Eurographics Italian Chapter Conference (2007)
Raffaele De Amicis and Giuseppe Conti (Editors)

Modelling and Animation of Theatrical Greek Masks in an Authoring System

P.A. Bertacchini[1], E. Bilotta[1], P. Pantano[2], S. Battiato[3], M. Cronin[1], G. Di Blasi[1], A. Talarico[1], A. Tavernise[1]

[1]Department of Linguistics, University of Calabria, Rende, Italy
[2]Department of Mathematics, University of Calabria, Rende, Italy
[3]Department of Mathematics and Computer Science, University of Catania, Catania, Italy

Abstract

This paper describes an authoring system for modelling Greek virtual masks that express emotions for synchronising their facial movements with pre-recorded speech files, and for creating theatrical performances. The system is related to parametric modelling and includes three interfaces: Editor, Recorder and Virtual Theatre. In the Editor interface it is possible to create several different 3D masks from a unique mask basic model. In the Recorder it is possible to import the models created in the Editor and select eight expressions (neutral, anger, surprise, sadness, fear, joy, disgust, attention), as well as to create and save little alterations in these expressions, and to synchronise the facial movements with speech on the time-line of the system. In the Virtual Theatre the animated Greek masks can be imported and can perform.

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Curve, surface, solid and object representation I.3.7 [Computer Graphics]: Animation

1. Introduction

Theatrical masks were the most important features of actors in Greek theatre; they covered the head and had a mouth opening to amplify the voice. Because of the fragility of the original masks, their features are known thanks to mosaic works, marble reproductions, pictures and small terracotta models. On Lipari (one of the Aeolian Islands), one of the last Greek settlements in Sicily, much evidence of terracotta production was found (little models of masks and little sculptures) [RID04]. These findings represent the three literary genres of Greek theatrical performances (tragedy, comedy and satire) and the largest documentary evidence of theatrical masks from the classical age [BBC01].
While there are many "real" reconstructions of Greek masks, "virtual" modelling of them is not so common and there is very little literature about their animation [TGB05]. Some experiments of modelling have been made using little models of terracotta masks from the Lipari collection held at the Kelvingrove Museum in Glasgow. These have been digitised and outputted by using Apple QuickTime VR object movies. Using a commercial non-contact photogrammetric system (Eyetronics Shapesnatcher, http://www.eyetronics.com/), the artefacts (typically 6-8 cm high) were captured and modelled by means of around 100,000 polygons, with a texture file of width up to 4,098 pixels. Rapid prototyping of replica objects had been undertaken by using a Z-corporation 3-dimensional printer [WIL04]. From these 3D models "real" masks were reconstructed and used in theatrical performances.
Regarding the masks' animation with facial expressions, starting from Parke's work [PAR74] much research has sought to realise a realistic modelling of facial movements in anthropomorphic interfaces. The task was one of the most difficult in Computer Graphics, as even the most minute changes in facial expression can reveal complex moods and emotions. Several authors explored a model-based approach for animating faces [PAR82], [CJH03], [MCBC00], [MLCP06], while other focused on an image-based approach, using either video of an actor [BV99], [BBPV03], [CJH03], or speech [BCS97], [BRA01], [EGP02] to drive the animation [ZSCS04]. 3D face tracking techniques addressed this problem by computing the deformation of a deformable 3D face model to a sequence of images [EBDP96], [PSS99], [BOP98], [DCM02], [BBPV03] or 3D marker positions [GGW*98].

However, the problem of accurately modelling facial expressions and other dynamic behaviour is still in its infancy: current shape capture technology, and in particular laser-scanners and most other high-resolution shape capture techniques do not operate effectively on fast moving scenes (a transition to a smile can occur in a fraction of a second) [ZSCS04]. In the case of the modelling of Greek masks, complexity increases, since the objects present caricatured grotesque shapes very far from human facial anatomy.

In this paper, we present an authoring system which allows fast transitions and that is related to parametric modelling and to the "responsive face" realised by Perlin [PER97]. In fact the "expressive character with minimal geometry" [PER97] was considered a good starting point to guarantee both realism and responsive facial expressions without using repetitive pre-built animations. Moreover, Perlin's system uses:

1. the "noise" function, that is random movements of head and eyes;
2. a mask model with a low resolution;
3. a small number of parameters to animate the facial expressions.

