De-Identified Feature-based Visualization of Facial Expression for Enhanced Text Chat

Shuo-Ping Wang[†], Mei-Ling Chen[†], Hao-Chuan Wang[†], Chien-Tung Lai[†], Ai-Ju Huang[‡]

[†]National Tsing Hua University, Taiwan [‡]HTC Creative Labs

Abstract

The lack of visibility in text-based chat can hinder communication, especially when nonverbal cues are instrumental to the production and understanding of messages. However, communicating rich nonverbal cues such as facial expressions may be technologically more costly (e.g., demand of bandwidth for video streaming) and socially less desirable (e.g., disclosing other personal and context information through video). We consider how to balance the tension by supporting people to convey facial expressions without compromising the benefits of invisibility in text communication. We present KinChat, an enhanced text chat tool that integrates motion sensing and 2D graphical visualization as a technique to convey information of key facial features during text conversations. We conducted two studies to examine how KinChat influences the de-identification and awareness of facial cues in comparison to other techniques using raw and blurring-processed videos, as well as its impact on real-time text chat. We show that feature-based visualization of facial expression can preserve both awareness of facial cues and non-identifiability at the same time, leading to better understanding and reduced anxiety.

Keywords: Enhanced chat, computer-supported collaboration, nonverbal cues, visualization and interaction techniques; motion sensing; de-identification.

Index Terms: H.5.3. Group and Organization Interfaces: Collaborative computing.

1 Introduction

Computer-based communication (CMC) tools introduce features and options unavailable in face-to-face (F2F) communication, such as whether there is visibility or not and whether one can identify the person who she or he is talking to. Properties of communication tools may afford different modes of interaction, which may subsequently affect the processes and outcomes of interpersonal communication (cf. [8]).

In this project, we are interested in investigating how to maintain both awareness of nonverbal cues and visual non-identifiability in communication. We note that the common dichotomy between visual communication tools (e.g., video conferencing) and text communication tools (e.g., instant messaging) may not meet the particular needs of users when people want to express and receive information about facial expressions for better communication, while not allowing others to see them to address technical or social constraints at the same time.

Video communication can convey rich visual cues, including gesture, facial expression, sound, and tone. These rich cues allow people to make naturalistic, face-to-face-like expressions in communication, which may reduce ambiguity of messages and support comprehension. However, one potential problem may occur is that the visibility of interlocutors may raise concerns of privacy when one prefers not to be seen or not to disclose information about the physical context (e.g., when commuting on a crowded bus). Also, there are situations where social bias may arise when communicators can see each other, such as personnel recruitment in organizations where hiring decisions may be influenced by how applicants look [11]. In cultures where individuals tend to be more aware of context information (e.g., nonverbal or peripheral cues available in video) [27], people may also prefer to have control of their visibility for social purposes like impression management.

Even in face-to-face social interactions, people may not always look at the face of the other person for socio-cultural reasons. First, staring at someone's face may be regarded as impoliteness in some cultures. Second, it may embarrass both interlocutors because watching on another's face is an intimate behavior unless you play a social role like interrogator, public performer or juror [12].

On the other hand, the lack of visibility in text communication can lead to ambiguity and misunderstanding. For example, the text message "Yeah, VERY good" can be interpreted as she or he really thinks that it is very good, or merely sarcasm. To complement the lack of nonverbal cues in text chat, one common approach is to provide emoticons such as ":)" as a device to express emotions or intimacy [10]. However, while emoticons are prevalent, and may achieve certain social consequences beyond what face-to-face communication can afford [26], emoticons are discrete and artificial in nature. Emoticons are still different from natural nonverbal cues, such as facial expressions made directly by people with their face muscles.

It is acknowledged that people are adaptive in social interactions and communication, which has the power to overcome the limitation of communication tools [26]. For example, close friends and family members may still be able to understand one another well enough when the communication channel does not provide rich cues, or is noisy and disrupted. The utility and usefulness of visual cues may also depend on social factors such as the pre-established relationships among interlocutors and the context of communication. Because of the wide variety of user needs and the adaptivity of users, we are not trying to replace one communication tool with another. Rather, our goal is to explore possible ways to de-identify interlocutors while still supporting their use of the facial expressions to communicate. We aim to offer a new option for users to handle situations where there is need to maintain both awareness of facial cues and invisibility of other personal or contextual information.

