Can Local Avatars Satisfy Global Audience? A Case Study of High-Fidelity 3D Facial Avatar Animation in Subject Identification and Emotion Perception by US and International Groups.*

Chang Yun, Zhigang Deng, and Merrill Hiscock[1]

Computer Science Department
University of Houston
Houston, TX, 77204, USA
http://www.cs.uh.edu

Technical Report Number **UH-CS-09-05**
January 31, 2009

Keywords: Subject identification, emotion perception, confidence rating, motion capture, facial emotion and expression

Abstract

This study investigates effectiveness of a local high-fidelity 3D facial avatar before the global audience by observing how US and International student groups differ in viewing non-verbal high fidelity 3D facial avatar animation embedded with motion data from 3 US individuals in subject identification and emotion perception. To synthesize the animated 3D avatars that convey highly-believable facial expressions, a 3D scanned facial model was mapped with high-fidelity motion capture data of three native US subjects as they spoke designated English sentences with specified emotions. Simple animations in conjunction with actual footages of the captured subjects speaking during the facial motion capture sessions were shown several times to both native US students and international students in similar settings. After a familiarization process, we showed them randomly arranged talking avatars without voice and asked them to identify the corresponding identity and emotion types, and to rate their confidences in terms of their selections. We found that US group had higher success rates in the subject identification although the related confidence ratings difference between two groups was not significant. The differences in emotion perception success rate and confidence ratings related to the emotion perception between two groups were not significant. The results of our study provide interesting insights to avatar-based interaction where national/cultural background of a person impacts the perception of the identity while it has little effect on the emotion perception. However, we observed that dynamics (i.e. head motion) could offset the cultural unfamiliarity disadvantage in the subject identification. We observed that both groups performed in nearly identical level in subject identification and emotion perception when they were shown the avatar animation with high expression and dynamics intensities. In addition, we observed that the confidence ratings had correlations with the accuracies of the subject identification but not with the accuracies of the emotion perception.

---

* This work is to appear on *ACM Computers In Entertainment (CIE), 7(2), 2009.*
[1] M. Hiscock is affiliated with the Department of Psychology, University of Houston.

Can Local Avatars Satisfy Global Audience? A Case Study of High-Fidelity 3D Facial Avatar Animation in Subject Identification and Emotion Perception by US and International Groups.

CHANG YUN, ZHIGANG DENG AND MERRILL HISCOCK
University of Houston

This study investigates effectiveness of a local high-fidelity 3D facial avatar before the global audience by observing how US and International student groups differ in viewing non-verbal high fidelity 3D facial avatar animation embedded with motion data from 3 US individuals in subject identification and emotion perception. To synthesize the animated 3D avatars that convey highly-believable facial expressions, a 3D scanned facial model was mapped with high-fidelity motion capture data of three native US subjects as they spoke designated English sentences with specified emotions. Simple animations in conjunction with actual footages of the captured subjects speaking during the facial motion capture sessions were shown several times to both native US students and international students in similar settings. After a familiarization process, we showed them randomly arranged talking avatars without voice and asked them to identify the corresponding identity and emotion types, and to rate their confidences in terms of their selections. We found that US group had higher success rates in the subject identification although the related confidence ratings difference between two groups was not significant. The differences in emotion perception success rate and confidence ratings related to the emotion perception between two groups were not significant. The results of our study provide interesting insights to avatar-based interaction where national/cultural background of a person impacts the perception of the identity while it has little effect on the emotion perception. However, we observed that dynamics (i.e. head motion) could offset the cultural unfamiliarity disadvantage in the subject identification. We observed that both groups performed in nearly identical level in subject identification and emotion perception when they were shown the avatar animation with high expression and dynamics intensities. In addition, we observed that the confidence ratings had correlations with the accuracies of the subject identification but not with the accuracies of the emotion perception.

Categories and Subject Description: H5.2 [User Interfaces]: Benchmarking, Evaluation/Methodology, Interaction Styles

General Terms: Avatar, Human Factor

Additional Key Words and Phrases: Subject identification, emotion perception, confidence rating, motion capture, facial emotion, facial expression, user study

1. INTRODUCTION

In recent years, we witness an explosive growth in creation and usage of high-fidelity 3D avatars in the entertainment industry where the avatars are used either as a representation of self or others characters. The primary reason for the growth in the usage of the high-fidelity avatars is due to growing expectation of audiences for more realistic avatars as the technology rises to be capable to meet the demand. As the technology pushes its limit further, it will be inevitable for the audience to have even higher expectation and demand for more realistic avatars that truly resemble human in near future. The high-fidelity 3D avatar development in entertainment and online environments becomes possible due to the usage of motion capture system that allows the developers to create 3D avatars and

embed them with very detailed behavioral realism. Creating the avatars that have high fidelity body motion is more prevalent throughout the industries due to relative ease to capture the human body motion using actors, process and implement them to the avatars. For example, most of the current 3D sports game developers capture real athletes' detailed athletic motions so that the avatars in the games also move with motion that resembles very closely to the realistic body motion of the athletes. On the other hand, the usage of the motion capture system to capture the human facial motion (and/or emotion) is less prevalent than the body motion capture due to higher cost, effort and time to capture the facial motion, and process and implement them to the face of the avatars. In movie industry, Polar Express by Warner Brothers Inc. (http://polarexpressmovie.warnerbros.com/) and Spiderman 3 by Columbia Pictures Inc. (http://www.sonypictures.com/homevideo/spider-man3/) are two examples where the motion capture system was used to capture the facial motion/emotion. In game industry, the demand to create more precise high-fidelity 3D avatars that conveys very detailed visual realism and behavioral realism of the face and the body is stronger than the movie industry. For example, Heavenly Sword from Sony Entertainment (http://www.us.playstation.com/heavenlysword/) and Grand Theft Auto 4 from Rockstar Games (http://www.rockstargames.com/IV/) create avatars with detailed body and facial motion capture data. Unlike movie, since the main avatars (or main characters) of games represent the users to interact with other avatars (or characters), the demand of the users for more realistic action of the avatars and interactions that resembles the human-to-human interaction closely are stronger in the game industry.

In terms of the high-fidelity 3D avatars that are created using facial motion capture data, there is one issue that needs everyone's attention – a disposition that the audiences (or customers) of most entertainment media in modern days are no longer confined to one country. In another word, a medium (particularly video game) that is created from one country is generally not for the audiences of the country exclusively. Rather, both the producers and the audiences of the entertainment media anticipate that the media are to be available worldwide. Based on this tendency, one must wonder on one seemingly obvious problem with the avatar: If a medium is created from one country, are the high fidelity avatars in the medium capable in expressing emotions precisely and correctly to the audiences from different countries? Or more precisely, can the audiences from different nations and/or cultural groups perceive the emotions of the avatars correctly as the avatars intended to deliver? Our study focuses on the ability of whether the high-fidelity 3D face avatar can deliver various types of emotions to different intercultural groups. For example, when motion capture data of an actor from a nation or culture expresses various types of emotion, we are interested in observing whether the audience from same nation or cultural background would perceive emotions more correctly than the audience from different nation or cultural background, or audiences from both groups perform in same level. The data would allow us to verify whether the national or cultural background of the face expressors affect the performance of the perceivers. In conclusion, we will determine whether the locally generated high fidelity 3D facial avatar is effective before the global audiences, accommodating the future entertainment industries in creating the avatars in more effective way before the global audiences.

Although there is no study available concerning the high fidelity 3D facial expression study in intercultural format, there are many studies available in regard to the

usage of avatars as Embodied Conversational Agents (ECAs) in human computer interaction (HCI) applications [Andre et al., 1998; Cassell et al., 2000; Deng et al., 2006; Fabri et al., 2002; Gratch and Marsella, 2006; Gratch et al., 2002; Lewis and Purcell, 1984; Nass et al., 1998; Marsella and Gratch, 2001; Rist et al., 1997; Ruttkay et al., 2002]. However, few studies have been conducted to measure the effectiveness of the avatars in intercultural settings where an avatar designed based on one cultural setting is presented to users who are from different cultural backgrounds. In particular, there is one area that lacks more vigorous research on the avatars' effectiveness measurement: *Emotional expressions generated by the avatars and the corresponding emotional perceptions and subject identifications by the users from the different cultural backgrounds.* There are limited (if any) research efforts among the existing research that investigated the effectiveness of the avatars using a realistic 3D facial avatar embedded with high-fidelity emotion expressions.

