In the previous chapters, we have reported on our efforts to empower end-users to browse a shared video collection based on personal interests and to create personalized, but still compelling, personal stories from it. In this chapter, we now shift our focus from the author to the recipient of the story.

Successful commercial video sharing systems have provided ample proof that video is a first-class Web object. Even social networks like Facebook, originally conceived for status updating, have become important distribution channels for both consumer and professionally generated video [73]. In these sharing systems, video content serves both as a medium for communicating a story (using implicit or explicit cinematic rules), and as a catalyst for communication between third-party viewers of that content [12][25][50].

---

1 This chapter is based on the following papers:


Chapter 5. Supporting Personalized End-User Comments within Third-Party Online Videos

Recent developments by video service providers have extended the means for third-party communication in ways that have never been possible with conventional broadcast or personal video systems. In addition to the base video content, a typical YouTube page also provides space for end-user generated comments (Figure 5.1). These include implicit forms of commentary (such as the number of views or anonymous ratings, e.g., ‘like’ or ‘dislike’), and explicit comments for interpreted viewers.

In the case of online video on demand, textual comments are usually statically placed underneath the media player. If desired, users need to make explicit any reference to a particular event that happens within the video object (e.g., “Look at that shiny, beautiful trombone at 1:56” in Figure 5.1). In YouTube, for example, when a user writes out a particular time code in the comment, it automatically turns into a ‘temporal hyperlink’, that when clicked takes the interested viewer to that part of the video. However, such comments do not reproduce the ‘commenting while watching’ activity people perform when consuming media together. In general, users cannot add comments that are synchronized with the video, unless the owner (who uploaded it) has given editing rights to the base video content.

Primarily, this chapter considers the scenario in which a recipient of the content – not necessarily the owner of the video or who created a personal video story, adds personalized comments that are synchronized to specific events within the video. By personalized we mean comments created to highlight a particular event that is interesting to, for instance, the end-user social circle. By synchronized, we mean that such comments will be rendered during video playback at the time such particular event happens, unlike the static comments displayed underneath, as in YouTube or Facebook. Supporting this functionality, which is aligned with the intimacy and reciprocity guidelines specified in Chapter 2, we expect to reproduce asynchronously the commenting experience people have when watching media together. In this direction, we have asked the following research question:

Question 1.5 Does the support for timed end-user commenting within pre-authored narratives provide an identifiable improvement over current media commenting approaches?

Motivated by a survey research on current media watching and commenting practices, this chapter reports on the design, implementation and user-centric evaluation of a video commenting paradigm for structuring synchronized
comments within media. Our results indicate that users appreciate the functionalities of our system and find it better to comment when compared to current video commenting tools.
Chapter 5. Supporting Personalized End-User Comments within Third-Party Online Videos

In order to realize our video commenting system, we also specify and describe a set of temporal transformations for multimedia documents. Our approach, unlike current solutions, allows end-users to create and share personalized timed text comments within third-party online videos. It also permits end-users to identify temporal navigation points by using hyperlinks within comments, and to associate contextual metadata (e.g., who wrote the comment and when). The benefit over current solutions lays in the usage of a rich commenting format that is not embedded into a specific video encoding format.

This chapter is structured as follows. Section 5.1 motivates our work, while Section 5.2 proposes a set of multimedia document transformations that allow end-users to add timed comments within third-party Web videos. Section 5.3 describes the design and implementation of a Web-based video commenting tool, which realizes such document transformations. In this section we also report on a predictive timing model for helping users to incidentally synchronize text comments with specific events within a video. Lastly, in Section 5.4 we present the results from the evaluation process, while in Section 5.5 we discuss the lessons learned and how these fit in the context of this thesis work.

5.1 Media Consumption and Commenting Practices

A sample group of 21 people were invited to participate in an evaluation process during the first quarter of 2012\(^2\). All participants were regular Internet users. Eighteen (18) people were in the 21-40-age range, while the other 3 were over 40 years old. Participants were from different nationalities including Brazilian, Chinese, Dutch, German, Hungarian and Irish.

