TECHNICAL REPORT

ISO/IEC TR 15938-8

First edition
2002-##-##

Information technology — Multimedia content description interface —

Part 8: Extraction and use of MPEG-7 descriptions

Technologies de l'information — Interface de description du contenu multimédia —

Partie 8: Extraction et utilisation des descriptions MPEG-7

PROOF/ÉPREUVE

Reference number
ISO/IEC TR 15938-8:2002(E)

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

```
   xpath="CameraMotion/CameraMotionSegment/FractionalPresence@ZOOM_IN_F"/>
      </OrderingKey>
      <MultimediaContent xsi:type="VideoType">
         <Video id="id1">
            <MediaTime>
               <MediaTimePoint>T00:00:00</MediaTimePoint>
               <MediaDuration>PT0M15S</MediaDuration>
            </MediaTime>
            <TemporalDecomposition>
               <VideoSegment id="id2">
                  <MediaTime>
                     <MediaTimePoint>T00:00:00</MediaTimePoint>
                     <MediaDuration>PT0M15S</MediaDuration>
                  </MediaTime>
                  <!-- CameraMotion descriptor value omitted -->
               </VideoSegment>
               <VideoSegment id="id3">
                  <MediaTime>
                     <MediaTimePoint>T00:00:10</MediaTimePoint>
                     <MediaDuration>PT0M10S</MediaDuration>
                  </MediaTime>
                  <!-- CameraMotion descriptor value omitted -->
               </VideoSegment>
               <VideoSegment id="id4">
                  <MediaTime>
                     <MediaTimePoint>T00:00:20</MediaTimePoint>
                     <MediaDuration>PT0M10S</MediaDuration>
                  </MediaTime>
                  <!-- CameraMotion descriptor value omitted -->
               </VideoSegment>
            </TemporalDecomposition>
         </Video>
      </MultimediaContent>
   </Description>
</Mpeg7>
```

3.5.9 Affective description

3.5.9.1 Affective DS

3.5.9.1.1 Affective DS examples

The following example demonstrates the use of the Affective DS to describe the story shape of a movie. The story shape represents the development of a story along time by assigning a score to each scene, which measures the degree of "story complication" in that scene. The peaks in the score usually correspond to climaxes in the story. Figure 3 shows an example of the story shape characterized for a certain action movie.

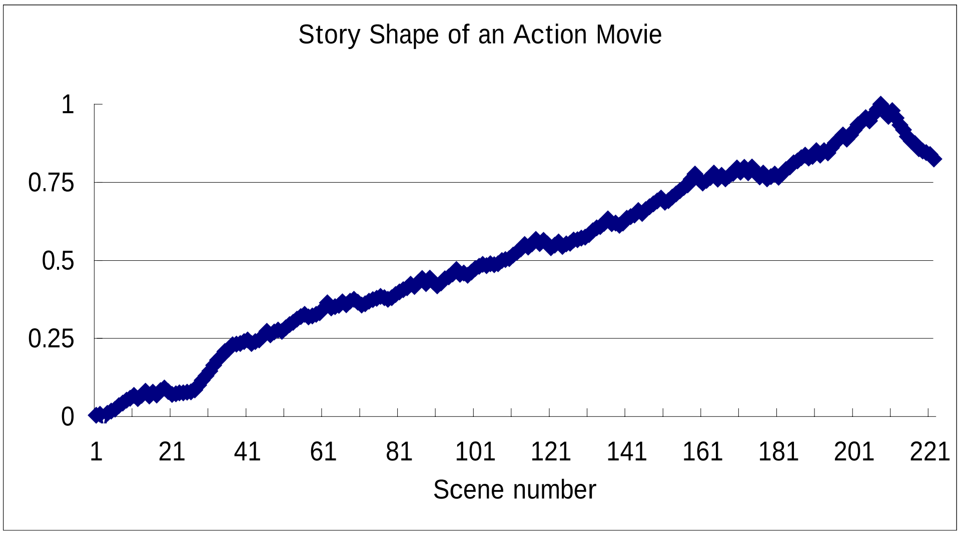


Figure 3 - Illustration of the story shape of an action movie.