In the field of psychology, whether facial expressions and corresponding perceptions are universal or culturally specific has been a debated issue for a long time, and noticeable research has been made to prove (or disprove) the validity of one or another. Some recent studies have shown culturally specific nature of the facial expressions and their perceptions by measuring the effectiveness of the emotion perception by different cultural groups [Elfenbein and Ambady, 2002, 2003; Elfenbein et al., 2002, 2007]. Their primary conclusion can be summarized in a simple phrase: "Familiarity breeds accuracy" [Elfenbein and Ambady, 2003]. In other words, they concluded that when an emotion expresser and an emotion perceiver have the same cultural background, the emotion recognition rate of the perceiver was higher than when they have different cultural backgrounds. However, other studies have shown the universality of facial emotion expression and perception [Ekman, 1994; Ekman and Friesen, 1971; Ekman et al., 1987; Matsumoto, 2002, 2007] where cultural background factor has little effect on the facial expression perception process. Although certain insights with regard to the relationship between emotion perception and cultural backgrounds are provided, these studies were conducted mainly with static photos that usually contains only single highlighted frame of the entire expression process rather than with animated avatars. Thus, it is still unclear whether the animated avatars will be able to produce similar results as when the experiments were conducted with the static photos.

In the meantime, in the field of computer science, a few studies investigated a relationship between emotion perception and cultural backgrounds using computer-generated avatars. In their study, Bartneck et al. [2004] observed that expression perceptions and culture of expression perceivers were generally independent when two simple iconic avatar animations with various emotions were shown to participants. However, they also identified some of the avatar animations that were perceived differently by the participants from the different cultures. On the other hand, the study by Koda and Ishida [2006] showed that the cultural backgrounds of the perceivers clearly affected how they perceived the expressed emotions by comic/anime type avatars (both human and non-human animations) with twelve different emotions. These studies, similar to the studies by the psychologists, still lack in validating whether the cultural background affects the perception of the emotions when realistic 3D avatars embedded with high-fidelity facial expressions are used as a primary stimulus instead of the non-realistic iconic or comic/anime drawn avatars.

In this study, we compare performance difference in subject identification, emotion perception and confidence rating among different cultural groups when they are shown clips of avatar animation where very realistic 3D avatars are embedded with highly-believable, and culture-specific emotions. Rapid advancements in avatar-based technologies will inevitably force the usage of the even more realistic 3D avatars with highly-believable emotions in the entertainment industry. It is our attempt to answer whether realistic avatars' expressed emotions can be universally perceived by the users from different cultural backgrounds. We intend to accomplish our goal by investigating whether culture-specific emotions must be embedded on the avatar among different cultural groups or the emotions from one cultural background can be used effectively to other cultural groups.

In addition to the emotion perceptions, we also conduct a subject identification study to look into whether the perceivers are able to identify the identity of the subjects who express the particular emotions embedded on the avatars. Like the emotion perception study, we aim to study whether the participant group from the same background as the subjects has a higher accuracy in identifying the subject than the groups with the different cultural backgrounds. We compare the accuracy of subject identification with the confidence ratings to look into any meaningful correlation between them.

Figure 1. A VICON motion capture system

In our experiment, we use a 3D facial model embedded with detailed, highly-believable emotions to prove (or disprove) whether the people of one cultural group can recognize expressed emotions of their own cultural group more successfully than when they attempt to recognize the expressed emotions of the other cultural groups. This is accompanied by a question asking them how confident they are with their choices on identifying each emotion expressed by the avatar animations. To conduct this experiment,

we created a realistic avatar using two principal components. The first component comes from a scanned high quality 3D facial model. The second component is high fidelity expressive facial motion acquired by an optical motion capture system (Figure 1). We captured the expressive facial motion of three chosen human subjects while they spoke three English sentences with five different emotions per each. First four emotions are anger, sadness, happiness, seriousness, and one of two emotions – surprise or disgust. After motion data acquisition, the captured expressive facial motion data were transferred to the pre-constructed static 3D face model to generate animated faces. In order to get familiarized with how the individuals uniquely express the emotions and how different emotions are expressed in different ways, selected video clips of the subjects during the facial motion capture process and the corresponding animated clips were shown simultaneously to two groups of participants. Then the groups were shown only the animated faces without voice to recognize the type of the emotion expressed in each case. We also asked the participants to rate the confidence level of their decisions in each case to study whether the intensity of the facial expressions would impact the participants' confidence ratings as well as the accuracy of the emotion perceptions.

In terms of experimental design consideration, we broadly categorized the students either as US students or International students instead of dividing them into more nation/culture-specific groups. We are aware that there are hundreds of nationalities/cultures available around the globe. However, we consider our study as a first step in this area of research. We believe our study will stimulate other researchers to follow the step.

The remainder of the paper is organized as follows: Section 2 briefly reviews recent efforts that are most related to our work. Section 3 describes setups and procedures of our experiment. Section 4 describes the results and corresponding conclusions. Then it is followed by discussion (Section 5) and future directions of our work (Section 6).

2. RELATED WORKS

A significant number of user studies had been conducted to evaluate a human-like computer interface for a broad variety of applications [Bailenson and Yee, 2006; Bente et al., 2001; Bonito et al., 1999; Busso et al., 2004; Guadagno et al., 2007; Hongpaisanwiwat and Lewis, 2003; Katsyri et al., 2003; Koda and Maes, 1996; Nowak and Rauh, 2005; Sproull et al., 1996; Walker et al., 1994]. Among these evaluation research studies, evaluating the avatars' usability and impact on task performances under various variations has been a hot topic [Bente et al., 2001; Pandzic et al., 1999]. For example, researchers performed user studies to evaluate the impact of using one's own face as the avatar in given tasks [Nass et al., 1998] or measure the performance based on the level of the visual realism of avatars [Pandzic et al., 1999]. Currently, one of most debated issues is how the degree of realism that avatars convey affects the usefulness of avatars in various applications. There are two possible outcomes in regard to this issue. First, the degree of realism avatars convey is independent to the usefulness of avatars. The study by Zanbaka et al. [2006] acclaimed that the similarity in persuasiveness of the avatars was observed regardless of the degree of visual realism of the avatars or even whether the avatars were human or non-human characters. Second, the degree of realism avatars convey affects the usefulness of avatars. A recent study based on a meta-analysis of 46 studies [Yee et al., 2007] concluded that the usefulness of more realistic human-like

representation in social interaction settings were greater than less realistic counterpart. Findings like Yee et al. [2007] provide an essential reason for researchers the importance in researching further on development of more realistic human avatars, their applications, and subsequent evaluations. And confirming the conclusion from Yee et al. [2007], Garau et al. [2003] and Kang et al. [2008] also emphasize in their study how combination of high quality visual realism and high fidelity behavioral realism of the avatar affect the perception and performance positively in their studies. Our study focuses on the evaluation of highly realistic 3D facial avatars embedded with high-fidelity facial emotions in subject identification, emotion perception and confidence ratings by the perceivers from different cultural backgrounds.

There have been many studies that investigated the embedment of emotions on animated avatars. Some studies such as the work by Fabri et al. [2002] used a generic face where the facial expressions were made by exaggerating or manipulating the size of eyes, mouth and eyelids from a generic facial model. On the other hand, the face designed sophisticatedly by applying expressive facial motion data to embed the human emotion [Cassell et al., 2000]. Researchers also looked into effectiveness studies using the expressive motion capture data to observe how people perceive visual emotions [Cassell et al., 2000; Deng et al., 2006]. Two particular areas in the emotion perceptions studies drew attentions to researchers: Emotion perception studies based on geometrical emotion intensity and intercultural backgrounds. There was a study to evaluate perceived emotion intensity based on the geometrical emotion intensity (or degree of the emotional expression intensity) and emotion recognition accuracy. Hess et al. [1997] investigated how the intensity of emotional expression (20%, 40%, 60%, 80% and 100%) affected the perception accuracy of four emotions (anger, disgust, sadness and happiness) using photographs as stimuli. The result revealed that the accuracies on each emotion were affected largely linearly by the emotional expression intensity. Similar study was conducted by Bartneck and Reichenbach [2005] while differentiating itself with the work of Hess et al. [1997] by using a cartoon-type human facial avatar as a stimulus. In this study, animated avatars with 5 different emotions (happiness, sadness, anger, fear and surprise) with ten different geometrical emotion intensity (10% - 100%) were shown to participants to rate the perceived intensity and the difficulty as well as the identification of the emotion. A curve-linear relationship between the geometrical intensity and the perceived intensity was observed. Meanwhile, the recognition accuracy was indifferent after the geometric intensity reached 30%. In the case of the difficulty study, the participants rated the task as difficult when the geometric intensity was low while they rated the task easy when the geometric intensity was high. In addition, they rated the fear as the most difficult emotion to identify. In our experiment, we create high-fidelity 3D avatars based on the captured emotions from three human subjects with three levels of facial expression intensity (less expressive non-actor, expressive non-actor, and amateur actor). Then we ask participants for the emotion identity and the rating based on how confident they are for their choices. We aim to uncover how the facial expression intensity can affect the accuracy of emotion perception.