1.1 To See or Not to See? Reconsidering Visibility in CMC System Design

It is increasingly clear from studies of CMC that the role of visibility in communication can vary depending on the tasks, users, and context of communication [8][9][24]. Visibility

[†]haochuan@cs.nthu.edu.tw

16-19 May, Edmonton, Alberta, Canada

Figure 1: Feature-based visualization of facial expression.

available in F2F and video communication can be effective and desirable in situations where nonverbal information matters, such as when the goal is to ensure understanding and to establish shared knowledge among interlocutors [8][9]. Visibility also entails accountability, which can improve participation in collaborative work [18].

On the other hand, invisibility as a feature of text communication does not necessarily lead to suboptimal communicative processes or outcomes. Long-term interactions can compensate for invisibility and the reduction of social cues in relational communication [25][26]. Invisibility affords non-identifiability, which is often a required feature for privacy. Also, invisibility may reduce the sense of social presence and social anxiety, and can support activities where self-expression and autonomy are important, such as group brainstorming [27] or self-disclosure [15].

Visibility and invisibility each appears to have pros and cons, and both can be useful properties when they are used appropriately. From the point of view of technology building and development, we argue that the gap between visibility and invisibility is too large to meet the nuanced, sophisticated goals that designers and users face today. More specifically, there is lack of options between the choices of visibility and invisibility. Both designers and users are in need of communication modalities in-between these two extremes. To investigate this design space, we consider the following question for the purpose of supporting text-based communication — Can we help people convey rich facial cues while non-visible during text chat?

The current work aims to make a system-oriented contribution by identifying and evaluating an interaction technique that reconciles the tension between visibility and invisibility. We present KinChat, a visualization tool that uses motion sensing to capture and track people's facial features (e.g., eyebrow, mouth etc.), and then visualize these features as animated graphic representations of faces to supplement text communication. Figure 1 shows a sample visualized face generated by KinChat. The design can supplement nonverbal facial cues in text chat by showing facial expressions and head movements in real-time without using the regular video channel.

We conducted two studies to examine 1)whether feature-based visualization helps preserve interlocutors' awareness of nonverbal facial expressions while safeguarding privacy at the same time, and 2) the impact KinChat has on interactive text chat. Results from the two studies confirmed that KinChat's approach can protect invisibility at the level of identity recognition and also provide an awareness of facial expression that improves the level of comprehension and reduces anxiety. At the end of the paper, we discuss the implications to future design and research in CMC.

2 DE-IDENTIFICATION AND AWARENESS OF NONVERBAL CUES

Previous interaction techniques and methods on balancing the tension between invisibility and awareness of nonverbal information, and de-identifying facial images can be divided into two categories, image-based or model-based approaches.

The image-based approach diminishes identifiable information of users with image processing techniques. For example, one may replace certain facial features [4] or add filters to the images of human faces [6].

Boyle et al. have studied the impact of different types and levels of image filtering on both awareness of other people and privacy maintenance. They found that blur filters can provide better awareness and privacy protection than pixelized filters [6].

The model-based approach, on the other hand, replaces interlocutors' faces with 3D-based human or non-human representations, such as avatars. The benefit of using avatar is that interlocutors can customize or select their own self-presentations [16]. However, avatars may also contain socially sensitive features like race, gender, and look that may result in unintended or uncontrolled influence on users' perception, experience and behaviors. Previous work also found that in online virtual games, avatars' appearance could shape players' behaviors [20][21].

Both approaches have the capability to successfully de-identify interlocutor's faces; however, they also appear to have their own constraints. Image-based approaches may block the recognition of facial expression and emotion when strong filters are applied to video [6]. Although the model-based approach using avatars can receive some extra benefits, such as increased user motivation [3] and satisfaction [22], the look and feel of avatars could influence communication in an unintended manner.

We aim to develop an alternative visualization technique to account for problems in existing techniques. Rather than considering ways that reduce visual cues with image filter or replace real faces with dissimilar 3D models, we adopt a constructive approach.

Following the design notion of social translucence [13], we focus on identifying which information or non-verbal features that a communicator conveys are crucial to guide smooth communication. We also stick to the principle of parsimony, avoiding adding unnecessary information or features to the communication channel to keep it as lean and succinct as possible. Therefore, model-based approaches like the use of avatars are outside the scope of our current purpose.