The description of the story shape shown in Figure 3 is given as follows:

```
<Mpeg7>
    <Description xsi:type="ContentEntityType">
        <!--
            Story Shape description by Affective DS.
            Note: Story shape is defined in the informative classification scheme
            identified by "AffectTypeCS".
        -->
        <Affective id="affective1">
            <Header xsi:type="DescriptionMetadataType">
                <Confidence>0.85</Confidence>
                <Version>1.0</Version>
                <Comment>
                    <FreeTextAnnotation>
                        Confidence is measured using a certain statistical
technique.
                        Noticeable experimental conditions are used to describe
                        the creation information.
                    </FreeTextAnnotation>
                </Comment>
                <PrivateIdentifier>example1</PrivateIdentifier>
                <Creator>
                    <Role href="urn:mpeg:cs:AffectBasedContentAnalysisCS:2001">
                        <Name>experimentalSubject</Name>
                    </Role>
                    <Agent xsi:type="OrganizationType">
                        <Name>N.Univ.</Name>
                    </Agent>
                </Creator>
                <Creator>
                    <Role href="urn:mpeg:cs:AffectBasedContentAnalysisCS:2001">
                        <Name>organizer</Name>
                    </Role>
                    <Agent xsi:type="PersonType">
                        <Name>
                            <GivenName>Yoshiaki</GivenName>
                            <FamilyName>Shibata</FamilyName>
                        </Name>
                    </Agent>
                </Creator>
                <CreationLocation>
                    <Region>jp</Region>
```

```
               <AdministrativeUnit>Tokyo</AdministrativeUnit>
            </CreationLocation>
            <CreationTime>1999-09-12T13:45:00+09:00</CreationTime>
            <Instrument>
               <Tool>
                  <Name>SemanticScoreMethod</Name>
               </Tool>
            </Instrument>
         </Header>
         <Type href="urn:mpeg:cs:AffectTypeCS:2001">
            <Name>storyShape</Name>
         </Type>
         <Score idref="scene1">0.00360117</Score>
         <Score idref="scene2">0.00720234</Score>
         <!-- : -->
         <Score idref="scene222">0.825006</Score>
      </Affective>
      <!--
         Description of AudioVisualSegment Decomposition
      -->
      <MultimediaContent xsi:type="AudioVisualType">
         <AudioVisual id="program" mediaTimeBase="./MediaLocator"
            mediaTimeUnit="PT1N30F">
            <MediaLocator>

   <MediaUri>http://www.mpeg.org/contents/TheMaskOfZorro.mpg</MediaUri>
            </MediaLocator>
            <!-- 0.0 .. 7669.5 sec for "program" -->
            <MediaTime>
               <MediaRelTimePoint>PT0S</MediaRelTimePoint>
               <!--   <MediaRelTimePoint>PT0S</MediaRelTimePoint> -->
               <MediaDuration>PT2H7M49S5N10F</MediaDuration>
            </MediaTime>
            <!-- Decomposition into scenes -->
            <TemporalDecomposition>
               <AudioVisualSegment id="scene1">
                  <!-- 0.0 .. 48.9 sec for "scene1" -->
                  <MediaTime>
                     <MediaRelTimePoint>PT0S</MediaRelTimePoint>
                     <MediaDuration>PT48S9N10F</MediaDuration>
                  </MediaTime>
               </AudioVisualSegment>
               <AudioVisualSegment id="scene2">
                  <!-- 48.9 .. 81.0 sec for "scene2" -->
                  <MediaTime>
                     <MediaRelTimePoint>PT48S9N10F</MediaRelTimePoint>
                     <MediaDuration>PT32S1N10F</MediaDuration>
                  </MediaTime>
               </AudioVisualSegment>
               <!-- : -->
               <AudioVisualSegment id="scene222">
                  <!-- 7639.8 .. 7669.5 sec for "scene222" -->
                  <MediaTime>
                     <MediaRelTimePoint>PT2H7M19S8N10F</MediaRelTimePoint>
                     <MediaDuration>PT29S3501N5000F</MediaDuration>
                  </MediaTime>
               </AudioVisualSegment>
            </TemporalDecomposition>
         </AudioVisual>
      </MultimediaContent>
   </Description>
</Mpeg7>
```

This description consists of two parts:

1. The instance of `Affective` DS with the `DescriptionMetadata` DS as a header, where each affective value is associated with a sub-segment (scene) by referencing its identifier ("`id`") using the `idref` attribute.
2. The hierarchical decomposition of the movie. The AV segment representing the entire movie ("program") is decomposed into 222 sub-segments called "scenes".

The "storyShape" is specified in the `Type` element by referencing a term defined in the informative Affective Type Classification Scheme (See Annex A). The information included in the `DescriptionMetadata` header indicates that the scores were obtained from students in N Univ. based on the Semantic Score Method. Note that the value of the `Confidence` element in the Header is 0.85, indicating that the score was obtained using a method that can provide a reliability for the score. In this case, the score is actually a composite score synthesized from the individual score of many individual students using statistical techniques.

Another example of using the `Affective` DS is to describe the affective response to the events and objects depicted in a soccer game video. Suppose that the video contains a goal event, where a goalkeeper "A" fails to capture a slow ball kicked by a forward "B", resulting in losing a point for "A's" team. The objects, the goalkeeper "A" and the forward "B", may be described using the `SemanticPerson` DS. Because a fan of "B's" team may feel angrier towards the goalkeeper, who fails to catch such an easy ball, than towards the forward who kicks the ball, his/her affective response to the object descriptions can be represented as:

```
<Affective>
   <Type href="AffectTypeCS:2001">
      <Name> anger </Name>
   </Type>
   <Score idref="Goalkeeper-id">0.8</Score>
   <Score idref="Forward-id">0.2</Score>
</Affective>
```

In this description, "`Goalkeeper-id`" and "`Forward-id`" reference the goalkeeper and the forward descriptions respectively, which are assumed to have been described elsewhere.

3.5.9.1.2 Affective DS extraction

The `Affective DS` does not impose any extraction method of the score. In other words, the `Affective DS` can be used to describe any kind of audience's affective information on multimedia content as long as it is represented as a set of scores of relative intensity. For readers' convenience, however, a couple of extraction methods are introduced in this subclause.

3.5.9.1.2.1 Semantic Score Method

The Semantic Score Method [Takahashi00-1] is a well-defined subjective evaluation of video based on Freytag's theory [Freitag98, Laurel93]. This method can be used to extract the story shape of video content.