How the expressive motions of non-realistic avatars can be perceived by participants with different cultural backgrounds were also studied [Bartneck et al., 2004; Koda and Ishida, 2006]. In the study by Bartneck et al. [2004], Japanese and Dutch participants were asked to rate the arousal and valence of the animations of two simple iconic avatars

– created by a Japanese professor – with 30 different sets of emotion. They found that, in general, the emotions are culturally independent in the perception study although some animations revealed that the culture clearly affects the perception. In the meantime, Koda and Ishida [2006] used 40 avatars (human figures, animals, plants, objects and imaginary figures) created in a comic/anime drawing style by Japanese designers to observe whether the participants from 8 different countries (Japan, Korea, China, France, Germany, UK, USA, and Mexico) would be able to identify 12 different facial expressions (happy, sad, approving, disapproving, proud, ashamed, grateful, angry, impressed, confused, remorseful and surprised). They found that the recognition accuracy of the facial expressions by Japanese participants was the highest and further acclaimed that the cultural backgrounds of the participants affect the perception of the avatars' facial expressions. In addition, they found that the participants from Korea had second-highest accuracy. The authors [Koda and Ishida, 2006] interpreted this as the result of cultural similarity between Japan and Korea due to the spatial proximity of the two. This study concluded that cultural backgrounds affect the perception of the avatars' facial expressions as it does on the human facial expression perception. In both studies conducted by Bartneck et al. [2004] and Koda and Ishida [2006], iconic and non-realistic human facial avatars were chosen as the primary stimuli to investigate whether cultural backgrounds affect the identification of the avatars' emotions. However, they did not provide enough conclusive evidence whether the perceived facial expressions of the avatars and the cultural backgrounds of the participants would be independent or not if high-fidelity 3D human facial avatars were used as primary stimuli. In our study, the participants will be divided into two groups: US and international. We will observe whether the emotion perception is culturally independent in general as the study from Bartneck et al. [2004] or is affected by the cultural background of the participants as the study from Koda and Ishida [2006].

Several studies were also conducted by psychologists to verify whether the accuracy of the emotion perception is affected by cultural backgrounds. Some of the recent studies have concluded that the cultural backgrounds of the facial expression expressers and perceivers do affect the emotion recognition accuracy [Elfenbein and Ambady, 2002, 2003; Elfenbein et al., 2002; Yuki et al., 2007]. When the cultural background of the emotion expresser matched the cultural background of the emotion perceiver, the resulting success rate was higher than when two had a mismatch. It was due to the fact that individuals from the same cultural background were exposed in a greater degree to the expressions that were generally accepted in that culture. Furthermore, greater exposure to the particular culture allowed the members of the culture to get familiarized with the socially-approved emotion. The study by Yuki et al. [2007] went a step further by identifying how Americans and Japanese perceive the emotions differently due to their tendency to emphasize the different facial expression cues. The study confirmed that Japanese weighted more heavily on the eye areas during the emotion perception process while Americans weighted more heavily on the mouth areas. Therefore, the study not only verified the cultural specificity in the facial emotion expression perceptions but also revealed the mechanism concerning people in different cultures expressing and perceiving facial emotions differently.

However, many studies also demonstrated that the emotion perception process was universal regardless of the cultural backgrounds of the facial expression expressers and

perceivers [Beaupre and Hess, 2005; Ekman, 1994; Ekman and Friesen, 1971; Ekman et al., 1987; Matsumoto, 2002, 2007]. In his study, Matsumoto [2002] pointed unwarranted nature of "in-group advantage in judging emotions across the culture" from the meta-analysis of Elfenbein and Ambady [2002] based on the fact that data were either under-qualified for methodological requirements to examine in-group advantage or did not demonstrate in-group advantage after reexamination. Again, in another study, he reconfirmed that the nationality of the expressers and the perceivers do not affect emotion perception process according to the observation that the perceivers did not alter their judgment on emotion perception process based on the nationality of emotion expressers that they judged [Matsumoto, 2007]. Beaupre and Hess [2005] observed whether there was in-group advantage among French-speaking native Canadians, and sub-Saharan Africans and Chinese immigrants in emotion expression and perception processes, and found that their data did not support the in-group advantage.

In all the studies mentioned above, photographs were used as stimuli for the emotion perception. Study by Katsuri [2006] concluded that the performance of emotion perception expressed by dynamics faces was higher than perception expressed in static photographs. In our study, high-fidelity 3D facial avatars are used as the stimuli to observe whether the high-fidelity 3D facial avatar have the same effect as the static photos of real people in terms of the emotion perception by different cultural groups.

The usage of the "forced-choice" experimental design was highlighted [Elfenbein and Ambady, 2003] for the cross-cultural emotion communication studies where the perceivers respond from a predetermined list after viewing the expressions. Frank and Stennet [2001] offered flexibility in the "forced-choice" experimental design where the perceivers have options to choose "inconclusive" if they cannot perceive the type of the emotion. Our study implements non-forced choice experimental design by providing an option "unidentifiable" to the perceivers.

Several studies demonstrated that dynamics such as head motions and emotion transitions were also one of vital components in person identification. Among them, two studies found that people were more accurate in identifying moving faces than still faces in person identification. In their study, Knight and Johnston [1997] showed participants two types of video footages where one was still faces and another moving faces to test whether the face movements were able to aid the participants in person identification. In addition, the authors also showed the videos both normal (positive image) and degraded (negative image) forms. They found that the participants were more successful in person identification when they were shown moving faces in degraded form. Lander et al. [2001] also observed similar results when they degraded the video footages by either pixelation or blurring and showed them to the participants for person identification. Regardless of degree of degradation by either pixelation or blurring, participants were more successful in identifying the identity of moving face than still face.

In the meantime, Katsyri and Sams [2008] performed basic emotion perception study using synthetic and natural faces as stimuli when they were shown either in static or dynamic form. They observed dynamic synthetic expressions were more successfully identified than static synthetic expressions while the natural expressions did not display any significant difference between dynamic and static faces. The stimuli of our study also include facial dynamics such as head motions and expression transitions from neutral

to specified emotion. We examine whether and how the dynamics affect in our subject identification and emotion perception experiments in the intercultural settings.

3. EXPERIMENTAL SETUP

The purpose of this study was to investigate whether national/cultural backgrounds of US and international participants affected their subject identification and emotion perception accuracies in a setting where they were shown a series of high-fidelity 3D facial avatar animation clips. In addition, we also designed our study to observe effects of expression and dynamics intensities of the avatar on the groups. In our study, we first captured data of three human subjects speaking in English and expressing designated emotions using a VICON motion capture system. After processed the data, we transferred them to a photorealistic 3D facial model and created a total of 45 facial animation clips. We chose 15 clips, arranged them randomly, and asked participants to complete a survey that includes the questions concerning subject identity, emotion perception and confidence level of their choices.

3.1. Facial Animation

To create high fidelity 3D avatars with realistic facial expressions, we decided to capture the expressive facial motions of three human subjects who were chosen based on following criteria:

Subject 1 (S1): Non-actor with low intensity of dynamics and expression
Subject 2 (S2): Non-actor with low dynamics intensity and medium expression intensity
Subject 3 (S3): Actor with high intensity of dynamics and expression

First two subjects, S1 and S2, were the computer science graduate students while the third subject (S3) was a senior majoring in theater/performing art at the University of Houston. Their ages ranged from 25 to 30 (M = 27.33, SD = 2.52). We categorized S1 and S2 as low and medium expressive non-actors based on the degree of their facial expression intensity in a natural setting. We chose the non-actors because we wanted to capture the facial expressions with low and medium degree of expression and dynamics intensities that they naturally made in normal case. In relative term, when high expression intensity is regarded as 100%, low expression intensity was ranging in 25%-35% while medium level intensity was ranging in 60%-70%. In case of dynamics intensity, we defined low intensity as low frequency and velocity head movements (or near-absent of the movements) while a high intensity as high frequency and velocity head movements. We excluded the verbal quality (i.e., the volume of the subjects' voice) from categorization criteria since our study is focused on non-verbal facial expressions.

