Generating Expressive Facial Mesh Animation : A Survey

HJW
Institution1 address
ykn@rtfm.moe

Abstract

With technology allowing for increasing realism in games and movies, facial animation is still a very challenging task.

"Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum."

1. Introduction

Facial animation can be applied to various fields.

Human tend to be very sensitive to facial motion psychologically. Slightest uncanniest in facial animation is directly lead to hurt overall experience of embodiment, and overall experience [3]. So, delivering natural expressive facial animation is a great interest in graphics field.

To achieve realistic 3D face animation naturally, high-quality animation is required. Animating high-quality expressive face is very labor-intensive job when done by animator. Another approach to animate face is to capture human face animation in 3D. Face capture is a well-understood field(cite here), yet such approach requires gigabytes of data from expensive capture system, and is hard to manipulate. Therefore, it is necessary to simplify such process.

To simplify such process, one can automatically generate facial animation or can simplify animating produce.

In this survey, I introduce and compare three research that animate expressive facial animation :

- JALI [1] and VisimeNet [6], a linguistic approach to lip-sync.
- MeshTalk [5], a deep learning method.
- D3DExpression [4], LSTM method which replicate facial expression.

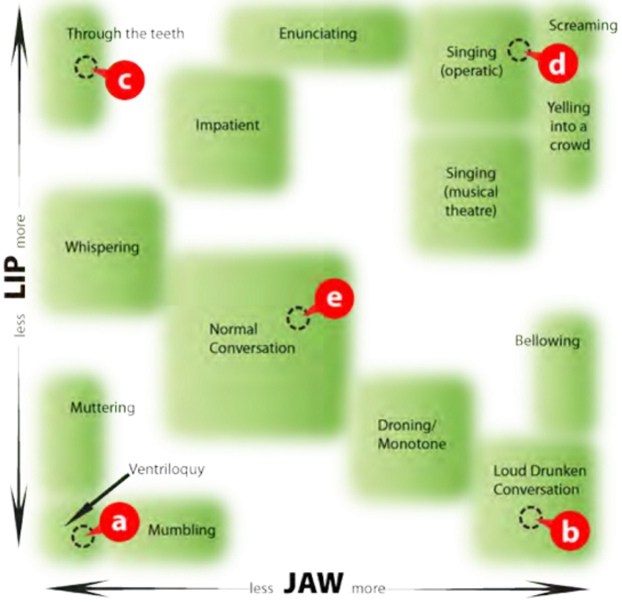

Figure 1. Speaking styles of JALI viseme field

2. Methods

2.1. JALI and VisimeNet

Lip-sync is traditionally done by linguistic approach which is mapping text to phonemes, then phonemes to visemes, then phonemes to the position of lip and jaw [2]. Phonemes to visemes is a complex many-to-many mapping.

JALI [1] is a state of the art viseme model. JALI takes jaw and lip activation multipliers into consideration, since jaw and lip is the most significant acoustic motion in face.

JALI many-to-one map phonemes to viseme. Then applied animated jaw-lip multipliers to the face.

As shown on Fig. 1, different speaking styles shows different jaw lip activation level multipliers. Which can be animated more intuitively.

JALI model requires manual labor such as aligning audio to plain text or phonemes. This lead to research to automate such process. This approach requires extracting viseme and jaw-lip model sequence from audio.

References

[1] Pif Edwards, Chris Landreth, Eugene Fiume, and Karan Singh. JALI: An animator-centric viseme model for expressive lip synchronization. *ACM Transactions on Graphics*, 35(4):1–11, July 2016. 1

[2] T. Ezzat and T. Poggio. MikeTalk: A talking facial display based on morphing visemes. In *Proceedings Computer Animation '98 (Cat. No.98EX169)*, pages 96–102, June 1998. 1

[3] David Hanson. Upending the Uncanny Valley. page 8. 1

[4] Rolandos Alexandros Potamias, Jiali Zheng, Stylianos Ploumpis, Giorgos Bouritsas, Evangelos Ververas, and Stefanos Zafeiriou. Learning to Generate Customized Dynamic 3D Facial Expressions. *arXiv:2007.09805 [cs]*, July 2020. 1

[5] Alexander Richard, Michael Zollhöfer, Yandong Wen, Fernando de la Torre, and Yaser Sheikh. MeshTalk: 3D Face Animation From Speech Using Cross-Modality Disentanglement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1173–1182, 2021. 1

[6] Yang Zhou, Zhan Xu, Chris Landreth, Evangelos Kalogerakis, Subhransu Maji, and Karan Singh. Visemenet: Audio-driven animator-centric speech animation. *ACM Transactions on Graphics*, 37(4):161:1–161:10, July 2018. 1