Facial Action Coding System for the Tongue

Rahma M. Tolba, Taha El-Arif, and El-Sayed M. El Horbaty

***Abstract*—** FACS (Facial Action Coding System) is an anatomically based system for describing all observable facial movements. FACS provides a very reliable description for the face upper parts but it does not for the lower parts of the face. That limits FACS from being the dominant technique in the Facial Animation field. In this paper, we proposed 12 AUs (Action Units) for the Tongue, based on tongue anatomy, following the same format Paul Ekman used in defining FACS AUs. We applied these AUs on a 3D human model using Daz Studio Pro. and compared the results with photos that have been captured for real humans performing the proposed AUs. The results were very analogous to the ones performed by the real humans. Then we used the proposed AUs to make tongue animation for a very popular tongue movement in Egypt which is called Zaghrouta. The resulted animation was very realistic and almost identical to the video taken for an Egyptian woman performing this movement. Now we are having anatomically defined and reliable action units to control the tongue and overcame FACS limitation in this area. The proposed AUs offer additional pose control dials, as add-on for the existing computer graphics software, which give the animator more control and flexibility over the tongue and new levels of dynamic movements.

***Keywords*—** FACS, Facial Action Coding System, Tongue, Tongue Animation.

I. Introduction

FACS (Facial Action Coding System) is an anatomically based system for describing all observable facial movements with Action Units or AUs. Each AU corresponds to an individual face muscle or muscle group and is identified by a number (AU1, AU4, AU27, …etc.). Despite FACS popularity, it does not cover all of the visible, distinguishable and reliable actions of the muscle controls in the lower parts of the face [20]. FACS used AD (Action Descriptor) not AU for a number of tongue actions because they haven't distinguished one tongue show from another, nor specified the muscular basis for these actions. It defined only three ADs for the tongue: AD 19 (Tongue Show), AD 36 (Tongue Bulge) and AD 37 (Lip Wipe). Ekman in [1] mentioned that there are other actions for the tongue apart from wiping the lip (AD 37) that tongue can do, such as curling the tongue, darting it in and out, moving it sideways and up/down which all scored till now as AD 19.

The human tongue is one of the most important yet least understood structures of the body. It is a mass of muscle that is almost completely covered with moist, pink tissue called mucosa. It is highly mobile and can be shifted into a number of different positions and also assume various shapes. Tongue simulation methods including a simple geometric tongue with rigid motion, a human sculpted tongue in keyframe positions, finite element modeling, and a highly complex model using soft objects. A geometric and kinematic model, describe the global shape and the predominant motions of the tongue, are discussed in [28]. The tongue shape is important for realistic synthesized mouth animation, even if only small portion of it is visible during the normal speech [4].

There are large and comprehensive packages for creating facial animation such as MAYA, 3Ds Max, DAZ Studio, Poser, Blender or LightWave 3D and also standalone systems that are specifically designed for creating facial animation such as Face Robot, Facefx, iClone 7 Faceware and Dio-matic Facial Studio; not to mention, hundreds of academic open source applications developed by researchers using various facial animation technologies [29]. Daz Studio is one of the top packages that offers separate controls especially for the tongue with very realistic results by using the Iray render engine which supports ray tracing to generate images and released by NVIDIA. In addition, it imports and exports various file formats to be used within or out of Daz Studio. One of the main differences between Daz Studio and other software applications such as Poser is that Daz 3D has also included support for its various generations of the "Genesis" technology which is used as the basis for its human figures [30].

The rest of this paper is organized as follows: Section II, briefly covers the related work in tongue modeling and animation. Our proposed AUs for the tongue are presented in Section III. The implementation of the proposed AUs is given in Section IV. Finally, Section V contains the conclusion and future work.

II. Related Work

Some of the first uses of the tongue in computer facial animation were in the creation of animated short films.

Facial Action Coding System for the Tongue.

Rahma M. Tolba, Computer Science Department, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt. (e-mail: rahma.tolba@cis.asu.edu.eg).

Taha El-Arif, Computer Science Department, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt. (e-mail: taha_elarif@cis.asu.edu.eg).

Rahma M. Tolba, Computer Science Department, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt. (e-mail: rahma.tolba@cis.asu.edu.eg).