The next three Sections explain how real Greek masks were analysed, 3D modelled and animated in our authoring system. The system architecture is summarised in Section 5.

2. Analysis of the Masks

We analysed the characteristics of the masks introduced in Menander's New Comedy. Here we present six masks selected from the terracotta models that were found on Lipari (Figure 1), representing some characters got from the ancient scholar Pollux' categorization [BB98]: a coaxer (*kólax*), a young man with fluttering hair (*oûlos neaniskos*), the main slave (*hegemòn therápon*), a second slave (*káto trichías*) and a girl (*kóre*). Because of the lack of old men's masks (*Eappos prates*), a mask from a prior age, that could represent the character, was chosen.

These masks were selected in order to play a comedy: we choose Plauto's *Miles Gloriosus*, derived from one of the most famous in Menander's fragmented work. The script of the comedy was simplified and adapted for better comprehension by the public at large and the facial expressions of the masks were selected on the basis of the speech. After that the speech of the comedy was performed by an actor and recorded.

We used the Facial Action Coding System (FACS) [EFH02] to analyse the basic facial expressions of the selected masks. In fact, FACS categorizes all possible movements of the facial musculature that produce a visible change in the face; in particular, each discrete movement is called Action Unit (AU) and the activation of different AU's results in a combination that makes up a facial expression [VAN98]. So we identified the Action Units in the masks' expressions with the FACS identification number (Figure 2).

Figure 1: *The masks of a coaxer, a young man with fluttering hair, the main slave, a second slave, a girl and an old man [BBC01].*

Using the AU identification, we detected the necessary vertices in the parametrized model, for the modelling of:

Figure 2: *Selected masks with the FACS identification number of the Action Units.*

1. the basic expression;
2. the facial expressions to be implemented (Table 1).

3. Modelling of Greek Masks

A generic basic model of a Greek mask (Figure 3), based on the coaxer's mask, was adopted as a starting point for the generation of every mask.

Using a series of heuristic strategies, a low number of vertexes (131) to reconstruct each of Menander's masks was set. The model was created in an external modelling system (3D Studio Max) and imported in the Editor interface (Face3DEditor) of the authoring system (Figure 4).

Some vertices of the basic model were changed and set up to create all the other models of Greek masks (Figures 5, 6,

Emotion	AUs
Joy	6+12+25
Sadness	1+4+15
Disgust	10+17+4
Anger	4+5+7+24
Surprise	1+2+5+26
Fear	1+2+4+5+20+25
Attention	4+25

Table 1: *AUs relative to the expressions to implement in the authoring system.*

Figure 3: *The basic model in the authoring system based on the coaxer's mask.*

7, 8, 9 show the original mask, the model in Editor interface and the corresponding virtual mask with a neutral expression). The colour in the Greek production of "real" masks (not in the terracotta models) was studied and the nearest to the original chosen.
All the virtual masks maintain one of the basic characteristic of the "real" ones, that make a mask powerful and lifelike: the asymmetries and the different moods that it presents from different angles or in different lights (Figure 15), which make it seem to change expression as it moves [WIL04].

4. Animation and Theatrical Performance of Masks

In the Recorder interface (Face3DRecorder) of the authoring system it is possible to select eight standard facial expressions (neutral, anger, surprise, sadness, fear, joy, disgust, attention) for every created mask (Figure 10), as well as to create and save small alterations in these expressions. The linear transition from a facial expression to another is decided by the user and in the future other "soft" transitions can be inserted. We tested the recognition of the eight standard facial expressions created by Face3DEditor. A mask with different facial expressions was imported in SuperLab LT®software and 15 subjects had the task to recognise the displayed emotions. The software recorded the correct recognition in the following percentage: neutral (87%), anger (80%), surprise (80%), sadness (93%), fear (80%), joy (67.7%), disgust (33.3%) and attention (40%). Moreover, reaction time showed the subjects' difficulty in the recognition of attention, while the shortest time was recorded for the recognition of fear [BBG*06].