We used semi-structured electronic questionnaires to collect users’ feedback. While multiple-choice questions allowed us to explore patterns and find trends (quantitative methods), open-ended questions aimed at capturing further insights into participants’ opinions and perceptions. The user study was divided in 3 parts. The first part, which is the focus of this section, consisted of a questionnaire to gather background information about respondents’ commenting practices when watching video content. Feedback answers were anonymous.

\(^2\) This was an independent study and it counted with a different set of participants from the ones involved in the evaluation process discussed in the previous chapters.
5.1.1 Survey Research

Figure 5.2 summarizes the results obtained in our survey about media consumption and commenting habits. As users’ practices were different, for each question we present the weighted average (colored column) and the respective standard deviation (bar). In Figure 5.2 the questions also have been clustered in two groups according to the consumption experience: synchronous (blue) and asynchronous (red) watching. For each scenario we also asked about participants’ conversational and commenting practices around media.

A wide range of TV watching habits has been reported by our participants. In average, our users watch TV every week (Q1.1). This was the second highest frequency score among our questions. Participants also reported having the habit (between occasionally and every week) of talking with family and friends about a TV show they have watched (Q1.2). When asked about the frequency they would converse about a TV program with collocated people while watching, the average answer was occasionally (Q1.3). The lowest score though was obtained in the question about how often they would send tweets related to a TV show they were watching (Q1.4). As reported elsewhere [13], this activity is becoming popular over the years and, in some cases, it can be used as an interactive return channel in which the audience can influence on live TV programs.

In the second media consumption scenario, we asked our participants about the habit of watching live video feeds on the Web (e.g., Justin.tv, Ustream.tv). The average of their feedback was around occasionally (Q1.5). Regarding the activity of commenting on the video event while watching, we asked how often they make use of the built-in open textual chat rooms generally available on those services. Again, rather small the frequency stayed between never and occasionally, which was slightly higher than the one reported for tweet messages (Q1.6).

Regarding on demand (asynchronous) watching, we asked our participants about the usage of YouTube, Facebook and SoundCloud. Validating previous research [73], YouTube was often (between every week and every day) used by our participants to watch online videos (Q1.7). However, posting comments to the video page did not seem to be a common practice among our participants (Q1.8). In conjunction with the use of YouTube, we also witnessed a fair high frequency of video viewing on social networking sites (Q1.9). In this case though, participants habitually comment more on videos when compared to the comments added in YouTube (Q1.10). One possible explanation for this behavior is that participants are more likely to post comments within their social circle than in the open.
Figure 5.2. Survey research about media consumption and commenting practices. Blue columns indicate synchronous watching and related conversational habits. In red, on demand (asynchronous) consumption and commenting.
Finally, we asked our users to report on their practices using online audio streaming services such as SoundCloud. In average, respondents occasionally listen to music on these platforms (Q1.11) and they do not have the habit of adding timed comments to songs (Q1.12).

In the second part of the survey we asked our participants about the possibility of adding timed text comments to particular events within a third-party online video (see Figure 5.3). Thirteen (13) out of 21 participants said they would possibly (yes or maybe) add timed comments to YouTube or Facebook videos if they could (Q1.13). One of the participants expressed that such a feature would be a “nice way to highlight sections/points of the video”. When asked whether they would share these synchronized comments within their social circle the number raised to 16 out of 21 (Q1.14), fairly higher than the number reported for sharing comments with everyone (Q1.15). In one case, a user justified by writing “I’m never interested in sharing my comments with the public... but to have a link that I just could send to friends”.

At last, we asked a question related to digital rights and ownership. In this case, only 2 participants expressed they would mind if other people could add timed comments to their videos (Q1.16). In these lines, one participant highlighted the necessity of having control over the commenting activity: "If everybody could add comments to any video it would become a real mess. Some people would use it
to damage the image of others”. This result seems to contradict the privacy issues discussed in the previous chapters, but it is important to keep in mind that here the videos were not necessarily personal (as opposed to the ones discussed in the extensive long-term evaluation process in the UK and the Netherlands).

