Eye, Lip And Crying Expression For Virtual Human

Mahardhika Candra Prasetyahadi, Itimad Raheem Ali, Ahmad Hoirul Basori, and Nadzaari Saari

Abstract—Eye movement, lip and crying expression are facial elements that able to give realistic expression on virtual human, especially on transferring emotion expression from human to virtual human. Eye expression can be described through standard approach such as MPEG4 and FACS, while lip synchronization needs Speech API to be described into appropriate meaning. Crying is another aspect of expression which is considered as extreme expression and it involve interaction between fluids(tears) with skin surface. In this paper, the critical analysis of the study will be discussed in detail based on previous works. The content is expected to give clear picture of research focus on eye, lip and Crying behavior in the future research.

Index Terms—Crying Expression, Lip Synchronization, Eye Behavior, Facial Expression, Facial Animation.

1 Introduction

THE animated virtual human faces have been used in many applications, such as movies, games, and embodied conversational agent's (ECAs). Eye movements, lip synchronization and crying expression all are able to give information about what we are saying and feeling, and help to communicate emotions. The crying and eye behavior are tightly coupled with human aware process. People focus on eye and lip expression to read human behavior. Emotion is proven to play an important role in human coherent facial behavior is very important to increase the credibility of the character [1]. Yet only few researches have explicitly considered a virtual human or agents capable of expressing its emotional states. With the development of facial animation platform, we found necessary for new research to generate the face expression automatically. Actually, it is very efficient to create visualize facial emotion and more realistic.

We observed some requirements to obtain a robust facial animation platform, a specific set of parameters get a satisfactory control of facial attributes for many faces, mesh deformation algorithms that produce behavior of virtual human. In our work, we describe an interactive facial animation framework that takes eye behavior and its crying, synchronized speech and facial expression to define the character action as a high-level description.

Displaying extreme expressions is a complicated task. It requires precise muscle modeling and skin deformation, and in the case of crying: real-time fluid simulation that realistically interacts with the face [2].

2 Related Work

2.1 Review Stage

Recently, there is an emerging number of making realistic facial animation but it's still one of the most challenging tasks in spite of extensive research. In literature we can efforts to produce a hingher-level parameter approach to construct facial animations in an effective way [1,3,4].

Yotusukura et al. [2003] aim to create a general purpose toolkit for building an easily customizable anthropomorphic an agent, and focus on the construction of an agent's face image synthesis. Then developed the face image synthesis module (FSM) that can be used by any skill level of users [3].

Shih et al. [2010] describes a real time a speaker's lip shape is synchronized with the corresponding speech signal, use sum of absolute difference (SAD) as vowel lip shape likelihood to cluster into categories an then adjust the source and destination pictures of lip shape in the transparent level using alpha blending for lip sync animation [4].

Aleksandra et al. [2010] presented an approach to implement a behavior realizer compatible with Behavior Markup Language (BML), the system based on hierarchical controllers which apply preprocessed behaviors to body modalities. Then described a novel solution to the issue of synchronizing gestures with synthesized speech using neural network, and propose improvement, to the BML specification [5].

Fagel and Bailly [2011], observe two goal, one of the experiment is to measure the change in behavior with respect to gaze when one interaction is wearing dark glasses and his/her gaze is not visible by other one, the second goal is to collect data on the multimodal behavior of one of the subjects by mean of audio recording, eye gaze head motion tracking in order to build a model than can be used to control a robot in comparable scenario in future experiments [6].

- *M.C. Prasetyahadi is with the UTM VicubeLab, Universiti Teknologi Malaysia, Malaysia, Johore 81310. E-mail: mahardhika@cs.its.ac.id.*
- *I.R. Aliis with the UTM VicubeLab, Universiti Teknologi Malaysia, Malaysia, Johore 81310. E-mail: weffee@yahoo.com.*
- *A. H. Basori is with the UTM VicubeLab, Universiti Teknologi Malaysia, Malaysia, Johore 81310. E-mail: uchiha.hoirul@gmail.com.*
- *N. Saari is with the UTM VicubeLab, Universiti Teknologi Malaysia, Malaysia, Johore 81310. E-mail: nadzaris@yahoo.com.*

Serra et al. [2012] present a modular visual speech animation framework by creating the first fully automatic system capable of generating visual speech for European Portuguese based on concatenation of visemes [7].