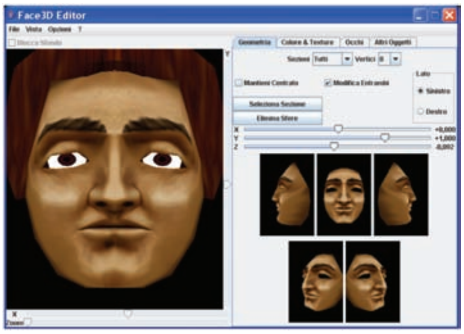

Figure 4: *The Editor interface in the authoring system.*

In the Recorder interface (Face3DRecorder) of the authoring system it is also possible to synchronise facial movements with speech on the time-line of the system. Through this time-line the recording of the theatrical text can be matched to the facial expressions chosen by the user (Figure 11). The result is very different from the text to speech technique.
In the Virtual theatre interface (Figure 12), the user can import the synchronizations obtained in Face3DRecorder, deciding the position of the masks and the time in which they start to talk to each other. In this way the user can edit the performances of different masks and the result can be transformed into movie files.

5. System Architecture

The software was implemented in Java2 Standard Edition 1.5.0 for the GUI components, in Java3D 1.3.1 for the 3D environment and uses Apple's QuickTime for Java (QT4J) API Library for the audio/video engine. The system was tested on several computers with different hardware and software configurations (PC Athlon64 with Windows XP and iBook G4 with MacOS X), giving in each configuration good performances.
The Java/Java3D choice gives us many interesting advantages:

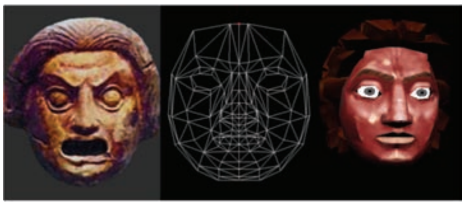

Figure 5: *The mask of a young man.*

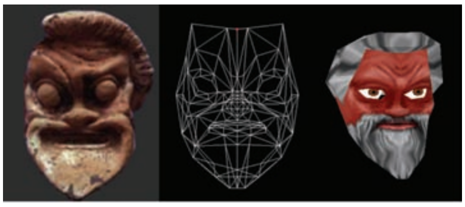

Figure 6: *The mask of the main slave.*

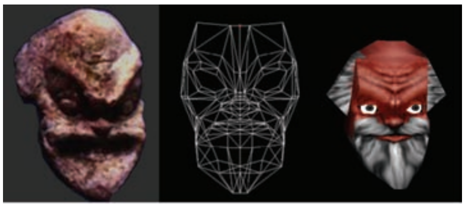

Figure 7: *The mask of a slave.*

Figure 8: *The mask of a girl.*

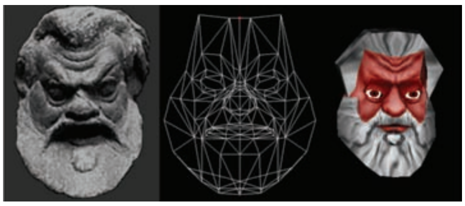

Figure 9: *The mask of an old man.*

- the software is ready for any platform implementing a JVM, the Java3D and the QuickTime for Java extensions;
- the software is able to run on any hardware configuration (desktop monitor, ultra-wall, CAVE, virtual reality environment, etc.) with few simple changes;
- the scene graph programming model provided by the Java3D library allows us a simple and easy to manage and maintain environment, ready for any kind of extension and application in several research and development field (virtual theatre, multi-modal interfaces, etc.).