Reeves et al. [2] described the use of a teardrop-shaped collection of 12 bi-cubic patches to model the tongue in "Tin Toy", Pixar's Academy Award-winning animated short film. Although the tongue was modelled, it was usually left in the back of the mouth. Cohen et al. [3] added a simple tongue model to increase intelligibility of the generated speech in their work with the deaf and speech reading.

A physically realistic tongue model is needed to have a natural looking animation. Realistic tongue model should be capable of approximating the shape of any tongue in any natural position because the shape of the tongue is influenced by contact with hard surfaces such as the hard palate and the teeth. To improve the visual realism of the tongue, surface and lighting details are added [4]. Tongue modeling falls into three types: statistical model [5], parametric model [6] and physiological model [7]. Yang et al. [22] proposed accurate tongue simulation method to produce 3D inner mouth animations. Kawai et al. [23] proposed a method that generates highly photorealistic 3D inner mouth animation using only frontal captured images and small-size databases.

To capture the tongue's movements, the following methods are used: ElectroPalatoGraphy (EPG), ElectroMagnetic Articulometer (EMA), Applied Diagnostic Cineradiography (CFG), VideoFluoroGraphy (VFG), X-RAY Microbeam (XRMB), UltraSonography (US) and Magnetic Resonance Imaging (MRI) [8]-[14]. Steiner et al. [15] created a kinematic 3D model of the tongue and teeth by adapting a conventional motion capture based animation paradigm which is driven directly by EMA data. Tongue animation is used to create realistic facial animations, speech production [16]-[18], articulatory training [19] and articulatory speech synthesis. Approaches for animating 3D tongue for visual speech include skeletal animation [5], [21] or morph-target animation. In Skeletal animation, the tongue is represented in two parts: Mesh and Rig. Tongue mesh represents tongue surface. Tongue rig is a hierarchical set of interconnected bones which is used to animate mesh's pose and keyframe. In morph-target animation, the tongue is first modelled in a base shape and a "target deformation" is then created for each other shapes. When the tongue is being animated, the animator can then smoothly morph (or "blend") between the base shape and one or several morph targets.

Although the tongue is not visually critical for distinguishing between all sounds, it is only critical in distinguishing a few sounds, such as /l/ and /th/ and the lack of a tongue is easily noted. For an accurate and realistic visual speech synthesis, collision detection should be considered into the simulated tongue. Rui Li et al. [24] proposed a method that is effective for generating a realistic tongue animation which includes collisions handling during speech, especially for Chinese speech. Li et al. [25] proposed Vowels visualization for lips and tongues in a speaker-independent manner. Changwei et al. [26] showed that realistic 3D tongue animation can be created by using tongue tracking results of the X-ray images which are captured when a human subject was pronouncing a set of Chinese vowels, consonants, and words. Musti et al. [27] provided an intelligible and coherent 3D Visual Speech Animation by adding the basic and necessary animation of the tongue which is simpler than simulation-based methods. Fabre et al. [13] presented a method that animate an articulatory tongue model automatically from ultrasound images.

III. Proposed FACS AUs for the Tongue

The tongue has a great ability to move in all directions. It runs in all three directions: from front to back, from the sides to the middle and from top to bottom. This allows the tongue to make the following movements. In FACS they did AD not AU for a number of tongue actions because they haven't distinguished one tongue show from another, nor specified the muscular basis for these actions. So, after studying the tongue anatomy [31] and know the muscles for each shape and move the tongue can do, we proposed 12 AUs for the tongue, following the same format Paul Ekman used in defining FACS AUs [1]. They are shown in Table I.

Table I. Proposed AUs for the Tongue.

Suggested AU Number	FACS Name	Muscular Basis
90	Tongue Up	Palatoglossus
91	Tongue Down	Genioglossus & Hyoglossus
92	Tongue to the Left	Right Genioglossus
93	Tongue to the Right	Left Genioglossus
94	Tongue Curl	Superior Longitudinal
95	Tongue Bend	Inferior Longitudinal
96	Tongue Retraction	Hyoglossus, Styloglossus
97	Tongue Darting out	Transverse
98	Tongue Protrusion	Genioglossus
99	Tongue Twist	Styloglossus, Genioglossus and Hyoglossus
100	Tongue Rolling	Styloglossus
101	Tongue Flattening	Vertical

AU 90 - Tongue Up is moving your tongue at the roof of your mouth, behind your teeth, touching hard palate, like you pronounce letter L. Compare Fig.1 (a) with (c) to see how the action is made. The action is subtler in Fig.1 (b) than in Fig.1 (c). It is always accompanied by AU 26 (jaw drop) so the tongue will be able to move upwards. **AU 91 - Tongue Down** is the action of the tongue returning from Up position to the neutral position by relaxing the tongue.

