Accuracy Performance and Potentiality of Real-Time Avatar Lip Sync Animation in Different Languages

N. H. Loh, S.S. Shaharuddin

Abstract: ***With the fast growth in computing power nowadays, the qualities of animation enable an extra layer of visually convincing realism. In lip sync animation, creation of realistic lip movement is arduous in getting the lip shape and position to synchronize with the speech sounds. To note, spending hours in manually generating every single lip movement can be a long and challenging task. Consequently, a comprehensive analysis on viseme based multiple phonemes in English, Bahasa Melayu and Mandarin was carried out, to develop an accurate and potential platform for real time talking avatar in multiple languages. The accuracy performance between human and avatar in real time were compared and evaluated. The findings revealed successful utilization of real time synchronization to drive the synthetic 3D avatars based on live speech input for multiple languages, with satisfied accurate lip motion result. This paper provides useful knowledge for multilingual solution which accurately predicts mouth movement on real human face, when a person is speaking and directs to lip sync process. It contributes to live performances and valuable in open-ended field with tons of potential, such as animation production industry, entertainment, gaming, digital marketing and media education.***

Keywords: Avatar; Human Lip Shape; Lip Sync Animation; Real Time; Speech Recognition.

I. INTRODUCTION

Real-time lip sync animation is an approach to perform the talking of a virtual computer-generated character known asavatar, which synchronizes an accurate lip movement and sound in live animation. To explained, lip synching is often a part of the post-production phase in the making of animation films. However, drawings, clay puppets and computer meshes do not talk; so, when the synthesised characters are required to say something, their dialogues have to be recorded and analysed first before animate them to speak. Therefore, lip synchronisation or 'lip-sync' is the technique of moving a mouth of an animated character in such a way that it appears to speak in synchronism with the sound track. Likewise, real time lip sync is a technique driven by human voice directly to generate an avatar to talk on the screen.

In this context, determination of human's lip pattern and movement showed the significance in generating natural speech. [1] emphasized that it is a necessary step in the process of mapping lip movements to the speech sound.

Revised Manuscript Received on April 15, 2019.

N. H. Loh, Faculty of Applied and Creative Arts, Universiti Malaysia Sarawak, Kota Samarahan, 94300 Sarawak Malaysia

S.S. Shaharuddin, Faculty of Applied and Creative Arts, Universiti Malaysia Sarawak, Kota Samarahan, 94300 Sarawak Malaysia

However, creating an accurate lip sync animation would be significantly more difficult especially in setting key frame value, as shown in Figure 1. In fact, it is particularly challenging in mapping the lip movements and sounds to be synchronized [2]. It is a time consuming process [3] especially in doing multilingual animation. To elaborate, the process is done manually through adjusting frame by frame that often needs several passes of fine tuning to match the sound [4]. As a result, it can be clearly seen that most of the animation films will choose not to redo the lip sync process when republish the animation with second language. The difficulty of the process causes heavy workload, time consuming and costly.

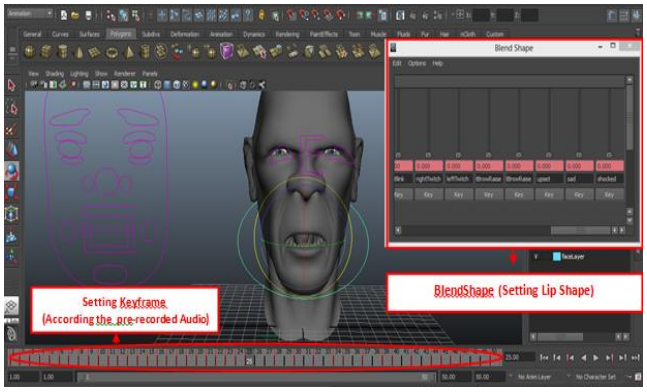

Fig. 1 Screen shot of setting keyframe to create lip sync animation in Autodesk MAYA

Therefore, the real time approach is needed to solve the difficulty of ordinary lip sync techniques and ensure realism in lip sync animation. In order to make character animation believable, correct lip shape corresponds to the sound is essential [5]. This paper provided automated digital speech model of viseme classification mapping to match the key phoneme sounds for English, *Bahasa Melayu* and Mandarin languages. Viseme is short for visible phoneme and refers to the shape of the mouth at the apex of a given phoneme [6]. Different categories of lip shapes in producing different phonemes sound had been analyzed very specifically using viseme categories. The approach reduces the difficulties of the ordinary lip sync technique which involves figuring out the speech timings and animating mouth positions manually to cohere with the sound. As for the making of real time animation which to be broadcasted with different language dialogues, it shortens the duration of production process and ensures an accurate outcome of lip sync.