer the questions. Users were also given simple and concise instructions for both parts of the experiment through the web interface.

The participants that undertook the experiment were 103 in number, 36 of who were male and 67 female. The age range varied from 17 to 35 years old and all of them were interested in cookery and The Greek Guide Association. All participants had experience in using the internet and video streaming services such as YouTube.

### 2.2 Materials

To design such an experiment, the open-source SocialSkip platform was modified according to the principles discussed in Section 1.1. Natively, the platform intentionally supports YouTube videos in order to expand the content available for experimentation.

The chosen video for the experiment presented herein was required to be interesting to a lot of people so as to motivate to be viewed by a satisfying number of participants. Additionally, it was required to be interesting to the participants, again to motivate users' interactions on the areas they deemed important.

Thus a set of criteria were identified for the video content selection process. To begin with, an important criterion on the selection of the content was the duration of the video. According to the creators' initial programming of the Socialskip the video should not be more than 10 minutes as otherwise it would be boring and users will not watch it until the end. Another key criterion was that the video should not be structured as the less structured it is, the more important the postdata for the future viewers would be [2]. Accordingly, the video required should be without montage, that is, without replay, slow motion or pause from the director. Moreover, the video should not have areas of increased importance too close to one another. Accordingly, a video about "Hot Wine" [2] was selected that lasts 6 minutes and 14 seconds was captured by The Greek Guide Association in Komotini, Greece.

### 2.3 Procedure

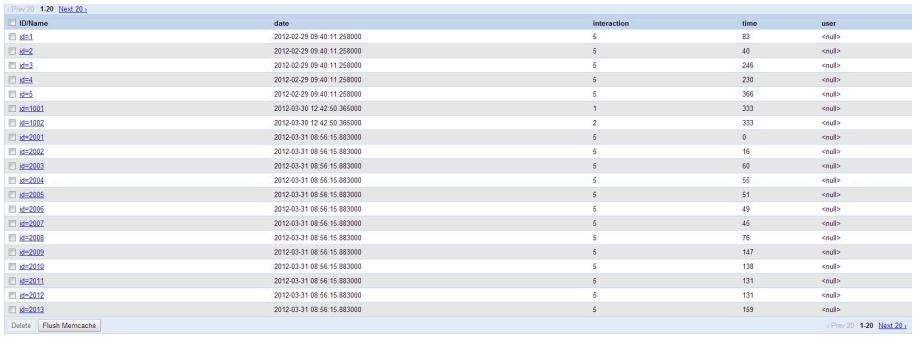
Before viewing a video, participants were offered a set of instructions as to the process of the experiment, stating that:

1. You may view this video as many times required but after selecting to reply to the questions associated, video viewing will not be possible. Thus, you are advised to memorise the parts of the video that you found interesting in order to describe these later,
2. You can use the slider while watching the video in any way you want (back, forward, pause),
3. When sure you have memorised the parts you found interesting press the link to move to the questions phase.

After the participants had selected to move to the questions phase they were given the following instructions: "*Describe the scenes that you consider important in the video (indicative number of scenes A, B, C, D, E). Write your description after you have given the letter of the scene. (ex. A the scene where the cooking instructions are given)*". Intuitively, the instructions were made in this way in order to gather from each user the most important scenes for them, while receiving unguided postdata.

[2] http://www.youtube.com/watch?v=XyZcS5GCF2k

Users' postdata were collected in web-documents supported by the Google docs [4] service. Every time a user initiated the experiment, a set of new records were created automatically in the web-documents, one per each interaction with the interface's buttons. This registration, as shown in Figure 1, has five fields: a unique id, the time-stamp the user began to view the content, the id of button interacted with, the time of the video in seconds the interaction lead to and the user's id as a number. In order to ensure anonymity of the participants, in both Figure 1 and the data user id's have been removed or replaced with random unique equivalent ids.



| ID/Name | date | interaction | time | user |
|---|---|---|---|---|
| id=1 | 2012-02-29 09:40:11.258000 | 5 | 83 | <null> |
| id=2 | 2012-02-29 09:40:11.258000 | 5 | 40 | <null> |
| id=3 | 2012-02-29 09:40:11.258000 | 5 | 246 | <null> |
| id=4 | 2012-02-29 09:40:11.258000 | 5 | 230 | <null> |
| id=5 | 2012-02-29 09:40:11.258000 | 5 | 366 | <null> |
| id=1001 | 2012-03-30 12:42:50.365000 | 1 | 333 | <null> |
| id=1002 | 2012-03-30 12:42:50.365000 | 2 | 333 | <null> |
| id=2001 | 2012-03-31 08:56:15.883000 | 5 | 0 | <null> |
| id=2002 | 2012-03-31 08:56:15.883000 | 5 | 16 | <null> |
| id=2003 | 2012-03-31 08:56:15.883000 | 5 | 60 | <null> |
| id=2004 | 2012-03-31 08:56:15.883000 | 5 | 55 | <null> |
| id=2005 | 2012-03-31 08:56:15.883000 | 5 | 51 | <null> |
| id=2006 | 2012-03-31 08:56:15.883000 | 5 | 49 | <null> |
| id=2007 | 2012-03-31 08:56:15.883000 | 5 | 45 | <null> |
| id=2008 | 2012-03-31 08:56:15.883000 | 5 | 76 | <null> |
| id=2009 | 2012-03-31 08:56:15.883000 | 5 | 147 | <null> |
| id=2010 | 2012-03-31 08:56:15.883000 | 5 | 138 | <null> |
| id=2011 | 2012-03-31 08:56:15.883000 | 5 | 131 | <null> |
| id=2012 | 2012-03-31 08:56:15.883000 | 5 | 131 | <null> |
| id=2013 | 2012-03-31 08:56:15.883000 | 5 | 159 | <null> |