We used a VICON motion capture system to record high-fidelity expressive facial motions of the three subjects at a 120 Hz sampling frequency while an off-the-shelf digital video camera was used to record the motion capturing process simultaneously as it focused on the subjects' frontal faces. A total of 10 VICON MX-40 cameras were used to capture detailed facial expressions. We placed 99 facial markers (each marker with approximately 5mm in diameter) on feature points of each subject to capture the unique facial expression as well as subtle differences between the subjects while they expressed

their facial emotion (Figure 2). The facial marker layouts used for the subjects were remained the same. We also placed 4 head markers (each marker with about 16mm diameter) on each subject's head to capture his/her head motions during the emotion capture sessions. We asked each captured subject to speak three pre-selected English sentences 5 times each with corresponding emotions:

Sentence 1: "*This was easy for us.*"
Emotion: Anger (ANG), sadness (SAD), happiness (HAP), serious (SER), surprise (SUR)
Sentence 2: "*Those thieves stole thirty jewels.*"
Emotion: Anger (ANG), sadness (SAD), happiness (HAP), serious (SER), disgust (DIS)
Sentence 3: "*She is thinner than I am.*"
Emotion: Anger (ANG), sadness (SAD), happiness (HAP), serious (SER), surprise (SUR)

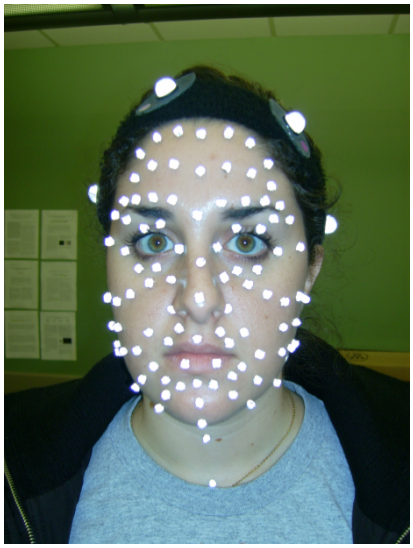


Figure 2. A captured subject with facial markers

We intentionally chose short sentences to aid the subjects to express their emotions as naturally and realistically as possible. The subjects, especially S1 and S2 in particular, had difficulties to express and maintain each emotion when the sentences were long.

In our study, we chose 6 emotions (anger, sadness, happiness, serious, surprise and disgust). In the exception of "serious" emotion, the emotions were based on the universal basic emotion classification (Ekman and Friesen, 1971).

Using the high-fidelity optical capture system such as the VICON system, it was possible to capture distinct and detailed dynamics and characteristics of facial expressions expressed by different subjects. In addition, as an effort to create a realistic avatar, we considered beforehand how the eye motion of the avatar must be handled throughout the motion capture process. Garau et al. (2001) demonstrated how eye

movement of an avatar in a conversational instance impacted the quality of communication. Here, the avatar with the gaze related to the conversation outperformed in communication compare to the avatar with the random gaze. In our study, since the motion capture process cannot capture the eye movement, we asked the subjects to gaze straightforward while they are speaking during the motion capture session. The fixation of the eye movement to the straightforward direction alleviated the avatar from a possible adverse effect where the realism of the facial animation of the avatar suffers negative impact due to uncoordinated eye and facial direction.

We created beforehand a high-quality 3D facial model of a middle-age American Caucasian female subject using a 3D scanner. We chose a Caucasian female 3D model for our study due to the following two reasons. First, in this study, we aimed to investigate the subject identity and the emotion perception of the 3D avatar speaking American English with American-style facial expressions. Therefore, we wanted our visual stimulus be as close to an American as the language and the facial expressions. Second, a previous study (Zanbaka et al., 2006) showed a cross-gender influence in persuasion where women were influenced more by male avatars and men were influenced more by female avatars. This cross-gender influence was observed regardless of whether the avatars were real human, virtual human or virtual character. In our study, we had a total of 12 female and 46 male participants. Therefore, to enhance the performance of the participants, we decided to use the female avatar instead of a male avatar.

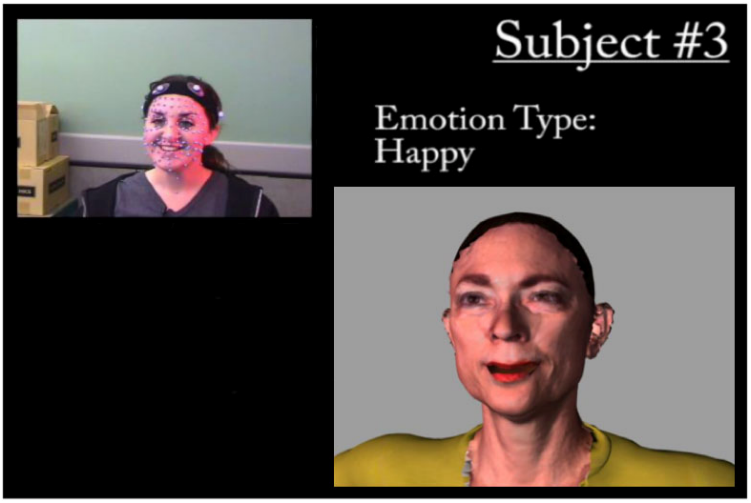


Figure 3. Video of a subject during the motion capture session along with the animated face from the subject's motion capture data used to let the survey participants to get familiarized with subjects' facial emotion expressions

We then transferred the captured facial motion data onto of the chosen 3D facial model to synthesize corresponding facial animations for each subject. This process allowed subjects' identities to be hidden behind the identical 3D model for later subject identification experiment. The process was similar to the studies from Knight and Johnston (1997) and Lander et al. (2001) where they masked true identities of the

subjects by processing their stimuli video clips. With the 3D model and the motion data, we created facial animation clips using Autodesk (formerly known as Alias) Maya software. Each clip contained a different subject speaking three sentences with different emotions. Similar to study from Katsyri and Sams (2008), the animation clips retained the dynamics information of both head motions and facial expression transitions (e.g. transition from initial neutral expression to designated expression in each data).

We emphasized on both the high quality visual realism and the behavioral realism of the avatar in creating process of the animation clips. Although the avatar suffered some degree of loss of visual realism during the process to create the avatar animation, the resulting animation still carries enough facial detail to be considered very realistic in visual realism. Using the facial avatar animations and the footages of the motion capture process of the subjects, we created two types of clips for the study. The first type of clip contained the animation of the facial avatar and the corresponding recorded video sequence of each subject speaking during the motion capture process in synchronization (Figure 3). This was designed to get the participants familiarized with both subtlety and distinction of the facial expressions in different emotion types expressed by different subjects. For example, the clip was able to show distinguishable characteristics when a single subject had the happy (HAP) facial expression and the angry emotion (ANG), or when two different subjects had the same surprise (SUR) facial expression. We showed movie clips of the motion capture process instead of static photos since people who went through learning faces using movies instead of still pictures prior to the person identification were more successful in identifying person (Roark et al. 2003). The second type of clip contained a non-voice animation clip where the avatar spoke one of the sentences with one of the studied emotions (Figure 4). Out of total 45 possible combinations, we chose 15 animation clips and randomly arranged them for the perception experiment.

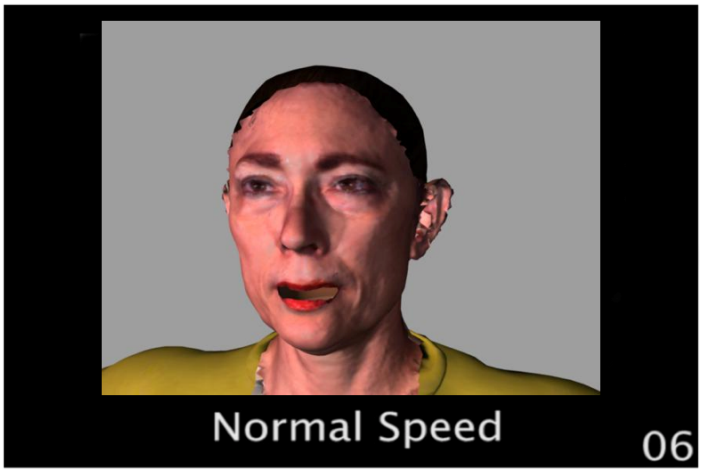


Figure 4. The animated face alone for the subject identification and the emotion perception

3.2. Study

The participants of the survey were asked to perform the following three tasks: *Identify the subject, recognize the emotion type and rate the confidence ratings*. For the survey, we choose two groups of people based on following status:

Group 1 (G1): US students (24)
Group 2 (G2): International students (34)

Fifty-eight participants of age between 19 and 37 attending the University of Houston volunteered for the study. The students from various disciplines (i.e. computer science, engineering, and business majors) were divided into two groups based on the nationality as Bartneck et al. (2004) did in their study: 24 US students and 34 international students. All the US students were born and lived in United States. They were composed with 16 Caucasian, 5 Hispanic, 2 Asian and 1 African-American ranging from 19 to 34 years of age (M = 24.92, SD = 4.59). Two out of 24 US students were females (8.38%). On the other hand, the international students are composed of 17 Chinese, 5 Indians, 3 Mexicans, 2 Tunisians, 1 each from Jordan, Korea, Russia, Ukraine, Iran, Nigeria, and Indonesia ranging from 22 to 37 years of age (M = 25.63, SD = 4.36). They resided in US between 6 months and 2 years prior to this study. Ten out of 34 international students were females (29.41%). We formed these two groups (G1 and G2) to study how they respond differently (or similarly) to the US avatar animation clips in terms of subject identity, emotion perception, and confidence rating.

To prevent any potential interaction between the two groups, they completed the survey in separate locations. They were also asked not to interact with one another within the group to obtain uninfluenced results from the participants.