In the rest of the paper, we describe our method of feature-based visualization of facial expressions. We present the first study that compared the proposed method to filtered videos on the awareness of facial expression and the protection of identifiability. We intentionally excluded model-based approaches in the evaluation study as the use of avatars would introduce external cues associated with the appearance and design of the avatars per se, which is not the main interest of the current work. In the second study, we then examined the impact of feature-based visualization of facial expression in aspects of communication.

3 VISUALIZATION OF FACIAL EXPRESSION

3.1 Features of Facial Expression

Research in psychology has shown that facial expressions play a crucial role in communication, especially the conveyance and recognition of emotions [2]. These basic emotions include happiness, sadness, anger, surprise, fear and disgust [12]. Facial expressions as mediators of basic emotions were shown to be cross-cultural and universal, and thus can be leveraged generally as a device of non-verbal communication.

This cross-cultural universality and robustness have been validated with populations who had no opportunity to receive information from modern mass media such as TV. Research with aboriginal people living in New Guinea showed that these people

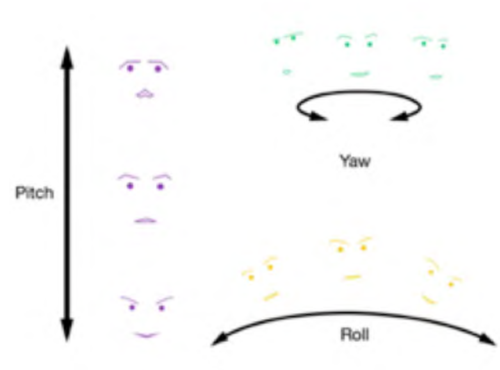

Figure 2: Visualization of rotation of the 3D Head Pose with respect to the three axes.

connect emotions and facial expressions in the same ways as people from the modernized cultural contexts [12].

To produce recognizable basic emotions, the movement of a number of key facial features has been shown to be important, while other features were considered less crucial [2][12]. For instance, for conveying happiness, fear, anger, and surprise the movement of facial features involve mainly the mouth. Nose is the main feature for conveying disgust, and in fact, it is a rare emotion that involves the use of nose. These findings imply that it is viable to succinctly track and visualize a number of key features to create a communication channel that mediates facial expressions.

3.2 Prototyping

We present one's face with 2D graphics (see Figure 1) but not with 3D models as in the use of avatars. This is because different avatars' appearance may still add untended effects on social interaction [20][28]. We are mainly interested in enriching communication by changing the way existing information is presented.

The visualization is implemented with the Microsoft Kinect motion sensor, which has the capability to track the movement of facial features and head position. More specifically, we used Kinect SDK to capture the information of 3D Head Pose and 2D Mesh and Points in our system.

3.2.1 3D Head Pose

The 3D Head Pose detection function in Kinect SDK can not only track the X, Y, and Z positions of the user's head, but also represent them as points in a right-handed coordinate system, where the Z-axis points toward the user and Y points up. With the coordinate positions of the 3D Head Pose, we can detect and visualize the distance between users' face and the Kinect sensor, with the head rotation angles along the three axes (see Figure 2).

3.2.2 2D Mesh and Points

For facial expression, Kinect can track 100 key points on one face. To represent users' facial expression parsimoniously, we selected to track and visualize only 24 of them. The points that we chose are considered as representative features sufficient for facial expression conveyance and recognition as inspired by earlier research in basic emotions. These key points are: four for each eyebrow, four for each eye and eight for mouth (see Figure 3). Another reason that we track key points is that it can save computational time and bandwidth when KinChat renders the visualizations of facial expressions in the chat client.

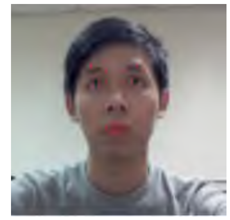

Figure 3: Video clips of facial expressions processed by different de-identification techniques. Top-left: original video; top-right: light blur filter; bottom-left: hard blur filter; bottom-right: feature-based visualization.

4 Study 1: De-identification and Awareness of Facial Information

To evaluate the proposed approach, we conducted two studies. The objective of the first study was to understand whether the visualization technique meets the design goal: the visualization can provides users with an awareness of facial expressions while still de-identifying personal information.

4.1 Hypotheses

We posited the two main hypotheses.

