Emotional Facial Expressions, Eye Behaviors, Lips Synchronization: Current and Future Direction

[1]Itimad Raheem Ali, [2]Ahmad Hoirul Basori, [3]Ghazali Sulong

[*1,] Corresponding Author PHD Researcher in UTM Vicube Lab, Faculty of Computing, Universiti Teknologi Malaysia, weffee@yahoo.com
[2]UTM Vicube Lab, Faculty of Computing, Universiti Teknologi Malaysia,
Interactive Media and Human Interface Lab., Department of Informatics, Institut Teknologi Sepuluh Nopember Surabaya, Indonesia, uchiha.hoirul@gmail.com
[3]UTM VicubeLab, Faculty of Computing, Universiti Teknologi Malaysia, Ghazali@spaceutm.edu.my

Abstract: In facial animation, the research trend is an eye gaze and voice modeling. However, the incorporation of eye movement, lip synchronization, and facial expression are necessary for a more realistic face model. This paper provides a thorough survey of facial animation with respect to eye movement, lip synchronization, and facial expression research. The significance of this survey paper is to provide a review of existing literature, giving insight to various approaches that have been proposed for problems in facial realistic modeling. More also, this survey is to open-up a new research direction in facial animation.
[Itimad Raheem Ali, Ahmad Hoirul Basori, Ghazali Sulong. **Emotional Facial Expressions, Eye Behaviors, Lips Synchronization: Current and Future Direction.** *Life Sci J* 2014;11(6):171-181]. (ISSN:1097-8135). http://www.lifesciencesite.com. 24

Keywords: Realistic modeling; facial expression; Eye movement; Lip synchronization; Emotion.

1. Introduction

Animated virtual character has been widely used in movies, games and embodied conversational agents (ECAs) in recent years to provide effective human computer interaction. Facial expression is very important to communicate the emotional manifestation of a character. New research on facial animation has led to an understanding of animated virtual character that helps to explore expressiveness, communication, and interactivity which are mainly focused on ECA development (Rossana B. Queiroz et al., 2009).The combination between eye movement, lip synchronization, gesture, emotional facial expression, and body orientation provide information about the flow of ideas, sequences of thoughts in decision making and depth of understanding and knowing. The gaze and saccadic eye movements tell a lot about the thinking process in a human mind. They are often referred to "window to the mind". Eye movements blended with the gaze conveyed significant nonverbal information and emotional intentions when a person speaks. Previous researchers have tried to realize the behavior of embodied expression such as BEAT (The Behavior Expression Animation Toolkit) which allows the virtual character to speak from input text written by animator through nonverbal expressive behaviors and synthesized speech (Cassell et al., 2001).

Realistic facial animation still remains a challenging task despite extensive research. This is because the description of emotion is an important process in human intelligence. It is important to generate virtual characters that can produce realistic utterances with correct synchronized facial animations involving close-to-nature lip movements and face expressions. For example, a frame containing a virtual character talking while at the same time looking at another object. Thus, creating interesting speaking and gazing scenes are important problem areas in animation research. This is a significant challenge as nonverbal behavior can be complex (Quené et al., 2012) (Schuller et al., 2013) (Gonseth et al., 2013). This paper clarifies the comprehensive survey to recent researches and emerging trendsin realistic of virtual human characters and its cover four important aspects in realistic modeling of virtual human character: (1) lip synchronization (2) eye behavior (3) emotion (4) facial expression. We surveyed different type of face synthesis model considered to image, video, and features information. This paper is divided into six sections. In section 2, important issues in the virtual reality of virtual human character are discussed. Section 3 survey in detail the facial animation and facial expression methods and its importance in areas like face detection and face recognition. Section 4 discussed the eye and lip synchronization for realistic expression. Section 5 discussed in details the facial modeling and animation techniques. Finally, important discussion and conclusions are drawn in section 6.

2. Virtual Reality

A realistic virtual human facial animation is yet to be realized, although virtual reality has existed since 1972. Understanding the human facial

expressions and emotions is a difficult process because it translates the complex human facial gestures and emotions to realistically model human conversations. Attempts have been made in the past to model the facial gestures and emotions in real time but the complexities with the facial emotions modeling of real characters is still an issue. In an effort to produce realistic virtual human facial animation, various researchers developed a face image synthesis model (FSM), which generates embodied agents based on multimodal dialog integration, speech synthesis, speech recognition and face image synthesis (Yotsukura et al., 2003).

Xface Open Source Project and SMIL-Agent Scripting Language are mostly used for creating and animating embodied conversational agents (ECAs). These set of tools (please refer to Figure 1) can be implemented, used and extended easily (Balcı et al., 2007).

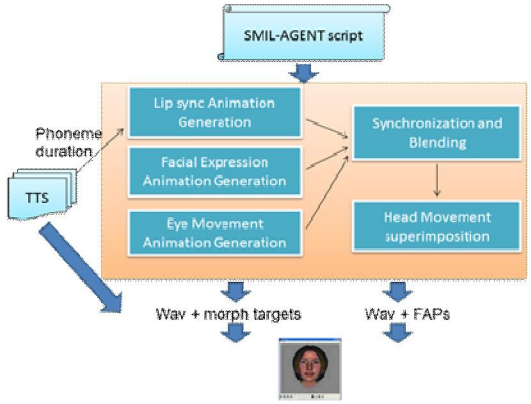

Figure 1. SMIL-Agent script processing (Balcı et al., 2007).

Queiroz et al (2009) developed a usable, extendable, and robust facial animation platform for MPEG-4 where parameterized face with a high-level description of facial actions and behaviors was achieved through an interaction between the user and the virtual character. Their method is diagrammatically shown in Figure 2[1]. Bailly et al. (2010) investigated the audio and visual face-to-face interaction between human to human and between human to virtual conversational agent. They aimed at determining the selection impact of award states and communicative functions (Bailly et al., 2010).

Using back propagation neural network (BNN), Cerekovic et al (2010) were able to model human body movements (Čereković et al., 2010). Gillies et al (2010) presented a real time multimodal interaction to the virtual character animation system in virtual reality setting (Marco Gillies et al., 2010). Lee et al (S. Lee et al., 2010) designed a lifelike responsive avatar framework (LRAF), to express avatar for real human process. They also analyzed the efficacy of the expressive avatars in real time (Sangyoon Lee et al., 2010). According to Shapiro (2011), a system for the movement of virtual characters should include a set of important aspects of the simulation character models and games. Through his findings obtained from human social communication research, he was able to put a high level of realism and control in his system for the movement of virtual characters (Shapiro 2011). Extending the work on body movement carried out in 2010, Cerekovic et al (2011) also applied neural networks (NN) for multiplatform real-actor animation system that mimics real time behavior like gestures and speech synchronization (Čereković and Pandžić, 2011). Real progress in multiplatform real-actor animation system was achieved when the production of fast performance realistic characters through human intervention became available (Kipp et al., 2010) (Rossana Baptista Queiroz et al., 2010) (Preda and Jovanova, 2013).