**Fig. 1.** Fields of the registrations of the users interactions.

During the video content's presentation, the slider was available just under the viewing area allowing the user to randomly navigate through the content at will. A "pause" button was also available for the participants' convenience (Figure 2) that freezed the video on the current scene it was pressed. The existence of both these buttons was dictated by the requirement to ensure realistic and familiar usage to the users. Furthermore, under the viewing area, the participants could see the total duration of the video as well as the relative current viewing video time.

Finally, to conclude the procedure participants had to press the "Submit" and "Exit" buttons as otherwise their answers and interactions would not register (Figure 2).

## 3 Results

Preliminarily results revealed that the "pause" button was not used a lot since users had the option to browse the content using the slider, an interface very close to real scenarios. Similarly, the "play" button is was not used a lot as it was only required to restart the video after pausing.

After having studied the interactions, a common viewing pattern emerged: at the beginning every user wanted to move the video forward to see the end.

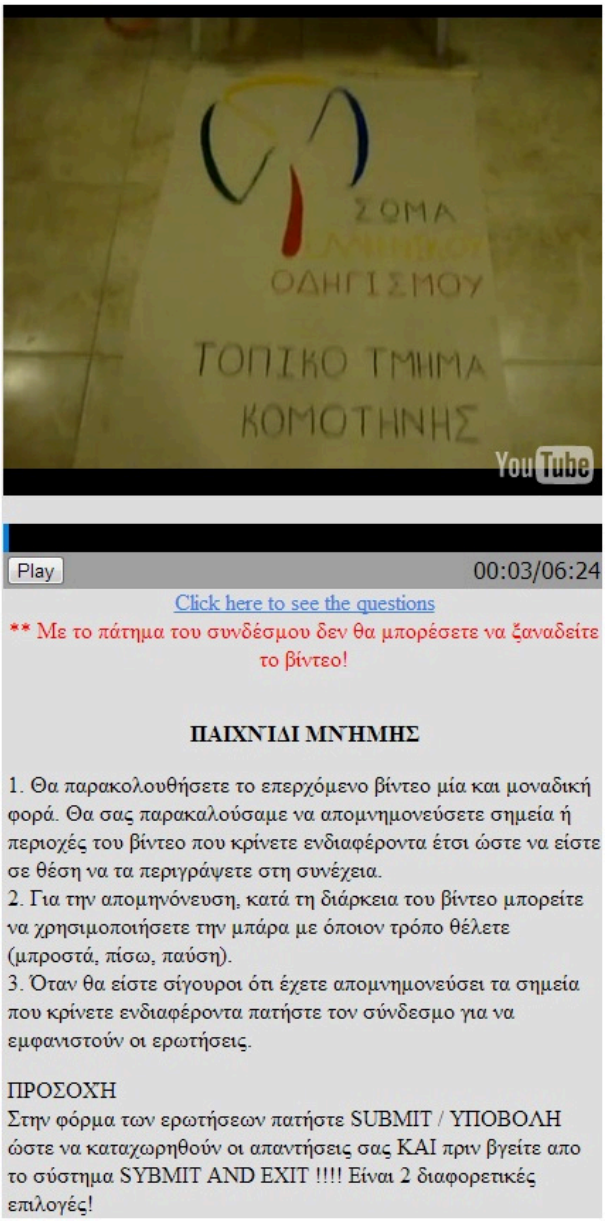



**Fig. 2.** Snapshot of socialskip.

Then, the participant would use the slider to go back to the parts thought of as interesting so that memorising potential answers to the questions given later would be possible. The common viewing pattern was additionally supported

by the fact that most users exhibited the same interactions. Moreover, most users's browsing approach included moving the video time back and forth as they were trying to identify a specific segment of the content. In any case most interactions were at the points of intuitive interest in the content. Figure 3 shows the cumulative users' interaction for the conducted experiment: the y-axis indicates the cumulative number of interactions while the x-axis shows the equivalent relative time of the content the interaction occurred. Three peaks (areas of interest) are easily distinguishable and while these are not even, it is obvious that some parts of the video prevail.