We conducted our study using a nonforced-choice method. The survey was conducted in a two-phase protocol. In the first phase, we handed out the survey and showed the first type of clip as the stimulus (containing the motion capture processes of three subjects and the corresponding avatars) to the participants. The participants were asked to pay close attention to unique subtleties and distinctiveness between different subjects as they expressed various emotions. The stimuli were shown three consecutive times for the groups to get familiarized with the distinctive differences between the facial expressions and the subjects' unique ways to express the emotions. We designed this familiarization process based on the study by Roark et al. (2003) where it demonstrated that more face learning experiences prior to the recognition experiment improved recognition results. In the second phase, they were shown the second type of clip containing randomly selected 15 animated avatar clips to answer a questionnaire composed with questions concerning subject identity, emotion type and confidence rating. First, after viewing each clip, they were asked to identify the subject in the animated clip. They were asked to choose from one of following four options: Subject 1 (S1), subject 2 (S2), subject 3 (S3) or unidentifiable. Second, they were asked to perceive the emotion type contained in each animated clip, and choose one from following seven options: Anger (ANG), sadness (SAD), happiness (HAP), serious (SER), disgust (DIS), surprise (SUR) or unidentifiable. And finally, they were also asked to fill out confidence ratings.

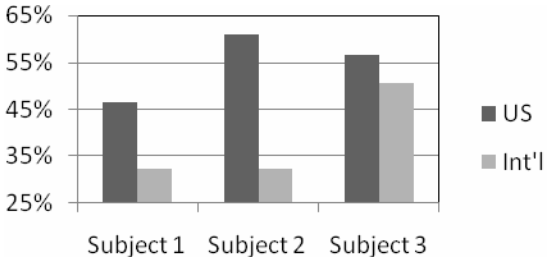
The rating was designed in a one-to-ten scale where one is considered as "absolutely unconfident" while ten was considered as "absolutely/superbly confident". To aid the groups for more detailed observation, each clip was played twice in a normal speed and then once in a 33% of normal speed. Afterward, a ten second pause was given to the participants to answer each survey question.

4. RESULTS AND CONCLUSIONS

4.1. Subject Identity and Confidence Rating

In subject identification study, we used a 2 (group) x 3 (subject) analysis of variance (ANOVA) to examine the accuracy with which groups identified the subject on which each avatar's expressions were based. We found that US students group (G1) had higher performance in subject identification experiment than International students group (G2), $F(1,159) = 9.00$, $p < .01$. When the avatar with embedded expressions of the non-actors (regardless of whether one was either low expressive non-actor (S1) or medium expressive non-actor (S2)) was shown in the subject identification experiment, G1 had higher success in identifying the right subject (46.38% to 32.29% in identifying S1 and 60.87% to 32.29% in identifying S2). Figure 5 shows how each group performed in overall subject identification experiment. On the other hand, the accuracies of G1 and G2 in identifying the actor (S3) were 56.52% and 50.69% where the difference was comparatively low (5.83%). Based on these results, we can draw a conclusion that the subject identification accuracy rates are higher when the cultural background (or nationality) of the subjects and observers is identical. However, this is applicable only when the facial expressions data of real, everyday people are embedded on the avatar. When the facial expressions of actors are used, cultural familiarity advantage (or in-group advantage) becomes irrelevant. This implies that the entertainment media using the avatars with high behavioral realism will not cause the global audience to suffer cultural unfamiliarity effect in the subject identification as long as the level of the facial expressions of the avatars can resemble that of the actors. In another word, cultural backgrounds of perceivers will not affect the accuracy of the subject identification if behavioral intensity of avatars' facial expressions is high enough to resemble the intensity of the facial expressions expressed by the actors. The difference in number of participants between G1 and G2 was justified based on results that the performances of two groups were compatible in identifying the actor while G1 performed with higher accuracy in identifying non-actors. There is an alternative way to explain why G1 and G2 performed in close level of accuracy in identifying S3 while G1 outperformed G2 in identifying S1 and S2. During the facial motion capture processes, we observed almost no dynamics from non-actors while the actor's dynamics were intense. The information related to the dynamics provided G2 sufficient information to identify S3 as well as G1 did. We conclude that the dynamics can also offset cultural unfamiliarity disadvantage that G2 had in the subject identification. On the other hand, since there were almost no dynamics in S1's and S2's motion data, the cultural unfamiliarity disadvantage influenced G2 more significantly in identifying S1 and S2. We verified this significant effect of subject (or more specifically the dynamics intensity of the subject) on the subject identification, $F(2,159) = 7.79$, $p < .01$. And there was also a significant

interaction between the group and the subject, $F(2,159) = 223.81$, $p < .01$. Therefore, in subject identification, we conclude that the cultural familiarity advantage affects greatly in the absence of the dynamics information. However, the presence of dynamics can effectively offset the cultural unfamiliarity disadvantage factors. Therefore, as long as it is possible to embed substantial intensity in dynamics and facial expressions, very realistic 3D avatars in the entertainment media can be effectively deployed for audiences regardless of their cultural backgrounds. This finding implies that when highly believable 3D avatars are created, it is not necessary to embed each avatar with culturally specific facial expressions to accommodate global audiences with diverse cultural backgrounds.



Subject	Subject 1	Subject 2	Subject 3
US	46.38%	60.87%	56.52%
Int'l	32.29%	32.29%	50.69%
Difference	14.09%	28.58%	5.83%

Figure 5. Successful 3 subject identification rate of US and International students

Although the confidence rating was rated based in emotion perception, we compared it with the subject identity accuracy to find any correlation between them. In general, the confidence ratings and the subject identification accuracies corresponded well one another. In other words, higher confidence rating corresponded to higher success rates in identifying the subject. Figure 6 shows how each group rated in overall confidence rating in conjunction with the subject identification. As expected, both G1 and G2 gave the highest confidence rating to S3 followed by S2 and S1 according to the degree of the emotional facial expression observed in the study by Bartneck and Reichenbach (2005). We also concluded that strong presence of the dynamics in S3 contributed to G1 and G2 to rate S3 highest in confidence rating. A 2 (group) x 3 (subject) ANOVA verifies this finding, $F(2,162) = 3.91$, $p < .05$. In the meantime, G1 rated the confidence ratings slightly higher than G2 in all cases. One might argue that G1 rated the confidence ratings higher since they were familiar with the socially agreed emotions in US culture (Elfenbein and Ambady, 2003), and concluded that less familiarity with the culturally accepted stereotypical emotional facial expressions led G2 to lower their confidence ratings as well as to perform less accurately in the subject identification. On the other hand, G1, equipped with familiarized culturally accepted stereotypical emotional facial expressions, rated the confidence ratings higher while performing better in subject identification compare to the international students. However, in our study, the

difference in confidence rating between G1 and G2 was not significant enough ($F(1,162) = 0.25$, p = ns) to support this argument. Finally, there was a high significant interaction between the group and subject ($F(2,162) = 260.21$, $p < 0.01$). With the confidence rating and the subject identification accuracy, we can tell that G2 had relatively difficult time identifying S2 and gave a lower confidence rating on S2 compare to G1 (the differences in ratings between G1 and G2 were 0.17 and 0.13 for S1 and S3 while it was 0.26 for S2).

Subject	Subject 1	Subject 2	Subject 3
US	5.15	5.58	6.21
Int'l	4.98	5.33	6.08
Difference	0.17	0.26	0.13

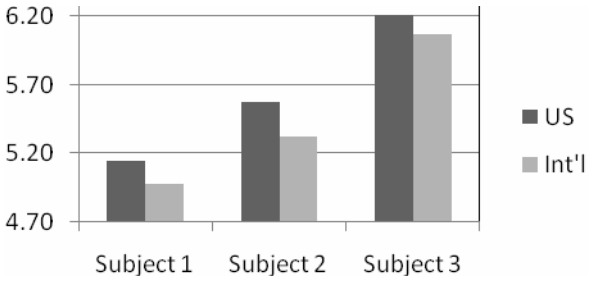


Figure 6. Confidence rating for 3 subjects

4.2. Emotion Perception and Confidence Rating

In emotion perception, the accuracy was assessed in a 2 (group) x 6 (emotion) ANOVA. We found that we had mixed results in the emotion perception experiment among US students (G1) and International students (G2), $F(1,336) = 0.88$, p = ns. Figure 7 shows how each group performed in overall emotion perception experiment. In perceiving sad (SAD), happy (HAP), serious (SER), and disgusted (DIS) emotions, G1 performed at least 7% more accurately than G2. Especially, in SER perception, G1 more accurately perceived than G2 by 18.75%. In perceiving anger (ANG) and surprised (SUR) emotion, G2 performed more accurately than G1 by 7.25% and 6.25% respectively. Based on this data, we drew a conclusion that the emotion perception accuracy rates were not necessarily higher in every emotion type when the cultural background (or nationality) of the subjects and observers is identical. In terms of emotion perception accuracy among different emotion types, we observed significant differences, $F(5,336) = 19.74$, $p < .01$. HAP emotion was most successfully perceived emotion while SAD emotion was least successfully perceived emotion. In addition, there was also a significant interaction between group and subject, $F(5,336) = 67.69$, $p < .01$. For example, other than HAP emotion, G1 was most successful in perceiving DIS emotion while G2 was most successful in perceiving ANG emotion.