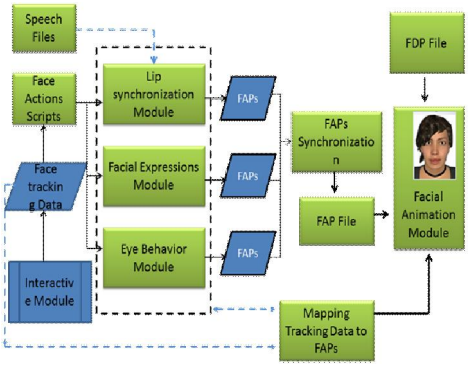

Figure 2. Overall architecture diagram of Queiroz's framework (Rossana B. Queiroz et al., 2009).

3. Facial Animation and Facial Expression

Facial expression comprises of human emotions, intent, and other verbal and nonverbal expressions. The facial expression synthesis and analysis is a vital aspect of computer science, which is realized with computer vision and graphics. Very little is done so far on a realistic model that incorporates speech gestures and facial expressions. Despite previous efforts in developing the coding system, realistic modeling in this area is still an issue. The methodologies that addresses the formation of an emotive 3D humanoid audio visual character that is believable to advance the interaction of human computer to a more realistic human-human interaction is yet to be formed (Fu et al., 2010). Facial Action Coding System (FACS) was the most exhaustive and the informed of the MPEG4 facial animation parameters that enables automated synthesis of facial actions and expression (Jeffrey F

Cohn, 2010). The previous efforts in this area were made possible in the development of psychology of coding systems to produce emotive speech and facial expression, which describes the facial action and behaviors (El Ayadi et al., 2011).

It is appropriate to make a virtual character similar to human communication in order to reflect emotive speech and facial expressions. Ekman et al proposed a theory, which is based on psychology for facial expression. Their approach is well-known and is the simplest approach that describes emotional state (Ekman and Friesen, 1978). Suwa presented a discussion on automatic facial analysis from a sequencing image (Suwa et al., 1978). Darwin worked with emotion, motivation, behavior or personality, sensations and cognitive processes (Darwin, 2005).

Ekman used a set of Action Units (AUs) for parameterization of facial expression (Ekman and Friesen, 1978) (Lien et al., 1998). This Action Units (AUs) defines facial expression in small regions. Most pervious approaches are based on 2D motion. However, it does not provide any control to emotion. According to Cao et al animating 3D face model requires the organization of the data structure so that efficient location of appropriate movement is determined. And this cannot be actualized without appropriate search algorithm such as the Support Vector Machine (SVM) for an automated detection of emotion of arbitrary input utterances. Conveying speech alongside gesture or expression is difficult, the question however is how can a realistic virtual human that incorporates emotions as well as gaze and speech behaviors be achieved (Cao et al., 2005). According to Brent Lance and Stacy Marsella gaze is the change in movement of physical parameters, such as the head speed in a gaze shift. Modeling gaze is a challenge since gaze is a complex behavior that includes eye movements, posture, head movements, all added together (Lance et al., 2007). These constitute the ECAs, which are meant to reflect the multimodal nature in human conversation that contains verbal and nonverbal expression. Modeling the nonverbal behavior such as personality, sex, emotion can be extremely complex (Stone et al., 2004) (J. Lee and S. Marsella, 2006) (Pelachaud, 2009).

Facial expressions have a diversified application importance in areas like face detection and face recognition. Facial expression analysis can be visualized for application in areas (Dulguerov et al., 1999) like face image compression and synthetic face animation (Avaro, 2000). In facial expression recognition, the classification of facial motions and features into groups that are simply based on visual information, human emotions such as voice, gestures, pose, gaze, and expressions are crucial for efficiency. However, facial expression consists of a lot of aspects that includes emotions, which often demands understanding of a given person status with regards to the expression (Dornaika et al., 2013)(Fasel and Luettin, 2003). Neji explored the possibility of the use of effective virtual communication (Ben Ammar et al., 2007). Oyarzun et al provided a 3D virtual presenter embedded in real time TV, it is a mixed realistic prototype called PUPPET (Oyarzun et al., 2010). Narendra Patel and Mukesh Zaveri generated a 3D face model from an image of a face that synthesizes expressions with aim to present a fully automatic, robust, and fast system (Patel and Zaveri, 2010). Yang proposed a robust expression transfer (NET) model and presents an animation method to transfer facial expressions extracted from video to the facial sketches for facial expression recognition (Yang et al., 2011).

4. Eye, Lip Synchronization for Realistic Expression

The eye gaze and lip synchronization is a necessary component in human communication, and thus plays a major role in the production of realistic conversations between human and virtual characters. The features in emotional facial expression are very important to capture the reality of the virtual characters. The voice carries much emotional information and the speech can be represented as a sequence of phones, each phone can be associated with a visual representation of the phoneme viseme. The animated visemes depend on the locality of the lip, jaw and tongue in a given phoneme. These techniques are popular for generating real time speech animation system that is able to project personality and interactive emotions. With the development of communication systems, it is now possible for people to interact directly with real time video devices. Such a communication system allows synchronization of the lip shape character with the corresponding speech in a real time voice driven mobile device designed by Shih et al (P. Shih et al., 2010).

The challenge is to maintain synchronization between the body movement and the expected normal human behavior. The eye movement is an important part of face to face conversation which carries the nonverbal information and emotional intent. Yarbus (1967) showed that pattern eye gaze accompanied by verbal instructions conveys intention to the observer through sequence of images (A.L. Yarbus, 1967). Cohen proposed an Eye Direction Detection (EDD) and an Intentional Detector (ID) as basic components of Shared Attention Mechanism (SAM) (Baron-Cohen et al., 1985).

The digital production of realistic eye movement will require an emotional eye movement animation scripting tool such as markup language (EEMML) which was developed by Zheng Li et al (Z. Li and Mao, 2011). Yotsukurai et al (2011) improved the animation scripting tool of Zheng Li et al when he created (FSM) Face image Synthesis Module which is a general toolkit for building an easily customize embodied agent based on multimodal integrated dialog, speech synthesis, speech recognition and face image synthesis (Salvati and Anjyo, 2011).