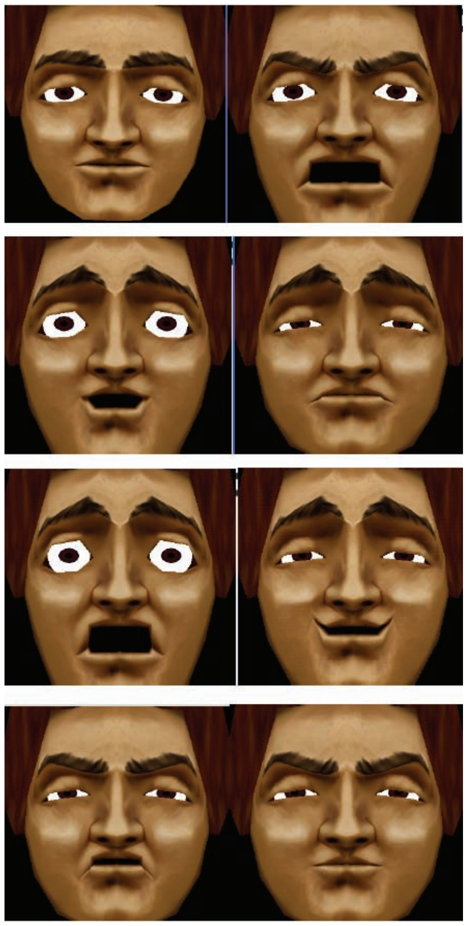

Figure 10: *The facial expressions to select in the Recorder interface.*

QuickTime for Java provides us with the power of the Apple's QuickTime (QT) engine for the audio and video components of the software. Unfortunately the QT engine is only available for the Windows and MacOS X platforms. Another choice for the audio/video engine can be the Java Media Framework, but this API is only available for the Windows and Linux platforms. We plan in the future the development of a platform independent version of the audio/video engine of our software.
The core of our system is the Face3D class; it provides all base functionality for the creation, manipulation and animation of any kind of virtual mask. In particular it provides the methods for:

- manipulating any single vertex of the mask as well as a set of vertices managed as a rigid body;

Figure 11: *The synchronisation of facial movements with the speech on the time-line of the system.*

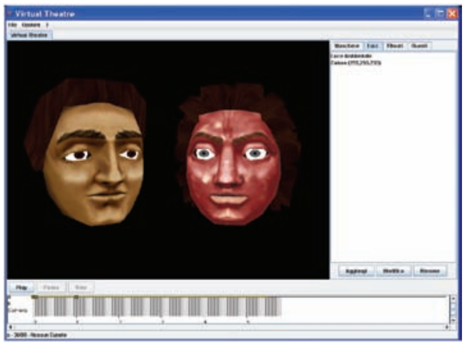

Figure 12: *The Virtual Theatre interface in the authoring system.*

- setting a colour, a texture image, a shading mode and the wire-frame visualization;
- activating a random noise [PER97], giving the mask more natural behaviour for eye and head movements;
- selecting a particular facial expression (joy, fear, sadness, and so on) by an array of floating values.

Starting from the Face3D base class, the user can easily extend it by implementing few abstract methods defining the various movements of the mask (such as mouth width and height, smile, sneer, etc.).
The second base class of our system is the Face3DUtils class; it provides a set of utility methods to simplify the Face3D creation and manipulation. For any extension of the Face3D class, the user has to implement an extension of the Face3DUtils class. The classes Face3DComponent, Face3DEditor and Face3DRecorder perform the visual manipulation and animation of the mask. Face3DComponent is the link between the logic component of the software (Face3D and Face3DUtils classes) and the GUI component. It provides a visual environment in which the user can manipulate the mask by moving the vertices in order to create his own kind of mask as well as a set of methods to interact with the user selection performed by the GUI (change of view and projection, resizing, etc.). The Face3DEditor and Face3DRecorder are the classes used to interact with the user.
The Face3DEditor allows the user to create a new mask starting from a predefined model. Technically it contains all the needed references to the Face3D class to manipulate it in an easy way. It allows one the vertices selection and movement by an easy GUI and colour/texture choice.
The main window of Face3DEditor shows the environment by which it is possible to perform the modelling. In the "Geometry" panel the user can modify the position of a vertex or a set of vertices; it is possible to select the vertices of a specific area of the face (eyes, mouth, etc.).
In the "Colour and Texture" panel it is possible to select the colour of the model and to load the textures (Figure 13). It is also possible to insert a background image (photo or drawing) as a reference image to modify the base model (Figure 14). Finally, some options allow the user to visualize the model in wire-frame mode and in "Flat" or "Gouraud" shading.
The Face3DRecorder is devoted to the setting or creation of facial expressions and to speech synchronization. The main window is subdivided into two parts: on the left we find the 3D model, on the right we find the commands for facial expressions.

Figure 13: *Examples of different textures.*

Figure 14: *An example of a background image in "Face3DEditor".*

Facial expressions are set by simple commands or created by a set of sliders indicating several face components such as eye, mouth, etc. It is possible to create a new expression by simply accurately moving these sliders.
Speech synchronization is performed by a recorder/player; the GUI shows the user a timeline panel indicating the audio file sequence. The user can insert at any point of the timeline a "key-frame" indicating a facial expression and the transition time between the previous expression and the next one. Speech simulation is performed by a pseudo-random movement of the mouth mask driven by some simple heuristic parameters chosen by the user. Moreover it is possible to activate the "noise" function to perform small movements (such as eye movement).
Finally, it is important to note that the software does not use any complex data structures or algorithms.
The Virtual Theatre environment allows the creation of a complete virtual comedy. In particular the user can select:

1. which masks to add to the virtual representation;
2. the audio files and the facial expressions for each added mask;
3. the lights used to illuminate the scene;
4. a random noise for the masks.

Technically the system uses the same metaphor already used in the Face Recorder, the software provides the user with a timeline panel indicating the sequences of all audio files involved in the comedy and the related buttons to start, pause and stop the representation. Moreover, the user can also easily provide a "key frame" in order to simulate the camera and mask moving inside the 3D scene.

6. Conclusions and Future Works

In this paper we have presented an authoring system for the modelling and animation of Greek masks to be used as "talking heads" in virtual theatrical plays for educational aims.
The system has three interactive windows, one to create the Greek masks, one to animate them and the last to edit a performance. In the Face3DEditor the user can model and save a mask, in the Face3DRecorder it is possible to animate it through a timeline, and in the Virtual Theatre interface the masks can interact like in a real performance; in particular:

1. the module Face3DEditor allows an interactive manipulation of the basic model with a fixed number of vertices. In fact, it is possible to modify the position of the vertices on the x, y and z axis, obtaining new faces;
2. the module Face3DRecorder does not generate automatic animations based on texts, but it is possible to synchronise facial expressions with pre-recorded speech files. The user selects the audio file and a window shows a timeline panel on which he can match facial expressions. Speech simulation is performed by a pseudo-random movement of the mouth mask, driven by some heuristic parameters, determined on the basis of physiological studies on human facial movements. In the time-line the user selects the step in which transition between two facial expressions begins and he can decide the duration of this transition. The selected facial expression is visualized in the timeline by a red point;
3. the module Virtual Theatre allows the editing of the performance and the creation of movies.

In conclusion, the presented authoring system represents:

1. the first system specifically developed for Greek mask modelling and animation;
2. an alternative style of cultural promotion.

Moreover, its main advantages are:

1. the use of a low number of vertices in the parameterization for the creation of Greek masks;
2. the ease of use of the environment.

During the development of the system we have found several possible improvements to the software, in particular:

1. the creation of an interactive system by which it is possible to reconstruct any kind of face, like an identikit tool;
2. the development of labial modelling for a more realistic facial animation;
3. the 3D virtual modelling of the body and the creation of avatars using ancient characters;
4. the creation of a virtual theatre system for representing Greek comedies and tragedies;
5. a "shape from shading" system for direct manipulation of a mask starting from a single photo;
6. the development of a speech synthesiser suitable for Greek tragedies starting from an XML file (SpeechML, for example), containing the information about the expressions;
7. a scripting system to allow the creation of batch animation sequences by a simple text file.

Finally, we foresee the use of the application in several fields (video games, multi-modal interfaces, didactics, and so on).

Acknowledgement

This work has been supported by the Research Program of National Interest (PRIN) project, co-financed by MIUR 2004 - prot. 2004111320. Furthermore, it is related to NETConnect (Connecting European Culture through New Technology) project, financed by Culture 2000 European program.

References

[BB98] BERNABO' BREA L.: *Le maschere ellenistiche della tragedia greca*. Centre Jean Bérard, Napoli, 1998.

[BBC01] BERNABO' BREA L., CAVALIER M.: *Maschere e personaggi del teatro greco nelle terracotte liparesi*. L'Erma di Bretschneider, 2001.

[BBG*06] BERTACCHINI P., BILOTTA E., GABRIELE L., SERVIDIO R., TAVERNISE A.: Il riconoscimento delle emozioni in modelli facciali 3d. In *Proc. AIP Congress - Italian Association of Psychology, Experimental Psychology Section* (2006).

[BBPV03] BLANZ V., BASSO C., POGGIO T., VETTER T.: Reanimating faces in images and video. In *Proc. EUROGRAPHICS2003* (2003).

[BCS97] BREGLER C., COVELL M., SLANEY M.: Video rewrite - visual speech synthesis from video. In *Proc. SIGGRAPH1997* (1997).

[BOP98] BASU S., OLIVER N., PENTLAND A.: 3d lip shapes from video - a combined physical-statistical model. *Speech Communication 26*, 1 (1998), 131–148.

[BRA01] BRAND M.: Morphable 3d models from video. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* (2001).

[BV99] BLANZ V., VETTER T.: A morphable model for the synthesis of 3d faces. In *Proc. SIGGRAPH1999* (1999).

[CJH03] CHAI J., JIN X., HODGINS J.: Vision-based control of 3d facial animation. In *Proc. of EUROGRAPHICS/SIGGRAPH Symposium on Computer Animation* (2003).

[DCM02] DE CARLO D., METAXAS D.: Adjusting shape parameters using model based optical flow residuals. *IEEE/TPAMI 24*, 6 (2002), 814–823.

[EBDP96] ESSA I., BASU S., DARRELL T., PENTLAND A.: Modeling, tracking and interactive animation of faces and heads using input from video. In *Proc. IEEE Computer Animation* (1996).

[EFH02] EKMAN P., FRIESEN W., HAGER J.: *The Facial Action Coding System*. Weidenfeld Nicolson, 2002.

[EGP02] EZZAT T., GEIGER G., POGGIO T.: Trainable videorealistic speech animation. In *Proc. SIGGRAPH2002* (2002).

[GGW*98] GUENTER B., GRIMM C., WOOD D., MALVAR H., PIGHIN F.: Making faces. In *Proc. SIGGRAPH1998* (1998).

[MCBC00] MASSARO D., COHEN M., BESKOW J., COLE R.: *Developing and Evaluating Conversational Agents*. CASSELL J., SULLIVAN J., PREVOST S., CHURCHILL E.: Embodied Conversational Agents. MIT Press, Cambridge, MA, 2000.

[MLCP06] MASSARO D., LIU Y., CHEN T., PERFETTI C.: A multilingual embodied conversational agent for tutoring speech and language learning. In *Proc. Ninth International Conference on Spoken Language Processing* (2006).

[PAR74] PARKE F.: A parametric model for human faces. *Technical Report UTEC-CSc-75-047* (1974).

[PAR82] PARKE F.: Parametrized models for facial animation. *IEEE Computer Graphics 2*, 9 (1982), 61–68.

[PER97] PERLIN K.: Layered compositing of facial expression. In *ACM/SIGGRAPH Technical Sketch, http://mrl.nyu.edu/ perlin/experiments/facedemo* (1997).

[PSS99] PIGHIN F., SALESIN D., SZELISKI R.: Resynthesizing facial animation through 3d model-based tracking. In *Proc. Int. Conf. on Computer Vision* (1999).

[RID04] RIDGWAY D.: The archaelogy of the aeolian islands. *The Classical Review 54*, 1 (2004), 214–215.

[TGB05] TAVERNISE A., GABRIELE L., BERTACCHINI P.: Simulazioni di agenti in un teatro greco. In *Proc. II Workshop Italiano di Vita Artificiale* (2005).

[VAN98] VANGER P.: Assessment of facial activity. *Methods of Psychological Research 3*, 1 (1998).

[WIL04] WILLIAMS R.: New comedy in performance project. *Digicult.Info - http://www.digicult.info/downloads/html/ 1071657904/1071657904-4.html 9* (2004), 4–7.

[ZSCS04] ZHANG L., SNAVELY N., CURLESS B., SEITZ S.: Spacetime faces: High resolution capture for modeling and animation. In *Proc. SIGGRAPH2004* (2004).

Figure 15: *The selected mask from different angles and in different lights.*