5.1.2 Requirements Gathering

In the study presented above we looked at media consumption and commenting practices of a group of Internet users using different applications. Even though the group was small and provided results with a high variance, we obtained strong indications that people consume media and doing so, they eventually comment and share such moments within their close circle and, sometimes, with the general public. Our respondents also appreciated the utility and usefulness of synchronized comments, as they would comment on particular events within media (Q1.13–Q1.15). These results led to the specification of the requirements described below. These were used as the basis for designing and validating the online video commenting system presented in the remaining of this chapter.

i. Retain base video integrity: viewers should not be able to alter the base video content, either in terms of adding embedded captions/comments or of providing visual overlays on the base content — this right is reserved to the content owner;

ii. Allow multiple-video aggregation: in certain occasions, end-users might watch a collection of videos that are played as a continuous playlist (e.g., a personalized video story or compilation, as shown in the previous chapter). In these cases, end-users should be able to create comments that would span across the multiple videos composing the playlist;

iii. Allow multiple-provider integration: the user should not be locked into a single video service provider (or source) for candidate content, but should be able to populate a playlist from multiple sources;

iv. Allow timed end-user comments: when watching an online video, viewers should be able to add comments that are time synchronized. This feature would reproduce (asynchronously) the watching and commenting activity people have when watching media collocated;

v. Allow micro-personalized timed comments: end-users should be able to create different sets of time-based comments for individual
users/communities, or share these as ‘broadcast’ comments (similar to existing approaches in YouTube and similar systems);

vi. **Allow selective end-user viewing:** end-users might be able to select and watch comments by specific individuals and/or user communities, by topic etc. This is important because some comments might be targeted to individual users while others might be intended to the general public; and

vii. **Allow timed end-user navigation:** end-user comments should be able to include direct navigation support via timed anchors in the text content. This will allow others to navigate to other interesting parts within the same collection or to link to external media.

### 5.2 Media Commenting meets Multimedia Document Engineering

To address the requirements discussed above, we modeled the problem of creating timed comments within online videos from a multimedia document engineering perspective, and thus identified a set of document transformations. By document transformations we mean manipulations that can be applied to add non-embedded, flexible temporal end-user comments. Video commenting has been dealt with in many ways, ranging from the usage of models that are not timed (e.g., HTML) or are unstructured (e.g., Flash) to standards such as MPEG-7 [1] and NCL [33]. Based on our analysis [62], we rely on SMIL 3.0 [17] as the basic framework that meets our requirements. First, we create a structured multimedia document around an input video(s). The document model of SMIL 3.0 retains the base video integrity, and it allows multiple-video aggregation and multiple-provider integration. Timed text content and temporal hyperlinks allow end-users to add synchronized comments and to include timed end-user navigation points. Contextual information allows targeting timed comments to different audiences. Finally, the structured underlying model enables selective viewing.

#### 5.2.1 Document Model

SMIL can integrate and compose a collection of audio, graphics, image, text, and video media items into a single presentation. As Web resources are distributed by nature – and might be very large in size –, in SMIL media objects are included by reference (using a URI - *Uniform Resource Identifier*). SMIL defines a single generic media object (*<ref> element*) that allows the integration of external...
media resources into a SMIL presentation. However, it is also possible to use more intuitive tags when referencing external media resources (e.g., the <video> element is a more specific alias for the generic SMIL media reference element). Note that as an implication of the use of references, the integrity of the base media is preserved, meeting requirement i.

In addition, SMIL provides a powerful hierarchical composition model from which individual presentation timelines can be generated. The main temporal structuring elements are the parallel (<par>) and sequential (<seq>) containers, each of which provides a local time base for scheduling media objects (e.g., external videos) or children time containers. By using such time containers, it is possible to combine videos and comments in different temporal ways, as illustrated in Figure 5.4. In this example, three videos, stored in different video servers, are rendered as a continuous video, while the comments span across the videos. The structured temporal container behavior satisfies requirements ii and iii.