Virtual character is achieved by combining the eye gaze, lip shapes, and expressions. The lip movements and voice should be synchronized to provide realistic lip-synchronization animation. Normally higher-level module provides the synchronization between the two modules to obtain the capabilities used in spoken dialog. Most systems use a set of visemes that are activated by a text-to-speech engine (TTS). The TTS engine translates an utterance in text format into a series of phonemes. This technique is used to generate a realistic speech animation without having to manually set the positions for a set of visemes (C. Lee et al., 2011).

Serra et al (2012) presented a visual speech animation module aimed at speeding up the dialog of virtual human and he assessed the quality of phonemes- to- viseme mappings devised for the English language (Serra et al., 2012). He discovered that the mechanism for lip synchronization could be carried out by decomposing the speech into a set of phonemes. These phonemes could then be represented as a set of visemes. The relationship between the phonemes in the signal and the visemes in the database is used to construct the appropriate lip shape, general functional diagram of a synthesizer shown in Figure 3. Cognitive demand has a significant effect on the human hearing. On the talking head system, Athanasopoulos et al (2011) proposed a talking head system using facial expressions which were generated from a Partial Differential Equation (PDE) (Athanasopoulos et al., 2011).

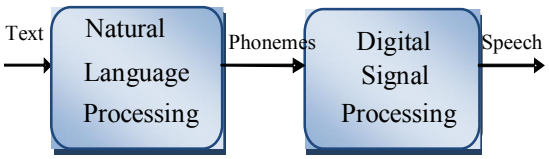

Figure 3.General function diagram of the TTS system (Athanasopoulos et al., 2011).

It is necessary to reduce the size of the data in virtual worlds as small as possible to allow real time rendering. To generate facial expressions it is necessary to store embedded emotional expressions in a database (Richard S. Wallace, 2004). Exchange of glances during speech is an important signal for continuity depending on the behaviors which can display hidden intentions (Tomasello et al., 2005).

From the eye gaze patterns during conversation, circumstances surrounding the event can be perceived and recognized. Most virtual characters are not provided with the features to derive meaning from the verbal and nonverbal communication gestures during conversation. Many types of animation can be based on animation models containing real time multimodal interaction with the virtual character animation system in a virtual reality setting (Kowler, 2011) (Marco Gillies et al., 2010). In the animation scene, Yun-Feng Chou and Zen-Cung Shih (2010) proposed an effective real time multimodal interaction model that included body movement, expressive facial animation synchronized with speech to generate virtual character from still images (Chou and Z.-C. Shih, 2010). This model also depicted the overall eye movement system architecture and animation procedure, as illustrate in Figure 4. Zhao (2012) investigated the characteristics of pre-saccadic shifts of attention in perception which can support shared attention mechanism (Zhao et al., 2012). With the extended of speech realistic visualization, Wang et al (2012) designed a talking head system with synthesized dynamics articulator at the phoneme level and explored the distinctions of producing the sounds using HMM-based synthesis to perform a Maximum Likelihood Parameter Generation algorithm for smoothing (Lan Wang et al., 2012).

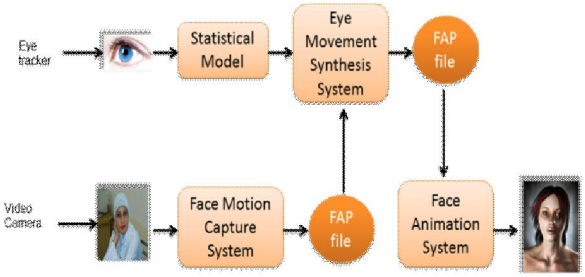

Figure4.Overall eye movement system architecture (Chou and Z.-C. Shih, 2010).

5. Facial Modeling and Animation Techniques

Realistic and credible real time automated animation is necessary in the creation of virtual human character. This type of animation is important in many present day applications which include games, virtual agents, and virtual reality and movie animations (Itimad et al., 2012).It is also essential in applications which require the interaction between the human and computer. The important factors which have to be considered to produce a realistic

face model include: face geometry, facial muscle behavior, emotional behaviors, eye gaze, lip synchronization, and texture synthesis. Methods for the simulation of face shape and facial muscle behavior are available in many publications (Stephen M. Platt Norman I. Badler., 1981)(Parke, 1982) (Waters, 1987) (Kalra et al., 1991).Many researchers developed a more realistic face models such as those needed in the design of digital games (Brand, 1999) (Guenter et al., 1998) (Frédéric Pighin et al., 1998) (Frédéric Pighin et al., 1998)(Seongah Chin, Chung yean Lee, 2011). Facial skin motion or behavior using motion capture is one of the automated geometric methods to model movements of faces in virtual human character (Dornaika et al., 2013) (Douglas DeCarlo et al., 1998) (Blanz and Vetter, 1999) (W. Lee and L. Magnenat-Thalmann, 2000) (Petajan, 1999). In the structuring of facial skin motion, the algorithm assumes that the skin comprised of several layers to give flexibility for the skin to move freely over the muscles and bones underneath. In automated geometric face motion model the skull determines the face shape proportions.

There is a growing interest in eye movement for face-to-face communication (Argyle, M. Cook, 1976). Pelachaud described that every communication gestures contain 60% gaze and 30% mutual gaze (Catherine Pelachaus, 1996). Colburn et.al (2000) and Garau et.al (2001) presented a simulation of eye gaze patterns of the interlocutor by analyzing the mutual gaze (Colburn et al., 2000) (Garau et al., 2001). In an effort to make virtual character realistic, Cassell et al. (1994) and Chopra Khullar et al. (2001) examined eye attachment during interactions or conversation (Cassell et al., 1994) (Chopra Khullar and N I Badler, 2001). D'Mello et al. at 2012 developed an intelligent tutoring system (ITS) with the help of commercial eye tracker (D'Mello et al., 2012). The eye movement is associated with behavior which is employed to bring attention in a variable environment (Roel Vertegaal et al., 2000). The gaze signal has been used to simulate the face-to-face communication to create dynamism in the conversation (R Vertegaal et al., 2001) (Kowler, 2011). Much of the studies focused on facial and eye movements of virtual human characters from the perspective of the user's perception. However, there is a big difference between dynamic and multiple gaze instances. The use of these gaze instances is important for multiple applications, acknowledgment, and attention.