Emotion	ANG	SAD	HAP	SER	DIS	SUR
US	43.33%	18.75%	66.67%	43.75%	54.17%	43.75%
Int'l	50.59%	8.82%	57.84%	25.00%	47.06%	50.00%
Difference	-7.25%	9.93%	8.82%	18.75%	7.11%	-6.25%

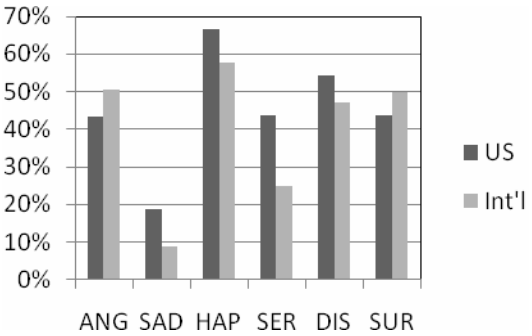


Figure 7. Successful 6 emotion perception level of US and International students

To draw more in-depth analysis, we checked the emotion perception performance data from G1 and G2 when the avatar with a specific emotion type was shown under three different subjects' expression data. Here, we use HAP emotion perception data to observe how the groups performed when observing avatars with different intensity of emotion and dynamics. Figure 8 shows each group's emotion perception data. One obvious trend was the rise of the emotion perception accuracy as the emotion and the dynamics intensity increased. As the groups saw higher intensity avatar animations, they had higher accuracy in HAP emotion perception. Another subtle disposition we found is the impact of the intensity in emotion perception accuracy on G2. When the intensity of the emotion was in either low or medium range, G2 performed worse than G1. However, when the intensity of the avatar's emotions was high, G1 and G2 performed almost identically. Figure 6 shows the difference in HAP emotion perception rates between G1 and G2. We conclude that G2 was more sensitive to the intensity of avatar's expression than G1. In another word, people outside the country of origin of avatar's motion data will be more sensitive in perceiving emotion type based on the intensity of the expression and the dynamics. In conjunction with the overall data, the comparison data between three different intensities of the expression and the dynamics demonstrated that overall higher emotion perception by G1 was due to G2's lower success rate in perceiving emotions when the avatar's intensities of the expression and the dynamics were set to low or medium.

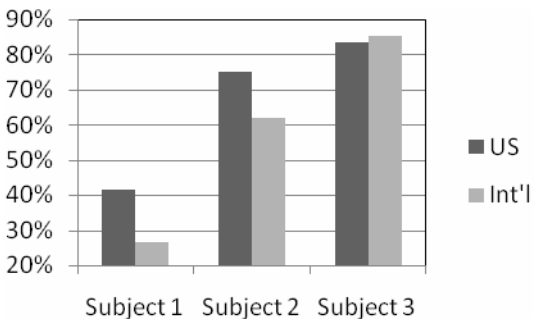


Figure 8. Successful HAPPY emotion perception level of US and International students

Emotion	ANG	SAD	HAP	SER	DIS	SUR
US	6.70	6.32	6.35	6.58	6.30	6.76
Int'l	6.05	5.74	5.63	5.32	5.81	6.30
Difference	0.65	0.58	0.73	1.25	0.48	0.45

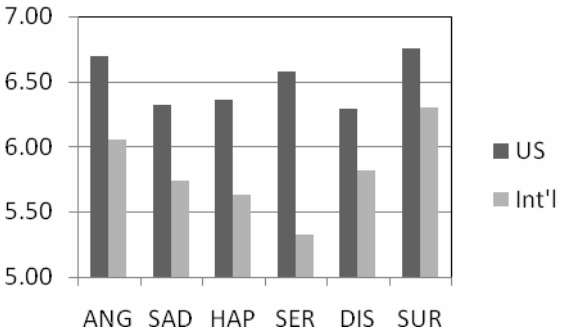


Figure 9. Confidence ratings for 6 emotion types

We surveyed the confidence rating of participants to observe how confident they were in their response to emotion perception. Here, 2 (group) x 6 (emotion) ANOVA test demonstrated that neither group ($F(1,324) = 0.51$, p = ns) nor emotion ($F(5,324) = 1.11$, p = ns) was a significant factor in how the confidence rating was rated. Figure 9 shows the participants' confidence ratings in 6 emotions. G1 gave higher confidence rating in all 6 emotion types. This indicates that G1 at least had higher confidence in their decision due to their familiarity with types of the expressions that the avatars were making. In contrary, G2 gave lower rating than G1 due to lesser familiarity of the expressions. However, the difference in confidence rating between two groups was not significant. On the other hand, there was a significant interaction between group and emotion, ($F(5,324) = 138.43$, $p < .01$). Both group rated ANG and SUR as top two highest confidence ratings. Other than these two emotions, G1 and G2 had different order of ratings: SER, HAP, SAD and DIS for G1 and DIS, SAD, HAP and SER for G2. Also

the degree of difference between highest rated confidence rating and lowest rated confidence rating was more significant in G2 (6.76(SUR) – 6.30(DIS) = 0.99) than G1 (6.30(SUR) – 5.32(SER) = 0.46).

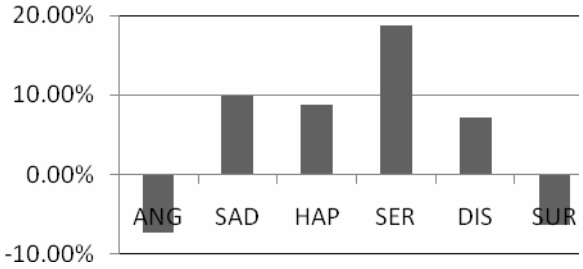


Figure 10. Difference in emotion perception accuracy between US and International students

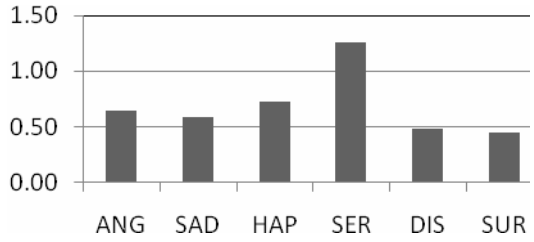


Figure 11. Difference in confidence ratings between US and International students

Comparing the data between the emotion perception and confidence rating, we can observe some matching disposition. For instance, by comparing the difference data from Figure 10 and 11, we can see that there is similarity in SAD, HAP, SER and DIS emotions. As the difference in emotion perception performance, SER emotion confidence rating had highest difference between G1 and G2. Likewise, DIS had lowest difference in both emotion perception and confidence rating among 4 emotions. However, we can see that G2 had higher accuracy in perceiving ANG and SUR emotions even though G2 had lower confidence ratings than G1. Furthermore, there are additional discrepancies between the emotion perception data and the confidence rating. First, G1 and G2 rated their confidence in identifying HAP emotion one of lowest among 6 emotions. However, the actual emotion perception data demonstrate that both group had highest accuracy in perceiving in HAP emotion. Second, both rated almost similar confidence ratings between SAD and HAP emotions. However, the results show that the groups were least successful in perceiving SAD emotion while most successful in perceiving HAP emotion. Therefore, we must conclude that discrepancy between emotion perception and confidence rating data is too obvious to support any disposition between them.

In the meantime, we also observed each group's confidence rating data when viewing the avatars embedded with different subjects' motion data. Here, we use HAP confidence rating data to observe how the groups performed when observing avatars with different intensity of emotion and dynamics. Figure 12 show each group's confidence

rating data. One obvious trend was the rise of confidence ratings as the emotion and the dynamics intensity increase. As the groups saw higher-intensity-data-based avatar animations, they rated confidence rating higher. Low intensity yielded slightly low confidence rating. Another subtle disposition we found was the impact of the intensity in confidence ratings on G2. When the intensity of the emotion was in either low or medium range, G2 rated lower than G1. However, when the intensity of the avatar's emotions was high, G2 gave slightly higher confidence rating than G1. We conclude that G2 is more sensitive to the intensity of avatar's emotion than the G1. In another word, if people outside the country where the avatar motion data is obtained will be more sensitive in confidence rating based on the intensity of the emotion and the dynamics. However, the confidence rating was not affected as severely as the emotion perception in regard to the sensitivity to the intensity since the differences between G1 and G2 in all three cases were no higher than 0.15.

