Enhancing audiovisual experience with haptic feedback: A survey on HAV(E)

Fabien Danieau, Anatole Lécuyer, Philippe Guillotel, Julien Fleureau, Nicolas Mollet, and Marc Christie.

Abstract—Haptics have been employed in a wide set of applications ranging from teleoperation and medical simulation to arts and design, including entertainment, aircraft simulation and virtual reality. As for today, there is also a growing attention from the research community on how haptic feedback can be integrated with profit to audiovisual systems. The objective of this paper is to present the techniques, formalisms and key results on the enhancement of audiovisual experience with haptic feedback. We first review the three main stages in the pipeline which are (i) production of haptic effects, (ii) distribution of haptic effects and (iii) rendering of haptic effects. We then highlight the strong necessity for evaluation techniques in this context and discuss the key challenges in the field. By building on technology and results from virtual reality, and tackling the specific challenges in the enhancement of audiovisual experience with haptics, we believe the field presents exciting research perspectives for which financial and societal stakes are significant.

Index Terms—haptic interfaces, multimedia, audiovisual, user experience

1 Introduction

In 1962, Heilig [1] introduced *Sensorama*, a system where one can watch a 3D movie, sense vibrations, feel wind and smell odors. This pioneering work opened the path for research in virtual reality, providing high-end interfaces that involve real-time simulations and interactions through multiple sensorial channels [2]. In this context, interactions that rely on the sense of touch (haptics) are central. Use of haptic technologies historically spreads way beyond the only field of virtual reality since numerous applications are found in medical, robotics and artistic fields. However, virtual reality has triggered the development and evaluation of numerous haptic interfaces to deliver and understand complex physical interaction with virtual objects [3] [4].

In contrast, research and technology developments for audiovisual entertainment remain essentially focused on improving image and sound. Although the stakes and potential industrial impact appears to be very significant, haptic feedback in a multimedia context, i.e. haptic feedback in combination with one or several media such as audio, video and text, remains underused. Only a few systems, known as "4D-cinemas", exploit this technology. However the number of contributions investigating the potential of haptic feedback for multimedia increases. In parallel, contributions in virtual reality, such as Reiner [5], have demonstrated that haptic feedback is a key factor in user immersion, a benefit of great interest to applications in entertainment.

Recent works defend this view. O'Modhrain et al. [6] have indeed demonstrated that the benefits of haptic feedback observed in virtual reality are applicable to multimedia applications. In addition, researchers observed that haptic feedback may open new ways to experience audiovisual content: the relation between users and audiovisual contents is not limited to a passive context where the user just listens and watches but could enable physical implication in a more immersive experience [7]. In addition to physically feeling the audiovisual content, the user could expect to receive a complementary piece of information or to feel an emotion through haptic interaction, a step behind immersion. Hence haptics is a complete medium and the combination of haptics and audiovisual content becomes a haptic-audiovisual (HAV [8]) content. As a result, the theme of haptic feedback for enhancing audiovisual experience is not properly part of the virtual reality field, but stands on its own with its specific requirements and scientific challenges.

Therefore many questions are introduced by this young field of study. How to bring haptic technology to the audiovisual field? To what extent can haptics affect the user's perception and understanding of the audiovisual content, and how can haptics be employed efficiently in conjunction with image and sound? What about acceptability of complex haptic interfaces for users? How to evaluate the quality of the user experience? Moreover, to what extent can a same haptic effect be experienced in different viewing scenarios (mobile tv, cinema or user living space, potentially shared) with possibly different devices?

The aim of this survey is to gather and classify the results obtained in this young field of research by identifying the key challenges. We will then build on these challenges to propose future paths for research that address our main questions.

- *F. Danieau is with Technicolor R&I, France and VR4I, INRIA France. e-mail: fabien.danieau@technicolor.com*
- *A. Lécuyer is with VR4I, INRIA, France.*
- *P. Guillotel, J. Fleureau and N. Mollet are with Technicolor R&I, France.*
- *M. Christie is with MimeTIC, IRISA, France.*

The paper is organized as follows. We will first propose and describe a general workflow for adding haptic effects to audiovisual content and build on this workflow to detail its three main stages: (i) production of haptic effects, (ii) distribution of haptic effects and (iii) rendering of haptic effects. We will then emphasize and discusss techniques for evaluating the Quality of Experience of users (QoE) in such systems. Finally we will conclude by discussing the rising research perspectives in the field.

2 A WORKFLOW FOR ADDING HAPTIC FEEDBACK TO AUDIOVISUAL CONTENT

In this survey, we propose to organize scientific and technical contributions along the typical workflow for video-streaming comprising three stages: (i) production, (ii) distribution and (iii) rendering (see Figure 1). We will use the term HAV to refer to haptic audio-visual, the property of a system comprising audiovisual content enhanced with an haptic experience (as introduced by [8]) and we will refer to the term "haptic effect" to designate the use of a haptic feedback in audiovisual content (a generalization of the term employed in the specific context of video viewing [6] [9] [10]).

The first stage in the workflow deals with the **production** of the content, i.e. how haptic effects can be created or generated in synchonization with the audiovisual content. Typically three techniques emerge from the literature: the capture and processing of data acquired from sensors, the automatic extraction from a component of the audiovisual content (image, audio or annotations) and finally the manual authoring of haptic effects. These production techniques and tools will be reviewed in Section 3.

The second stage in the workflow deals with the **distribution** of haptic effects. Given that current technologies allow massive distribution of media over the networks, there is a strong requirement in distributing haptic effects with the content, which in turn raises questions on formalizing haptic effects. Referred to as haptic broadcasting (a term introduced by Cha et al. [11]), the term defines the synchronized transmission of haptic effects over networks for which models, formalizations and techniques are reviewed in Section 4.

Finally the last stage refers to the rendering of the content, more precisely how an encoded haptic effect can be rendered on a specific haptic device and experienced by the user. A range of techniques and devices have been proposed in the literature and the Section 5 offers a broad overview of contributions classified by the type of device (wearable, handheld, desktop or chair).

A key aspect, transversal to production, distribution and rendering, is the evaluation of the user experience. While approaches have strongly focused on the technical aspects of these three stages, there is a clear necessity to measure the quality of audiovisual experiences enhanced with haptics and provide common tools and metrics for such evaluations. We will refer to the term **Quality of Experience** (QoE see [12]) and provide an overview in Section 6.

3 PRODUCTION

Production is the task of creating haptic effects in order to enhance an audiovisual content. Three methods have been reported in the literature: (i) capturing haptic effects from the real world using physical sensors, (ii) generating haptic effects by an automated analysis of audio and/or visual contents, and (iii) manually synthesizing haptic effects from scratch or by editing effects obtained with the previous methods. Before detailing all three methods, this survey proposes a classification of haptic effects based on their perceptual characteristics (tactile, kinesthetic, and proprioception).

3.1 Haptic effects for audiovisual content

A few contributions have previously reported classifications of haptic effects, of which the most exhaustive is proposed by Waltl [13]. The author detailed several sensory effects such as taste, smell and haptic. Haptic effects reported were temperature, wind, whole body vibration, water sprayer, passive kinesthetic motion and force (the user simply holds a force-feedback device), active kinesthetic (the user can explore actively the content thanks to a force-feedback device), tactile and rigid body motion (the whole body of the user is moved as in motion simulators). This classification was built in a way each effect was directly linked to a specific device.

In contrast, the classification we propose is based on haptic perceptual capabilities. In the haptics community, haptic feedback is often separated into two categories: tactile and kinesthetic feedback. The literature reports three types of tactile stimuli : perception of vibrations, of pressure [14] and of temperature [15]. In a similar way, two types of kinesthetic stimuli may be defined [16]: perception of movements (and positions of the users limbs) and the perception of forces. A last type of haptic perception may be the one resulting from the motion of the users own body [17]. Both the vestibular system and the haptic system (movement of limbs and of internal organs, i.e the proprioception) contribute to the perception.

We build upon this classification to propose a table summarizing haptic effects 1 in HAV systems in which each category is mapped to contributions from the literature (each of which will be discussed along this paper). The reader may also refer with profit to the guidelines provided by the haptic community to design vibrotactile effects [18] or haptic feedback in multimodal environments [19]. Obviously, these unitary effects can be combined to create more complex effects (the haptic effect associated to an explosion may be defined through the combination of temperature and vibrations).

Interestingly, haptic effects are mostly used to represent physical events which occur in the scene (cf.