5.2.2 Timed Text Content

Unlike most text formats [15], text content in SMIL is not only constrained by its style and layout capabilities, but also by the temporal context of the presentation. For instance, text must be rendered simultaneously with related objects, and it must be hidden when these are finished.

Authors can define small amounts of lightly formatted text containing embedded temporal markup within the context of a SMIL presentation. Such text may be used for labels within a presentation or for incidental comments or foreign-language subtitles. It is also possible to use large amounts of structured text (with or without temporal markup), but in this case it is recommended the use of SMILText as a text media object, or the use of objects encoded in formats such as XHTML or DFXP (Distribution Format eXchange Profile) [27].

The SMILText also define a set of additional elements and attributes to control timed text rendering (see Figure 5.4). All SMILText content is processed in a manner consistent with other SMIL media. The SMILText profile also allows SMILText to be used as an external format. Moreover, since the smilText elements and attributes are defined in a series of modules, designers of other markup languages may reuse these modules when they wish to include a simple form of timed text functionality into their language. SMILText, as a text container with an explicit content model for defining timed text, meets requirement iv.
Figure 5.4. SMIL document model and temporal containers.
Figure 5.5. Timed text content and temporal hyperlinks.
5.2.3 Temporal Hyperlinks

SMIL 3.0 Linking Modules define SMIL 3.0 document attributes and elements for navigational hyperlinking. SMIL hyperlinks may be triggered by user interaction or other events, such as temporal events. SMIL 3.0 provides only inline link elements. Links are limited to unidirectional single-headed links (i.e. all links have exactly one source and one destination resource).

As with styled time-based text comments, adding temporal hyperlinks via text content can enrich the content viewing experience for end-users and for their social circle. This association makes SMIL meet requirement vii.

It is important to highlight that our document model allows links to be added to content without violating the legal rights of any party. This is possible because navigation points within the video are encoded as a series of content events in the SMIL document. Two classes of links can be provided as illustrated in Figure 5.5:

- **Intra-video Navigation Link**: a text link that takes the viewer to another location within the active video; and
- **Inter-video Navigation Link**: a text link that takes the viewer to another piece of content, outside the active video.

5.2.4 Contextual Information

Current Web-based video solutions provide limited support for including metadata related to the comments. For example, they do not allow end-users, at authoring time, to create different views on the comments, depending on the target audience. As discussed in the previous chapter, one might not send the same set of comments to her family and for to singing teacher.

SMIL 3.0 allows associating meta-information to any element within the document body, including timed text comments. This makes it possible to provide information with semantic intent within the presentation information, by binding relevant nodes with meta-information.

As mentioned before, SMILText allows text comments to be described as single structured units that can be targeted to different audiences. Therefore, we can consider each comment entry as the smallest unit that can be tagged. In order to share a video with comments, we should add contextual metadata, such as who has created the comment, when, why, how, and to whom [60]. Support for targeted
comments might increase the authoring overhead, but it provides a level of personalization that is lacking in common Web environments.

SMIL can tackle the contextual problem, requirement v, by allowing metadata to be associated with timed text comments. Figure 5.6 illustrates this process. Here we see a master comment stream that has been composed by Dick specifically targeted for all viewers within his social circle.

### 5.2.5 Selective Viewing

One shortcoming of current video captioning/commenting systems – whether closed captions or stream of comments on a Web page – is that every user visualizes the same collection of information. It is doubtful that even the most interested reader will go through the dozens of comments created by unknown individuals – but there is a much stronger incentive to view the 20 or so comments that are likely to be generated by family members or close personal friends (as indicated by the results presented in Section 5.1).

In order to deal with this problem, video commenting tools can make use of the structured nature of SMIL to selectively present content, requirement vi. Video commenting tools can enable users to — besides the traditional turn on/off all comments — select and watch comments created by a certain individual or community, about specific topics, or created on a certain day. Moreover, aggregated comments and metadata can be used for generating diagrams of hotspots within videos. All of this is possible thanks to the document model — structured text comments can be analyzed — and to the contextual information associated to the comments. Figure 5.6 illustrates a scenario in which a viewer is interested in a certain category of comments.