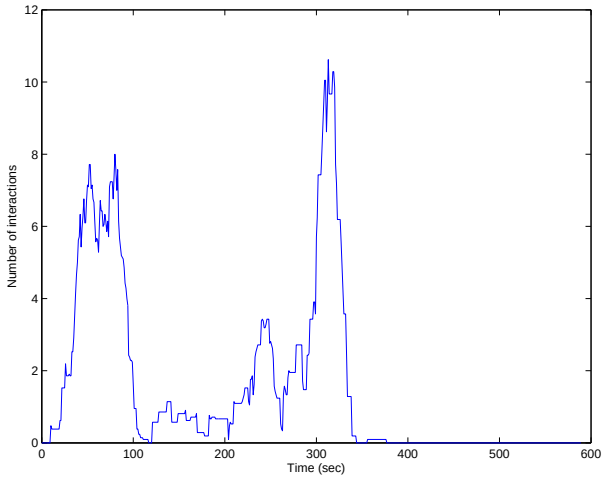



**Fig. 3.** The collected interaction signal.

The first part, where several interactions occurred, was the part where the ingredients are given. In this scene the first secret for the recipe is revealed. In the following picture we can see a part of that scene. (Figure 4)

In this scene there is a detailed presentation of the ingredients which will be used to make the "hot wine". It is one of the most important scenes because when a video is about a recipe the part where the ingredients are presented is important. Moreover, the sequence the ingredients will is presented, also a important for the success of the recipe. Thus, according to the users' interactions it was apparent that this scene was one of the most important.

Continuing further, based on users' interactions as shown in Figure 3, the second part found important was that of a description of how to make an ingredient used in the recipe, Figure 5. In this scene the actors divulge how to make the spice which should be added in the recipe in a specific way to ensure successful blending with the rest ingredients and aroma of the recipe. Thus, it is one of the key parts of the content as well as of the recipe as it is crucial to

**Fig. 4.** The part where the ingredients are given.

its success. Furthermore, the actors prefaced the scene by announcing that "the second secret ..." thus pointing out the importance of the scene.

**Fig. 5.** The scene of how to make an ingredient used in the recipe.

In order to test further the notion of what may be deemed as interesting in video content by collective intelligence, although the video is clearly about a recipe for making "hot wine", at the end of it, a scene not related to the recipe was intentionally included. In this scene, one of the two actors re-appears dressed as Santa Claus, as shown in Figure 6. The scene registered as the third more significant part in the content based on users' collective intelligence. Although a scene that is unrelated to the content of the video is expected not to receive

special attention, the experimental results showed that most interactions had appeared in this part of the video, thus indicating that users thought of it as equally important. This result challenges common notions as to what should be thought of as important.

**Fig. 6.** The entrance of Santa Claus.

## 4 Conclusion

In this work, a controlled user-experiment was conducted in order to identify the segments of video content participants thought of as most important by means of registering their interactions with the video interface. The experiment was designed and implemented in order to achieve high degree of realism to the typical contemporary scenario of web video-streaming services.

Initial results on the data collected showed high correlation of the areas (video time segments) that received more interactions from the users with their free-text submitted replies on what they thought of as important. In other words, the most important scenes according to the participants of the experiment, that form the so called "ground truth", highly coincide with the patterns emerged in users' interaction time signal.

The existence of a signal (here the signal counts how many times the slider was located at a specific second) that can carry the information about the most important video scenes could be a valuable tool for Wed applications. Moreover, the existence of the aforementioned users' signal can provide the basis for new series of metrics in order to study new characteristics of users' interactions. Indeed, the identification of most important scenes is just a first order video characteristic. On a higher level, one could search for second order characteristics: what is the duration of each important scene, how popular each scene is

(there could be more than one popular scenes in each video but what is their gradation popularity) and how one can predict the most important scenes from low quantity early data of users' interactions. The above issues can be addressed in a very rigorous manner since these features can be mapped to well known functions in standard signal processing theories. These aspects will be addressed in a future work.

## References


1. Avlonitis, M., Karydis, I., Chorianopoulos, K., Sioutas, S.: Treating Collective Intelligence in Online Media, chap. Semantic Multimedia Analysis and Processing. CRC Press / Taylor & Francis (2013)
2. Chorianopoulos, K., Leftheriotis, I., Gkonela, C.: Socialskip: pragmatic understanding within web video. In: Proceddings of the 9th international interactive conference on Interactive television. pp. 25–28. EuroITV '11 (2011)
3. Gkonela, C., Chorianopoulos, K.: VideoSkip: event detection in social web videos with an implicit user heuristic. Multimedia Tools and Applications pp. 1–14 (2012)
4. Google: Google docs - online documents, spreadsheets, presentations. (2013), https://docs.google.com/
5. Karydis, I., Avlonitis, M., Sioutas, S.: Collective intelligence in video user's activity. In: Artificial Intelligence Applications and Innovations (2). pp. 490–499 (2012)
6. SocialSkip: User-based video analytics. (2013), https://code.google.com/p/socialskip/
7. Vimeo: Your videos belong here. (2013), https://vimeo.com/
8. YouTube: Share your videos with friends, family, and the world. (2013), http://www.youtube.com/
9. YouTube: Statistics (2013), http://www.youtube.com/t/press_statistics