Li, Z. and Mao, X. [2012] proposed an Emotional Eye Movement Markup Language (EEMML) which is an emotional eye movement animation scripting tool that capable the authors to generate emotional eye movement in virtual agents. This system designed to interact large human-agent or agent-agent[8].

Cassell J. et al. [2001] proposed a Behavior Expression Animation Toolkit (BEAT) allows animators to input typed text that wish to be spoken by an animated human figure, and obtain as output a synchronized nonverbal behaviors and synthesized speech in a form that can be sent to animation system [9]. Marriott A. et al. [2002], proposed a Virtual Human Markup Language (VHML), allow to gathers several language such as gesture markup language, emotion markup language and facial animation markup languages providing tag structure for facial and body animation, gesture, speech, emotion and so on. This language is XML-based and coves different abstraction levels [10].

In psychology, research has been done on why people cry. This research is important if we want to identify the parameters of a situation in which someone cries. Being able to describe the circumstances of someone who is crying can be used for several things. The current emotion of a person, for example, provides a lot of information about their current facial expression. Also, reproducing a certain situation through the use of parameters can be used as a more elaborate control mechanism for crying. There have been several studies that investigate why people cry, and these have been summarised by Vingerhoets et al. [11]. In particular, we would like to mention Borgquist [12] who did a study among students, which pointed out three mood states in which crying occurs: anger, grief or sadness, and joy. He also pointed out accompanying physical states such as fatigue, stress and pain.

2.2 Facial Animation

As we have discussed in the previous section, emotions are an important part of conveying a realistic crying face. The most important visual effect of an emotion is the facial expression. To generate a facial expression we need at least a deformable model of a face. This can be either a model based on splines or be a traditional polygonal mesh. The latter is more conventional, although there are facial animation methods that are based on spline deformation [13]. Usually, computer games use a polygonal mesh to represent models. These models used to be animated by morph vertex animation (or per-vertex animation). This meant that an animation was stored in the mesh as a series of vertex positions, and was played frame by frame like a movie. These different positions were often created by attaching a skeleton to the mesh, and then creating the different key frames by putting the skeleton in different poses. This system was computationally very inexpensive, since it just had to play an animation frame by frame. These days, skeletal animation is more common to deform meshes in video games. A system of rigid bones is attached to the model and stored with the model. Each bone has a weighted influence on some vertices, and new positions of a vertex are calculated by a weighted blending of the bones affecting a vertex. This is shown in Figure 2.1. The bones can be animated with skeletal animation files, describing the motion of each bone. This method is more flexible and memory effcient, but is computationally more expensive. There exist some methods that use spline-based skeletal animation, as opposed to a rigid skeletal system, such as [14, 15].

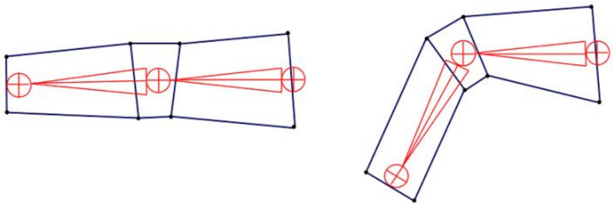

Figure 2.1: A simple model of an arm displaying bone animation. The bones have a weighted influence on the vertices. Note that the vertices of the elbow are affected by both bones [2].