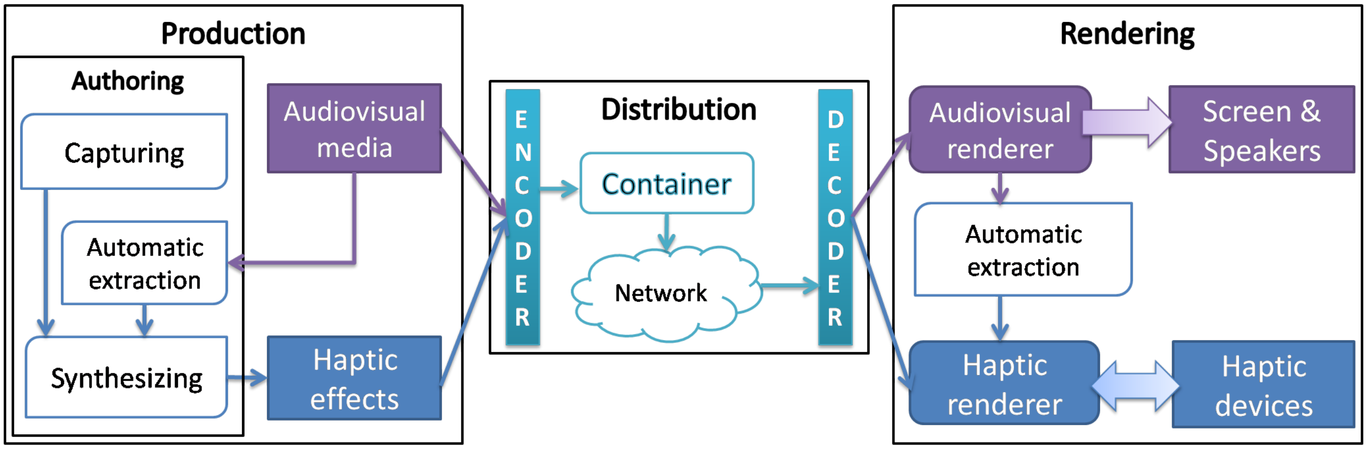

Figure 1. Workflow for adding haptic effects to audiovisual content. In this survey, we propose to consider haptic effects as a component of a multimedia content. Effects are typically produced, distributed and rendered in the user living space in parallel to the audiovisual content.

Table 1). The user perceives stimuli which are directly related to the audiovisual content (e.g. bumps when driving offroad), therefore augmenting the physical event and augmenting the sense of "being physically present". However other aspects of an audiovisual content can be enhanced such as ambiance (see Kim etal. [20]). Actually, one can draw the parallel between the role of haptic effects in audiovisual contents and the one of audio in movies: audio is used for increasing the realism (sound effects) as well as for creating ambiances (music). In movies, a clear separation is drawn between diegetic sounds (a sound for which the source is visible on the screen) and non-diegetic sounds (a sound for which the source is not visible nor implied in the action, typically such as a narrator comment or mood music). Therefore, there is a strong potential in non-diagetic haptic effects to augment non-visual content by contributing to enhancing information perceived by the user, ambiance or emotion.

Enhancing information, ambiance or emotion with haptic effects is not straightforward. With regard to emotional aspects of a video, the haptic effect designer may explore results from research on affective haptics: recent works attempt to communicate affect with haptic feedback [21] while others trigger user's emotion with the help of haptic patterns [22] [23].

Haptic Perception	Haptic Effect	Ref.
Tactile	Temperature	[24] [25]
	Vibration	[26] [27] [23] [13] [28] [20] [25]
	Pressure	[29] [25] [30] [31]
Kinesthetic	Movement	[32] [33]
	Force	[34] [9] [10] [33] [35]
Proprioception	Body Motion	D-Box[9] Mediamation[10]

Table 1
List of potential Haptic Effects for audiovisual content. Unitary effects can be combined to create complex effects.

3.2 Capturing haptic effects from the real world

A first approach for creating haptic effects is to capture haptic effects related to an object or actor of a scene. Piezo-electric sensors can also be used to capture forces [6] or vibrations but, most of the time, accelerometers are used to record accelerations and deduce forces applied to the targeted object. Brady et al. [36] equipped a radio-controlled car to capture accelerations on X,Y and Z axes. These recorded data were then directly transmitted and rendered to the user's control device. Recorded accelerations on X and Y axes control an embedded 2DoF force-feedback device and acceleration on the Z-axis drives a vibration device. Similarly, Danieau et al. [35] placed a camera together with an accelerometer on an actor's chest to capture a first-person point-of-view video and the associated motion. Different scenarios were used to capture different kinds of movements: riding a bike, riding a horse, being in a car braking or turning, and the videos were replayed with haptic effects of force generated from the recorded accelerations. Kuchenbecker et al. [37] followed a database-driven approach by recording haptic events in a database to replay parts later. The authors recorded accelerations resulting from the impact of a stylus on different materials (wood, foam). These accelerations are transduced into forces and replayed by a force-feedback device when the user touches virtual materials.

The second approach consists in capturing haptic effects related to the whole scene. Solutions have explored the use of depth cameras (or 2.5D cameras) to build touchable images [10]. A more precise result could be obtained with 3D trackers [38] but these devices are more expensive and the analysis of the scene would be longer. The problem of capturing haptic effects remains strongly constrained by the available hardware. In comparison to video and sound a limited number of devices exist, mainly accelerometers and 3D cameras with considerable variations in precision and cost.

3.3 Automatic extraction of haptic effects from audiovisual content

Haptic effects can also be created automatically by extraction. The key idea is to generate haptic effects which are consistent with media content in order to highlight specific aspects. For example a scene showing an explosion could be enhanced by haptic feedbacks such as vibrations and heat. In such case, video and sound analysis may be used to detect explosions and then automatically add haptic effects.

Automatic extraction can occur in the Production stage or in the Rendering stage (cf. Figure 1). In the Production stage, haptic effects are automatically generated and can be modified by the creator. In the Rendering stage, haptic effects are automatically generated on the client side.

In the following paragraphs we review generation from visual, audio and metadata content.

3.3.1 Generation from visual content

A classical way to extract content from an audiovisual media consists in using video analysis techniques. Typical algorithms rely on features detectors to extract points of interest inside an image to build evolved information (e.g. object identification) [39]. These algorithms proposed important variations in the features they propose (robustness to light variations, motion, computation cost). Some specific algorithms are dedicated to detect specific features such as faces [40] or motion [41]. Detecting events is also possible. Video abstraction [42] and video data mining [43] have been both used for event detection but is restricted to specific topics like sport games where the range of events is limited and a priori known. Once the targeted event is detected in the audiovisual content, the haptic effect could be generated. For instance, Réhman et al. [27] have shown how to automatically extract events from a soccer game video and to display them with a vibrotactile device. Five vibration patterns were designed to represent the position of the ball on the field, to the team leading or to the goals. However the contribution mainly focused on how to render the effects rather than on the video analysis. Earlier works were conducted in the context of sensory substitution [44], but the aim was to use haptic feedback to replace visual information rather than use haptics as a complement of these data.

The difficulty to directly extract haptic information from a video was pointed out by Mc Daniel et al. [38]. To simplify the problem, the authors built a database which maps visual information (picture of an object) to haptic information (3D shape of the object). They rely on the database to generate appropriate haptic feedback for an object identified from visual information.

Even if computer vision provides a broad range of tools, most techniques have not been explored in detail to analyze and generate haptic feedback. Moreover the interested reader will have to deal with the typical issues of the field (robustness, adaptability of the detection algorithm [39]).

3.3.2 Generation from audio content

Haptic effects can also be created from the audio content within the audiovisual media. The main approach consists in transducing the audible signal into a signal suitable for vibration motors. Chang and O'Sullivan [45] used a band-pass filter to isolate frequencies compatible with a targeted vibration motor and then amplify and render the output signal on this device. This system was developed for mobile phones which then vibrate according to ringtones. The MOTIV[1] development platform from Immersion is a similar commercially available system. Furthermore a module called "Reverb" allows to automatically add haptic effects to any application using the output audio stream.

The approach selected by Nanayakkara et al. [46] is even more direct and does not require any processing of the audio stream. The authors developed a chair for deaf people which renders music and vibrations. The sound is played by specific speakers attached to the seat, which are designed to propagate vibrations to the surface they are attached to.

Most research follows this straightforward technique which tries to transduce audio with vibrations. To extend this approach, one could attempt to represent information conveyed by the audio stream. Then audio analysis techniques to extract specific features would be useful. For example the system proposed by Zhang and Kuo [47] allows to identify music, speech and environmental sound in an audio signal.