Fig.1: (a) Neutral position (b): Subtle Tongue Up (C): Tongue Up

Table II showing an example for each of the following proposed AUs: 92, 93, 94, 95 and 96. The table is showing a picture for the neutral position and another picture for the proposed AU to be easy to compare between them and notice the performed action.

Table II. Examples of Proposed AUs: 92, 93, 94, 95 and 96.

AU Number	FACS Name	Action Picture
92	Tongue to the Left	
93	Tongue to the Right	
94	Tongue Curl	
95	Tongue Bend	
96	Tongue Retraction	

AU 97 - Tongue Darting Out is when the vertical and horizontal fibers contract at the same time, the tongue becomes narrower and longer as shown in Fig.2. **AU 98 - Tongue Protrusion** indicates that the tongue can be seen and protrudes beyond the boundaries of the teeth as shown in Fig.3. **AU 99 - Tongue Twist** is turning your tongue over in your mouth, either clockwise or counterclockwise as shown in Fig.4. **AU 100 - Tongue Rolling** is rolling the lateral edges of the tongue upwards into a tube shape as in Fig.5. **AU 101 - Tongue Flattening** is the tongue width getting wider than normal, as shown in Fig.6.

Fig.2: AU 97 – Tongue Darting Out

Fig.3: AU 98 - Tongue Protrusion

Fig.4: AU 99 - Tongue Twist

Fig.5: AU 100 - Tongue Rolling

Fig.6: AU 101 - Tongue Flattening

We will show a number of Action Unit combinations which involve the proposed FACS AUs for the tongue. Not all the possible combinations of these AUs will be shown and described, but some of the most common ones:

1) Tongue lateralization = Tongue out (AU 97) + Tongue to the Left (AU92) … as shown in Fig.7
   or Tongue Out (AU 97) + Tongue to the Right (AU 93) … as shown in Fig.8.
2) Tongue Elevation= Tongue Out (AU 98) + Curling Tongue (AU 94) … as shown in Fig.9.
3) Tongue Depression = Tongue Out (AU 98) + Bending Tongue (AU 95) … as shown in Fig.10.

Fig.7: Tongue Left lateralization

Fig.8: Tongue Right lateralization

Fig.9: Tongue Elevation

Fig.10: Tongue Depression

IV. IMPLEMENTATION OF THE PROPOSED FACS AUS

DAZ Studio is designed to allow users to manipulate "ready to use" models and figures as well as other supporting 3D content. It is aimed at users who are interested in posing human and non-human figures for illustrations and animation. It was created as an alternative to Poser, the industry-leading software in use for character manipulation and rendering [30]. It includes various generations of the "Genesis" technology which is used as the basis for its human figures and add-on content like clothing, hair, and additional morphs. Daz Studio Pro 4.10 has 7 default controls for the tongue with add-on option that offers a set of morphs designed to give the character's tongue new levels of dynamic movement. We used Tongue Control extension to the default tongue controls of Genesis 2 which is based on our proposed FACS AUs as shown in Fig. 11. In Table III we demonstrated each of the proposed AUs on a 3D human model with highly realistic appearance.