3.5.9.1.2.1.1 Freytag's Triangle

Gustav Freytag, a German critic and playwright, suggested in 1863 that the action of a play could be represented graphically when the pattern of emotional tension created in its audience was evaluated. Since tension typically rises during the course of a play until the climax of the action and falls thereafter, he modeled this pattern in a triangle form, referred to as "Freytag's Triangle" shown in Figure 4. In this figure, the left side of the triangle indicates the rising action that leads up to a climax or turning point while the right side of the triangle is the falling action that happens from the climax to the conclusion. The horizontal axis of the graph is time; the vertical axis is *complication*. According to Brenda Laurel [Laurel93], the complication axis of the Freytag's triangle represents the *informational attributes* of each dramatic incident. An incident that raises questions is part of the rising action, which increases *complication*, while one that answers questions is part of falling action, resulting in decreasing the complication, i.e., *resolution*.

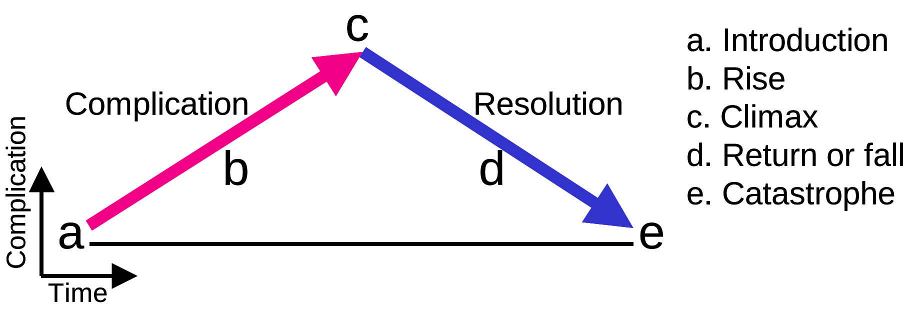


Figure 4 - Freytag's triangle [Laurel93].

In reality, however, things are more complicated than in Freytag's idealized model. One dramatic incident may raise some questions and answer others simultaneously. Hence, the *degree of complication* is introduced to specify the net increase of complication caused by an incident. The degree of complication is represented by a positive value when complication caused by an incident's raising questions overwhelms resolution by answering questions and a negative value vice versa. The *cumulative complication* for an incident is defined as the cumulative sum of the degree of complication for all incidents preceding this incident, representing the net intensity of complication for this incident. Note that because of a fractal-like property of a play where whole story is composed of several sub-stories that can be further divided into various dramatic incidents, the shape of a practical play is characterized in more irregular and jagged form than shown in Figure 4.

In order to help readers' understanding, an example from [Laurel93] is reproduced. Assume the following background situation: a group of strangers have been invited by an anonymous person to spend the weekend in a remote mansion. During the night, one member of the group (Brown) has disappeared. Some of the remaining characters are gathered in the drawing room expressing concern and alarm. The butler (James) enters and announces that Brown has been found. The following are conversations made among those people.

> **James:** I'm afraid I have some rather shocking news.
>
> **Smith:** Spit it out, man.
>
> **Nancy:** Yes, can't you see my nerves are absolutely shot? If you have any information at all, you must give it to us at once.
>
> **James:** It's about Mr. Brown.
>
> **Smith:** Well?
>
> **James:** We've just found him on the beach.
>
> **Smith:** Thank heavens. Then he's all right.
>
> **James:** I'm afraid not, sir.
>
> **Smith:** What's that?
>
> **James:** Actually, he's quite dead, sir.
>
> **Nancy:** Good God! What happened?
>
> **James:** He appears to have drowned.
>
> **Smith:** That's absurd, man. Brown was a first-class swimmer.

The informational components raised in the above dialog are summarized as:

> James has shocking news.
>
> The news concerns Brown.
>
> Brown has been found.
>
> Brown is dead.
>
> Brown has drowned.
>
> Brown was a good swimmer.

Then, each component is evaluated based on the degree of complication (between 0 and +/-1). Possible scoring result is shown in Table 7.

Table 7 - Complication/Resolution based evaluation

Informational Component	Degree of Complication	Cumulative Complication
a. James has shocking news.	+0.4	0.4
b. The news concerns Brown.	+0.5	0.9
c. Brown has been found.	-0.7	0.2
d. Brown is dead.	+0.9	1.1
e. Brown has drowned.	-0.4	0.7
f. Brown was a good swimmer.	+0.8	1.5

In this table, the component **c** and **e** are evaluated as *negative* complication (resolution). The former provides an answer to the puzzle that "Brown had disappeared", while the latter gives an answer to the question that "how Brown died" raised in the component **d**. The third column in the table denotes the cumulative sum of the degree of complication from the component **a**. Assume that each component in the table is a dramatic incident occurring sequentially. Then, since the degree of complication evaluated at each incident indicates the increase of complication at each incident, the cumulative complication in the table reflects the net complication at each moment resulting from preceding incidents since the initial one. The cumulative complication is then used to visualize the story shape for the dialog as shown in Figure 5.