3.3.3 Generation from metadata

Metadata can contain information about movements or physical properties of objects within the media. Yamaguchi et al. [9] extracted data from a Flash[2] animation to compute force feedback while the user explores the content. Since this format allows to access the geometry and position of elements within the 2D animation, it is possible to compute a force-feedback focused on one of the objects in the scene. The authors defined a virtual mass for the targeted object and they computed a force-feedback relative to the acceleration and mass of this object. This technique can be applied to computer animations where the 3D model of the scene is available. But the system remains specific to animations and is not suitable for classical videos. However some data formats allow to describe audiovisual contents. For example the MPEG-7 standard focuses on the description of multimedia content and can contain a description of movement

1. http://www.immersion.com/products/motiv/
2. http://www.adobe.com/products/flash.html

within a scene [48], which opens many possibilities for the generation of haptic effects.

3.4 Graphical authoring tools for synthesizing haptic effects

Although haptic effects can be created automatically, there is a demand to author them before their integration into an audiovisual content, or to edit original effects which cannot be generated in an automated way.

Two main categories of graphical authoring tools have been designed. The first one allows users to specify the behavior of one or several actuators. In this case the designer has to use the same device as the end-user. In the second category the designer edits haptic cues that the user will perceive without referring to specific hardware. Various data formats and graphical tools are summarized in Table 2.

3.4.1 Device-oriented effects

The first category allows to define behavior of actuators. Typically the behavior of an actuator can be controlled by specifying a curve representing the amplitude of the stimulation (vibration or the force in time). Hapticons editor [49] was created to edit trajectory patterns called "haptic icons" on a 1DOF force feedback device (a knob). Ryu et al. [50] have considered the editing of vibrotactile patterns and created the posVib Editor where the behavior of the vibration motor is also represented by a curve. This tool is already used in the industry. The aforementioned MOTIV[1] development platform provides a curve editor for designing vibrotactile patterns for various devices (mobile phones, gamepads, etc.).

Graphical interfaces are quite different when they propose to edit the behavior of an array of motors. User has to specify the behavior of each motor in time. A representative example was developed by Rahman et al. [28] and Kim et al. [20].

3.4.2 User-oriented effects

The second type of graphical tool focuses on describing what the user should feel instead of defining how actuators should behave. This implies that the haptic rendering is handled by dedicated software.

The MPEG Rose Annotation tool was designed to associate sensory effects to multimedia contents [13] (cf. section 3.1). It allows the designer to tune sensory effects all along a movie. One or several effects can be added on a timeline which determines when they start and when they finish.

A different approach consists in describing material properties of objects within a scene. It implicitly determines what the user feels when he touches the object. This type of tool looks like 3D editors where the author sees directly the 3D object he is manipulating. This time the author does not edit visual properties but haptic ones (friction, stiffness). We refer the readers to the presentation of the K-Haptic Modeler [51] as well as the HAMLAT tool [52] which is a graphical editor for HAML (cf. section 4.1.1).

4 DISTRIBUTION

The second stage consists in formalizing haptic effects into data to be synchronized, stored and transmitted with the audiovisual media. Even if the range and nature of haptic effects is not yet well defined, there are several attempts in provinding formalizations. These formats are summarized in Table 2 which displays, when available, the associated authoring tools (cf. section 3.4), and solutions to transmit haptic effects over the network (cf. Video Container column of Table 2).

4.1 Data formats for haptic effects

Several works proposing a format for describing haptic feedback for audiovisual contents were identified, but also others more generic which can be used in this context. We will detail contributions based on XML, a versatile description language, on CSV a simple format to store data and on VRML, a language dedicated to the description of 3D worlds.

From these studies we can stress two kinds of formats for denoting haptic effects as identified in the Graphical Authoring Tools (cf. section 3.4). First category is device-oriented and defines actuators behavior, and second category is user-oriented and describes data from the user's point of view. Choosing a format influences the way the Rendering stage has to be handled. Both approaches are summarized in Table 2.

The issue of formalizing haptic effects is solved by companies such as D-Box[9] or Immersion[1] which propose commercial solutions for rendering haptic effects to audiovisual contents. For instance, D-Box created a proprietary language to add haptic effects to a movie, called D-Box Motion Code™. However details of these formats are not currently available and the edition of haptic effects is not allowed by the end-user.

4.1.1 XML-based

The first way to formalize haptic feedback relies on XML language. The HAML language [54], standing for Haptic Application Meta-Language, is a generic format for describing haptic feedback which contains information about haptic device, haptic rendering and visual rendering (cf. Listing 1). The aim of this format is to be able to use any haptic interface with any virtual world. This way the system adapts the haptic rendering to the capabilities of the haptic interface used. This language is dedicated to virtual reality applications but it could be used to describe scenes in an audiovisual content: objects and their location, geometry, haptic properties (stiffness, damping, friction), etc. This format respects the MPEG-7 standard which yields standardized tools to structure and organize descriptions of multimedia contents [48].

Type of Effect	Format	Data Content	GUI	Video Container	Ref.
User-oriented	MPEG-V (XML)	Description and organization of sensory effects in a multimedia content	Yes (MPEG RoSE Annotation Tool)	MPEG-2 TS	[13] [53]
	MPEG-7 (XML)	Description of a 3D scene, haptic device and haptic rendering	Yes (HAMLAT)	n/a	[54] [52]
	XML	Haptic properties of a 3D scene: friction, stiffness, etc. of objects	Yes (K-HapticModeler)	n/a	[51]
	VRML	Description of 3D objects and associated haptic rendering methods	No	n/a	[55]
	MPEG-4 BIFS (VRML)	Information about depth, sifftness, friction of a scene	No	MPEG-4	[10]
	CSV	Information about motion into a scene	No	n/a	[35]
Device-oriented	CSV	Trajectory patterns	Yes (Hapticon Editor)	n/a	[49]
	XML	Vibration patterns	Yes (PosVibe Editor)	n/a	[50]
		Description of haptic devices properties and description of how they are activated	Yes (TouchCon)	n/a	[56]
		Vibration patterns of a tactile array	Yes	n/a	[28]
	MPEG-4 BIFS (VRML)	Vibration patterns of a tactile array	Yes	MPEG-4	[20] [33]

Table 2
Overview of existing formats to edit and store haptic effects. Two types of haptic effect can be described: effects focused on what the user will perceive (User-oriented), and effects focused on how the actuators will behave (Device-oriented). Most of the time a graphical user interface is designed to easily edit data. Some formats are compatible with a container allowing to embed both audiovisual and haptic contents and to be distributed via streaming platforms.

```
<HAML>
...
    <SceneDS>
        <Object>
            <Type>Mesh</Type>
            <Name>Cube</Cube>
            ...
            <Tactile>
                <Stiffness>0.8</Stiffness>
                <Damping>0.9</Damping>
                <SFriction>0.5</SFriction>
                <DFriction>0.3</DFriction>
            </Tactile>
        </Object>
    </SceneDS>
</HAML>
```

Listing 1. Example of a HAML file (xml-based) [8]. Here haptic properties (stiffness

In close relation with video viewing, the Sensory Effect Description Language described by Waltl [13] also relies on XML. This language is designed to add sensory effects to any multimedia content: movies, video games, web, etc. Users can create groups of effects and synchronize them with another media (cf. section 3.1 for the list of effects). For each effect the designer can specify at least its intensity and duration. However devices and techniques to render the effect are not specified. If converting an intensity into vibrations is simple, the rendering of a forward movement over 2 meters with an acceleration of $30cm.s^{-2}$ is less straightforward (cf. Listing 2). At the time of this paper, this language is close to be standardized by the MPEG working group as the MPEG-V format [57].

```
<sedl:SEM>
    <sedl:Effect xsi:type="sev:RigidBodyMotionType" activate="true" si:pts="1593000">
        <sev:MoveToward distance="200" acceleration="30"/>
    </sedl:Effect>
    <sedl:GroupOfEffects si:pts="1647000">
        <sedl:Effect xsi:type="sev:VibrationType" activate="true" intensity-range="0 100" intensity-value="10"/>
        <sedl:Effect xsi:type="sev:WindType" activate="true" intensity-range="0 100" intensity-value="5"/>
    </sedl:GroupOfEffects>
</sedl:SEM>
```

Listing 2. Example of a MPEG-V file (xml-based) [13]. Here a “Move Toward Effect” is defined followed by a group of effects combining a “Wind Effect” and a “Vibration Effect”.

In an approach dedicated to instant messaging applications, Kim et al. [56] developed an XML-based format to exchange haptic feedback: “TouchCons”. This allows users to send haptic messages (vibrations patterns or thermal effects for instance). Two main files compose this system. First the Library XML details a list of haptic messages and how they should be rendered (device used, intensity, duration) and second the Device XML, describes the available devices and associated capabilities. Then to send a message the user picks one in the Library XML and when he receives a message, it is rendered according to the capabilities of his devices listed in Device XML. This framework could be used to described haptic effects (instead of TouchCons) and to send them to the end-user. The effects would be then rendered according to the user's devices configuration.

