LuciaWebGL: A New WebGL-Based Talking Head

Alberto Benin, Piero Cosi, Giuseppe.Riccardo Leone, Giulio Paci[1]

ISTC-CNR, UOS Padova

{ alberto.benin, piero.cosi, riccardo.leone, giulio.paci }@pd.istc.cnr.it

Abstract

In this DEMO we present the first worldwide WebGL implementation of a talking head (LuciaWebGL), and also the first WebGL talking head running on iOS mobile devices (Apple iPhone and iPad).

Index Terms: WebGL, talking head, TTS, MPEG-4, mobile, Android, iOS

1. Introduction

Face-to-face communication is the main element of human-human interaction because both acoustic and visual signal simultaneously convey linguistic, extra-linguistic and paralinguistic information. Therefore, facial animation is a research topic since the early 70's and many different principles, models and animations have been proposed over the years [1]. There are many ways to control a synthetic talking face. Among them, geometric parameterization [2], morphing between target speech shapes [3], muscle and pseudo-muscle models [4],[,5]. In the late 90's a specification for efficient coding of shape and animation of human face was included in the MPEG-4 international standard. The focus was extended from traditional audio and video coding to other multimedia context including images, text, graphics, 3D scenes, animation and synthetic audio. Concerning Facial Animation MPEG-4 standard defines the shape of the model (FDP) and a set of actions (FAP); the animation is obtained by specifying a stream of numbers that is for each frame the values of the Facial Animation Parameters. Many implementations of this standard were born in the last decade [6] as stand alone applications built for research purpose. At ISTC-CNR of Padua we developed LUCIA, a talking head based on an open source facial animation frame-work [7].[8],[9].

More Recently, WebGL ("*Web Graphics Language*") was introduced [10], and currently supported by the major web browsers development companies. WebGL aims to provide web developers the same types of capabilities that are provided to developers of rich 3D desktop and mobile applications, all within the context of a browser. It exposes unprecedented access to the graphics processing hardware available on most platforms today, directly to web developers. What's even better is WebGL isn't a proprietary plugin like Flash or Silverlight that users must install, but rather it's an open standard just like HTML and CSS, and it is essentially a standard programming library to develop 3D computer graphics for the web. Using classical web-programming techniques web developers relied solely on the processing power of the CPU (central processing unit) on a user's computer, in fact, HTML, CSS, and JavaScript processing all happened in the same shared area of the CPU, making it difficult to create immersive 3D interactive experiences. WebGL solves this scalability problem by opening up access to the computer's GPU (Graphics Processing Unit), a piece of hardware specifically tuned to handle the types of mathematical challenges posed by high-performance 3D applications.

In other words, the introduction of WebGL, extending the capability of the JavaScript programming language to generate interactive 3D graphics within any compatible web browser, opens the possibility to bring all these applications to the home computers of a very large number of persons and to burst this natural way of interaction with the machines. The WebGL Working group (including Mozilla, Google, Apple, Opera and more recently Microsoft) started in early 2009; two years after, on February 2011, they released the version 1.0 of WebGL. WebGL is based on OpenGL Embedded System 2.0, the Graphic Library developed for mobile devices. This is a real revolution because it brings the power of 3D graphics directly into web browsers without installing any plug-ins or customized and maybe dangerous software. Any application can run in all the platforms that support the new standard. The easy integration of this technology in any website offers promising future uses for WebGL Avatars: on-line personal assistant, storyteller for web-book, digital tutor for hearing impaired are only few examples.

2. LuciaWebGL

LuciaWebGL [11] is a 3D Italian expressive talking avatar based on the new WebGL technology.

The typical WebGL application is composed by three parts: the standard html code, the main JavaScript program and a new shading language section. The html section is intended mainly for user interaction: check-box, input values and often a debugging zone for errors or warning messages during the execution. The JavaScript part is the core of the application: the graphic library itself, all the matrix manipulation, support and utility functions take place here. The input from the user is connected with JavaScript variables via ad-hoc event- driven procedures. The novelty is the third part, which is the Shading Language code. This software runs on the Video Card. It is called GLSL and it is derived from the C programming language. Actually, these are the instructions that calculate every pixel colour value on the screen whenever the drawing function is called in the JavaScript main program. To be able to change the values of the GLSL variables from the JavaScript WebGL Application Program Interface implements special methods to connect them with JavaScript objects, arrays, and variables. During the initialization of the WebGL page the shader code is compiled and copied to the memory of the Video Card ready to be executed on the Graphic Processing Unit. At the beginning of the connection model parts data are sent on the Internet using the lightweight data-interchange format JSON (JavaScript Object Notation) [12]. This is the only moment where you can wait for a while because of the amount of the data to transmit.

[1] Authors are in alphabetic order.

Lucia is a female talking head working on standard MPEG-4 Facial Animation Parameters (FAPs) [13], speaking with the Italian version of Festival [14] or Mary TTS [15], and it is totally based on true real human data [16]. The 3D kinematics information are used to create lips articulatory model and to drive directly the talking face, generating human facial movements.

We developed a facial animation system based on this library to interact with the user in a bimodal way. The overall system is a client-server application using the http protocol: we have a client (a browser or an app) and a web server. No software download and no plugin are required. All the software reside on the server and the visualization player is delivered inside the html pages that the client ask at the beginning of the connection. On the server side a software called AudioVideo Engine generates the phonemes and visemes information needed for the animation.