It is a difficult to classify the facial modeling and animation techniques, since there is no delimiter that separates techniques used so far in virtual animation. With that regards, we consider discussing the various existing techniques on an approach based methodology.

5.1. FACS approach

Facial action coding system (FACS) was first developed by Ekman (1978) to describe all facial movements. It was not initially intended for use in animation. Action units (AUs) are the basic actions in FACS. AUs describe the 46 points in the face where contraction of facial muscles or group of muscles can occur, and the combination of these units generates facial expressions. FACS has found a wide range of use in virtual characters realization (Kopp et al., 2011), for instance in the work by Ari Shapiro, they used FACS to describe a system for the movement of virtual characters, which included simulation character models and games for social research where a high level of realism and control is achieved (Shapiro, 2011). Čereković et al. at 2010 designed a real actor (RA) system by considering embodied conversational agents (ECA) to achieve multimodal behavior realization. They implemented a solution in their design for speech concurrency with gesture using neural networks and manual adjusting of the face animation model to synthesize face expressions (Čereković et al., 2010). Lijuan Wang et al. (2010) unified the multimodal human behavior in generating ECA (Lijuan Wang et al., 2010). The current state-of-the-art for facial description (either FACS itself or muscle-control versions of FACS) has two major weaknesses:

- The action units are purely local spatial patterns. Real facial motion is rarely completely localized; Ekman himself has described some of these action units as an "unnatural" type of facial movement.
- There is no time component of the description, but only a heuristic one. It is known that most facial actions occur in three distinct phases: which are the phases describing the state of application, release, and relaxation.

Other limitations of FACS include the inability to describe fine eye and lip motions, and the inability to describe the co-articulation effects found most commonly in speech. Although the muscle-based models used in computer graphics have alleviated some of these problems, they are still too simple to accurately describe real facial motion.

Using lexicons of human perception Zhang et al. (2013) was able to connect facial expression to intention and to dialogue. Since human perception is normally structured semantically, it can be broken down into different recognizable elements with the help of Latent Semantic Analysis which allows the analytical system to use semantic structures and linguistic restriction for the identification of embedded conversation (L. Zhang et al., 2013).

Figure 5 shows some of the more simple action units of the main codes observed on the face.

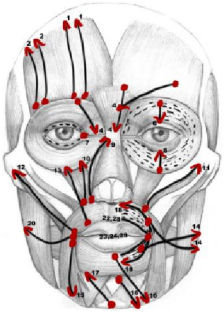

Figure 5. Some of the more simple action units main codes shown on the face. The red arrows directions are the directions of the AU movement while the red points are the center of mass of what's being moved (Ekman and Friesen, 1978).

5.2. Deformation approach

A face model without the capabilities of deforming is passive and has little role in virtual humans. The deformation approach considers that facial movements which are deformations of the facial features that appear naturalistic and realistic. The surface of the facial mesh often produces high-quality movement that exhibits facial expressions. This is possible since the facial mesh can be manipulated using a function which operates on a small elemental mesh unit. For the manipulation to be effective the facial muscle structure is ignored. The hierarchical splines provide local improvements to the surface of B-spline and add new patches with the selected area to create expressions. Hierarchical B-splines are a compact and economical way to represent the surface of spline and achieve high rendering speed. Muscles combined with the hierarchical spline surface are able to create hierarchical puffy skin surface and a variety of facial expressions (Thiebaux et al., 2008).

5.3. Physics muscle approach

Facial muscles are thin, voluntary and move according to the movement of the subcutaneous tissues. Different region of the face have different subcutaneous tissues. There are three types of facial muscles in terms of their actions: linear/ parallel muscles, elliptical/ circular sphincter muscles and sheet muscles. Platt and Badler (1981) used the human face structure to create muscle modeling. The muscle arc was applied to flexible meshes in order to generate many facial expressions (Stephen M. Platt Norman I. Badler., 1981). Martino (2007) used the context dependent viseme to evaluate the efficiency of conveying speech information to a speech synchronize facial animation system. The quality of the viseme representation in this approach is to some extent dependent on the number and distribution of fiduciary points (Martino, 2007).

5.4. Performance driven approach

This method is basically used where the animation of the human face is hindered by inaccurate movement tracking, which causes difficulty in controlling the facial animation. This problem is significant in real time scenarios such as movies, where it is vital to create interactivity of animations in a way that combines motion and expressions naturally. Zhigang and Neumann (2008) synthesized a system for expressive speech animation with controls at the phoneme level. They generated visualization and interaction phoneme clusters that are composed of several facial motion frames (Deng and Neumann, 2008). Marcus Thiebaux et al (2008) introduced smart body (SB) which is an open source modular platform for animating ECAs in real time while Gregor Hofer et al (2008) introduced a novel technique to automatically synthesize lip motion trajectories (Thiebaux et al., 2008) (Hofer et al., 2008) . Zoric et al (2010) improved the work of Gregor Hofer by generating animation trajectories for desired lip animation in speech using speech signals (Zoric et al., 2010).

5.5. MPEG-4 approach

Moving picture expert group (MPEG-4) is a method introduced in 1998 and is used in geometry coding and for transmitting animation parameters. The three dimension (3D) face model has many definition and animation parameters, MPEG-4 specifies the animation by defining these parameters named, which are face definition parameters (FDP) and facial animation parameters (FAP). FDPs include information for building 3D face geometry, and FAPs are designed to encrypt the animation face emotions, expressions, and speech pronunciation. Usually the face contains 68 parameters and these parameters are classified into 10 groups depending on the parts of the face. FAPs represent most of the facial expression. The parameter set contains the viseme and the expression. The units that are responsible for the face parameters animation is called face animation parameters units (FAPU).

When the face is in the neutral state FAPUs can be calculated from the distances between major facial features. KorayBalci et al. (2007) created the embodied conversational agents (ECA) Xface tools using MPEG4 and key-frame driving with the aid of SMIL- Agent scripting language. The ECA developed by KorayBalci is able to mimic facial expressions associated with emotions where gesturing is absent (Balcı et al., 2007). It was necessary for Zoric et al (2010) to incorporate facial gesturing in real time using MPEG4 to animate speech automatically and realistically [80]. However,

the system developed by Zoric does not display emotions. To overcome this problem, LoicKessous et al. (2009) suggested an automatic emotion recognition which can be used in scenes where speech is in progress. Although Kessous used Baysian classifier to identify and recognize different types of emotions for different gesturing, the animation produced was still unable to produce human-like face movements (Kessous et al., 2009). Arsov et al (2010) introduced several human-like face movements by implementing a real time interaction system that is able to offer practical systems for learning and practice of Cued Speech (CS) for online environments (Arsov et al., 2010). Figure 6 shows Part of the facial feature points defined in MPEG-4.