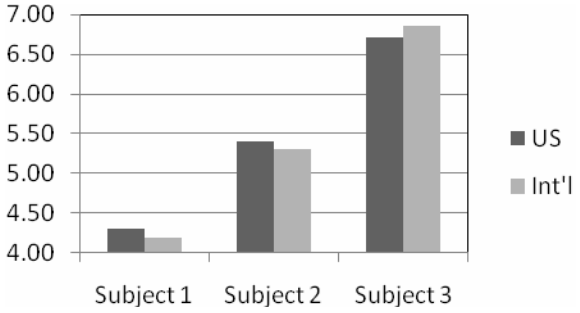


Figure 12. Confidence ratings of HAPPY emotion

5. DISCUSSION

Based on our results, we suggest few guidelines for the high fidelity 3D avatar design and development that can accommodate the global audiences.

5.1. High emotion/dynamics intensity

Most of current entertainment industry captures expression data from professional actors in creating high fidelity 3D avatars, guaranteeing the avatar animation to retain high intensity expression and dynamics. **Having the avatars with high intensity expression and dynamics allows both local audience and global audience to perform in approximately equal level in subject identification and emotion perception of the avatar animations.** It is noteworthy to mention the importance of the dynamics: High intensity of dynamics contributes to aid the global audience in the recognition of subject and emotion type since the dynamics also are parts of emotion expression process. Removal of the dynamics can cause lowering of the global audience's ability to recognize the subject and the emotion types. Therefore, creating the high fidelity 3D avatar animation embedding both high intensity expression and dynamics is a right direction in avatar creation process since embedding the expression data alone will reduce the ability of the global audience in subject identification and emotion perception.

5.2. Low and medium emotion/dynamics intensity

We selected non-actors to create the avatar animation with low/medium intensity expression and low-intensity dynamics data since theirs are more natural than the actors' while maintaining low or medium level of intensity. In terms of the subject identification, the international group (G2) performed equally regardless of the different expression intensity (low in S1 vs. medium in S2) in identifying the non-actors while the intensity of dynamics for both was equally low. This led us to infer that the difference in expression intensity alone does not contribute in change of success rate in subject identification by the global audience assuming the range of intensity remains in low/medium level. Therefore, increasing the dynamics intensity is necessary to assist the global audiences in correctly recognizing the subject with low and/or medium level expression intensity.

In emotion perception, the result was mixed among different emotion types between G1 and G2. G1 was had higher accuracies in emotion perception in SAD, HAP, SER, and DIS emotions while G2 in ANG and SUR emotions. By observing the example of how each group performed in recognizing HAP emotion based on different subject type provided us a hint for the reason for mixed results in emotion perception. We found that G1 and G2 had almost identical level of success in emotion perception when viewing the avatar animations embedded with the actor (high expression/dynamics intensity). On the other hand, G1 had higher emotion perception accuracies than G2 when viewing the avatar animations embedded with the non-actors (low to medium expression intensity and low dynamics intensity). Comparing two non-actors demonstrate how low dynamics intensity suppresses performance of G2 in the emotion perception more severely than G1. Therefore, in designing the avatar animations, it is necessary to include higher level of dynamics intensity to accommodate the global audience in subject identification and emotion perception processes.

5.3. Implementation of emotion into the game

Under current non-player characters (NPCs) design, due to either time, financial, or technical restriction to add facial expressions to the NPCs, alternative methods are being exploited to add emotions to the NPCs. Zubek and Khoo (2002) added a capability to the NPCs to chat with the player where they expressed different types of emotions based on results of fights between the NPCs and the player by exchanging chat messages in a shooting game. In their synthetic character design framework, Kline and Blumberg (1999) designated emotion as one of four major components that influences believable synthetic characters in processes of making realistic decisions and actions. In the meantime, Freeman (2004) introduced "emotioneering" technique where combination of scenario, storyline, gameplay, characteristics and character-to-character relationship-based techniques were used to add emotions into the game. In particular, techniques such as NPC interesting techniques, NPC deepening techniques, NPC character arc techniques, and NPC rooting interest techniques were used to introduce and develop the emotions for the NPCs in core-level NPC designing process. Most of computer and video games follow above rules in implementing the emotions to the NPCs. Therefore, the emotions are usually expressed by the NPCs in terms of the reactions such as immediate actions (i.e. attack the player on sight), relationships (i.e. become close friends), and attitudes (i.e. fearful) toward the player character (PC) rather than in terms of the facial expressions.

In addition, the author also suggested usage of cinematic movie clips as a technique to add (or reveal) the emotions for both PCs and NPCs. Games such as Heavenly Sword and Grand Theft Auto 4 are good examples that apply this technique with detailed facial expressions for both PCs and NPCs. However, only few games are capable to let the NPCs to make the facial expressions. Even for the existing games that implement facial expressions use only generic form of facial expressions. For example, the Elder Scroll IV: Oblivion, one of Bethesda Softworks' RPG games for Microsoft Xbox 360 video game console and 2006 Game of the Year (http://www.bethsoft.com/eng/games/games_oblivion.html), implemented a dialog system with NPCs where, depends on topics the player choose to speak, the NPCs either love, like, dislike, or hate the player. These four emotions are revealed in forms of respective facial expressions. However, these four expressions were low-detail expressions. Furthermore, the differences between two favorable expressions (love and like) and two unfavorable expressions (hate and dislike) were designed by controlling the emotion intensity. If we regard love expression as a generic happy expression with 100% intensity, we can regard like expression as a generic happy expression with about or less than 50% expression. Likewise, if we regard hate expression as a generic anger expression with 100% intensity, we can regard dislike expression as a generic anger expression about or less than 50% expression (or sometimes it more resembles a medium-intensity sad expression). In future we are interested in designing the NPCs with detailed facial expressions in a computer game and investigate whether the results from this study is valid during the game play as well.

6. FUTURE WORKS

Contrary to previous studies done by Elfenbein et al. (2002) and Elfenbein and Ambady (2003), our results revealed that the group of international students performed as accurately as the group of US students in the category of the emotion perception. The result agrees with the study done by Bartneck et al. (2004) that the emotional recognition is culturally independent. To further confirm our findings, we plan to conduct more comprehensive user studies under the similar setting with an increased number of US and the international students while the number of US and the international participants is set to equal. We also plan to equalize the number of the male and the female participants to observe how different genders perform in the study. Under the revised conditions, we will observe whether the unexpected results that we obtained in this study can be reproduced (i.e., the participants rated the confidence rating of HAP emotion relatively low although they were most successful in perceiving this emotion). In addition, we plan to implement our study into a video game environment where the characters (both player and non-player characters) are capable of making their facial expressions during the game. Second, we are interested in splitting the group of international students into several subgroups based on their specific cultural backgrounds (i.e. Indians, Asians, English speaking Europeans and non-English speaking Europeans) under controlled duration of residency in US (i.e. 6 months to a year) to observe whether we can reproduce the same results from this study as the work done by Elfenbein and Ambady (2003). This will provide us how nationality/culture-specific audiences perform differently. Finally, we are interested in performing facial motion capture from a diversified cultural group other than the US (e.g., Chinese). Then we want to use this dataset to conduct the similar

experiment to several national/cultural subgroups (i.e. Americans, Chinese, non-Chinese Asians and Europeans).

REFERENCES

ANDRE, E., RIST, T., MULLER, J., 1998. Guiding the User through Dynamically Generated Hypermedia Presentations with a Life-like Character, *IUI '98*, 21-28, San Francisco, CA.

BAILENSON, J. N., Yee, N., 2006. A Longitudinal Study of Task Performance, Head Movements, Subjective Report, Simulator Sickness, and Transformed Social Interaction in Collaborative Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 15(6), 699-716.

BARTNECK, C., TAKAHASHI, T., KATAGIRI, Y., 2004. Cross-Cultural Study of Expressive Avatars, *Proceedings of the Social Intelligence Design*, 21-27.

BARTNECK, C., REICHENBACH, J., 2005. Subtle emotional expressions of synthetic characters, *International Journal of Human-Computer Studies*, 62(2), 179-192.

BEAUPRE, M. G., HESS, U., 2005. Cross-Cultural Emotion Perception among Canadian Ethnic Groups. *Journal of Cross-Cultural Psychology*, 36(3), 355-370.

BENTE, G., KRAMER, N. C., PETERSON, A., de RUITER, J. P., 2001. Computer Animated Movement and Person Perception: Methodological Advances in Nonverbal Behavior Research. *Journal of Nonverbal Behavior*, 25(3), 151-166.

BONITO, J. A., BURGOON, J. K., BENGTSSON, B., 1999. The Role of Expectations in Human-Computer Interaction. *GROUP '99: International Conference on Supporting Group Work*, 229-238, Phoenix, AZ.

BUSSO, C., Deng, Z., YILDIRIM, S., BULUT, M., LEE, C. M., KAZEMZADEH, A., LEE, S., NEUMANN, U., NARAYANAN, S., 2004. Analysis of Emotion Perception using Facial Expressions, Speech and Multimodal Information. *Proceedings of ACM 6th International Conference on Multimodal Interfaces (ICMI 2004)*, 205-211.

CASSELL, J., SULLIVAN, J., PREVOST, S., CHURCHILL, E. F., 2000. Embodied Conversational Agents. *MIT Press*.