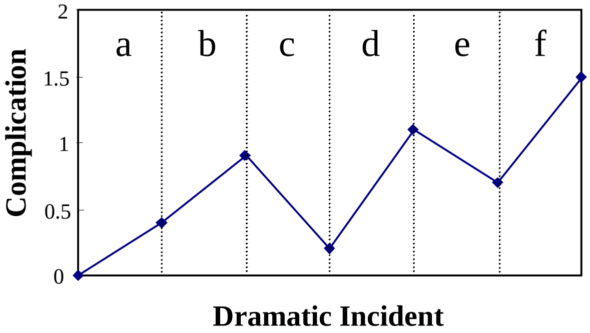


Figure 5 - The story shape for the dialog example.

3.5.9.1.2.1.2 Semantic Score Method

Based on the Freytag's play analysis, a subjective evaluation method for storied video, called "Semantic Score Method", is proposed [Takahashi00-1]. According to Brenda Laurel [Laurel93], an implicit assumption was made in the Freytag analysis that there is a direct relationship between what we *know* about the dramatic incident and how we *feel* about it. The method, however, mainly focuses on the former aspect, i.e., the method is developed as an analytical tool for a subjective video evaluation. In short, the evaluators are asked to give a positive (negative) value to each pre-determined scene according to the degree of complication (resolution) they perceived. The evaluators are expected to *interpret* what happens in the scene and *analyze* dramatic incidents involved in the scene in order to characterize the scene with a single value (called Semantic Score) from the complication/resolution viewpoint.

In order to obtain reliable data from general audiences, it is useful to provide the following items as supporting tools of the method [Takahashi01]:

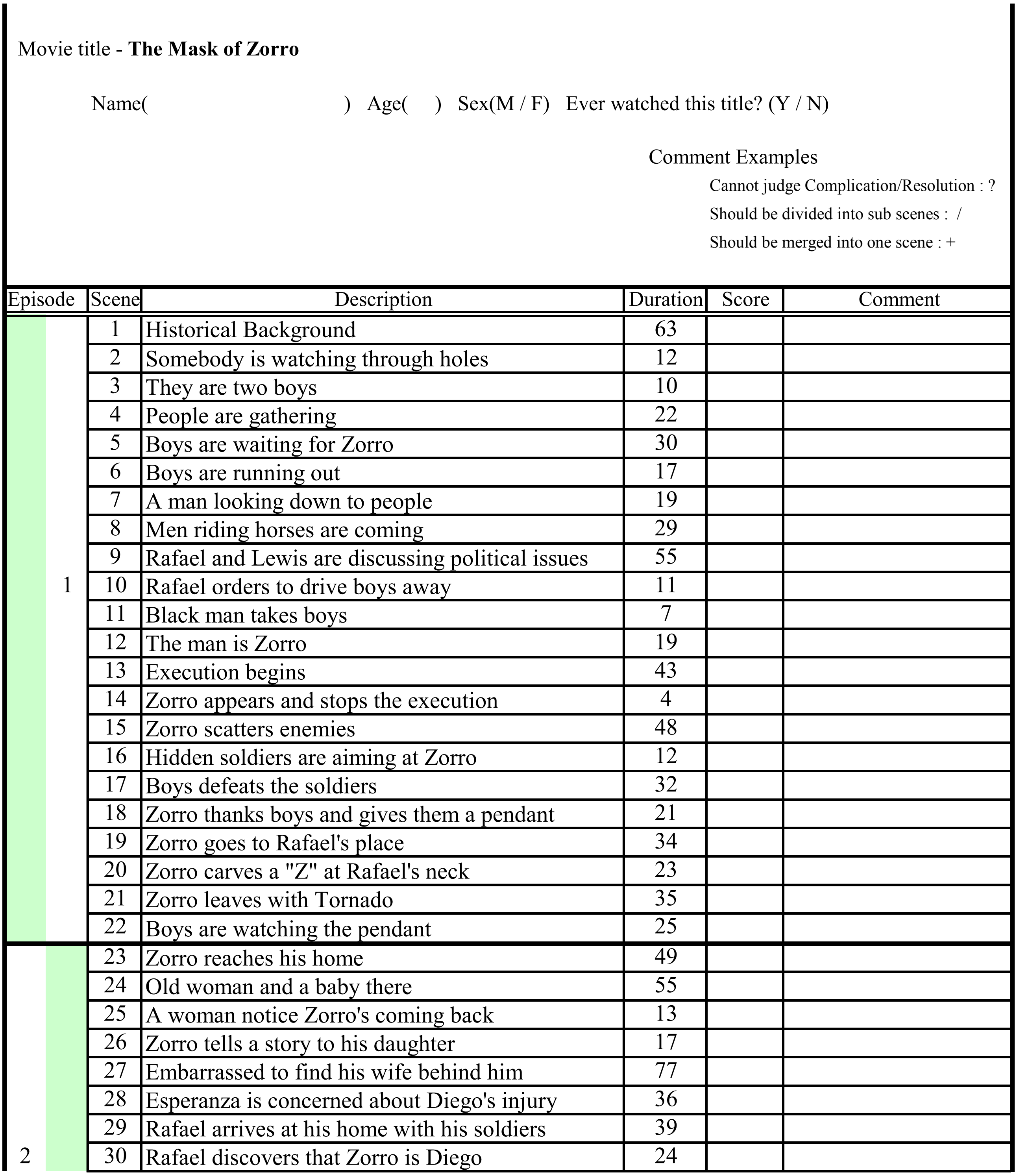
- Instruction video and booklet
- Test material (target movie)
- Specially designed score sheet

The instruction video and booklet are used to explain the purpose of the evaluation, the evaluation procedure and the evaluation criterion (the complication and the resolution). The instruction video can also include a concrete evaluation example: A demonstration of the scene scoring using a short storied video provides evaluators with a common yardstick with which how a certain scene is to be scored. The score is typically assigned within a range between –5 and 5 by steps of one. If necessary, however, it should be allowed to score a scene with a fractional value or beyond the range as well.