Finally the XML representation can be used to determine the behavior of actuators directly. For example, Rahman et al. [28] described vibration patterns of a vibrotactile array: the vibration intensity of each motor is indicated in a XML-file. This approach is simple but the effects described can be rendered only by a specific device.

4.1.2 CSV-based

Comma Separated Values (CSV) is a file format where data are stored in a simple text file (and usually separated by commas). Enriquez et al. [49] relied on this format to store positions of a knob. This direct approach is simple but dedicated to command a specific device. Danieau et al. [35] also used this type of format but the authors stored information about the motion embedded with a video (acceleration in $m.s^{-2}$ on 3 axes for each instant t). The motion effect is then rendered by the user's haptic device.

4.1.3 VRML-based

A third method to describe a haptic content uses VRML/X3D. Basically this language serves to represent 3D worlds and contains information needed by visual rendering systems. Sourin and Wei [55] proposed an extension of this language by adding haptic rendering techniques. One motivation of this language is to transmit virtual objects and their associated haptic rendering algorithms over the web. In a similar way to HAML this solution allows to describe an audiovisual scene and the associated rendering techniques.

The two techniques presented hereafter are based on the MPEG-4 BIFS format, also known as MPEG-4 Part 11 [58]. BIFS, which stands for Binary Format for Scenes, is a scene description protocol based on VRML. Cha et al. [10] extended this format to add haptic properties to a video. The authors built a "touchable" movie, i.e. a movie in which spectators can feel the depth of the images using a force-feedback device. For each frame of the video the authors associated textures properties (stiffness, static friction and dynamic friction; cf. Listing 3).

```
Shape{
  appearance Appearance {
    texture ImageTexture {
      url "color_image.jpg"
    }
    hapticSurface HapticTextureSurface {
      stiffnessRange 0.1 10
      staticFrictionRange 0.2 0.9
      dynamicFrictionRange 0.3 0.9
      maxHeight 1.0
      hapticTexture ImageTexture{
        url "haptic_image.jpg"
      }
    }
  }
  geometry Depth {
    focalLength 6.983
    pixelWidth 0.00123
    nearPlace 10
    farPlane 200
    texture ImageTexture {
      url "depth_image.png"
    }
  }
}
```

Listing 3. Extended MPEG-4 BIFS (VRML-based) [10]. This file describes haptic properties of a visual scene (color_image.jpg). The depth map and associated friction are provided.

This modified BIFS format can also be used to store vibrotactile patterns used to drive array of vibration motors. In Kim et al.'s work [20] a pattern is encoded into a grey-scale image where each pixel represents an actuator and the intensity of the pixel corresponds to actuator activation intensity: from black (0) for idle to white (255) for maximal vibration. In a similar way, vibrotactile patterns can be associated to frames of a video (cf. Listing 3: instead of "haptic_image.jpg" a "tactile_pattern.jpg" would be associated to the visual scene). Thus the MPEG-4 BIFS format extended by Cha et al. can both describe a 3D scene and/or contain data to drive vibrotactile arrays. These two possibilities have been implemented by Kim et al. [33] for adding haptic textures effects or vibration effects to educational videos.

4.2 Haptic-video containers

A container is a meta-file format that can hold several files in a single one which makes the distribution of files easier. This stage is depicted on Figure 1 where all components of the content are compressed and synchronized into one container and thus can be easily transmitted over the network [59]. These containers are mainly used in multimedia applications to store into one file both the audio and the visual contents.

If several containers embedding audio and video exist (ogv, avi, mp4, etc.), those containing audiovisual and haptic contents are less common. A simple solution could consist in directly embedding the file containing the haptic data into the container if this one allows the attachment of external files (for instance with the mkv container). O'Modhrain and Oakley [34] relied on the Flash standard to distribute videos enhanced with haptic effects. They integrated haptic feedback in their home-made animation and the media was played by a web browser embedding the Immersion Web plug-in. This alternative is suitable for distribution purposes although limited to the rendering capability of the plug-in and to a specific audiovisual content (animation).

To take advantage of streaming platforms, a solution is to develop formats for haptic effects compatible with video containers allowing to be played while they are downloaded. Some formats (cf. section 4.1) where designed to support this streaming feature. Modified MPEG-4 BIFS [10] can be embedded into a classical MPEG-4 container. In a similar way MPEG-V is compatible with the MPEG-2 TS container [53]. This streaming challenge has been identified as **haptic broadcasting** by Cha et al. [11]. This is a specific challenge different from the classical transmission of data for teleoperation [60]. The purpose is not to remote control a device but to send multimedia containing audio, video and haptic contents. The two presented formats are in an early stage of development but demonstrate the possibility of haptic broadcasting.

5 Rendering

Once the haptic content has been transmitted to the user, the haptic device needs to decode and render the content to provide the appropriate effect (in the same way video is displayed on the screen or audio is rendered on the speakers, cf. Figure 1). In the following, we review a list of haptic interfaces proposed for "enhanced" video viewing.

We classified these devices into four categories which are wearable devices, handheld devices, desktop devices and haptic seats. We synthesized the results presented hereafter in Table 3.

5.1 Wearable devices

Wearable devices are designed to be worn by the user while he experiences the audiovisual content. Typically they are composed of several vibrotactile actuators embedded into clothes.

Exploring the idea of enhancing live sports experience, Lee et al. [26] have proposed a device with vibrotactile sensation through an assembly of 7x10 vibrotactors attached to the user forearm. This prototype was used to render movements of the ball on the field during a soccer game. The tactile array was mapped to the field and vibrations were triggered at ball locations. According to the authors this device allows the user to better understand ambiguous game situations.

Kim et al. [20] [33] designed a tactile glove for immersive multimedia. It contains 20 tactile actuators per glove (4 per finger). The gloves are wireless-controlled and produce vibrotactile patterns while the user watches a movie. These patterns were previously authored and synchronized with the video.

In pursuing the goal of integrating tactile devices in everyday clothes, Rahman et al. [28] reported how easily vibrotactile arrays can be integrated into clothes such as jackets or arm bands. Actually this topic has been intensively studied for virtual reality purposes [63] and a lot of similar devices have been designed.

A tactile jacket was also developed by Lemmens et al. [23]. They explored the influence of tactile devices on spectators emotional level and focused on the design of a tactile jacket with 16 segments of 4 vibration motors covering the torso and the arms. Motors are activated following patterns related to specific emotions. For example, the feeling of love is enhanced by firing the motors in the stomach area in a circular manner.

In the idea of extending the range of devices in a wearable, Palan et al. [25] presented a vest embedding vibration motors, solenoids and Peltier elements. The vest was designed to display three haptic effects as realistic as possible: gunshot, slashing and blood flow. The motivations driving this research is improving experience in playing video games. Following similar motivations, a jacket proposed by TNGames[3] produces effects such as explosions, gunshots or accelerations using 8 air cells (the jacket is commercially available).

While the embedded devices do not yield a significant change in weight or wearability of clothes (most are composed by simple vibrotactile actuators), the range of haptic effects which can be produced is rather limited.

5.2 Handheld devices

The second type of device corresponds to the handheld devices. In this case the user experiences haptic feedback through a portable device held in his hand.

Vibrotactile technology appears well-suited for portable devices. For years, the gaming industry uses vibrating joypads to enhance immersion video games. Moreover mobiles devices (phones and tablets) are now equipped with vibration motors which may be used to enhance multimedia contents[4]. Using this technology, Réhman et al. [27] relied on a mobile phone equipped with a vibration motor to display haptic cues related to a soccer game. Alexander et al. [31] developed a prototype of a mobile TV providing tactile feedback thanks to ultrasounds. The device is a screen with a 10x10 array of ultrasonic transmitters set on its back. This way the user holds the device to watch the audiovisual content and to experience haptic feedback on his fingers.

The remote control developed by O'Modhrain and Oakley [34] is a different sort of handled device providing force-feedback. A gaming joystick was actually rehoused into a device looking like a remote control. In a similar way Yamaguchi et al. [9] used a computer mouse with a 2DOF force-feedback joystick.