### 5.3 A Timed Text Video Commenting System

Based on the temporal transformations discussed in the previous section, we designed and implemented a video commenting system as an independent application. Our solution allows end-users to easily add timed text comments to particular events within third-party online videos. In the remaining of this section we detail the technical aspects of such commenting system.
A Timed Text Video Commenting System

Figure 5.6. Contextual information and selective viewing.

```xml
<smil>
  <head>...</head>
  <body>
    <par>
      <seq>...</seq>
      <smilText region="c1"...>
        ...
        <div>
          <metadata name="author" content="Dick">
          <metadata name="audience" content="all">
            <clear begin="t1-5s">That's an awesome guitar solo!</clear>
            <clear begin="t1+15s"/>
          </div>
          <div>
            <metadata name="author" content="Dick">
            <metadata name="audience" content="his daughter">
              <clear begin="t3-15s">Do you remember you played the same song last year?</clear>
              <clear begin="t3+5s"/>
            </div>
          ...
        </smilText>
    </par>
  </body>
</smil>
```
5.3.1 Infrastructure

The high-level workflow of our video commenting system is illustrated in Figure 5.7. The interaction starts when a user requests a video. For that we make use of the YouTube Data API (Application Programming Interface), which provides programmatic access to the videos stored in YouTube. It allows us to retrieve a set of videos matching a user-specified search term or retrieve standard feeds (e.g., most viewed today). The data is requested using AJAX (Asynchronous JavaScript and XML) and returned in the XML (eXtensible Markup Language) format, then parsed and presented to the end-user.

For video playback we use the YouTube Player API, which is exposed via JavaScript. It allows us to control not only the ‘Look and Feel’ of the player, but also the playback behavior of the videos in our Web application. With the current YouTube infrastructure, the client Web browser must be HTML5 compliant or have Flash Player 10.1, or higher, installed. Most importantly, the Player API provides the necessary time information to synchronize the text comments within a video. This feature is obtained by listening to specific events, which are fired accordingly (e.g., time update event). A similar infrastructure would be necessary for making the commenting tool available for videos hosted in different providers. In this case, the YouTube Data and Player APIs should be replaced and the interfaces of the new provider adapted accordingly.

Since the viewer has no rights to add comments to the base video, the timed comments are stored separately on our Web server. As mentioned previously, the actual format used to encapsulate the multimedia presentation (base video plus a layered collection of timed comments) is W3C’s SMIL 3.0. In fact, timed comments are specified in SMILText, the embedded text format for use within SMIL 3.0. SMIL allows us to respect the video owner’s rights and to keep a provider-agnostic enriched video. As such, comments can be shared, modified and analyzed independently.

For the synchronized playback of end-user comments we implemented a SMILText JavaScript engine that runs on the client’s Web browser. Its API allows us to embed SMILText functionalities in Web pages and have the presentation controlled by an external source, in this case the YouTube video player. The SMILText engine has reasonably complete coverage of the features defined in the SMIL 3.0 SMILText External Profile. The API also provides a number of other utilities for adding and manipulating timed text content, making possible the creation of applications such as the commenting tool presented in this chapter.
5.3.2 User Interface

In order to allow end-users to comment the videos we need a user interface that hides all the complexity from them. This is achieved with our video commenting system, which wraps the video content and all the timed text comments in a multimedia presentation. The commenting interface (Figure 5.8) is composed of a video rendering area (1), a rendering space for comments (2), an input area (3) and the sidebar controls (4). In most cases, relative passive end-users simply will watch a piece of video content forwarded to them. If the content itself has embedded comments, these can be selectively turned on or off via the sidebar controls.
interface. During playback users may also choose to insert new comments in the input area (Figure 5.8 (3)).