Another method of animating a mesh is by using pose animation [16]. This actually resembles vertex animation in that it stores the movement of a vertex as a key frame in the mesh. But in pose animation a key frame, called a pose, describes the offset of one or more vertices to their original positions. When creating several poses, these poses can be blended because they are offsets of the same set of vertices. For example, a mesh can store a two poses of the same face, one with an angry expression and one with a sad expression. An example of pose blending can be seen in Figure 2.2. These poses can then be blended to show an expression of an emotion that is a mixture between sad and angry. This method is already used for facial animation [16].

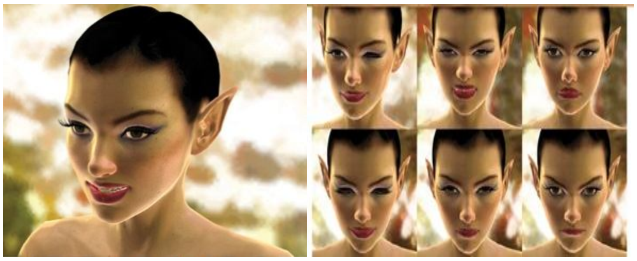

Figure 2.2: Pose Blend Animation. The face on the left is the neutral pose. The next two faces express anger and sadness. The face on the right is a mixture between anger and sadness. Notice the ”angry” eyebrows and the ”sad” mouth [2].

3 CRITICAL OBSERVATION OF FACIAL ANIMATION

Virtual world, games and embodied conversational agents (ECAs) importance of nonverbal behaviors. Eye movement, lip synthesize combined with the gesture facial expression and body way all give realization human manner, so cohesive facial behavior increasing the veracity of the character.

3.1 Eye Behavior in Facial Animation

The interested work is done by lance and Marsella, who make a model for emotionally expressive head and body movement during eye movement based on Gaze Warping. Transformation (GWT), which is combining between temporal scaling and spatial transformation parameter that describe the emotional expressive eye movement shift [17].

Lance and Marsella then proposed a model of virtualization emotion expressive eye behavior that builds by GWT by improving the implementation [18].

3.2 Lip Synchronization for Facial Animation

The interested work is done by lance and Marsella, who make a model for emotionally expressive head and body movement during eye movement based on Gaze Warping. Transformation (GWT), which is combining between temporal scaling and spatial transformation parameter that describe the emotional expressive eye movement shift [17].

It is important to construct a model that can produce the ability of utterances to more communication between humans including speech and nonverbal behaviors. Speech is represented as sequence of discrete sounds or phone [19], not only composition of sound but also a specific articulatory movement facial expression. But to date, the speech animation is not done well because the position and orientation of the visible part comprising the lips, teeth, jaw, tongue and cheeks. All articulators can affect the production of a given phone but not all change is visible. Must of systems uses a set of visems that activated by a text-to-speech engine (TTS). TTS engine convert an utterance in text format into a series of word sounds and poses.

Shapiro [2011] used a simple system scheme by creating a one-to-one mapping between phonemes and visemes. Each phoneme would referring to its corresponding viseme and be phase in and phase out by overlapping the phase out period of one viseme with the phase-in period of a second viseme. The system merge a set of the Facial Action Coding System (FACS) units, which can be used to emotional expression and display facial movements unrelated to speech [10].

Serra et al [2012] presented a modular visual speech animation framework aimed to speeding up and casing the visual speech animation. They are develop automatic visual speech automation system European Portuguese based on the concatenation of visemes. Then they present the results evaluation that was bringing to estimate the quality of two different phoneme-to-viseme mapping devised for the language.

The main contribution of researcher is the creation of automatic system capable of generating visual speech for European Portuguese. This is challenging to generate a 3D character facial movement spoken like virtual human [Serra 2012].

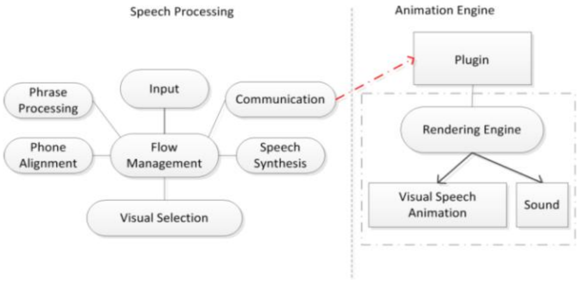

Figure 3.1: Framework architecture overview. The framework is divided into the speech processing component (left) and the plug-in embedded in the animation engine (right) [7].