Similar to clothes-based devices, handheld devices cannot embed heavy actuators and therefore only a restricted range of haptic effects can be rendered. However,

3. http://tngames.com/
4. http://www.immersion.com/markets/mobile/products/

Type of interface	Device	Actuator	Haptic Effect	Ref.
Wearable	Vibrotactile Arm Band	7x10 vibration motors	Vibrations (related to position of a ball during a soccer game)	[26]
	Vibrotactile Glove	20 vibration motors (4 per finger)	Vibrations	[20]
	Vibrotactile Arm Band or Jacket	Array of vibration motors (variable size)	Vibrations	[28]
	Vibrotactile Jacket	16x4 vibration motors	Vibrations (related to user's emotions)	[23]
	Vibrotactile Vest	Vibration motors + solenoids + Peltier Elements	Pressure (gunshot), Temperature (blood flow), Vibrations (slashing)	[25]
	Vibrotactile Vest	8 air cells	Vibrations and Pressure (gunshots, acceleration, explosion)	TNGames[3]
Handheld	Mobile Phone	Vibration motor	Vibrations (related to status of soccer game)	[27]
	Mobile Phone	Vibration motor	Vibrations	Immersion[1]
	Remote Control	2DOF Joystick	Force	[34]
	Computer mouse	2DOF Joystick	Force	[9]
	Portable TV	10x10 array of ultrasound transducers	Pressure	[31]
Desktop	Force-feedback device	3DOF motorized arm	Movement	[32]
	Phantom[5]	6DOF motorized arm	Movement	[33]
	Novint Falcon[6]	3DOF motorized arm	Force (texture of an image)	[10]
	Novint Falcon[6]	3DOF motorized arm	Force (motion in the video)	[35]
	n/a	Array of 324 ultrasound transducers	Pressure	[30]
	Air receiver	Array of air-jets	Pressure	[29]
	Philips[7] AmBX	Vibration motor + 2 Fans (+ 2 LED spotlights)	Vibration (+ Wind & Light)	[13]
Haptic Seat	Vibrotactile Blanket	176 vibration motors	Vibrations (related to user's emotions)	[61]
	Vibrotactile Chair	3x4 vibration motors	Vibrations	[62]
	Couch	Vibration motor	Vibrations (of the whole seat)	Guitammer[8]
	Moving Chair	4 compressors under chair legs	3DOF Body Motion (pitch, roll, heave)	D-Box[9]

Table 3
Overview of existing haptic displays used for enhancing audiovisual content. Wearable devices (wearable or handheld) typically provide vibrations while ground-based devices (desktop or seats) can generate strong forces.

the use of a common device in the user living space (remote control, mobile phone) seems popular with regards to acceptance.

5.3 Desktop devices

Force-feedback devices are mainly used in virtual reality to interact with virtual objects. The user can feel and often modify the content. In a video viewing context their use is different in the sense that the user cannot modify the content. He receives haptic cues, sometimes while exploring actively the content, but the audiovisual content will not change. For example in the solution devised by Gaw et al. [32], the user hold a force-feedback device and is guided along a prerecorded path while viewing a movie. The same technique was used by Kim et al. [33] to enhance educational videos with a Phantom[5] device. In a similar way, Danieau et al. [35] used a force-feedback device to make the user feel the global movement in a video.

These devices have also been tested in the task of "touching" images in a video [10]. In this case the user can explore actively the video content and receives haptic feedback through a Novint Falcon device[6].

Others desktop devices are designed to convey haptic feedback to the user without contact. The main advantage of these interfaces is that the user does not manipulate a complex device, so the interaction is more ecological. An example is a fan which generates air

5. http://www.sensable.com
6. http://www.novint.com

streams and simulates the haptic effect of wind. Associated to a thermal device, fans may be used to create temperature variations [24]. Fans providing wind effects are commercially available such as The Philips amBX system[7]. It provides wind effects but also light effects and add vibrations to a keyboard.

This concept can be also used to allow contact with virtual objects without handling a device. Hoshi et al. [30] used ultrasounds to exert pressure remotely on a user's skin. Their prototype is composed of an array of 324 airborne ultrasound transducers. The device is able to exert a force of 16mN at a focal point of a diameter of 20mm over a 180x180mm surface. This invisible surface is placed at 200mm above the device. Combined with a 3D display system, the author succeeded to create touchable floating images. A similar system was previously developed by Suzuki and Kobayashi [29] which relies on air jets.

5.4 Haptic seats

A fourth type of device relates to haptic seats. While seated on a modified chair, the user passively feels haptic effects.

Vibrotactile actuators have once again been used in a number of approaches. The tactile blanket [61], for example, is equipped with 176 actuators and displays vibration patterns designed to enhance the user emotion. This is the blanket's version of the Lemmens' Jacket [23].

More recently Israr and Poupyrev [62] embedded an array of 12 vibrotactile actuators in the back of a chair and proposed an original way to control it. They relied on tactile illusions to make the user feel a continuous stimulus while actuators are at discrete locations.

Interestingly, several commercial products of this category are available. An example is the "couch shaker" from The Guitammer Company[8]. This type of device relies on actuators to shake the couch or sofa. Actually this operates like a subwoofer by propagating low-frequency vibrations to the couch instead of playing sounds. Some seating devices attempt to provide more complex effects such as motion. Typically such seats are fixed on actuators or motion platforms. The D-Box[9] seat is one of them and features 3 DOF: pitch, roll and heave.

Haptic seats are commonly encountered in theme parks or amusement arcades where they are typically used as motion simulators. Some of them even embed several devices to provide a wide range of effects (water spray, air blast, leg ticklers, etc. See MediaMation[10] company.) But these devices are not adapted to the end-user living space and the cost is prohibitive for mass market. In contrast, the D-Box[9] seat is a product adapted to consumers and it is designed to be integrated into a living room but remains expensive. The alternative is the use of devices based on vibrotactile arrays but the range of tactile effects which can be rendered is quite restricted.

7. http://www.ambx.philips.com
8. http://www.thebuttkicker.com
9. http://www.d-box.com
10. http://www.mediamation.com

6 Quality of Experience

Haptic effects aim at enhancing the audiovisual experience. This means that the Quality of Experience (QoE) of a video viewing session is expected to be higher with haptic feedback than without. But how to assess this hypothesis? Jain [64] discussed the necessity of capturing the QoE for evaluating a system, but he also underlined the difficulty to identify and measure factors characterizing this metric due to its subjective nature.

Nevertheless Hamam et al. [65] [8] proposed a first model to evaluate QoE in multimedia haptics and identified four factors based on the rendering quality and on user-centered measures: physiology, psychology and perception. The rendering quality relies on the quality of the visual, audio and haptic feedback. Perception measures describe the way the user perceives the system depending on his own experience, his fatigue and other factors which can alter his perception. Physiological measures identify how the system modifies the user's biological state and psychological measures highlight changes of his state of mind. The authors have also detailed an exhaustive list of parameters related each factor (e.g. respiration rate, body temperature or blood pressure for physiological measures). This provides a taxonomy of the different characteristics influencing the Quality of Experience, but techniques to evaluate them are not presented.

In this section we detail classical techniques to measure the QoE of audiovisual systems enhanced with haptic effects. The typical approach found in the literature is a subjective measure based on questionnaires. The second set of techniques we present is based on the capture of biosignals. Here the user's emotion is inferred from his physiological state, thereby providing a less biased measure.

6.1 Subjective measures: questionnaires

Most contributions on haptic feedback for multimedia rely on simple questionnaires to evaluate the impact of haptic feedback on the Quality of Experience. Participants are usually asked to answer questions on a Likert-scale. For example, Kim et al. [20] studied the benefits of vibrotactile feedback for enhancing movies by using 4 general questions (is this more interesting than movies? is the tactile content easy to understand? is the tactile content related to scene movies and does the tactile content support immersion?). Ur Rhéman et al. [27] covered the same aspects through a more detailed questionnaire while some approaches limit their analysis to the only aspect of user satisfaction [66].

A more elaborate approach consists in characterizing the Quality of Experience with several factors. Hamam et al. [67] evaluated 5 factors (extracted from their model described above): Realism, Usefulness, Intuitivism, Fatigue and QoE. Danieau et al. [35] identified 4 factors: Sensory, Comfort, Realism and Satisfaction. "Sensory" characterizes how the haptic feedback contributed to the immersion. "Realism" describes how much the simulation is realistic and consistent with the user's representation of the real world. "Comfort" measures the global confort of the user experiencing the system (a proxy for acceptance). "Satisfaction" determines how well the user enjoys the system. These 4 factors were combined into one QoE measure.

Each of these contributions has developed a questionnaire to evaluate the quality of experience, with obviously strong overlaps (e.g. satisfaction). This highlights a need for a standardized questionnaire to evaluate and compare different systems. The identification of factors to be measured is a difficult task, but several are already evaluated in a systematic way: comfort, interest, acceptance and satisfaction. They can serve as a base to build a subjective measure of the QoE.

6.2 Objective measures: physiological data

A second approach to evaluate the Quality of Experience consists in measuring changes of user's physiological signals. The QoE cannot be directly determined from this measure but it can be used to infer the user's emotional state which contributes to the QoE. To our best knowledge, no work relied on this technique in the context of HAV systems. However, interesting results can be found in topics related to this survey.