One key feature of our video commenting system is its ability to (semi-) automatically compute the temporal alignment of user-generated comments. To explain how this functionality works, consider the example in Figure 5.8. In a usage scenario, we assume users will interact after a certain moment of interest has passed (e.g., after seeing the trombone on screen). In this case, comments need to be synchronized in such a way as to avoid situations in which the comment – “Look at that shiny, beautiful trombone!” – appears after the instrument is not longer visible. Our approach for this use case is as follows. When an end-user indicates s/he wants to add a comment, the video playback is paused and the input area gains focus (Figure 5.8 (2)). As the interaction is performed right after listening to or watching an event of interest, we assume that the current moment ($t_{\text{now}}$) is the end of the comment ($t_{\text{end}} = t_{\text{now}}$). As an initial guess, we consider that the start time of the comment ($t_{\text{start}}$) is equal to the current time ($t_{\text{now}}$) minus a minimal duration ($\text{MinDur}$) that a comment should stay on screen for being effectively read ($t_{\text{start}} = t_{\text{guess}} = t_{\text{now}} - \text{MinDur}$).
Based on our prediction model and its parameters – e.g., the number of words in a comment \( (N) \), the average duration of a character/phoneme of a word in a specific language \( (\alpha) \) and the average duration of pauses \( (\beta) \) – \( t_{\text{guess}} \) is recalculated, being \( t_{\text{start}} \) then determined by the maximum value among \( t_{\text{guess}} \), the end

\[
\begin{align*}
    t_{\text{start}} &= \max\{0; t_{\text{guess}}, t_{\text{end}}\} \\
    t_{\text{end}} &= t_{\text{now}} \\
    t_{\text{guess}} &= t_{\text{now}} - \max\left\{\frac{\text{MinDur}}{\alpha} \left( \sum_{i=1}^{N} \text{length}(i) \right) \right\}
\end{align*}
\]

Figure 5.9. Predictive timing support for end-user comments.
of the previous existing comment (t_{end}) and zero. Figure 5.9 illustrates scenarios in which t_{start} assumes different values. In the example of Figure 5.8, the start and end time computed for “Look at that shiny, beautiful trombone!” stayed around 3min57s. When the user saves the comment, the video playback is resumed. The predictive timing functionality often provides coarse temporal support; users may fine-tune the timing if desired. In our experience, such fine-tuning is not necessary unless tightly coupled subtitles are being created.

5.4 Evaluation

As mentioned before, the survey discussed in Section 5.1 was followed by 2 other experiments. First, participants were instructed to interact with the prototype system presented in Section 5.3 and then, fill in a questionnaire to report their experiences. In the second and last part, they were asked to further explore the commenting activity by close captioning a sample video (approx. 7 minutes duration) and fill in another questionnaire. Table 5.1 summarizes the number of participants involved in each part of the evaluation process presented in this chapter. In the next sections we present the results and discuss the findings from the evaluation of our online video commenting tool.

<table>
<thead>
<tr>
<th>Evaluation Questionnaire</th>
<th>Number of Respondents</th>
<th>% of Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Current video watching and commenting practices</td>
<td>21</td>
<td>100.0%</td>
</tr>
<tr>
<td>2. Commenting on videos with our prototype system</td>
<td>18</td>
<td>85.7%</td>
</tr>
<tr>
<td>3. Captioning videos using our prototype system</td>
<td>12</td>
<td>57.1%</td>
</tr>
</tbody>
</table>
5.4.1 Commenting on Videos

In general, participants’ feedback regarding our video commenting tool was very positive (see Figure 5.10). When asked how much they liked the service (Q2.1), 13 out of 18 answered some or a lot. All respondents reported that our video commenting tool is helpful for adding synchronized comments to YouTube videos (Q2.2). Some expressed such appreciation by saying that “synchronization is much better” and “I can easily add comments to specific moments in the video. In Facebook I think I can’t. In YouTube I can but I have to type the time moment in the comment”.

When compared to regular comment threads in YouTube or Facebook, 9 users said our tool is better or much better (Q2.3). A user justified his/her answer by saying that “the possibility to comment on a specific moment in the video adds a lot of functionality. Instead of saying, ‘after 16 seconds he does this’, you can just comment at that moment. This also works quite well on SoundCloud as far as I have seen”. On the other hand, 5 participants said they were unable to judge. One of them explained: “I have never added comments to Facebook nor YouTube. However, the way to add comments in this (video commenting) tool is intuitive”.