4.1.1 H1: De-Identifying Facial Images

Unlike video or face-to-face, the feature-based visualization does not reveal users' facial outline, hair, nose, ears, and skin. It also simplifies eyebrows, eyes, and mouth to lines. Therefore, it is more difficult to identify a face with the visualization than with videos.

4.1.2 H2: Awareness of Facial Expression

Because of reduced richness of facial features, we expect that there will be lower awareness of facial expression by using our visualization than videos. Because we highlight key facial features, we also expect that there will be greater awareness of facial expression by using our visualization than using image-based de-identification techniques, of which all facial features are blurred unanimously without differentiation.

4.2 Design of the Study

To test our hypotheses, we designed a video content rating study. First, we prepared videos of actors performing certain facial expressions and then processed the videos using different de-identification techniques. Then we asked participants to watch the de-identified videos and respond to a survey that assesses participants' awareness of the semantic content of the videos, and their abilities to identify the actors performing in the videos. As an initial feasibility study for comparing the effects of different techniques, the design of the study is cost-efficient and also can control the influence of other factors present in interactive communication.

4.2.1 Conditions of Information Filtering

We compared four conditions: original video, feature-based visualization, light-blur filtering, and hard-blur filtering. For image-based filtering, we chose to use the blur filters rather than other filters (e.g., pixelization) as blurring was identified to be a more effective technique for performing de-identification than other image processing techniques [6]. See Figure 4 for examples of the four conditions compared in the study.

Figure 4: Video clips of facial expressions processed by different de-identification techniques. (top-left: original video; top-right: light blur filter; bottom-left: hard blur filter; bottom-right: feature-based visualization)

To prepare the videos, we chose four facial expressions for four actors to perform. The four facial expressions include *frowning*, *eyebrow-lifting*, *smiling*, and *surprise*. Originally we intended to ask our actors to perform major facial expressions from the basic emotion categories, including smile, surprise, sadness, anger, fear and disgust. However, because our actors were not professional, they could not perform all the emotions naturally. Thus in addition to two basic emotions they can perform well (smiling and surprise), we asked them to make two expressions that are not among the basic categories of emotion, frowning and eyebrow-lifting. They did not report any problems performing the four expressions we selected. Note that we recruited regular users rather than professional actors to perform the expressions as a way to simulate the types of facial expressions that regular users would see during everyday communication.

We recruited two males and two females as actors. Each actor was asked to perform all four facial expressions. We videotaped all the expressions first. The researchers then applied three image-processing filters on the original video: light blur (Gaussian blur filter with radius 30), hard blur (Gaussian blur filter with radius 80) and the feature-based visualization of facial expression. Therefore, there were a total of 64 videos generated: 4 actors × 4 facial expressions × (3 filters + 1 original video).

We compiled a computerized questionnaire based on the generated videos. There were 64 versions of the questionnaire, one for each of the 64 videos generated. The questionnaires have identical questions and differ only in the video stimulus being used.

4.2.2 Participants

We recruited 64 participants from the campus of a major university and its surrounding community in north Taiwan. 90% of the participants were students pursuing undergraduate or graduate degrees in the university. Females accounted for 26.6% of the participants (73.4% as male). Participants were randomly assigned to complete one of the 64 versions of the questionnaire on a computer.

4.3 Procedure

The questionnaire consisted of four steps. In the first step, participants were instructed to watch a video clip sampled from the pool of 64 videos (as described above) and then recognize which person was in the video by selecting from the photos of the four actors. We evaluated the accuracy of character identification based on the responses obtained in this step.

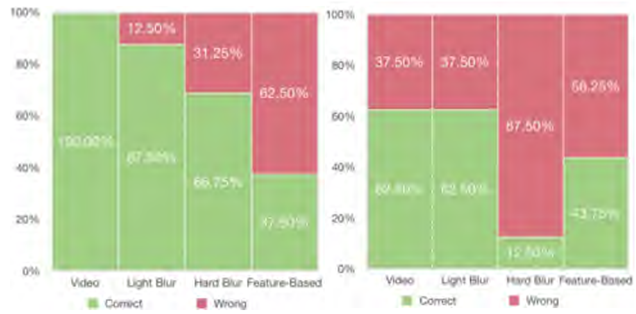

Figure 5: Left: Accuracy of Character Identification. Right: Accuracy of Facial Expression Identification.