In the context of Virtual Reality, Meehan et al. [68] captured the heart rate, skin conductance and skin temperature of subjects in virtual stressful environment. These measures helped to determine the user's feeling of "presence" and were compared to subjective users' self-reports (see [69] for a survey on "presence"). From this work it appeared that the heart rate is correlated to the feeling of presence, while changes in the skin conductance were less powerful and changes of the temperature were not significantly strong enough. Haptic feedback significantly improved presence.

Mandryk et al. [70] observed biosignals of video games players to determine their user experience. The skin conductance, heart rate, facial muscle activity and respiration rate were captured. The authors concluded that, for most participants, playing against a friend is funnier than playing against the computer. The physiological measures were significantly consistent to the self-reported measures.

In a video viewing context, Fleureau et al. [71] studied the potential of physiological signals for detecting emotional events. Participants were simply watching several videos while their heart rate, skin conductance and facial muscle activity were recorded. A detector based on machine learning techniques was designed, and from a user's biosignals, it was robustly able to determine whether he was experiencing an emotional event and if this event was positive or negative.

The selected physiological signals in these works are mostly the same: heart rate, galvanic skin response facial muscle activity. All yield interesting results despite the different contexts (virtual reality, video games, video viewing). This opens interesting possibilities for evaluating video viewing enhanced with haptic effects, and closed-loop forms could furthermore be proposed in which physiological signals could control nature and intensity of events to better adapt the haptic effects to the users.

7 Overview and Perspectives

We have presented an overview of how haptic effects can enhance audiovisual content. Contributions along the stages of haptic production, distribution and rendering were reported. Some of these works propose solutions addressing all stages and may be seen as implementations of the generic workflow displayed in Figure 1. These general approaches are summarized in Table 4.

While solutions clearly demonstrate how haptic effects can be used with audiovisual content using tactile or kinesthetic feedback, reported contributions do not explore combinations of effects (e.g. kinesthetic and tactile). This is mostly due to the devices which generally embed one type of actuator. As a consequence, a wide range of effects cannot be displayed and conjunction of effects is rarely explored nor evaluated, despite significant potential benefits. Furthermore, there appears to be a gap between the use of portable haptic interfaces (wearable or handheld), conveying poor effects, and complex devices (motion simulators) not adapted to the user living space. As a consequence, there is a clear necessity in designing new haptic devices dedicated to audiovisual enhancement. But this implies a better understanding of needs in HAV systems which seem to differ significantly from needs in virtual reality systems.

Indeed, further research on user perception has to be conducted to determine relevant haptic stimuli for effective and appropriate audiovisual entertainment. So far the link between haptic stimuli and user experience is not well established. Haptic effects are mainly used in a similar way haptic feedback is used in virtual reality: to immerse the user physically in the audiovisual scene. However a few works relied on haptic effects to enhance non-diegetic aspects of a video such as the ambiance or emotions. This appears as a key challenge of this young field of study.

The distribution stage also requires research efforts. Each solution proposes a different technique to formalize haptic effects, obviously due to the absence of a common definition for haptic effects. Moreover only half of the contributions proposed a way to transmit the media to a remote display device. But several techniques allowing haptic broadcasting are emerging. Multimedia containers embedding audiovisual and haptic effects are currently developed and standardized (MPEG-V, MPEG-4 BIFS). Regarding the distribution stage, the MPEG-V format is a promising solution which is going to be standardized by the MPEG group. The reader interested in standards for haptic effects must refer to this ongoing work. A list of haptic effects is proposed as well as a XML-based method to describe them. This format is also designed to be compatible with streaming technologies. However this new standard has to follow the evolution of this emerging field of study. New haptic effects and new devices will probably be created and should be accounted for in an evolutive standard.

In most of the solutions haptic effects are synthesized. Authors manually create and synchronize haptic effects to the audiovisual contents. Each solution proposes a different technique to edit haptic effects, while some general editing tools could be proposed with the advent of new standards. For haptic effects not synthesized by authors, the contributions have proposed to extract haptic cues automatically from the visual content. But they are limited to a specific audiovisual content: soccer games following pre-defined rules and animations where the position and geometry of objects is known beforehand. The automatic extraction of haptic effects for any audiovisual content remains a very complex task. Efforts are necessary to adapt current algorithms to this new purpose. Extraction can be facilitated by metadata that describe the content of the media, but extracting haptic effects from videos is a new challenge for which new specific techniques need to be designed.

One last aspect to be discussed in this survey is the quantification of the benefits brought by haptic effects to audiovisual contents. Some of the presented works have conducted user evaluations mostly based on questionnaires. If most of them show that haptic effects enhance the user experience, the different contributions are hardly comparable together. There is pressing need for common and robust tools to evaluate this Quality of Experience.

8 Conclusion

In this survey we explored the possibilities provided by haptic feedback for enhancing audiovisual content. This field, which is referred to as HAV, is a young field of study where various trends are emerging. We arranged the presentation of contributions against a generic workflow and identified the key challenges pertained to this new way of experiencing videos.

The first stage, related to production of haptic effects, is the identification and generation of haptic effects which must be applied on the user during the display of the media. We detailed different formats to store and synchronize haptic effects to the audiovisual media, from a simple text-based representation to standardized XML formats. The key issue is the creation of haptic feedback. While a number of authoring tools are available, these effects may also be captured from physical sensors or generated from the other part of the media (video, audio or metadata).

Once the media is enriched with haptic effects, it has to be sent to the user. In the current context, media are often distributed through streaming platforms to distant users. This stage depends on the way haptic data are stored. If these issues are already solved for audiovisual media, there are only few standards for media with haptic effects. However some pioneer contributions demonstrated the feasibility of this approach.

In the last stage the user perceives the media thanks to dedicated haptic devices. These devices are haptic interfaces generally designed for this purpose which display haptic cues during video viewing.

Altogether, the results of our survey suggest that research efforts need to focus on the design of data formats and technology for spreading HAV contents. The development of authoring tools is also necessary to allow the creation of such media. This may lead to a new type of professional activities in the cinema industry. As the 3D movies now need “stereographers”, these new haptic audiovisual contents would require “hapticgraphers”. Moreover the development of tools to evaluate the quality of experience and the acceptance of such systems are mandatory. There is no doubt that the next decade will lead to exciting novel research results emerging from this young but promising field of study, yielding new tools and displays for adding haptics to multimedia for a more compelling user experience with audiovisual content.

References

[1] M. Heilig, “Sensorama simulator,” Aug. 1962. U.S. Patent 3050870.

[2] G. C. Burdea and P. Coiffet, *Virtual Reality Technology*. Wiley-IEEE Press, 2003.

[3] S. J. Biggs and M. Srinivasan, “Haptic Interfaces,” *Handbook of virtual Environments*, pp. 93 – 116, 2002.

[4] V. Hayward, O. R. Astley, M. Cruz-Hernandez, D. Grant, and G. Robles-De-La-Torre, “Haptic interfaces and devices,” *Sensor Review*, vol. 24, no. 1, pp. 16–29, 2004.

[5] M. Reiner, “The Role of Haptics in Immersive Telecommunication Environments,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, pp. 392–401, Mar. 2004.

[6] S. O’Modhrain and I. Oakley, “Touch TV: Adding feeling to broadcast media,” in *European Conference on Interactive Television: from Viewers to Actors*, 2003.

[7] N. Magnenat-Thalmann and U. Bonanni, “Haptics in virtual reality and multimedia,” *IEEE Multimedia*, vol. 13, no. 3, pp. 6–11, 2006.

[8] A. El Saddik, M. Orozco, M. Eid, and J. Cha, *Haptic technologies - Bringing Touch to Multimedia*. Springer, 2011.

[9] T. Yamaguchi, A. Akabane, J. Murayama, and M. Sato, “Automatic generation of haptic effect into published 2D graphics,” in *Eurohaptics*, 2006.

[10] J. Cha, M. Eid, and A. E. Saddik, “Touchable 3D video system,” *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 5, pp. 1–25, Oct. 2009.

[11] J. Cha, Y.-S. Ho, Y. Kim, J. Ryu, and I. Oakley, “A Framework for Haptic Broadcasting,” *IEEE Multimedia*, vol. 16, pp. 16–27, July 2009.

[12] K. Kilkki, “Quality of experience in communications ecosystem,” *Journal of universal computer science*, vol. 14, no. 5, pp. 615–624, 2008.

[13] M. Waltl, *Enriching Multimedia with Sensory Effects: Annotation and Simulation Tools for the Representation of Sensory Effects*. VDM Verlag Saarbrücken, Germany, 2010.

[14] K. Shimoga, “A survey of perceptual feedback issues in dexterous telemanipulation. II. Finger touch feedback,” in *IEEE Virtual Reality Annual International Symposium*, pp. 271–279, IEEE, 1993.