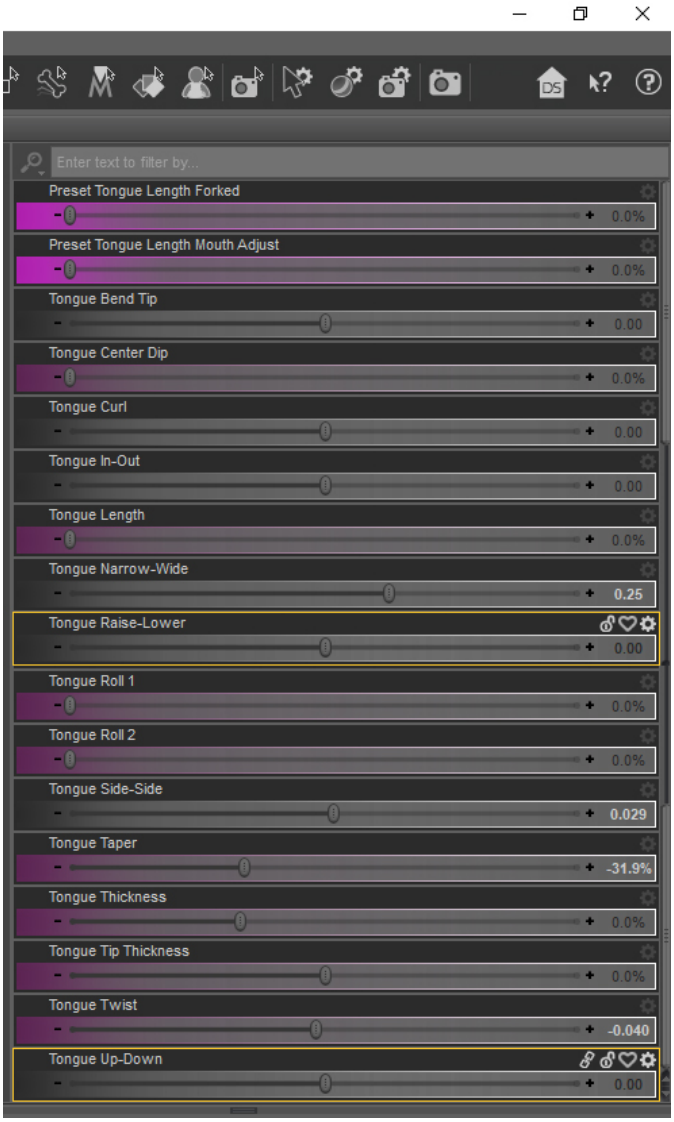
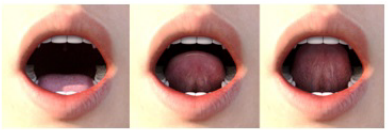
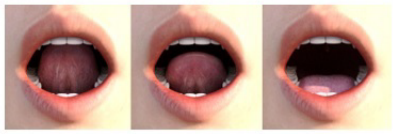
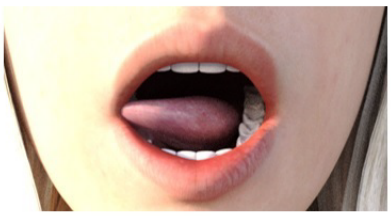
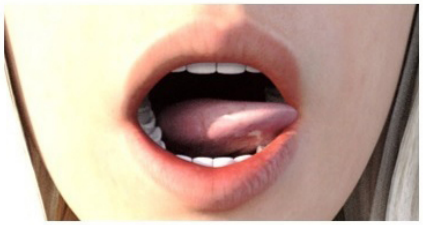
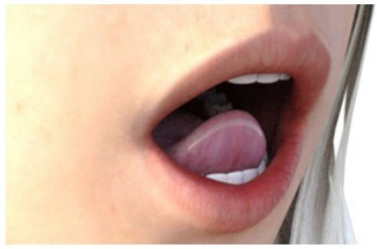
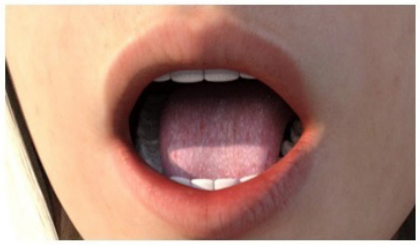
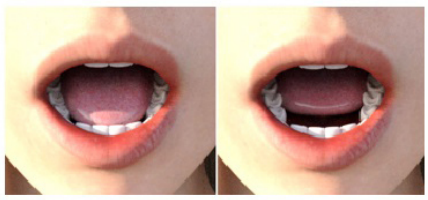
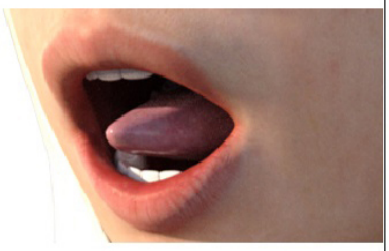
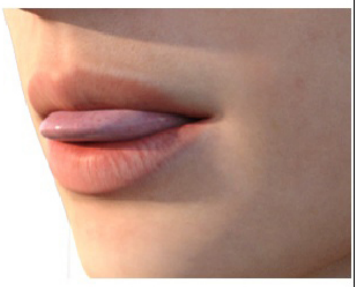
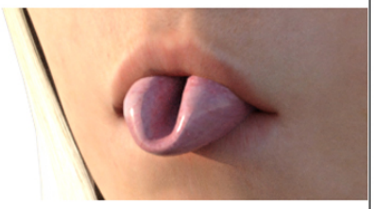


Fig.11: Daz Studio 4.10 Pro Tongue Control Sliders.

