Beyond Digitalisation: Facial Motion Capture for Mak Yong through the Perspective of Aesthetic Experience and Uncanny Valley

Muhammad Zaffwan Idris and Naimah Musa
Creative Multimedia Department, Universiti Pendidikan Sultan Idris, Perak, Malaysia
Email: {zaffwan, naimah.m}@fskik.upsi.edu.my

Abstract—In recent years, study on movement has risen among industry professionals and scholars. This also apply to the dance area where movement and kinesthetic play the major role. However, to measure the dance movement and kinesthetic is not complete without considering several element that affects the dance performance such as emotion, knowledge, and technical parts. The aim of this research is to observe both the potentials and challenges in digitising facial expressions by the dancer in expressing different type of emotions as part of preserving a Malay folkdance known as Mak Yong. The intention is to use facial motion capture (MoCap) device to record the facial expressions then applying the data in different form of visual representations. Further discussions are based on theory and model known as the Aesthetic Experience (AX) and Uncanny Valley.

Index Terms—facial motion capture (MoCap), Aesthetic Experience (AX), uncanny valley, Mak Yong

I. INTRODUCTION

This paper is in its preliminary stage. However, as part of a bigger research project, this paper highlights on the potentials and challengers that might occur during the process of digitising Mak Yong focusing on facial expression using MoCap. Mak Yong is complex form of artistic expression as it combines dance and drama components as part of the whole performance. The research on preservation of Intangible Cultural Heritage (ICH) in Mak Yong as in management [1] as well as in other ICH using MoCap, for examples, as in Cypriot folk dance [2] and Malaysia such as Zapin [3]. Thus, preservation of Mak Yong using MoCap can be considered novice. Mak Yong emphasizes on the expressions through body movement and facial expression. Since Mak Yong can be segmented into dance and drama, the former consisted of dance routine, musical instrument and song, while the later consisted of storytelling, dialogue and character. The process of digitising the whole Mak Yong performance could become an overwhelming procedure. To begin with, one of the segments in Mak Yong that needs to be highlighted is the facial expression. This is because the facial expression plays an important role in both dance and drama component especially because of the storytelling and different characters found in Mak Yong.

II. LITERATURE REVIEW

Many scholars such as Skeat, Cuisinier, Sheppard and Malm have mentioned Mak Yong in the 20th century as one of the unique form of cultural expression. However, a scholar named Mubin Sheppard, claimed that Mak Yong was introduced into Kelantan from Pattani about 2000 years ago as a form of entertainment. Prior to 1920, the Sultan of Kelantan and royalties were responsible in maintaining groups of Mak Yong performers. Unlike any other traditional dance theatre, Mak Yong is unique as it combines all the performing arts elements such as dance, vocal, song, instrumental music, acting, the word that spoke, ritual and poetry [4].

Repertoire in Mak Yong has expanded into numerous phases, begins with Dewa Muda, which is regarded as the story that explains the origins of Mak Yong, and has expended into twelve of classical stories. There are Dewa Muda, Dewa Pechil (with Dewa Samadaru as a variant title/version), Dewa Sakti (or Rajah Sakti), Dewa Indera-Indera Dewa, Dewa Panah (or Anak Rajah Panah), Anak Raja Gondang, Gading Bertimang, Raja Tangkai Hati, Raja, Muda Lakseng, Rajah Muda Lembek, Raja Besar Dalam Negeri Ho Gading, and Bedara Muda.

A. Pak Yong

Pak Yong is portrayed as a main actor in Mak Yong to carry several characters in different stories. Basically, Pak Yong can either performed by a male or female actor. However, female actors have been playing these roles nowadays. Besides being portrayed as a hero, Pak Yong is also being visualised as a King or a Prince. In Mak Yong repertoire, one or more actors may play the Pak Yong roles. Therefore, there could be a young Pak Yong and an older Pak Yong.

Most of the Pak Yong actors are known for his or her wide experiences in Mak Yong as Pak Yong is the leading actor in Mak Yong performance (Fig. 1). Basically, the actor or actress will undergo several other characters before being appointed as Pak Yong based on skills, stage presence, and appearance [5].

As Mak Yong and many other folkdances are facing huge threats through modernization and lack of interest...
from younger generation, digital preservation is deemed as one of the best solutions in order to ensure its existence [6]. Plus, with the advancement of digital technology, apart from preserving the authenticity of Mak Yong, a new interpretation and presentation of Mak Yong like 3D animation, digital games, and hologram exhibition are made possible. The following topics are some of the key discussions in order to support the intention.

III. MOTION CAPTURE

Motion Capture or better known as MoCap is a system that is able to capture the movement of actors or objects digitally for animation purposes. By placing the markers onto the actor or the object, the movement and trajectory can automatically be traced and recorded. MoCap is able to extract and measure the body movement including several parameters such as speed, position, and angle during the process of archiving, while the data collected may be applied to manage instruments and scientific investigation [7]. Currently, MoCap has been widely used in many areas such as in digital game and movie productions such as The Lord of the Rings, Rise of the Planet of the Apes, Avatar, Ironman and Marvel Avengers.