Audiovisual Content		Haptic Effect	Production	Distribution	Rendering	Ref.
Category	Details					
Sport	Soccer game (3D simulation)	Vibrations (ball position)	[Automatic extraction] The system traces the ball during soccer game	n/a	Vibrotactile array embedded into an arm band	[26]
	Soccer game (simulation)	Vibrations (ball position, goals, team leading)	[Automatic extraction] Video analysis of events from a soccer game (not implemented, events are received from the simulation)	n/a	Mobile phone equipped with vibration motor	[27]
Animation	Animation (home-made with Flash)	Force (related to an object of the animation)	[Automatic creation] Force-feedback is computed from the position and geometry of the object	Flash	Mouse with a joystick (2DOF force feedback)	[9]
	Cartoon (home-made with Flash)	Force (related to onscreen character)	[Synthesis] Force-feedback is defined during edition of the cartoon	Flash	Remote control with a joystick (2DOF force feedback)	[34]
	Cartoon / Movie	Movement (user's hand is guided according to a trajectory)	[Capturing] Trajectories recorded from force feedback device	n/a	Force-feedback device	[32]
Movie	Movie	Force (user touches the image)	[Synthesis / Capturing] Material properties for each frame (depth, stiffness, etc.) stored into MPEG-4 BIFS	MPEG-4	Novint Falcon (3 DOF force-feedback)	[10]
	Movie (from Youtube)	Vibrations	[Synthesis] Vibration patterns stored into XML file	XML file on a web server	Vibrotactile array embedded into arm band or jacket	[28]
	Movie	Vibrations	[Synthesis] Vibration patterns stored into MPEG-4 BIFS	MPEG-4	Vibrotactile array embedded into gloves	[20]
	Movie	Vibrations and Wind	[Synthesis] Sensory effects stored into MPEG-V file	MPEG-2 TS	Philips amBX system	[13]
	Educational video	Vibrations or Force (user touches the image)	[Synthesis] Haptic effects (vibrations or haptic properties) stored into MPEG-4 BIFS	MPEG-4	Vibrotactile gloves or Phantom device (6DOF force-feedback)	[33]
	Movie	Force (related to the motion in the video)	[Capturing] The motion is captured by accelerometers	n/a	Novint Falcon (3DOF force-feedback)	[35]

Table 4
Summary of existing schemes for adding haptic effects to audiovisual contents. Each system proposes a solution for synchronizing and rendering haptic feedback within an audiovisual content. Some schemes specify ways to distribute the media over the network.

[15] L. A. Jones and M. Berris, "The psychophysics of temperature perception and thermal-interface design," in *Proceedings of the 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pp. 137–142, IEEE, 2002.

[16] L. A. Jones, "Kinesthetic Sensing," *Human and Machine Haptics*, pp. 1–10, Mar. 2000.

[17] A. Berthoz, *The brain's Sense of Movement*. Harward University Press, 2000.

[18] J. B. van Erp, "Guidelines for the use of vibro-tactile displays in human computer interaction," in *Eurohaptics*, vol. 2002, pp. 18–22, 2002.

[19] K. Hale and K. Stanney, "Deriving haptic design guidelines from human physiological, psychophysical, and neurological foundations," *IEEE Computer Graphics and Applications*, vol. 24, no. 2, pp. 33–39, 2004.

[20] Y. Kim, J. Cha, J. Ryu, and I. Oakley, "A Tactile Glove Design and Authoring System for Immersive Multimedia," *IEEE Multimedia*, pp. 2–12, 2010.

[21] J. Smith and K. MacLean, "Communicating emotion through a haptic link: Design space and methodology," *International Journal of Human-Computer Studies*, vol. 65, no. 4, pp. 376–387, 2007.

[22] D. Tsetserukou, A. Neviarouskaya, H. Prendinger, N. Kawakami, and S. Tachi, "Affective haptics in emotional communication," in *IEEE International Conference on Affective Computing and Intelligent Interaction and Workshops*, no. 1, pp. 1–6, 2009.

[23] P. Lemmens, F. Crompvoets, D. Brokken, J. V. D. Eerenbeemd, and G.-j. D. Vries, "A body – conforming tactile jacket to enrich movie viewing," *Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pp. 7–12, 2009.

[24] J. Dionisio, "Projects in VR Virtual Hell : A Trip Through the Flames," *IEEE Computer Graphics And Applications*, pp. 11–14, 1997.

[25] S. Palan, R. Wang, N. Naukam, L. Edward, and K. J. Kuchenbecker, "Tactile Gaming Vest (TGV)," 2010. http://iroboticist.com/2010/03/26/tgv/, last access on December 2011.

[26] B. Lee, J. Lee, J. Cha, C. Seo, and J. Ryu, "Immersive Live Sports Experience with Vibrotactile Sensation," *Human-Computer Interaction*, pp. 1042–1045, 2005.

[27] S. ur Rehman, "Turn Your Mobile Into the Ball: Rendering Live Football Game Using Vibration," *IEEE Transactions on Multimedia*, vol. 10, pp. 1022–1033, Oct. 2008.

[28] M. A. Rahman, A. Alkhaldi, and J. Cha, "Adding haptic feature to YouTube," *International conference on Multimedia*, pp. 1643–1646, 2010.

[29] Y. Suzuki, M. Kobayashi, and A. J. Interface, "Air jet driven force feedback in virtual reality," *IEEE computer graphics and applications*, vol. 25, no. 1, pp. 44–7, 2005.

[30] T. Hoshi, M. Takahashi, T. Iwamoto, and H. Shinoda, "Noncontact Tactile Display Based on Radiation Pressure of Airborne Ultrasound," *IEEE Transactions on Haptics*, vol. 3, no. 3, pp. 155–165, 2010.

[31] J. Alexander, M. T. Marshall, and S. Subramanian, "Adding haptic feedback to mobile tv," in *ACM Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1975–1980, 2011.

[32] D. Gaw, D. Morris, and K. Salisbury, "Haptically Annotated Movies: Reaching Out and Touching the Silver Screen," *Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pp. 287–288, 2006.

[33] Y. Kim, S. Park, H. Kim, H. Jeong, and J. Ryu, "Effects of different haptic modalities on students' understanding of physical phenomena," in *IEEE World Haptics Conference*, pp. 379–384, 2011.

[34] S. O'Modhrain and I. Oakley, "Adding interactivity: active touch in broadcast media," in *IEEE Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pp. 293–294, 2004.

[35] F. Danieau, J. Fleureau, A. Cabec, P. Kerbiriou, P. Guillotel, N. Mollet, M. Christie, and A. Lécuyer, "A Framework for Enhancing Video Viewing Experience with Haptic Effects of Motion," in *IEEE Haptics Symposium*, 2012.

[36] A. Brady, B. MacDonald, I. Oakley, S. Hughes, and S. O'Modhrain, "Relay: a futuristic interface for remote driving," in *Eurohaptics*, 2002.

[37] K. Kuchenbecker, J. Fiene, and G. Niemeyer, "Event-Based Haptics and Acceleration Matching: Portraying and Assessing the Realism of Contact," *Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pp. 381–387, 2005.

[38] T. L. McDaniel, K. Kahol, P. Tripathi, D. P. Smith, L. Bratton, R. Atreya, and S. Panchanathan, "A methodology to establish ground truth for computer vision algorithms to estimate haptic features from visual images," *IEEE International Worksho on Haptic Audio Visual Environments and their Applications*, no. October, pp. 94–99, 2005.

[39] T. Tuytelaars and K. Mikolajczyk, "Local Invariant Feature Detectors: A Survey," *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2007.

[40] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.

[41] W. Hu, T. Tan, L. Wang, and S. Maybank, "A Survey on Visual Surveillance of Object Motion and Behaviors," *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, vol. 34, pp. 334–352, Aug. 2004.

[42] B. T. Truong and S. Venkatesh, "Video abstraction," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 3, Feb. 2007.

[43] X. Zhu, X. Wu, A. K. Elmagarmid, Z. Feng, and L. Wu, "Video data mining: semantic indexing and event detection from the association perspective," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, pp. 665–677, May 2005.

[44] P. Bach-y Rita and S. W. Kercel, "Sensory substitution and the human-machine interface," *Trends in Cognitive Sciences*, vol. 7, no. 12, pp. 541–546, 2003.

[45] A. Chang and C. O'Sullivan, "Audio-Haptic Feedback in Mobile Phones," *Extended Abstracts of the Computer Human Interaction Conference on Human factors in Computing Systems*, pp. 1264–1267, 2005.

[46] S. Nanayakkara, E. Taylor, L. Wyse, and S. Ong, "An enhanced musical experience for the deaf: Design and evaluation of a music display and a haptic chair," in *ACM International conference on Human factors in computing systems*, pp. 337–346, 2009.