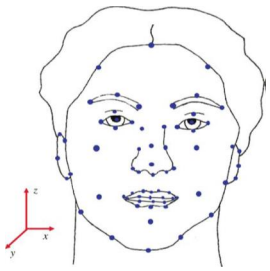

Figure 6. Part of the facial feature point defined in the MPEG-4 standard (Shapiro, 2011)

The face object specified by MPEG-4 is a representation of the human face structured in a way that the visual manifestations of speech are intelligible while the facial expressions allow recognition of the speaker's mood so that the animation is as good as the movement of a real speaker. To fulfill these objectives, MPEG-4 specifies three types of facial data.

1) Facial Animation Parameters (FAPs): FAPs allow one to animate a 3D facial model at the receiver. The way by which this model is made available at the receiver is not relevant. FAPs allow the animation of key feature points in the model, independently or in groups, as well as the reproduction of visemes and expressions.

2) Facial Definition Parameters (FDPs): FDPs allow one to configure the 3D facial model to be used at the receiver, either by adapting a previously available model or by sending a new model. The new or the adapted model is then animated by means of FAPs.

3) FAP Interpolation Table (FIT): FIT allows one to define the interpolation rules for the FAPs that have to be interpolated at the decoder. The 3D model then is animated using the FAPs sent and the FAPs interpolated according to the FIT.

5.6. Visual speech approach

Human speech and facial expressions go hand in hand because when a person is speaking the face expression is in motion. For instance, the mouth movement upon speech adds to the facial expression but these expressions are difficult to imitate or model especially for a recorded acoustic speech. The reason being that the different languages comprises of large vocabulary, large number of phonemes and also speech co-articulation. For accurate modeling it is vital that there is synchronism between vocabulary, phonemes, and speech co- articulation (Frederic and Keith, 2008) (Skantze and Moubayed, 2012).

Jose Mario De Martino and Fabio Violaro evaluated how the efficacy of speech synchronizes with facial animation when the speech information is conveyed unto the animation model (Martino, 2007). Gabriel Skantze and Samer AlMoubayed introduced an IrisTK – toolkit specifically for rapid development of real time systems for multi-party face-to-face interaction (Skantze and Moubayed, 2012). Berger and Hofer developed a general purpose ECA with modular architecture aptly named SAIBA which represents Situation, Agent, Intention, Behavior, and Animation (SAIBA). SAIBA framework defines three levels of abstraction that starts from computation of the agent's communicative intention up to behavior planning and realization. It allows fewer possible customizations when applied to different agent technologies and on a variety of media (Bevacqua et al., 2011). They developed a form of human computer interaction which allows communication of thoughts, intention, and knowledge via nonverbal sounds like hissing and buzzing. They termed the combination of speech and computer facial animation "carnival" (Berger and Hofer, 2011). The aim of carnival is to integrate real time speech processing with facial animation in an object oriented environment (Vinayagamoorthy et al., 2006).

5.7. Facial gesture approach

Facial gesture constitutes facial movements, which is brought about by the changes in the muscles of the face, the movements of the face through the movement of the head, and miscellaneous facial expressions (Iverson and Goldin-Meadow, 1998). John R. Leigh (2006) developed virtual agent platform to coordinate robot gestures with speech that displayed the full gestures in human communication. He called this platform "GRETA" (Greta is the core of an MPEG-4 decoder and is compatible with standard "Simple Facial Animation Object Profile") (John R. Leigh, 2006).

GorankaZoric presented a real time method for automatic speech driven facial gesture using MPEG-4 approach (Zoric et al., 2010). Aleksandra

Cerekovic and Igor S. Pandzic studied realistic ECA behaviors associated with speech and nonverbal behaviors using the RealActor multi-platform animation system. Their solution was implemented using neural networks for a more adaptive face animation of synchronized gestures and speech .Anh et al. (2011) used gesture and speech modeling to communicate human expressive ideas and discovered through their experiments that this model resulted in remarkable speech information transfer and understanding (Anh and Pelachaud, 2011).

5.8. Eye animation approach

Although the eye plays an important role in the interpretation of verbal communication behavior, it has not been researched on extensively. The eye is an embodiment of continuity in conversation and emotional expressions. For this reason, gaze is an important interaction between humans (in real life) and between virtual human characters (in virtual reality). Previous research work focused on the perception of the user or on face-to-face communication instead of establishing methods for conversational agent modeling. For example, saccades which depict fast movements of the eyes from one position to another became the model interest (S. P. Lee and Norman I Badler, 2001). Brent Lance and Stacy C. Marsella (2008) initiated a method that explores an observer's attribution of emotional state to gaze based on the gaze warping transformation (GWT) to find a model between emotion and physical manner of gaze that allows for the generation of believable emotionally expressive gaze (Lance et al., 2008). Gerard Bailly et al. (2010) investigated the agent which manipulates the audio and visual real life and virtual face-to-face interactions. The agent is based on mutual gaze patterns during human interactions and the selective measurements and the impact of award states and communicative functions (Bailly et al., 2010). Unfortunately, very few literatures have delved on eye movement which incorporates saccades

Ibbotson implemented an eye movement model of saccades and statistical models of eye-tracking data. The eye animations with saccades were designed to exhibit natural and effective communication (Ibbotson and Krekelberg, 2011). Lopatovska and Arapakis (2011) introduced emotional theories and methods, and their role in human information behavior (Lopatovska and Arapakis, 2011).

5.9. Facial expression approach

Facial expression is a natural phenomenon which humans use to communicate emotions. It shows the level or degree to which a human being tries to clarify and emphasize certain points to support comprehension, disagreement, intentions and to regulate interactions under any situation and environment (Torre and Jeffrey F. Cohn, 2011). These facts highlight the importance of automatic facial behavior analysis, including facial expression of emotion and facial action unit (AU) recognition. Research in this area has been of interest since the past twenty years (Gunes and Pantic, 2010). Until recently, most of the available data sets for expressive faces were limited in size. They contain only deliberately posed affective displays, mainly of the prototypical expressions of six basic emotions (i.e. anger, disgust, fear, happiness, sadness, and surprise), recorded under highly controlled conditions. Recent efforts focused on the recognition of complex and spontaneous emotional phenomena rather than on the recognition of deliberately displayed prototypical expressions of emotions (Zeng et al., 2007) (Nicolaou et al., 2011) (Vinciarelli et al., 2011).