MoCap is an advanced tool to track, record and digitised the complexity of body movement especially in dance as it is efficient, produces accurate data and able to assess the movement and gesture [8], [9]. This process can help to simplify the process of animating a 3D character, which normally takes longer time to complete. Two type of MoCap are known as marker based and markerless [10]. However, the optical marker based is most widely used and can be divided into active (LEDs) or passive markers (balls coated with reflective material). Passive markers are able to give the highest accuracy and response in a short time. [11]. Recently, MoCap has also been used in the digitisation of folkdance [12], [6]. (Fig. 2) market demand especially in entertainment and animation. Mayya et al. [13] denoted that there are a few methods used to recognize facial expression, such as, by using the support of Vector Machine classifier with boosted CBP features, expression recognition can be achieved and also technique known as Deep Convolutional Neural (DCN) framework, by extracting the features of the images.

B. Facial Recognition

In 2007, Woodward et al. [14] used the stereo web-cameras to create the performance of a virtual character. However, this technique needs three modules, marker tracking, face animation, and the virtual performance. According Wójcik, [15], there are six ways or techniques to capture and analyse the facial regonition; i) Classical face recognition algorithms ii) Artificial neural networks iii) Gabor wavelets iv) Face descriptor-based method v) 3D based face recognition vi) Video based recognition. One of the popular techniques is EMOTE, a facial recognition SDK, a system that is able to integrate the target emotions without needing an open source computer vision (OpenCV), a library of programming function that mostly used for real time computer vision.

Basically, the most common approach used in the facial recognition is by using these three steps; i) face detection and tracking ii) feature extraction iii) classification of expression. In the end of process, facial changes can be identified as prototypic emotions or facial action units [16]. This technique is less advanced or also known as refined tracking, as it unable to fully capture three-dimensional (3D) motions such as head rotation.

C. Facial Motion Capture

Facial MoCap is a process of electronically tracking and converting the movements of a person's face into a high resolution digitally by using tools such as cameras or laser scanners to track the expressions (Fig. 3). There are two types of facial MoCap existing in the market, which are i) markerless and ii) marker based. For marker based, Edward et al. [17] used Vicon MoCap system with the Nexus 1.7.1 software where 100 markers with a diameter of 1.5 mm and 21 markers with a diameter of 3mm are used to capture the expressive of facial motion on the face.

Meanwhile, markerless is used to capture and analyse thoroughly the face's surface including the texture and geometry. The advantage of using the markerless for facial MoCap is it does not require any markers to be placed on the face, making it more convenient for the actor during the recording process (Fig. 4).
V. AESTHETIC EXPERIENCE (AX)

The AX is fundamentally a form of communication between the media, individuals and environment [19]. So, there is no doubt that the notion of ‘aesthetic experience’ is current in contemporary aesthetics as AX is a basis of modern culture that has been aestheticised process. At the early stage, AX started as ‘regressive state’, revives the early developmental phases (for example, symbiotic fusion with caregivers) [20]. By that, Hagman believed AX is special in form of relationship with the object. It means that audience members, without prejudice are willing to give themselves over the experience and approach with compassion neutrality.

AX may appear through body and emotional engagement along with the work of art by a combination process of the perception-action and motion-emotion loops [21].

A. Emotions

Certain visual cues can evoke emotional response while the visual of a person’s social background along with emotional experience will become a continuous trigger of experience [22]. Therefore, movements especially in dance are intimately linked to emotional experience, the connection between movement and emotions is believed to be bidirectional where the human body may be main influence or gain certain emotional experiences [23]. The emotions are believed to be the signs of values as emotion charges the AX through a pragmatist perspective [24].

Emotions however are human psychological states that are important in personal relations, experience and also everyday life, as there are hundreds of emotions identified [25]. As mentioned by Falip et al. [26], emotions have lots of definition and categorisation. Thus, with facial expression, emotion will be differ and can be measured (Fig. 5). Basic emotion also known as primary, fundamental or prototype is hard to define as scholars and theorist have not agreed on what kind and how much emotions are considered as basic, but at the same time, most of the numbers are between three to eleven [27]. [28] argues facial emotional can be divided into six basic emotion, which are anger, disgust, fear, surprise, joy, and sad (Table I).