[47] T. Zhang and C.-C. Kuo, "Audio Content Analysis for Online Audiovisual Data Segmentation and Classification," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 4, pp. 441–457, 2001.

[48] S. Chang, T. Sikora, and A. Purl, "Overview of the MPEG-7 standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 688–695, 2002.

[49] M. Enriquez and K. MacLean, "The hapticon editor: a tool in support of haptic communication research," *Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pp. 356–362, 2003.

[50] J. Ryu and S. Choi, "posVibEditor: Graphical authoring tool of vibrotactile patterns," *IEEE International Workshop on Haptic Audio visual Environments and Games*, pp. 120–125, Oct. 2008.

[51] Y. Seo, B.-c. Lee, Y. Kim, J.-p. Kim, and J. Ryu, "K-HapticModeler : A Haptic Modeling Scope and Basic Framework," *International Workshop on Haptic Audio Visual Environments and their Applications*, 2007.

[52] M. Eid, S. Andrews, A. Alamri, and A. El Saddik, "HAMLAT: A HAML-based authoring tool for haptic application development," *Haptics: Perception, Devices and Scenarios*, pp. 857–866, 2008.

[53] K. Yoon, B. Choi, E. Lee, and T. Lim, "4-D broadcasting with MPEG-V," in *IEEE Multimedia Signal Processing*, no. 2, pp. 257–262, 2010.

[54] F. El-far, M. Eid, M. Orozco, and A. Saddik, "Haptic Applications Meta-Language," *IEEE International Symposium on Distributed Simulation and Real-Time Applications*, pp. 261–264, 2006.

[55] A. Sourin and L. Wei, "Visual immersive haptic rendering on the web," *ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, p. 1, 2008.

[56] Y. Kim, H. Shin, and M. Hahn, "A bidirectional haptic communication framework and an authoring tool for an instant messenger," in *IEEE Advanced Communication Technology*, vol. 3, pp. 2050–2053, 2009.

[57] MPEG Group, "ISO/IEC FDIS 23005-3 Information technology – Media context and control – Part 3: Sensory Information." 2011.

[58] ISO/IEC 14496-11, "Information technology – Coding of audio-visual objects – Part 11: Scene description and application engine," 2005.

[59] D. Wu, Y. T. Hou, W. Zhu, Y.-Q. Zhang, and J. M. Peha, "Streaming video over the Internet: approaches and directions," *IEEE Transac-*

tions on Circuits and Systems for Video Technology, vol. 11, pp. 282–300, Mar. 2001.

[60] P. F. Hokayem and M. W. Spong, "Bilateral teleoperation: An historical survey," *Automatica*, vol. 42, pp. 2035–2057, Dec. 2006.

[61] E. Dijk, A. Weffers-Albu, and T. De Zeeuw, "A tactile actuation blanket to intensify movie experiences with personalised tactile effects," in *International Conference on Intelligent Technologies for Interactive Entertainment*, pp. 2–3, 2010.

[62] A. Israr and I. Poupyrev, "Tactile Brush: Drawing on Skin with a Tactile Grid Display," *Computer Human Interaction*, pp. 2019–2028, 2011.

[63] R. Lindeman, Y. Yanagida, H. Noma, and K. Hosaka, "Wearable vibrotactile systems for virtual contact and information display," *Virtual Reality*, vol. 9, pp. 203–213, Dec. 2006.

[64] R. Jain, "Quality of experience," *IEEE Multimedia*, vol. 65, no. 1, pp. 1–2, 2004.

[65] A. Hamam, M. Eid, A. E. Saddik, and N. D. Georganas, "A quality of experience model for haptic user interfaces," in *Proceedings of the 2008 Ambi-Sys workshop on Haptic user interfaces in ambient media systems*, pp. 1–6, 2008.

[66] M. Waltl and C. Timmerer, "Increasing the user experience of multimedia presentations with sensory effects," in *11th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, 2010.

[67] A. Hamam, N. Georganas, and A. El Saddik, "Effect of haptics on the Quality of Experience," in *IEEE International Symposium on Haptic Audio-Visual Environments and Games (HAVE)*, pp. 1–6, IEEE, 2010.

[68] M. Meehan, B. Insko, M. Whitton, and F. P. J. Brooks, "Physiological measures of presence in stressful virtual environments," *ACM Transactions on Graphics*, pp. 645–652, 2002.

[69] M. Schuemie, P. Van Der Straaten, M. Krijn, and C. Van Der Mast, "Research on presence in virtual reality: A survey," *CyberPsychology & Behavior*, vol. 4, no. 2, pp. 183–201, 2001.

[70] R. L. Mandryk, K. M. Inkpen, and T. W. Calvert, "Using psychophysiological techniques to measure user experience with entertainment technologies," *Behavior & Information Technology*, 2005.

[71] J. Fleureau, P. Guillotel, and Q. Huynh-Thu, "Physiological-based affect event detector for entertainment video applications.," *IEEE Transactions on Affective Computing*, 2012.

Fabien Danieau is a Ph.D student at Technicolor R&I (France) and in the INRIA VR4I Team (France). His research interests include haptic interactions, multimedia and user experience. In 2010, he received the engineering degree in Cognitics from the Ecole Nationale Supérieure de Cognitique (France).

Anatole Lécuyer is Senior Researcher at the French National Institute for Research in Computer Science and Control (INRIA), Rennes, France. His main research interests include virtual reality, haptic interaction, and brain-computer interfaces. He is/was involved in various collaborative projects related to haptics such as European project "NIW" (08-11) on Augmented Walking, and European Network of Excellence "INTUITION" (05-08) on Virtual Reality in which he leaded the Working Group on "Haptic Interaction". Anatole Lécuyer serves as expert in VR and haptics for public bodies such as European Commission (EC), European Research Council (ERC) and French National Research Agency (ANR). He is secretary of the French Association for Virtual Reality and former secretary of the IEEE Technical Committee on Haptics. He is associate editor of ACM Transactions on Applied Perception (ACM TAP) and International Journal of Human-Computer Studies (IJHCS).

Philippe Guillotel is Distinguished Scientist at Technicolor R&I (France), in charge of Video Perception, Processing & Coding research. He received the engineering Degree in electrical engineering, telecommunications and computer science from the Ecole Nationale Supérieure des Télécommunications (France) in 1988, and the M.S. degree in electrical engineering and information theory from University of Rennes (France) in 1986. He has worked in the Media & Broadcast Industry for more than 20 years. In addition to his work on video coding systems and compression algorithms, his work at Technicolor has covered video processing, video delivery as well as human perception and video quality assessment studies. He is the author of more than 50 patents and 30 papers.

Julien Fleureau is Researcher at Technicolor R&I (France) in the Video Processing & Perception laboratory. He received the engineering and M.S. degrees in electrical engineering and computer science from Ecole Centrale de Nantes (France) in 2005, and the Ph.D. degree from University of Rennes (France) in 2008. His Ph.D works in the fields of image processing and biomedical modeling were completed at the Laboratoire de Traitement du Signal et de l'Image (INSERM, France) during a two-year postdoc. His main research interests deal with signal processing, datamining and machine learning applied to biomedical and physiological signals analysis.

Nicolas Mollet is Researcher at Technicolor R&I (France). His scientific interest is focused on human-machine interactions, multi-entities and informed environment approaches. He defended his PhD in 2005 (INRIA Rennes), in which he proposed a full authoring environment of virtual maintenance training sessions on military equipment, based on dynamic and behavioral models of objects and scenarios. He joined in 2007 the IIT laboratory (Genova, Italy), where he led a team working on Virtual Reality applied to multi-robots, through distributed information and behavioral models. He was involved in the European Network of Excellence "INTUITION" (05-08), with a focus on Virtual Reality for training. He is a founding member of the European Association of Virtual Reality (EuroVR). He joined Technicolor In 2011, where he is currently leading activities on Augmented Reality, Immersion, and Presence, aiming to create new Technicolor applications and services.

Marc Christie is an Associate Professor at Rennes 1 University and working in the INRIA Mimetic research team. His main center of interest is Interactive Virtual Cinematography. He defended his Ph.D in 2003 (Nantes University) in which he proposed a declarative approach to camera control using interval-based constraint solving techniques and explored aspects related to camera path planning. He contributed to publications in both computer graphics (Eurographics, SCA) and constraint solving (CP series). With Patrick Olivier (Newcastle University), he authored the first state of the art report on the topic (Eurographics 2006), organized a session on camera control at the Smartgraphics 2008 conference, and recently presented a course at Siggraph Asia 2009. Marc Christie is currently working on visibility computation, screen composition editing and interactive virtual and real cinematography.