Shen Zhang proposed a framework to synthesize the emotional facial expressions for an MPEG4 compliant talking avatar based on the three dimensional PAD model. This system enhances the emotional expressivity of talking avatar (S. Zhang et al., 2010). BinbinTu, Fengqin Yu proposed a recognition algorithm that can identify emotional states more accurately (B. Tu and Yu, 2011). Angela Tinwell presented a study of how facial expression in the lower face region effects on emotion and the uncanny valley (UV) phenomenon in realistic virtual characters (Tinwell et al., 2011).

6. Discussion and Conclusion

The eventual goal for research in facial modeling and animation is a system that creates real time realistic animation and automated as much as possible with adaptation to individual faces. There was an attempted to generate realistic facial modeling and animation. The most inspiring attempts were performed to create face modeling and rendering in real time. Human facial anatomy is complex. Human possess inherent sensitivity to facial appearance. Therefore no real-time system can generate subtle facial expressions and emotions realistically for virtual characters.

Creative animation of virtual characters requires complex effective communication tool to produce facial animation. The elements of the face such as facial muscles, facial bones, emotional facial expression, gaze and synchronization of the lips in conversation is a challenge to configure in high quality animation. Good virtual human characters require efforts coupled with a sufficient time for skilled animators. It has been suggested that good dialogue animation that provides the objective of what is being said can be produce by designing the

technical conversation between mouth shape and the phoneme of the spoken communication such that it has synchronism with the facial expression; thus realistically mimicking the emotion of the speaker. Most emotion of the human can be recognized through their facial expression. According to Jeff Wilson, as the face is a primary tool in the understanding of emotion, realistic facial animation is one important issue in computer graphic research area. This paper provided a thorough survey of facial animation with respect to eye movement, lip synchronization, and facial expression research. The existing literatures on various approaches proposed for problems in facial realistic modeling were reviewed. The review embodies the issues associated with facial animation approaches, which draws on the similarities between methods and the subdivision of these methods into models.

References

1. A.L. Yarbus, (1967). Eye movements during perception of complex objects. Eye Movements and Vision 2, 171–196.
2. Anh, L.Q., Pelachaud, C. (2011). Expressive Gesture Model for Humanoid Robot 224–231.
3. Argyle, M. Cook, M. (1976). Gaze and Mutual Gaze. Combridage University Press. London.
4. Arsov, I., Jovanova, B., Preda, M., Preteux, F. (2010). On-Line Animation System for Learning and Practice Cued Speech.
5. Athanasopoulos, M., Ugail, H., Gonz, G. (2011). On the Development of a Talking Head System Based on the Use of PDE-Based Parametic Surfaces 56–77.
6. Avaro, O. (2000). MPEG-4 Systems: Overview. Signal Processing Image Communication 15, 281–298.
7. Bailly, G., Raidt, S., Elisei, F. (2010). Gaze, conversational agents and face-to-face communication. Speech Communication 52, 598–612.
8. Balcı, K., Zancanaro, M., Pianesi, F. (2007) . Xface Open Source Project and SMIL-Agent Scripting Language for Creating and Animating Embodied Conversational Agents. Proceedings of the 15th international conference on Multimedia. ACM 1013–1016.
9. Baron-Cohen, S., Leslie, A.M., Frith, U. (1985). Does the autistic child have a “theory of mind”? Cognition 21, 37–46.
10. Ben Ammar, M., Neji, M., Ammar, M. Ben (2007). Agent-based Collaborative Affective E-learning System. Proceedings of the ImmersCom 5, 123 – 134.
11. Berger, M.A., Hofer, G. (2011). Graphically Speaking.
12. Bevacqua, E., Paristech, T., Paristech, C.T., Looser, J., Pelachaud, C.(2011). Cross-media agent platform 1, 11–20.
13. Blanz, V., Vetter, T. (1999). A morphable model for the synthesis of 3D faces. Proceedings of the 26th annual conference on Computer graphics and interactive techniques SIGGRAPH 99 pp, 187–194.
14. Brand, M. (1999). Voice puppetry. computer graphics(SIGGRAPH Proc.) 21–28.
15. Cao, Y., Tien, W.C., Faloutsos, P., Pighin, Frédéric (2005). Expressive speech-driven facial animation. ACM Transactions on Graphics 24, 1283–1302.
16. Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., Stone, M. (1994). Animated conversation, Proceedings of the 21st annual conference on Computer graphics and interactive techniques SIGGRAPH 94. ACM Press.
17. Cassell, J., Vilhjálmsson, H.H., Bickmore, T. (2001). BEAT: the Behavior Expression Animation Toolkit. In: Fiume, E. (Ed.), Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. ACM, pp. 477–486.
18. Catherine Pelachaus, N.I.B. and M.S. (1996). Generating Facial Expressions for Speech. Department of Computer and Information Science. University of Pennsylvania. University of Pennsylvania.
19. Chopra Khullar, S., Badler, N I. (2001). Where to look? Automating attending behaviors of virtual human characters. Autonomous agents and multiagent systems 4, 9–23.
20. Chou, Y.-F., Shih, Z.-C. (2010). A nonparametric regression model for virtual humans generation. Multimedia Tools and Applications 47, 163–187.
21. Cohn, Jeffrey F. (2010). Advances in Behavioral Science Using Automated Facial Image Analysis and Synthesis 128–133.
22. Colburn, R.A., Cohen, M.F., Drucker, S.M. (2000). The Role of Eye Gaze in Avatar Mediated Conversational Interfaces. Microsoft Research Report 81, 2000.
23. Darwin, C., 2005. the Expression of the Emotions in Man and Animals, Second edi. ed. Hazell, Watson and Viney, Aylesbury, UK.
24. DeCarlo, Douglas, Metaxas, D., Stone, M.(1998). An anthropometric face model using variational techniques. Computer 98, 67–74.
25. Deng, Z., Neumann, U. (2008). data-driven 3D facial Animation, first edit. ed. image processing. Springer.
26. Dornaika, F., Moujahid, A., Raducanu, B. (2013). Facial expression recognition using tracked facial actions: Classifier performance analysis. Engineering Applications of Artificial Intelligence 26, 467–477.
27. Dulguerov, P., Marchal, F., Wang, D., Gysin, C. (1999). Review of objective topographic facial nerve evaluation methods. The American journal of Otology 20, 672–678.
28. D’Mello, S., Olney, A., Williams, C., Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. International Journal of Human-Computer Studies 70, 377–398.
29. Ekman, P., Friesen, W. (1978). Facial Action Coding System Consulting.
30. El Ayadi, M., Kamel, M.S., Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. Pattern Recognition 44, 572–587.
31. Fasel, B., Luettin, J. (2003). Automatic facial expression analysis: a survey. Pattern Recognition 36, 259–275.
32. Frederic, P., Keith, W. (2008). Computer Facial Animation, second edi. ed. Welleseley.
33. Fu, Y., Tang, H., Tu, J., Tao, H., Huang, T.S. (2010). Human-Centered Face Computing in Multimedia Interaction and Communication. Intelligent Multimedia Communication Springer 280, 465–505.
34. Garau, M., Slater, Mel, Bee, S., Sasse, M.A. (2001). The impact of eye gaze on communication using humanoid