DENG, Z., BAILENSON, J., LEWIS, J. P., NEUMANN, U., 2006. Perceiving Visual Emotions with Speech. *Proceedings of the 6th International Conference on Intelligence Virtual Agents (IVA) 2006*, 4133, 107–120.

ELFENBEIN, H. A., AMBADY, N., 2002. Is There an In-Group Advantage in Emotion Perception? *Psychological Bulletin*, 128(2), 243-249.

ELFENBEIN, H. A., AMBADY, N., 2003. When Familiarity Breeds Accuracy: Cultural Exposure and Facial Emotion Perception. *Journal of Personality and Social Psychology*, 85(2), 276-290.

ELFENBEIN, H. A., LEVESQUE, M., BEAUPRE, M., HESS, U., 2007. Toward a Dialect Theory: Cultural Differences in the Expression and Recognition of Posed Facial Expressions. *Emotion*, 7(1), 131-146.

Elfenbein, H. A., Mandal, M. K., Ambady, N., Harizuka, S., Kumar, S., 2002. Cross-cultural patterns in emotion perception: Highlighting design and analytical techniques. *Emotion*, 2(1), 75–84.

EKMAN, P., 1994. Strong Evidence for Universals in Facial Expressions: A Reply to Russell's Mistaken Critique, *Psychological Bulletin*, 115(2), 268-287.

EKMAN, P., FRIESEN, W. V., 1971. Constants Across Cultures in The Face and Emotion. *Journal of Personality and Social Psychology*, 17(2), 124-129.

EKMAN, P., FRIESEN, W. V., O'SULLIVAN, M., DIACOYANNI-TARLATZIS, I., KRAUSE, R., PITCAIM, T., SCHERER, K., CHAN, A., HEIDER, K., LcCOMPTE, W. A., RICCI-BITTI, P. E., TOMITA, M., 1987. Universal and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, 53(4), 712-717.

FABRI, M., MOORE, D. J., HOBBS, D. J., 2002. Expressive Agents: Non-verbal Communication in Collaborative Virtual Environments. *Proceedings of Autonomous Agents and Multi-Agent Systems (Embodied Conversational Agents)*, Bologna, Italy.

FRANK, M. G., STENNETT, J., 2001. The Forced-Choice Paradigm and the Perception of Facial Expressions of Emotion. *Journal of Personality and Social Psychology*, 80(1), 75-85.

FREEMAN, D., 2004. Creating Emotion in Games: The Craft and Art of Emotioneering. 1st Ed., New Riders Publishing.

GARAU, M., SLATER, M., BEE, S., SASSE, M., 2001. The Impact of Eye Gaze on Communication using Humanoid Avatars, Proceedings of SIGCHI conference on Human Factors in Computing Systems, 309-316. Seattle, WA.

GARAU, M., SLATER, M., VINAYAGAMOORTHY, V., BROGNY, A., STEED, A., & SASSE, M. A. 2003. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. *Proc. CHI'03*.

GRATCH, J., MARSELLA, S., 2005. Evaluating a Computational Model of Emotion. *Journal of Autonomous Agents and Multi-Agent Systems*, 11(1), 23-43.
GRATCH, J., RICKEL, J., ANDRE, E., CASSELL, J., PETAJAN, E., BADLER, N., 2002. Creating Interactive Virtual Humans: Some Assembly Required. *IEEE Intelligent Systems*, 17(4), 54-63.
GUADAGNO, R. E., BLASCOVICH, J., BAILENSON, J. N., MCCALL, C., 2007. Virtual Humans and Persuasion: The Effects of Agency and Behavioral Realism. *Media Psychology*, 10(1): 1-22.
KANG, S., WATT, J. & ALA, S. 2008. Communicators' perceptions of social presence as a function of avatar realism in small display mobile communication device. *Proc. Hawaii International Conference on System Science*.
HESS, U., BLAIRY, S., KLECK, R. E., 1997. The Intensity of Emotional Facial Expressions and Decoding Accuracy. *Journal of Nonverbal Behavior*, 21(4), 241-257.
HONGPAISANWIWAT, C., LEWIS, M., 2003. Attentional Effect of Animated Character. *Proceedings of IFIP INTERACT03: Human-Computer Interaction*, 423-430.
KATSYRI, J. 2006. Human Recognition of Basic Emotions from Posed and Animated Dynamic Facial Expressions, Ph.D. *Dissertation at Helsinki University of Technology*.
KATSYRI, J., KLUCHAREV, V., FRYDRYCH, M., SAMS, M., 2003. Identification of synthetic and natural emotional facial expression. *Proceedings of International Conference on Auditory-Visual Speech Processing (AVSP'2003)*, 239-244.
KATSYRI, J., SAMS, N., 2008. The effect of dynamics on identifying basic emotions from synthetic and natural faces. *International Journal of Human-Computer Studies*, 66(4), 233-242.
KLINE, C., BLUMBERG, B., 1999. The Art and Science of Synthetic Character Design. Proceedings of the AISB 1999 Symposium on AI and Creativity in Entertainment and Visual Art, Edinburgh, Scotland.
KNIGHT, B., JOHNSON, A., 1997. The Role of Movement in Face Recognition. *Visual Cognition*, 4(3), 265-273.
KODA, T., MAES, P., 1996. Agents with Faces: The Effect of Personification. *Proceedings of the 5th IEEE International Workshop on Robot and Human Communication (RO-MAN'96)*, 189-194.
KODA, T., ISHIDA, T., 2006. Cross-cultural Study of Avatar Expression Interpretations. *Proceedings of the 2006 International Symposium on Applications and Internet (SAINT 2006)*, 130-136, Phoenix, AZ.
LANDER, K., BRUCE, V., HILL, H., 2001. Evaluating the Effectiveness of Pixelation and Blurring on Masking the Identity of Familiar Faces. *Applied Cognitive Psychology*, 15(1), 101-116.
LEWIS, J., PURCELL, P., 1984. Soft Machine: A Personable Interface. *Proceedings of the Graphics Interface*, 223-226.
MARSELLA, S., GRATCH, J., 2001. Modeling the Interplay of Plans and Emotions in Multi-Agent Simulations. *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, Edinburgh, Scotland.
MATSUMOTO, D., 2002. Methodological requirements to test a possible in-group advantage in judging emotions across the cultures: Comment on Elfenbein and Ambady (2002) and evidence. *Psychological Bulletin*. 128(2), 236-242.
MATSUMOTO, D., 2007. Emotion judgments do not differ as a function of perceived nationality. *International Journal of Psychology*. 42(3), 207-214.
NASS, C., KIM, E. Y., LEE, E. J., 1988. When My Face is the Interface: An Experimental Comparison of Interacting with One's Own Face or Someone Else's Face. *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 148-154.
NOWAK, K. L., RAUH, C., 2005. The Influence of the Avatar on Online Perceptions of Anthropomorphism, Androgyny, Credibility, Homophily, and Attraction. *Journal of Computer-Mediated Communication*, 11(1), 153-178.
PANDZIC, I. S., OSTERMANN, J., Millen, D., 1999. User evaluation: Synthetic talking faces for interactive services. *The Visual Computer*, Springer-Verlag, 15, 330-340.
RIST, T., ANDRE, E., MULLER, J., 1997. Adding animated presentation agents to the interface. *IUI '97: Proceedings of the 2nd International conference on intelligent user interfaces*, 79-86.
ROARK, D. A., O'TOOLE, A. J., ABDI, H., 2003. Human Recognition of Familiar and Unfamiliar People in Naturalistic Video. *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'03)*, 36-41.
RUTTKAY, Zs., DORMANN, C., NOOT, H., 2002. Evaluating ECAs – What and how? *Proceedings of the AAMAS02 Workshop on Embodied Conversational Agents*.
SPROULL, L., SUBRAMANI, M., KIESLER, S., Walker, J. H., Waters, K., 1996. When the Interface Is a Face. *Human-Computer Interaction*, 11(2), 97-124.

WALKER, J. H., SPROULL, L., SUBRAMANI, R., 1994. Using a Human Face in an Interface. *Proceedings of the SIGCHI conference on human factors in computing system: Celebrating Interdependence*, 85-91, Boston, MA.

YEE, N., BAILENSON, J. N., RICKERTSEN, K., 2007. A Meta-Analysis of the Impact of the Inclusion and Realism of Human-Like Faces on User Experiences in Interfaces. *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 1-10.

YUKI, M., MADDUX, W. W., MASUDA, M., 2007. Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States. *Journal of Experimental Social Psychology*, 43, 303-311.

ZANBAKA, C., GOOLKASIAN, P., Hodges, L., 2006. Can a Virtual Cat Persuade You? The Role of Gender and Realism in Speaker Persuasiveness. *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 1153-1162, Montreal, Quebec, Canada.

ZUBEK, R., KHOO, A., 2002. Making the Human Care: On Building Engaging Bots. AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment.