

# Chapter 4.11

## Face Animation: A Case Study for Multimedia Modeling and Specification Languages

**Ali Arya**

*University of British Columbia, Canada*

**Babak Hamidzadeh**

*University of British Columbia, Canada*

### **ABSTRACT**

This chapter will discuss the multimedia modeling and specification methods, especially in the context of face animation. Personalized Face Animation is and/or can be a major user interface component in modern multimedia systems. After reviewing the related works in this area, we present the ShowFace streaming structure. This structure is based on most widely accepted industry standards in multimedia presentations like MPEG-4 and SMIL and extends them by providing a higher level Face Modeling Language (FML) for modeling and control purposes and by defining image transformations required for certain facial movements. ShowFace establishes a comprehensive framework for face animation consisting of components for parsing the input script, generating and splitting the audio and video “behaviors,” creating the required images and sounds, and eventually displaying or writing

the data to files. This component-based design and scripted behavior make the framework suitable for many purposes including web-based applications.

### **INTRODUCTION**

Specifying the components of a multimedia presentation and their spatial/temporal relations are among basic tasks in multimedia systems. They are necessary when a client asks for a certain presentation to be designed, when a media player receives input to play, and even when a search is done to retrieve an existing multimedia file. In all these cases, the description of the contents can include raw multimedia data (video, audio, etc.) and textual commands and information. Such a description works as a Generalized Encoding, since it represents the multimedia content in a form not necessarily the same as the playback format,

## Face Animation

and is usually more efficient and compact. For instance a textual description of a scene can be a very effective “encoded” version of a multimedia presentation that will be “decoded” by the media player when it recreates the scene.

Face Animation, as a special type of multimedia presentation, has been a challenging subject for many researchers. Advances in computer hardware and software, and also new web-based applications, have helped intensify these research activities, recently. Video conferencing and online services provided by human characters are good examples of the applications using face animation. Personalized Face Animation includes all the information and activities required to create a multimedia presentation resembling a specific person. The input to such a system can be a combination of audio/visual data and textual commands and descriptions. A successful face animation system needs to have efficient yet powerful solutions for providing and displaying the content, i.e., a content description format, decoding algorithms, and finally an architecture to put different components together in a flexible way.

Although new streaming technologies allow real-time download/playback of audio/video data, bandwidth limitation and its efficient usage still are, and probably will be, major issues. This makes a textual description of multimedia presentation (e.g., facial actions) a very effective coding/compression mechanism, provided the visual effects of these actions can be recreated with a minimum acceptable quality. Based on this idea, in face animation, some researches have been done to translate certain facial actions into a predefined set of “codes.” Facial Action Coding System (Ekman & Friesen, 1978) is probably the first successful attempt in this regard. More recently, MPEG-4 standard (Battista, et al., 1999) has defined Face Animation Parameters to encode low-level facial actions like jaw-down, and higher level, more complicated ones like smile.

Efficient use of bandwidth is not the only advantage of multimedia content specifications

like facial action coding. In many cases, the “real” multimedia data does not exist at all and has to be created based on a description of desired actions. This leads to the whole new idea of representing the spatial and temporal relation of the facial actions. In a generalized view, such a description of facial presentation should provide a hierarchical structure with elements ranging from low-level “images,” to simple “moves,” to more complicated “actions,” to complete “stories.” We call this a Structured Content Description, which also requires means of defining capabilities, behavioural templates, dynamic contents, and event/user interaction. Needless to say, compatibility with existing multimedia and web technologies is another fundamental requirement, in this regard.

Having a powerful description and specification mechanism, also is obviously powerful in search applications that currently suffer when looking for multimedia content. MPEG-7 standard (Nack & Lindsay, 1999) is the newest arrival in the group of research projects aiming at a better multimedia retrieval mechanism.

Considering three major issues of Content Delivery, Content Creation, and Content Description, the following features can be assumed as important requirements in a multimedia presentation systems (Arya & Hamidzadeh, 2002):

1. Streaming, i.e., continuously receiving/displaying data
2. Structured Content Description, i.e., a hierarchical way to provide information about the required content from high-level scene description to low-level moves, images, and sounds
3. Content Creation (Generalized Decoding), i.e., creating the displayable content based on the input. This can be decoding a compressed image or making new content based on the provided textual description.
4. Component-based Architecture, i.e., the flexibility to rearrange the system components, and use new ones as long as a certain interface is supported.

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/chapter/face-animation-case-study-multimedia/27131](http://www.igi-global.com/chapter/face-animation-case-study-multimedia/27131)

## Related Content

---

### Semi-Supervised Multimodal Fusion Model for Social Event Detection on Web Image Collections

Zhenguo Yang, Qing Li, Zheng Lu, Yun Ma, Zhiguo Gong, Haiwei Pan and Yangbin Chen (2015). *International Journal of Multimedia Data Engineering and Management* (pp. 1-22).

[www.irma-international.org/article/semi-supervised-multimodal-fusion-model-for-social-event-detection-on-web-image-collections/135514](http://www.irma-international.org/article/semi-supervised-multimodal-fusion-model-for-social-event-detection-on-web-image-collections/135514)

### Data Hiding Schemes Based on Singular Value Decomposition

Victor Pomponiu, Davide Cavagnino, Alessandro Basso and Annamaria Vernone (2010). *Advanced Techniques in Multimedia Watermarking: Image, Video and Audio Applications* (pp. 254-288).

[www.irma-international.org/chapter/data-hiding-schemes-based-singular/43475](http://www.irma-international.org/chapter/data-hiding-schemes-based-singular/43475)

### Stream Dependent Caching

Phillip K.C. Tse (2008). *Multimedia Information Storage and Retrieval: Techniques and Technologies* (pp. 341-367).

[www.irma-international.org/chapter/stream-dependent-caching/27023](http://www.irma-international.org/chapter/stream-dependent-caching/27023)

### SSIM-Based Distortion Estimation for Optimized Video Transmission over Inherently Noisy Channels

Arun Sankisa, Katerina Pandremmenou, Peshala V. Pahalawatta, Lisimachos P. Kondi and Aggelos K. Katsaggelos (2016). *International Journal of Multimedia Data Engineering and Management* (pp. 34-52).

[www.irma-international.org/article/ssim-based-distortion-estimation-for-optimized-video-transmission-over-inherently-noisy-channels/158110](http://www.irma-international.org/article/ssim-based-distortion-estimation-for-optimized-video-transmission-over-inherently-noisy-channels/158110)

### A Layered Multimedia Presentation Database for Distance Learning

Timothy K. Shih (2002). *Multimedia Networking: Technology, Management and Applications* (pp. 151-171).

[www.irma-international.org/chapter/layered-multimedia-presentation-database-distance/27031](http://www.irma-international.org/chapter/layered-multimedia-presentation-database-distance/27031)