5.4.2 Close Captioning Videos

The last experiment was the most time consuming one, and for this reason, only 12 participants committed to complete it. Users were kindly requested to close caption a 7 minutes speech video. This task was first performed using our video commenting tool, and later using a standard video player and a text editor. This experiment allowed us to evaluate the effectiveness and usefulness of the time prediction algorithm provided in our commenting tool.

Using our tool, participants spent in average 61 minutes (Standard Deviation: SD = 48 minutes), and other 101 minutes (SD = 33 minutes) without it. The utility of our commenting system has also been reflected in the answers to the questionnaire (see Figure 5.11). When asked how much easier it was to add close captions with our system compared to the other method, all respondents said it was much easier or easier (Q3.1). A similar feedback has been obtained in the question regarding participants’ appreciation for the predictive synchronization of captions/comments (Q3.2). In this case, 7 users reported to have liked a lot. In one case, one participant mentioned that “most captions were synchronized nice to the video, and the prediction algorithm does work. It saves a lot of time having not to
fine tune the start and end points, as you have to do with the SRT format”. And another user added: “the prediction works really good, the captions are usually where they are supposed to be!”.  

Figure 5.10. Results from the evaluation of our commenting tool.
Although our primary objective was not to provide a close caption authoring tool, the point we want to make here is that video commenting systems like ours should not only allow users to add timed comments, but also help them by offering automatic processes that make the commenting task simpler and more intuitive.

Figure 5.11. Results from the close captioning activity using our prototype system.
Chapter 5. Supporting Personalized End-User Comments within Third-Party Online Videos

5.5 Discussion

In this chapter we presented our efforts in supporting personalized content enrichment. Motivated by a survey on media watching and commenting practices, we introduced and evaluated a video commenting paradigm that follows the intimacy and reciprocity guidelines introduced in Chapter 2. Results from the evaluation process show that that users appreciated the functionalities of our system and would potentially use it to communicate with their close circle (requirements on intimacy and reciprocity) and, also, with the general public.

The survey research about media watching and commenting practices represents the first major contribution of this chapter. While this study is relevant to analyze user behavior; it is even more to motivate our work. Do people want to add timed comments within videos? Our results provide evidences that regular Internet users would add synchronized comments while consuming video on demand if they had the appropriate tools for doing that.

From a document model perspective, all the requirements presented in Section 5.1 are met by using a structured multimedia language like SMIL. In this work we focused on text, but a similar approach could be used for other types of user-generated enrichments [56][61]. The video commenting tool reported in this chapter also addresses the functional requirements. The transformation process starts when a video URL is given as an input. Next, our video commenting system applies a document model transformation, which respects the owners’ rights by retaining the video integrity (requirement i) and allow compilations that include video clips from multiple sources (requirement iii). Timed text content is applied as soon a user clicks the input area (requirement iv). This means that given a multimedia document, our tool adds a parallel container that synchronizes comments with a particular video. Whenever a new comment entry is inserted, implicit metadata is automatically added (requirement v). As these comments can be targeted to different audiences, they can be selectively rendered (requirement vi). Multiple-video aggregation and timed end-user navigation (requirements ii and vii, respectively) can be met by integrating the personalized narratives presented in the previous chapter.

The evaluation of our video commenting system represents the second major contribution of this chapter. It shows that this paradigm brings a measurable increment over existing commenting systems. It also shows that the burden of synchronizing comments can be minimized by the use of predictive timing. These results answer our research question. Finally, we do not claim synchronized
comments should replace traditional ones, but rather be complementary. Regular comments are targeted to a fundamentally different use case than the ones offered by our system. On the one hand, in Facebook or YouTube, people can comment about a video, but also give feedback to the author or start a conversation about something unrelated. On the other hand, our video commenting system can be used to highlight interesting things for other viewers, maybe to make a point about a particular event within the video. Such textual comments should be preferably simple; otherwise viewers will have problems to read while watching a video.