The test material is a movie whose story shape is to be characterized. Since evaluators are asked to score scenes one by one, the video should be modified from its original form. For example, by marking the end of each scene with the final frame as a still (the scene number superimposed) for a few seconds, evaluators can recognize each scene easily, resulting in smooth evaluation of the scenes.

One of the issues in the method is the scene definition. A scene is typically defined as a video segment that has minimum semantics as a story component with monotonic complication or resolution. The boundary between scenes is identified when a situation is drastically changed. Here, the situation includes time, place, character, context (e.g. in dialog), particular dramatic incident, and so on. This implies that one scene may be composed of several shots or that a long shot may be divided into several scenes. For example, when a long video shot has both the question raising and answering sequentially, the shot should be divided into two concatenated scenes. Based on the scene definition, one movie is typically divided into 100 - 250 scenes, resulting in each scene lasting 30 – 60 seconds. The scene length depends on its genre: an action type movie tends to have shorter scenes while the one regarded as a love story tends to have longer scenes than other genres.

In addition, a specially designed score sheet is useful to record the scores the evaluators assigns. Figure 6 shows a part of an example of the score sheet. In this sheet, each row corresponds to a scene, which is composed of the scene number, a short scene description, duration of the scene, and a cell to be filled with the complication value. Supplemental information is provided for the sake of evaluators' convenience, i.e., evaluators can easily recognize where they are evaluating at any moment. In addition, several consecutive scenes are grouped to form an episode, the second level story component that can be identified without ambiguity. In this score sheet, the boundary of the episode is represented with a thick solid line. Although the yardstick evaluators keep in mind may vary during the evaluation, a request to keep the consistency of scoring within whole video is often hard to achieve. It is therefore practical to ask evaluators to at least keep the consistency within the episode.



Movie title - **The Mask of Zorro**

Name(   ) Age(   ) Sex(M / F) Ever watched this title? (Y / N)

Comment Examples

Cannot judge Complication/Resolution : ?

Should be divided into sub scenes : /

Should be merged into one scene : +

Episode	Scene	Description	Duration	Score	Comment
1	1	Historical Background	63		
	2	Somebody is watching through holes	12		
	3	They are two boys	10		
	4	People are gathering	22		
	5	Boys are waiting for Zorro	30		
	6	Boys are running out	17		
	7	A man looking down to people	19		
	8	Men riding horses are coming	29		
	9	Rafael and Lewis are discussing political issues	55		
	10	Rafael orders to drive boys away	11		
	11	Black man takes boys	7		
	12	The man is Zorro	19		
	13	Execution begins	43		
	14	Zorro appears and stops the execution	4		
	15	Zorro scatters enemies	48		
	16	Hidden soldiers are aiming at Zorro	12		
	17	Boys defeats the soldiers	32		
	18	Zorro thanks boys and gives them a pendant	21		
	19	Zorro goes to Rafael's place	34		
	20	Zorro carves a "Z" at Rafael's neck	23		
	21	Zorro leaves with Tornado	35		
	22	Boys are watching the pendant	25		
2	23	Zorro reaches his home	49		
	24	Old woman and a baby there	55		
	25	A woman notice Zorro's coming back	13		
	26	Zorro tells a story to his daughter	17		
	27	Embarrassed to find his wife behind him	77		
	28	Esperanza is concerned about Diego's injury	36		
	29	Rafael arrives at his home with his soldiers	39		
	30	Rafael discovers that Zorro is Diego	24		

Figure 6 - Score Sheet for the Semantic Score Method.

3.5.9.1.2.1.3 Evaluation Procedure on Semantic Score Method

Using the supporting tools introduced above, typical evaluation procedure based on the Semantic Score Method can be described as follows:

- Instruct evaluators using the instruction video and booklet
- Ask evaluators to watch the designated title in a normal fashion.
- Ask evaluators to re-watch the test material and to evaluate it using the score sheet.
- Ask evaluators to answer some questionnaires and interview.

It should be noted that it is useful to ask evaluators to watch the assigned title before actual evaluation (Step 2). The reason for this is to let evaluators know the content in advance so that evaluators can evaluate the title calmly. What the method aimed at is not an identification of exciting part of a video where evaluators might lose themselves but to characterize how a story develops. Thus, excitation caused by unexpected development of a story and/or audiovisual effects should be carefully controlled. In other words, if evaluators were really excited with watching the title, they might even forget the evaluation itself.

Figure 7 shows a typical example of the story shape for four evaluators based on the Semantic Score Method. The story shape, called “Semantic Graph” in this method, can be obtained by integrating the complication/resolution value (given by evaluators) with respect to the story timeline.