Table III. The Proposed Tongue AUs on 3D Human Model.

AU Number	FACS Name	Action Picture on 3D Human Model
90	Tongue Up	
91	Tongue Down	
92	Tongue to the Left	
93	Tongue to the Right	
94	Tongue Curl	
95	Tongue Bend	
96	Tongue Retraction	
97	Tongue Darting out	
98	Tongue Protrusion	
99	Tongue Twist	

AU Number	FACS Name	Action Picture on 3D Human Model
100	Tongue Rolling	
101	Tongue Flattening	

We used the proposed AUs to make tongue animation for a very popular tongue movement in Egypt which is called Zaghrouta. The Egyptian Zaghrouta is a loud voice that started by lifting the Tongue Slightly Up (AU 90) accompanied with a rapid movement of the tongue from the Left (AU 92) to the Right (AU 93) and ended with Tongue Slightly Up (AU 90), Tongue Retraction (AU 96) and Lip Corner Puller (AU 12). This action is shown in Fig.12 as a sequence of images. The resulted animation was almost identical to the real action performed by a real Egyptian woman.

Fig.12: Sequence of images simulates the Egyptian Zaghrouta.

V. Conclusion and Future Work

Several researchers recognized the need to represent the tongue in a facial animation system. Tongue visual information along with lips' movements can enhance speech perception in noisy environment or when one or more talkers have hearing disorder. They are also effective hearing aids in second language learning and teaching hearing impaired people how to speak. In order to improve one of the FACS drawbacks which related to defining action units to control tongue muscles, we proposed 12 AUs for tongue movements based on tongue anatomy. This proposal will help animators to represent and build realistic tongue movements in a very easy way through tongue control sliders as it already offered as add-on in Daz Studio and Poser.

In future work, one can define and implement all possible combinations of the proposed AUs in different fields such as tongue animation in teaching the second language.

References

[1] P. Ekman, W. Friesen and J. Hager, "Facial Action Coding System Manual", Research Nexus division of Network Information Research Corporation, 2002.

[2] W. Reeves, "Simple and complex facial animation: Case studies", SIGGRAPH '90 Course Notes 26: State of the Art in Facial Animation, pages 88-106, Aug. 1990.

[3] M. Cohen and D. Massaro, "Modeling coarticulation in synthetic visual speech. In Nadia Magnenat-Thalmann and Daniel Thalmann, editors, Models and Techniques in Computer Animation", Springer- Verlag, pages 139-156, Tokyo, 1993.

[4] S. King, "A Facial Model and Animation Techniques for Animated Speech", PHD thesis, The Ohio State University, 2001.

[5] J. Woo, F. Xing, J. Lee, M. Stone and J. Prince, "A spatio-temporal atlas and statistical model of the tongue during speech from cine-MRI", Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization Journal, March 2016.

[6] M. Ilie, C. Negrescu and D. Stanomir, "An efficient parametric model for real-time 3D tongue skeletal animation", 9th Communications (COMM) International Conference, Romania, June 2012.

[7] K. Honda, H. Hirai and J. Dang, "A Physiological Model of Speech Production and the Implication of Tongue-Larynx Interaction", Third International Conference on Spoken Language Processing (ICSLP 94), Japan, 1994.

[8] C. Qin and M. Carreira-Perpinan, "Reconstructing the full tongue contour from EMA/X-ray microbeam", ICASSP 2010.

[9] O. Engwall, "Combining MRI, EMA and EPG measurements in a three-dimensional tongue model", Speech Communication, 2003.

[10] P. Badin, G. Bailly, L. Rev_eret, M. Baciu, C. Segebarth and C. Savariaux, "Three-dimensional linear articulatory modeling of tongue, lips and face, based on MRI and video images", J. Phon. 2002.