Currently, official support has been given for main platforms such as Google Chrome, Firefox, Opera and more recently Microsoft Internet Explorer on desktop PCs and also on Android-based mobile devices. Starting from iOS 5, WebGL is enabled only for the advertisement library class (which is intended for placing ad-banners in applications), and we have been able to use this feature to visualize and animate our WebGL talking head [17], [18].

3. Conclusions

Even if, while nearly all desktops and most smartphones on the market today have dedicated graphics hardware and the ability to support a technology like WebGL, there are still some hurdles to overcome. As with any new open standard, like HTML5, the biggest obstacle is often cross-browser support and compatibility. Thankfully, there are already a handful of browsers with great WebGL support. Chrome, Firefox, and Opera all have good support for WebGL. Safari has support but it hasn’t been enabled quite yet. On the contrary Microsoft had expressed concerns with implementing support for WebGL in their Internet Explorer browser in the past, but in the last IE 11 version WebGL is finally supported.

We have presented here LuciaWebGL, an Italian MPEG-4 standard FAPs driven talking head, which is the first attempt to port talking head technology directly to the web via WebGL. LuciaWebGL, similarly to his predecessor Lucia, has a high quality 3D model and a fine co-articulation model, which is automatically trained by real data, used to animate the face. LuciaWebGL runs on any WebGL compatible browser and now, being the first 3D talking head in the world with such a capability, on iOS mobile devices (iPhone and iPad).

4. Future Work

In the near future, we will integrate speech recognition to have double input channel and we will work on dynamic mesh reduction to enhance the frame rate on slow computational power hardware.

5. Acknowledgements

Parts of the research reported on in this paper were performed in the context of the Italian project “MIUR-FIRB, WIKIMEMO.IT: il Portale della Lingua e della Cultura Italiana”, and “MISE-2015, SPEAKY ACUTATTILE: una nuova piattaforma inclusiva di assistente intelligente vocale multicanale”.

6. References

[1] F. I. Parke, F.I. and Waters, K., “Computer facial animation”. Natick, MA, USA: A. K. Peters, Ltd., 1996.

[2] Massaro, D.W. et al., “Developing and Evaluating Conversational Agents. Embodied Conversational Agents”, MIT Press, Cambridge, USA, 287-318, 2000.

[3] Bregler, C.et al., “Video Rewrite: Driving Visual Speech with Audio”, in Proceedings of SIGGRAPH ’97, 353-360, 1997.

[4] Terzopoulos, D. et al., “Realistic Face Modeling for Animation”, in Proceedings of SIGGRAPH ’95, 55-62, 1995.

[5] Vatikiotis-Bateson, E. et al., “Physiology-Based Synthesis of Audiovisual Speech”, in Proceedings of 4th Speech Production Seminar: Models and Data, 241-244, 1996.

[6] Pandzic, I.S. and Forchheimer, R. Eds., “MPEG-4 Facial Animation: The Standard, Implementation and Applications”, New York, NY, USA: John Wiley & Sons, Inc, 2003.

[7] Cosi, P., Fusaro, A. and Tisato, G., “Lucia a new italian talking-head based on a modified cohen-massaro labial co-articulation model”, in Proceedings of Eurospeech 2003, September 1–4, Geneva, Switzerland, Vol. III, 2269-2272, 2003.

[8] Cosi, P. and Perin, G., “Labial co-articulation modeling for realistic facial animation”, in Proceedings of ICMI ‘02, 4th IEEE International Conference on Multimodal Interfaces, October 14-16, Pittsburgh, PA, USA, 505-510, 2002.

[9] Leone, G.R., Paci, G. and Cosi, P., “LUCIA: An open source 3D expressive avatar for multimodal h.c.i”, in Camurri, A., Costa C., Volpe G.(eds.) Proceedings of INTETAIN2011, Abstract, vol. CD, n.1, 2011.

[10] WebGL, http://www.khronos.org/webgl/.

[11] Leone, G.R. and Cosi, P., “Lucia-WebGL: a web based Italian MPEG-4 Talking Head”, in Salvi G., Beskow J., Engwall, Al Moubayed S. (Eds.), Proceedings of AVSP 2011, Aug 31-Sep 3, Volterra, Italy, 123-126, 2011.

[12] JSON, http://www.json.org/.

[13] Ekman, P. and Friesen, W., “Facial Action Coding System”, in Consulting Psychologist, 1978.

[14] P. Cosi, P., Tesser, F, Gretter, R. and Avesani, C., “Festival speaks Italian!”, in Proceedings of Eurospeech 2001, September 3-7, Aalborg, Denmark, 509-512, 2001.

[15] Schröder, M. and Trouvain, J., “The German text-to-speech synthesis system MARY: A tool for research, development and teaching”, International Journal of Speech Technology, October 2003, Volume 6, Issue 4, pp 365-377, 2003.

[16] Ferrigno, G. and Pedotti, A., “Elite: A digital dedicated hardware system for movement analysis via real-time tv signal processing”, in IEEE Trans Biomed Eng., Nov 32(11), 943-50, 1985.

[17] Benin A., Cosi P., Leone G.R., “A webGL talking head for mobile devices”, in Proceedings of AVI 2012, Advanced Visual Interfaces International Working Conference, May 21-25, Capri Island (Naples), Italy, 788-789, 2012.

[18] Benin A., Leone G.R., Cosi P., “A 3D talking head for mobile devices based on unofficial iOS WebGL support”, in Proceedings 17th International Conference on 3D Web Technology, August 04-05, 2012, Los Angeles, CA – USA, 2012.