avatars. Proceedings of the SIGCHI conference on Human factors in computing systems CHI 01 309–316.
35. Gillies, Marco, Pan, Xueni, Slater, Mel (2010). Piavca: a framework for heterogeneous interactions with virtual characters. Virtual Reality 14, 221–228.
36. Gonseth, C., Vilain, A., Vilain, C. (2013). An experimental study of speech/gesture interactions and distance encoding. Speech Communication 55, 553–571.
37. Guenter, B., Grimm, C., Wood, D., Malvar, H., Pighin, F. (1998). Making Faces. In: Cohen, Michael (Ed.), British Dental Journal. ACM Press, pp. 55–66.
38. Gunes, H., Pantic, M. (2010). Automatic, Dimensional and Continuous Emotion Recognition. International Journal of Synthetic Emotions 1, 68–99.
39. Hofer, G., Yamagishi, J., Shimodaira, H. (2008). Speech-driven Lip Motion Generation with a Trajectory HMM 2314–2317.
40. Ibbotson, M., Krekelberg, B. (2011). Visual perception and saccadic eye movements. Current opinion in neurobiology 21, 553–8.
41. Itimad, R., Hoirul, A., Mahardika, C., Farhan, M., Nadz, S. (2012). Eye, Lip and Crying Expression for Virtual Human. In: ICIDM. ICIDM, malaysia.
42. Iverson, J.M., Goldin-Meadow, S. (1998). Why people gesture when they speak. Nature.
43. John R. Leigh, D.S.Z. (2006). The Neurology of Eye Movements: Book-and-DVD Package (Contemporary Neurology Series, 70), Edition, F. ed, Biomedical Engineering. Oxford University Press, USA.
44. Kalra, P., Mangili, A., Magnenat-Thalmann, N., Thalmann, D. (1991). SMILE: A Multilayered Facial Animation System. Proceedings IFIP conference on Modelling in Computer Graphics 189–198.
45. Kessous, L., Castellano, G., Caridakis, G. (2009). Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis. Journal on Multimodal User Interfaces 3, 33–48.
46. Kipp, M., Heloir, A., Schr, M. (2010). Closing the Gap between Behavior Planning and Embodied Agent Presentation 57–63.
47. Kopp, S., Krenn, B., Marsella, S., Marshall, A.N. (2011). Towards a Common Framework for Multimodal Generation : The Behavior Markup Language.
48. Kowler, E. (2011). Eye movements: the past 25 years. Vision research 51, 1457–83.
49. Lance, B., Marsella, S.C., Rey, M. Del, (2007). Emotionally Expressive Head and Body Movement During Gaze Shifts 72–85.
50. Lance, B., Marsella, S.C., Rey, M. Del (2008). The Relation between Gaze Behavior and the Attribution of Emotion : An Empirical Study 1–14.
51. Lee, C., Lee, Sangyong, Chin, S. (2011). Multi-layer structural wound synthesis on 3D face 177–185.
52. Lee, J., Marsella, S. (2006). Nonverbal Behavior Generator for Embodied Conversational Agents. In: Gratch, J., Young, M., Aylett, R., Ballin, D., Olivier, P. (Eds.), Intelligent Virtual Agents. Springer, pp. 243–255.
53. Lee, S.P., Badler, Norman I (2001). Eyes Alive.
54. Lee, Sangyoon, Carlson, G., Jones, S., Johnson, A., Leigh, J., Renambot, L. (2010). Designing an Expressive Avatar of a Real Person. In: Intelligent Virtual Agents. pp. 64–76.
55. Lee, W., Magnenat-Thalmann, L. (2000). Fast head modeling for animation. Image and Vision Computing 18, 355–364.
56. Li, Z., Mao, X. (2011). EEMML: the emotional eye movement animation toolkit. Multimedia Tools and Applications 60, 181–201.
57. Lien, J.J., Kanade, T., Cohn, J F, Li, C.-C.L.C.-C., (1998). Automated facial expression recognition based on FACS action units, Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition. IEEE Comput. Soc.
58. Lopatovska, I., Arapakis, I. (2011). Theories, methods and current research on emotions in library and information science, information retrieval and human–computer interaction. Information Processing & Management 47, 575–592.
59. Martino, J.M. De, (2007). Benchmarking Speech Synchronized Facial Animation Based on Context-Dependent Visemes.
60. Nicolaou, M.A., Gunes, H., Pantic, M. (2011). Continuous Prediction of Spontaneous Affect from Multiple Cues and Modalities in Valence-Arousal Space, IEEE Transactions on Affective Computing. IEEE.
61. Oyarzun, D., Mujika, A., Álvarez, A., Legarretaetxeberria, A., Arrieta, A., Carretero, P. (2010). High-Realistic and Flexible Virtual Presenters 108–117.
62. Parke, F.I., 1982. Parameterized Models for Facial Animation, IEEE Computer Graphics and Applications.
63. Patel, N., Zaveri, M. (2010). 3D Facial Model Construction and Expressions Synthesis using a Single Frontal Face Image. international Journal of Graphics 1, 1–18.
64. Pelachaud, C.(2009). Studies on gesture expressivity for a virtual agent. Speech Communication 51, 630–639.
65. Petajan, E. (1999). Very low bitrate face animation coding in MPEG-4. Encyclopedia of telecommunications 17, 209–231.
66. Pighin, Frédéric, Hecker, J., Lischinski, D., Szeliski, R., Salesin, D.H. (1998). Synthesizing realistic facial expressions from photographs. Proceedings of the 25th annual conference on Computer graphics and interactive techniques SIGGRAPH 98 2, 75–84.
67. Preda, M., Jovanova, B. (2013). Avatar interoperability and control in virtual Worlds. Signal Processing: Image Communication 28, 168–180.
68. Queiroz, Rossana B., Cohen, Marcelo, Musse, Soraia R. (2009). An extensible framework for interactive facial animation with facial expressions, lip synchronization and eye behavior. Computers in Entertainment 7, 1.
69. Queiroz, Rossana Baptista, Braun, A., Moreira, J.L., Cohen, Marcelo, Musse, Soraia Raupp, Thielo, M.R., Samadani, R. (2010). Reflecting User Faces in Avatars. Springer-Verlag Berlin Heidelberg 420–426.
70. Quené, H., Semin, G.R., Foroni, F. (2012). Audible smiles and frowns affect speech comprehension. Speech Communication 54, 917–922.
71. Richard S. Wallace (2004). The Anatomy of A.L.I.C.E. Artificial Intelligence Foundation, Inc.
72. Salvati, M., Anjyo, K.. (2011). Developing Tools for 2D / 3D Conversion of Japanese Animations 4503.
73. Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Müller, C., Narayanan, S. (2013). Paralinguistics in speech and language—State-of-the-art and the challenge. Computer Speech & Language 27, 4–39.
74. Seongah Chin, Chung yean Lee, S.L. (2011). Multi-layer Structural Wound Synthesis on 3D face. computer animation and virtual worlds 22, 177–185.