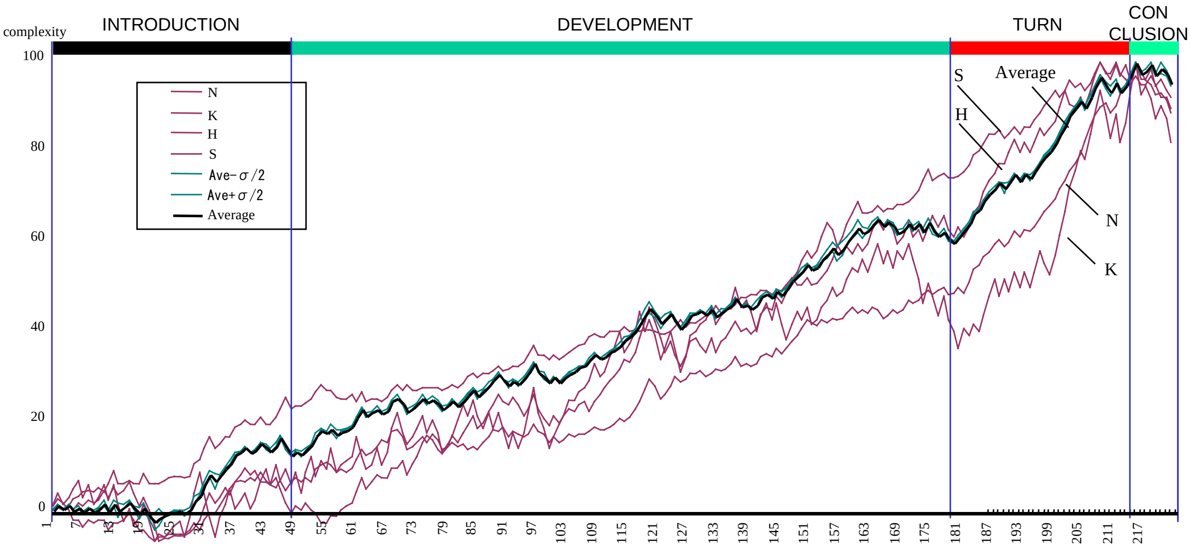


Figure 7 - Semantic Graph of “THE MASK OF ZORRO.”

In Figure 7, the vertical axis denotes the accumulated complication while the horizontal axis is the scene number (instead of time stamp of the whole movie). All data are normalized with respect to their maximum peak value so that direct comparison among them is available. It is also noted that a whole story is divided into four regions based on a conventional Japanese story model known as Ki-Sho-Ten-Ketsu, where Ki, Sho, Ten and Ketsu correspond to Introduction, Development, Turn, and Conclusion, respectively.

The thick line in the graph of Figure 7 is obtained by combining the four Semantic Graphs. Note that a simple averaging operation does not work well in this analysis because there are cases where a scene is scored with both high positive and negative values. Although the case clearly suggests that evaluators recognize something in the scene, a simple averaging operation may diminish the information. For the thick line in Figure 7, a special averaging is used [Takahashi01], where the magnitude and the sign of the combined score are determined by averaging the absolute value of the original scores and by taking the majority decision of them, respectively. With this averaging technique, the dulling of the graph shape can be successfully avoided.

3.5.9.1.2.2 Physiological Measurements and Analysis

Other possible methods to extract the score for the Affective DS instantiation include the physiological measurements and analysis.

Recent developments of sensor technology and brain science have shown the potential to reveal human emotional and/or mental states through physiological measurement and analysis. This suggests that the physiological approach can be a promising tool for multimedia content evaluation, which provides valuable information that is hard to detect through other approaches. Specifically, the measurement and analysis of physiological responses from audience watching multimedia content will characterize the content from the viewpoint on how audience feels interested and/or excited. Furthermore, since some response may reflect specific emotions such as happiness, anger, sadness, and so on, it can also describe how the audience’s emotion changes during his/her watching the content.

Comparing the evaluation method described in the last subclause, the physiological approach has the following advantages:

Can evaluate a content in real-time,

Can evaluate a content in automatic way with an appropriate apparatus, Can obtain information of high-resolution in time for a content such as video and audio.

Furthermore, it is possible to obtain a response that is not consciously influenced by the audience.

In the following, a couple of trials on the movie evaluation using the physiological measurements and analysis are introduced.

Figure 8 shows time dependent Electromyogram (EMG) signals obtained from three audiences for a certain scene in “THE MASK OF ZORRO”. The EMG signal is the electrical signal of muscles recorded by an electromyograph. In this measurement, electrodes are placed on the forehead of audiences, and the voltage difference between the electrodes is continuously recorded. Hence, when a muscle gives a particular movement, then the movement is indirectly detected as a change of the electric signal.

As is seen in Figure 8, there is a spike in the EMG signals and, more notably the spikes in the EMG activities from three audiences coincide. In fact, this is the moment when all audiences smile at the bang sound in the movie. In order to explain what happens in the scene, two images are extracted before and after the bang sound and shown under the graph. At the image before the bang sound, Zorro attempted to jump off a wall onto his waiting horse (see left image). But just before he lands, the horse moves forward and Zorro ends up on the ground. The bang sound occurs at this moment. Since he was supposed to mount the horse successfully, he is embarrassed after the bang sound (see right image). Audiences also expected that Zorro could mount the horse smoothly, the unexpected happening leads them to smile.