[11] K. Hiiemae and J. Palmer, "Tongue Movements in Feeding and Speech", College of Engineering and Computer Science at SURFACE, Critical Reviews in Oral Biology & Medicine, pages 413-429, Nov. 2003.

[12] O. ENGWALL, "Combining MRI, EMA and EPG measurements in a three-dimensional tongue model", Speech Communication, 41(2-3):303–329, Oct. 2003.

[13] D. Fabre, T. Hueber, L. Girin, X. Pineda and P. Badin, "Automatic Animation of an Articulatory Tongue Model from Ultrasound Images of the Vocal Tract", Speech Communication, vol. 93, pp. 63-75, Aug. 2017.

[14] D. Fabre, T. Hueber and P. Badin, "Automatic animation of an articulatory tongue model from ultrasound using Gaussian mixture regression", InterSpeech ISCA, Singapore, Sep. 2014.

[15] I. Steiner, K. Richmond and S. Ouni, "Speech animation using electromagnetic articulography as motion capture data", 12th International Conference on Auditory-Visual Speech Processing, pp. 55-60, arXiv:1310.8585, Oct. 2013.

[16] M. Stone and M. Epstein, "Functional Segments in tongue Movements", Clinical Linguistics and Phonetics, ISSN: 0269-9206 print / ISSN: 1464-5076 online, 2004.

[17] K. Iskarous, "Patterns of Tongue Movement", Journal of Phonetics 33, pp. 363-381, 2005.

[18] S. King and R. Parent, "A 3D parametric tongue model for animated speech", The Journal of Visualization and Computer Animation, pp. 107–115, 2001.

[19] Y. Chen, F. Liao, J. Wu, H. Yang and C. Wu, "Mixmodel Driven 3D Facial Synthesis to Computer-Aided Articulation Training", IFMBE proceedings, Jan. 2008.

[20] M. Alkawaz, D. Mohamad, A. Basori and F. Mohamed, "A Crucial Investigation of Facial Skin Colour Research Trend and Direction", International Journal of Multimedia and Ubiquitous Engineering, Vol. 10, pp. 295-316, No. 1, 2015.
[21] M. D. Ilie, C. Negrescu, and D. Stanomir, "An efficient parametric model for real-time 3D tongue skeletal animation," in 9th International Conference on Communications (COMM), pp. 129–132, doi:10.1109/ICComm.2012.6262577., Bucharest, Romania, June 2012.
[22] Yang, Y., Guo, X., Vick, J., Torres, G.L. and Champbell, T., "Physics-Based Deformable Tongue Visualization", IEEE Trans. VCG, Vol.19, pp.811–823, 2013.
[23] M. Kawai, T. Iwao, A. Maejima and S. Morishima, "Automatic Generation of Photorealistic 3D Inner Mouth Animation only from Frontal Images", Journal of Information Processing Vol.23 No.5, 693–703, Sep. 2015.
[24] R. Li, J. Yu, C. Jiang, C. Luo and Z. Wang, "A Mass-spring Tongue Model with Efficient Collision Detection and Response During Speech", IEEE 9th International Symposium Chinese Spoken Language Processing (ISCSLP), Singapore, 2014.
[25] H. Li, M. Yang and J. Tao, "Speaker-Independent Lips and Tongue Visualization of Vowels", Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference, Canada, May 2013.
[26] L. Changwei, L. Rui, Y. Lingyun, Y. Jun and W. Zengfu, "Automatic Tongue Tracking in X-Ray Images", Chinese Journal of Electronics Vol.24, No.4, Oct. 2015.
[27] U. Musti, S. Ouni, Z. Zhou and M. Pietikäinen, "3D Visual Speech Animation from Image Sequences", ICVGIP '14, Bangalore, India, Dec. 2014.
[28] Catherine Pelachaud, C. W. A. M. van Overveld, Chin Seah, "Modeling and Animating the Human Tongue during Speech Production", Proceedings Computer Animation '94, pages 40-49, May 1994.
[29] Chen Liu, "An Analysis of the Current and Future State of 3d Facial Animation Techniques and Systems", Master thesis, University of China, 2006.
[30] DAZ Studio, Wikipedia 31 January 2018 edition: https://en.wikipedia.org/wiki/DAZ_Studio.
[31] Crispian Scully CBE, "Oxford Handbook of Applied Dental Sciences", Oxford University press, 2002.