<table>
<thead>
<tr>
<th>No.</th>
<th>Emotions</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Anger</td>
<td>Combination of inner eyebrow pulled lower along with the eyes wide opened, tightened lid and tightened lip or opened (exposed the teeth)</td>
</tr>
<tr>
<td>2</td>
<td>Disgust</td>
<td>Combination of the relaxed eyebrow and eyelid, wrinkled nose, lip corner depressor, and lower lip depressor (asymmetrically)</td>
</tr>
<tr>
<td>3</td>
<td>Fear</td>
<td>Combination of inner/outer brow raiser, lowered brow, upper lid raiser, tightened lid, stretched lip and jaw, while eyes are tense and alert</td>
</tr>
<tr>
<td>4</td>
<td>Happy/Joy</td>
<td>Relaxed eyebrow. Cheek raiser, pulled corner of lip back toward ears and mouth opened are combined</td>
</tr>
<tr>
<td>5</td>
<td>Sad</td>
<td>Combination of inner brow raiser, slightly eyes closed, lowered brow and lip corner depressor while mouth is relaxed</td>
</tr>
<tr>
<td>6</td>
<td>Surprise</td>
<td>Combination of inner / Outer brow raised, upper lid raiser and opened while the lower relaxed, and jaw drop opened</td>
</tr>
</tbody>
</table>

VI. UNCANNY VALLEY

Highly realistic computer games, immersive application and computer-generated characters would sometimes tick off the eeriness sensation or even disgust the user. The negative sensation due to the highly realistic character had already been discussed by Masahiro Mori, a roboticist who found the reveal of negative sensation when human responded to robots and prosthetics that are not quite human-like [30], in a 1970 article in the Japanese journal Energy and it was not completely translated into English until the year of 2012.

Mori [31] posits a graph with the human likeness on the x-axis and the level of affinity on the y-axis. Referring to Fig. 6, people have greater affinity when the robots become more human-like for instance the prosthetic hand that has achieved a degree of resemblance to the human form. However, once people realise that the real-looking hand is actually artificial, people will lose their sense of affinity or the level of affinity will turn to negative. As for this case, the hand becomes what is known as ‘uncanny valley’.

<table>
<thead>
<tr>
<th>Affinity</th>
<th>Human Likeness</th>
<th>Prosthetic Hand</th>
</tr>
</thead>
</table>
Similar to other computer-generated character, when user perceives that the character is not really alive, the sensation of uncanny valley will make the user see them as dead corpse or zombie. The situation is getting worse when the uncanny valley character is animated to move [32] and the details regarding the affinity of human towards the moving character.

Additionally, the uncanny valley in modeling and animation is also fundamentally affected by a series of inconsistencies, such as i) Inconsistency of realistic modeling and creepy animation ii) Inconsistency of the animation of one part of the face and another iii) Inconsistency of the 3D face animation and the expressions of our own faces [33].

Tinwell et al. [34] indicate that once the character reaches high levels of human likeness, exhibits human-like appearance and behavior, the audience will place a greater scrutiny. Factor like facial expression is often the cause that makes a character looking uncanny and appearing life-less, especially when the character’s expression appear odd or unnatural. Makarainen et al. [35] add that there are two specific ways that make the character look uncanny: They appear either i) distorted or ii) lifeless. Uncanny due to the distorted is referring to situations in which one can detect something abnormal and the perception of the abnormality sense has disturbed the audience. For instance, a face with abnormally large eyes will come to be uncanny even the degree of realism is higher. As for the perspective of lifeless, uncanny sensation would not be able to escape from the audience when they look at a perfectly human-like robot. This feeling will appear when the audience aware of the robot is not human.

As for the case as seen from the context at above, Tinwell et al. [34] have been outlined many guidelines to advise the character designer on how to go across the uncanny valley. The guidelines have included factors like facial feature and proportion as well as level of detail in skin texture. Designer can achieve more positive result if the character can be designed with non-human feature but have the ability to emote like a human [36]. On the other hand, changing a character’s appearance to more cartoon-like can eliminate the uncanny and boost positive feeling among the audience [37].

Geller [38] also agrees that a good way to go across uncanny valley is alter a character’s proportions and structure outside the ‘human’ range. An example is given in his article, saying that why Gollum is so successful is due to he has big eyes and the shape of his face is not quite human. The audience will unanimously deem the character as not a human. Thus, they don’t have to judge the character by the same rule as if it is. But if designer portrays human, audience can figure out what is missing.

VII. CONCLUSION

The advancement of digital technology including computer graphics has made a significant change on how we deal with cultural preservation. With the perfect computer renderings, that is almost impossible for us to make a distinction from reality. Undeniable, increasing the accuracy of recording motion and realism of 3D modeling environment can notably increase user immersion, but conversely improving the virtual characters realism seems to impede the user’s acceptance as been highlighted in the uncanny valley. As for the context of Mak Yong digital preservation, the digital technology may improves significantly, but in the same time, aesthetic experience must be well considered in order to assure the type of emotions intended to be portrayed are accurate representation and not derail during the process of digitalisation as mentioned in the uncanny valley. Therefore, this research is crucial in order to provide a clear guideline for future researcher, creative industry practitioner and local cultural organisation.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHOR CONTRIBUTIONS

All authors had involved in writing the paper and approved the final version.

ACKNOWLEDGMENT

This research was supported in part by Ministry of Education Malaysia (MOE) through Fundamental Research Grant Scheme, (FRGS/1/2017/WAB04/UPSI/02/1).

REFERENCES

[10] P. Nogueira, “Motion capture fundamentals: A critical and comparative analysis on real-world applications,” in Proc. the 7th