Lip synchronization module follows the methodology described in Rossana [2009]. Basically the module develops a robust facial animation platform. The module use MPEG-4 facial animation standard for face an animation parameterization. MPEG-4 defined 84 feature points placed on the character head. The restriction information about 3D face model is called facial definite parameters (FDPs). The adapted model is animated by the meaning of FAPs [11]

Every module has FAPV and every FAP value relevant by a corresponding FAPV construct an independent model for MPEG-4 PLAYER [11]

X-face as the MPEG-4 FA engine used by [Rosana 2009] play the FAP animation generated by animation engine module. X-face toolkit is incorporating four piece of software, first the X-face core library, which enables researcher to embed 3D facial animation in their application. Second, X-face Ed authoring tool generate MPEG-4 parameters meshes. Third, X- face player, a sample application that quantized the toolkit in action. Fourth, X-face client, Which allows remote network control of the X-face player.

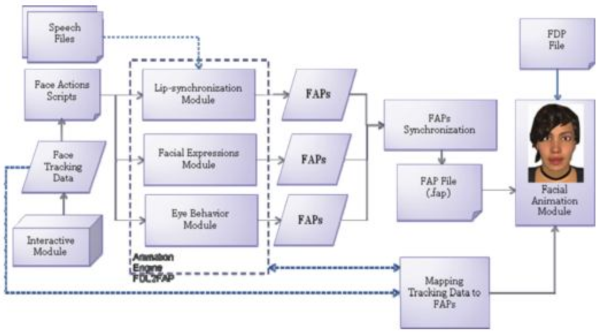

Figure 3.2: Overall architecture diagram of the framework [1].

This framework solves three types of conflicts: viseme expression, eye expression and Head control. This module proposed an interactive application “the virtual mirror” by creates the animation directly instead.

X-face it is a set of open source tools for creating talking head using MPEG-4 and key frame based animation driven by SMIL Agent scripting language by [Balci 2007] aimed to support free and open source tools for research. Then actually, for each face model. It’s necessary to prepare asset of key meshes with the different facial expression and visemes automatically through facial movements scripts with lip synchronization [Rodrigues 2007]

A lip sync system based on voice driven method is designed to animate the realistic spoken face given the text based on the input text as [Shih 2012] do three parts in the experiment single words recognition via SVM before and after phonemes classification. Then estimate the word error rate in real time system.

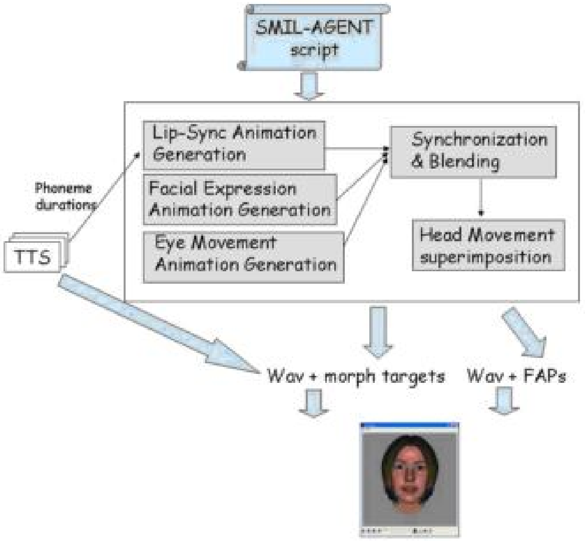
SMIL-AGENT script
Lip-Sync Animation Generation
Facial Expression Animation Generation
Eye Movement Animation Generation
Synchronization & Blending
Head Movement superimposition
Phoneme durations
TTS
Wav + morph targets
Wav + FAPs

Figure 3.2: SMIL-Agent script processing [20].

The use of kernel-based month shape clustering algorithm inspired based on one class support vector machines. This system provide an information to an talking face in order to more realistic smoothed lip motions in real time voice driven talking face lip sync systems.

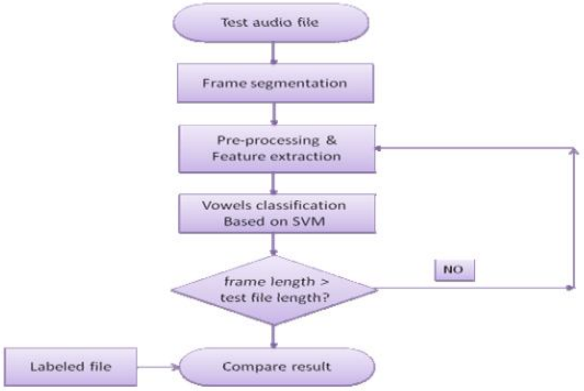
Test audio file
Frame segmentation
Pre-processing & Feature extraction
Vowels classification Based on SVM
frame length > test file length?
NO
Labeled file
Compare result

Figure 3.3: Block diagram of experimental setup [4].

6 DISCUSSION AND CONCLUSION

In this paper we describe the important aspect of virtual characters animation, these including eye movements, speech synthesis, nonverbal behaviors, crying and eye saccade. Our paper intends to implement a character animation yield high levels of realism and control.

It’s challenging to generate 3D character based on speech or eye behavior and obtaining a character like virtual human behavior.

In the future study we intend to improve the animation by finding a solution to the difficulty in implementation of articulation problem and representation of the realism tears added to the eye movements.

ACKNOWLEDGMENT

This research is supported by the Ministry of Science and Technology (MOSTI) and collaboration with Research Management Center (RMC), Universiti Teknologi Malaysia (UTM). This paper is financial supported by E-Science Grant Vot. No.: R.J130000.7911.4S026

REFERENCES

[1] Queiroz, R. B., M. Cohen, et al. (2010). "An extensible framework for interactive facial animation with facial expressions, lip synchronization and eye behavior." Comput. Entertain. 7(4): 1-20.

[2] Tol, W. and A. Egges (2009). Real-Time Crying Simulation. Proceedings of the 9th International Conference on Intelligent Virtual Agents. Amsterdam, The Netherlands, Springer-Verlag: 215-228.

[3] Yotsukura, T., S. Morishima, et al. (2003). Model-based talking face synthesis for anthropomorphic spoken dialog agent system. Proceedings of the eleventh ACM international conference on Multimedia. Berkeley, CA, USA, ACM: 351-354.

[4] Shih, P.-Y., J.-F. Wang, et al. (2010). Kernel-Based Lip Shape Clustering with Phoneme Recognition for Real-Time Voice Driven Talking Face. Advances in Neural Networks - ISNN 2010. L. Zhang, B.-L. Lu and J. Kwok, Springer Berlin Heidelberg. 6064: 516-523

[5] Aleksandra, #268, et al. (2010). A controller-based animation system for synchronizing and realizing human-like conversational behaviors. Proceedings of the Second international conference on Development of Multimodal Interfaces: active Listening and Synchrony. Dublin, Ireland, Springer-Verlag: 80-9.

[6] Fagel, S., G\, et al. (2011). Speech, gaze and head motion in a face-to-face collaborative task. Proceedings of the Third COST 2102 international training school conference on Toward autonomous, adaptive, and context-aware multimodal interfaces: theoretical and practical issues. Caserta, Italy, Springer-Verlag: 256-264.

[7] Serra, J., M. Ribeiro, et al. (2012). A Proposal for a Visual Speech Animation System for European Portuguese. Advances in

Speech and Language Technologies for Iberian Languages. D. Torre Toledano, A. Ortega Giménez, A. Teixeiraet al, Springer Berlin Heidelberg: 267-276.

[8] Li, Z. and X. Mao (2012). "EEMML: the emotional eye movement animation toolkit." Multimedia Tools and Applications 60(1): 181-201.

[9] Cassell, J., H. H, et al. (2001). BEAT: the Behavior Expression Animation Toolkit. Proceedings of the 28th annual conference on Computer graphics and interactive techniques, ACM: 477-486.

[10] Marriott A, Stallo J (2002) Vhml—uncertainties and problems, a discussion. In: Proceeding of AAMAS'02 workshop on ECA-Let's specify and evaluate them. Italy.

[11] A. J. J. M. Vingerhoets. Adult Crying, A Biopsychosocial Approach. Psychology Press,2001.

[12] A. Borgquist. Crying. American Journal of Psychology, 17:149–205, 1906.

[13] M. Hoch, G. Fleischmann, and B. Girod. Modeling and animation of facial expressions based on b-splines. The Visual Computer, 11,(2):87–95, May / June 1994.

[14] E. Chuang and C.Bregler. Performance driven facial animation using blendshape interpolation. 2002.

[15] S. Forstmann and J. Ohya. Skeletal animation by spline aligned deformation on the gpu.IEICE technical report. Image engineering, 106(608):47–52, 2007.

[16] S. Forstmann, J. Ohya, A. Krohn-Grimberghe, and R. McDougall. Deformation styles for spline-based skeletal animation. In SCA '07: Proceedings of the 2007 ACM SIG-GRAPH / Eurographics symposium on Computer animation, pages 141–150, Aire-la-Ville,Switzerland, Switzerland, 2007. Eurographics Association.

[17] Lance, B. and S. C. Marsella (2007). Emotionally Expressive Head and Body Movement During Gaze Shifts. Proceedings of the 7th international conference on Intelligent Virtual Agents. Paris, France, Springer-Verlag: 72-85.

[18] Lance, B. J. and S. C. Marsella (2008). A model of gaze for the purpose of emotional expression in virtual embodied agents. Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems - Volume 1. Estoril, Portugal, International Foundation for Autonomous Agents and Multiagent Systems: 199-206.

[19] Ostendorf, M.: Moving Beyond the "Beads-on-a-String" Model of Speech. In: Proc. IEEE ASRU Workshop, pp. 79-84. IEEE Press (1999).

[20] Balci, K., Not, E., Zancanaro, M., and Pianesi, F. 2007. Xface: Open source project and smil-agent scripting language for creating and animating embodied conversational agents. In Proceedings of the 15th International Conference on Multimedia (MULTIMEDIA'07). ACM, New York, 1013–1016.

[21] Rodrigues,P.S.L. 2007. UmSistema de Geracao de Express˜ Faciais 3D com Processamento de Fala. Ph.D. dissertation, Pontif´Rio de Janeiro.

M.C. Prasetyahadi obtained his bachelor degree in Informatics Engineering from Institut Teknologi Sepuluh Nopember Surabaya, Indonesia, Currently; he is a Masters by research student at Software Engginering, Faculty of Computing, Universiti Teknologi Malaysia. He is a member of ViCube Lab research group at, Faculty of Computing, Universiti Teknologi Malaysia. His research interests include facial animation and computer vision.

I.R. Ali She is a Masters by research student at Software Engginering, Faculty of Computing, Universiti Teknologi Malaysia. She is a member of Vicube Lab research group at, Faculty of Computing, Universiti Teknologi Malaysia.

A.H.Basori received the B.Sc. from Department of Informatics, Institut Teknologi Sepuluh Nopember Surabaya, Indonesia (2004). Currently, he is a PhD student at Department of Computer Graphics and Multimedia, Faculty of Computer Science and Information System, Universiti Teknologi Malaysia. He is supervised by Dr. Abdullah Bade and Dr. Mohd Shahrizal Sunar. His research interests include avatar, human emotion and haptic tactile rendering.