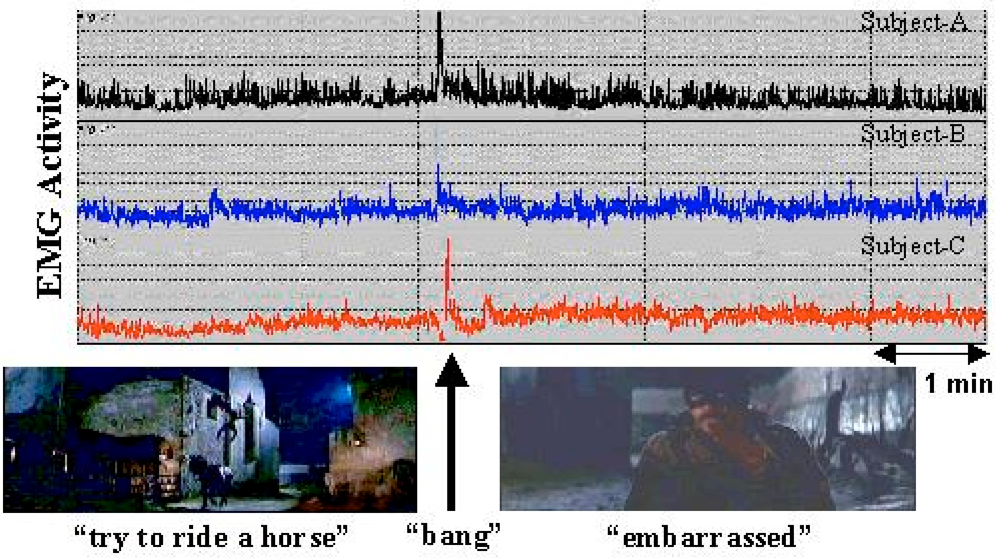


Figure 8 - Spikes in Electromyogram (EMG) caused by smiling in “THE MASK OF ZORRO.”[1]

As is demonstrated, the EMG measurement can be used to detect smiles of audiences. Strictly speaking, what is detected is a particular muscle movement at the forehead, and it could happen not only for smile but also for other emotions. Therefore, electromyography is a promising tool to capture some emotions of human being through his/her muscle activity.

Another example shown in Figure 9 concerns the highlight scene detection through the analysis of non-blinking periods. Video image of audience’s eye is captured and analyzed using image-processing technology to extract the eye-blinking points in time. Then non-blinking periods are measured as a time difference between two eye-blinking points. Figure 9 shows non-blinking periods along the entire movie. In the graph of Figure 9, the horizontal axis denotes time over whole the movie while the vertical axis is non-blinking period in second. This graph is created as follows: when a non-blinking period is given, then a regular square whose height (the vertical coordinate) is the same as the period is aligned on the horizontal period. Hence, the length of non-blinking periods can be easily seen from the graph.

According to the graph, it is observed that there are several long periods of non-blinking. The notable point again is that these long non-blinking periods correspond well to the highlight scenes in the movie. Here, the highlight scene is defined as the one audience pays special attention to. In order to show what happened at each highlight scene, an image is extracted from each highlight scene and shown around the graph with an

[1] "The Mask of Zorro".

© 1998 Global Entertainment Productions GmbH & Co. KG. Courtesy of TriStar Pictures. All Rights Reserved.

arrow pointing to the corresponding non-blinking period. Simple text annotation was also attached to each frame to describe the scene.

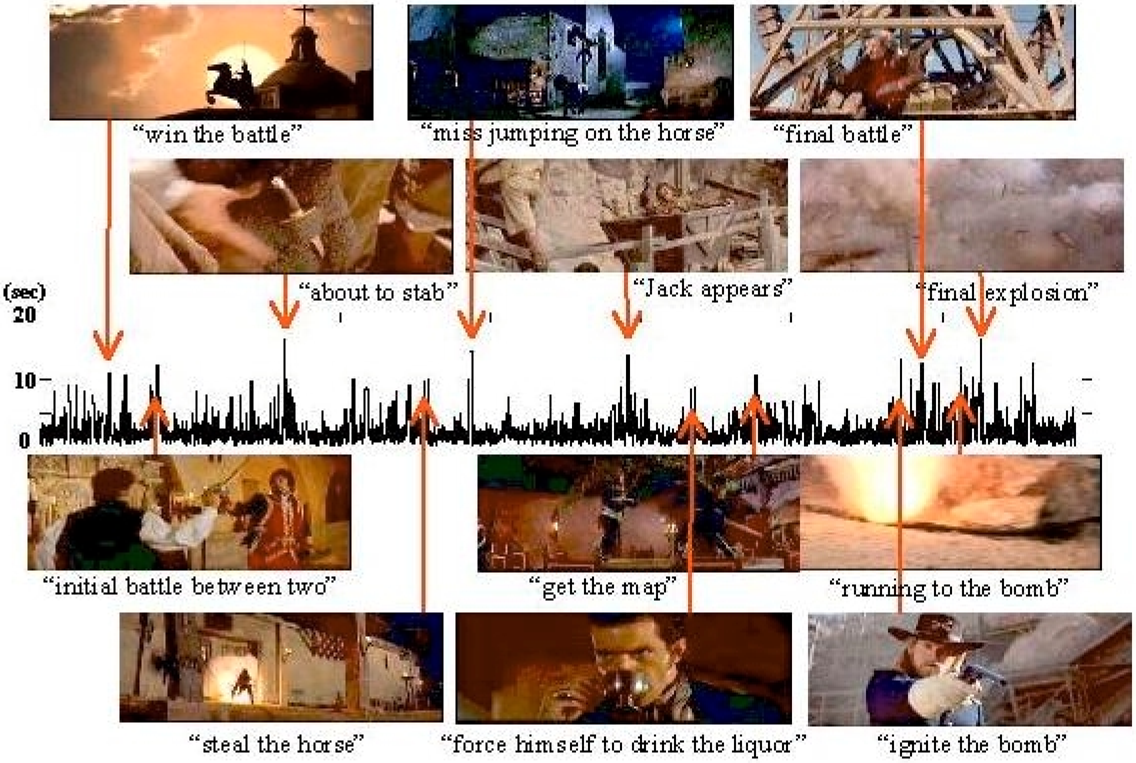


Figure 9 - Highlight scenes detected by non-blinking periods in "THE MASK OF ZORRO."[1]

Figure 9 clearly indicates that the detection of non-blinking period can be a tool to identify the highlight scene in a movie. This is qualitatively explained by the fact that, as a natural property of human being, we tend to open our eyes wide without blinking when we watch something that attracts our attention.

3.5.9.1.3 Affective DS use

The description using the `Affective DS` provides high-level information on audience's interpretation as well as perception on multimedia content and therefore can be used in various ways. One of the examples is given as a preprocessing for video summary: In the case of story shape for example, one can obtain a video summary that reflects the story development. Furthermore, the highlight video summary can be obtained by selectively concatenating high score video segment notably when the Type element takes a value of, e.g., "excited". The description can also be used as fundamental data in high-level multimedia content analysis. For example, since the patterns of the story shape strongly depends on the genre of audiovisual content, it may be used to classify the content into its genre [Takahashi00-2].

In the following, the use of the story shape to analyze a trailer creation [Takahashi00-3] is demonstrated.

A trailer is a short movie clip consisting of small pieces of video segment mainly taken from an original movie. It is used to advertise a new movie and therefore a trailer often includes a video segment, telop, narration, and so on, that does not appear in the original move in order to enhance its effectiveness. Strictly speaking, a trailer is not a so-called video summary: it is rare that we can grasp the outline of a movie by just watching its trailer, but should be attractive enough to make many people feel like to watch the movie. Although the trailer creation itself is a highly refined artistic work, it is interesting to investigate how a skilled and talented creator creates an attractive trailer from the viewpoint of video segment selection.

Using the story shape description based on the Semantic Score Method obtained from various movies together with their originally created trailers, the analysis reveals a strategy on which scenes are to be chosen for an attractive trailer. Borrowing the conventional Japanese story model, the scene selection strategy is summarized as follows:

- **Introduction (Ki):**

  Choose both complication and resolution scenes whose absolute Semantic Score are higher than a given threshold,

  Choose scenes at local peaks in the Semantic Graph (story shape) and the following scene,

- **Development (Sho):**

  Choose complication scenes whose Semantic Score are higher than a given threshold,

  Choose scenes at local peaks in the Semantic Graph,

- **Turn (Ten):**

  Same as those in the Development,

- **Conclusion (Ketsu):**

  No scene should be chosen.

In order to simulate a practical trailer creation, further strategy is needed because the scenes used in the Semantic Score Method typically last for thirty to sixty seconds and thus simply concatenating selected scenes gives a long video clip which is too long to be a trailer. A study [Takahashi00-3] reveals the strategy on how to identify a shot within the selected scene. According to the strategy, the following criteria should be taken into consideration:

- Upper body image of main actor/actress
- Whole body image of main actor/actress
- Visual effect (CG, telop, dissolve, etc.)
- Sound effect (climax of BGM, explosion, scream, etc.)
- Speech
- High activity (of visual object and/or camera work)
- Shot length (slow motion more than several seconds)
- Camera zoom-in/out

By assigning an appropriate weighting value to each shot within a scene according to the criteria listed above, a shot candidate for a trailer can be determined at each scene.

According to the evaluation of the *simulated* trailer created based on the strategy introduced above, the resulting trailer can gain more than 60 points compared with an original trailer having 100 points as a basis. Note that the simulated trailer is created only using video segments in the original movie, i.e., none of extra technique such as taking special video segments and/or advertising narration that is not included in the original movie is taken into account. Hence, this result clearly indicates that the strategy introduced above actually reflects a certain essence of an attractive trailer creation. Although it is obvious that a trailer created in such way cannot overwhelm the one created by a skilled and talented creator, it is expected that this sort of approach will bring some reference materials that stimulate his/her creativity.

3.5.10 Phonetic description.

3.5.10.1 PhoneticTranscriptionLexicon header

Information on extraction and use is not provided.

3.6 Media description tools

3.6.1 Introduction

This clause specifies tools for describing the media features of the coded multimedia data. The coded multimedia data may be available in multiple modalities, formats, coding versions, or as multiple instances. The tools defined in this clause allow the description of an original instance of coded multimedia data and the multiple variations of the original instance. The following tools are specified in this clause:

Table 8 - Overview of Media Information Tools.

Tool	*Functionality*
Media Information Tools	Tools for describing the media-specific information of the multimedia data. The `MediaInformation DS` is centered around an identifier for the content entity and a set of coding parameters grouped into profiles. The `MediaIdentification DS` allows the description of the content entity. The different `MediaProfile DS` instances allow the description of the different sets of coding parameters values available for different coding profiles.