75. Serra, J., Ribeiro, M., Freitas, J., Orvalho, V. (2012). A Proposal for a Visual Speech Animation System. Springer-Verlag Berlin Heidelberg 267–276.
76. Shapiro, A. (2011). LNCS 7060 - Building a Character Animation System 98–109.
77. Shih, P., Wang, J., Chen, Z. (2010). Kernel-Based Lip Shape Clustering with Phoneme Recognition for Real-Time Voice Driven Talking Face 516–523.
78. Skantze, G., Moubayed, S. Al. (2012). IrisTK : a Statechart-based Toolkit for Multi-party Face-to-face Interaction 69–76.
79. Stephen M. Platt Norman I. Badler. (1981). Animating Facial Expressions. SIGGRAPH 15, 245 – 252.
80. Stone, M., DeCarlo, Doug, Oh, I., Rodriguez, C., Stere, A., Lees, A., Bregler, C. (2004). Speaking with hands: creating animated conversational characters from recordings of human performance. In: SIGGRAPH 04 ACM SIGGRAPH 2004 Papers. ACM Press, pp. 506–513.
81. Suwa, M., Sugie, N., K. Fujimora (1978). A Compliminary note on pattern recognition of human emotional expression. proceeding of the fourth International Joint Conference on Pattern Recognition 408–410.
82. Thiebaux, M., Rey, M., Marshall, A.N., Marsella, S., Kallmann, M. (2008). SmartBody : Behavior Realization for Embodied Conversational Agents 12–16.
83. Tinwell, A., Grimshaw, M., Nabi, D.A., Williams, A., (2011). Facial expression of emotion and perception of the Uncanny Valley in virtual characters. Computers in Human Behavior 27, 741–749.
84. Tomasello, M., Carpenter, M., Hobson, R.P. (2005). V. JOINT INTENTIONS AND ATTENTION. Monographs of the Society for Research in Child Development.
85. Torre, F. de la, Cohn, Jeffrey F. (2011). Facial Expression Analysis. Springer.
86. Tu, B., Yu, F. (2011). Bimodal Emotion Recognition Based on Speech 691–696.
87. Vertegaal, R, Slagter, R., Van Der Veer, G, Nijholt, A., (2001). Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In: Proceedings of the ACM Conference on Human Factors in Computing Systems. ACM, pp. 301–308.
88. Vertegaal, Roel, Veer, Gerrit van der, Vons, H. (2000). Effects of Gaze on Multiparty Mediated Communication. In Proceedings of Graphics Interface 95–102.
89. Vinayagamoorthy, V., Gillies, M, Steed, A., Tanguy, E., Pan, X, Loscos, C., Slater, M. (2006). Building Expression into Virtual Characters. Building 76, 21–61.
90. Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F., Schroeder, M. (2011). Bridging the Gap Between Social Animal and Unsocial Machine: A Survey of Social Signal Processing. IEEE Transactions on Affective Computing 3, 1–20.
91. Wang, Lan, Chen, H., Li, S., Meng, H.M. (2012). Phoneme-level articulatory animation in pronunciation training. Speech Communication 54, 845–856.
92. Wang, Lijuan, Qian, X., Han, W., Soong, F.K.(2010). Synthesizing Photo-Real Talking Head via Trajectory-Guided Sample Selection 446–449.
93. Waters, K.(1987). A muscle model for animation three-dimensional facial expression. ACM SIGGRAPH Computer Graphics 21, 17–24.
94. Yang, Y., Zheng, N., Liu, Y., Du, S., Su, Y., Nishio, Y. (2011). Expression transfer for facial sketch animation. Signal Processing 91, 2465–2477.
95. Yotsukura, T., Morishima, S., Nakamura, S. (2003). Model-based talking face synthesis for anthropomorphic spoken dialog agent system. ACM Multimedia 351–354.
96. Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S. (2007). A survey of affect recognition methods: audio, visual, and spontaneous expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence 31, 39–58.
97. Zhang, L., Jiang, M., Farid, D., Hossain, M. a. (2013). Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot. Expert Systems with Applications.
98. Zhang, S., Wu, Z., Meng, H.M., Cai, L. (2010). Facial Expression Synthesis Based on Emotion Dimensions for Affective Talking Avatar, T. Nishida. ed, Springer-Verlag Berlin Heidelberg.
99. Zhao, M., Gersch, T.M., Schnitzer, B.S., Dosher, B. a, Kowler, E. (2012). Eye movements and attention: the role of pre-saccadic shifts of attention in perception, memory and the control of saccades. Vision research 74, 40–60.
100. Zoric, G., Forchheimer, R., Pandzic, I.S. (2010). On creating multimodal virtual humans—real time speech driven facial gesturing. Multimedia Tools and Applications 54, 165–179.
101. Čereković, A., Pandžić, I.S. (2011). Multimodal behavior realization for embodied conversational agents. Multimedia Tools and Applications 54, 143–164.
102. Čereković, A., Pejša, T., Pandžić, I.S. (2010). A Controller-Based Animation System for Synchronizing and Realizing Human-Like Conversational Behaviors 80–91.

3/30/2014