In the second step, they watched the same video again and described the actor's facial expression using their own words. Thus, this was an open-ended question, and we did not give them any hints on how to produce the descriptions. We evaluated the accuracy of facial expression identification based on data collected in this step. To analyze the data, two coders reviewed each description and verified whether the description as correct or not (Kappa = .90).

In the third step, we presented a web page showing four video clips, one for each of the de-identification conditions, of the same actor performing the same facial expression. The content of these four clips were the same– an actor performing the same facial expression as the previous steps but processed using different filters (i.e., original video, hard blur, light blur, and feature-based visualization of facial expression). Then participants were asked to rank the four clips by how difficult it was to identify who is the actor from the hardest to the easiest. The ranking result reveals how well a condition may protect identifiability of personal information. The more difficult it is to identify a particular person, the higher the level of protection.

In the last step, participants watched the four clips again and ranked them in terms of their ability to distinguish facial expressions. In any of the steps, if participants were not sure about their answers, they were allowed to answer that they did not know. They were not forced to make a guess to ensure the validity of data collected.

Every participant was randomly assigned to a set of video clips in which the actor and the facial expressions shown across the four steps were controlled. For example, if the video clip in the first two steps was actor A smiling with light blur filtering, the four video clips in the last two steps were of A smiling with light blur, hard blur, feature-based visualization, and the original video.

4.4 Results

4.4.1 Accuracy of Character Identification

The left chart of Figure 5 shows the accuracy rates of character identification (e.g., identifying who is the actor) in different conditions. In the original video condition, all of the participants were able to identify the character correctly (100% accurate). This is not the case for processed videos. The accuracy rates were 87.50% in light blur condition, 68.75% in hard blur condition, and only 37.50% in feature-based visualization. A likelihood ratio chi-square test was performed to determine whether the accuracy of character identification in different conditions were equal. The accuracies were not equal across conditions, $\chi^2(3) = 20.99$, $p < .0001$. It is the most challenging to identify who is the actor with our proposed visualization technique.

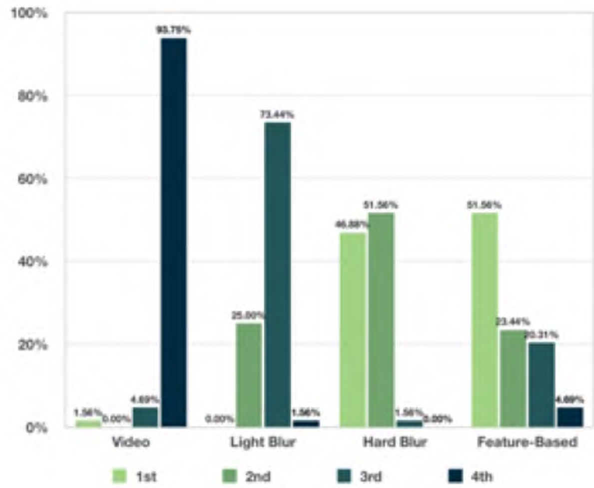

Figure 6: Ranking of Perceived Difficulty to Distinguish Facial Expressions.

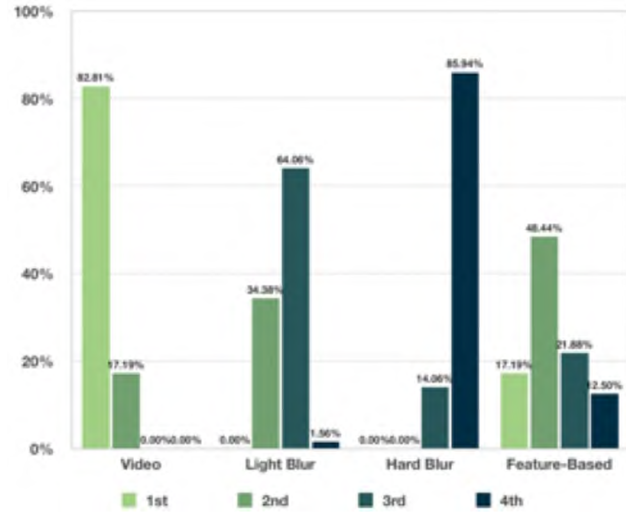

Figure 7: Ranking of Perceived Easiness to Identify Individuals.

4.4.2 Accuracy of Facial Expression Identification

The right chart of Figure 5 shows the accuracy rates of facial expression identification obtained from the coding of participants' open-ended descriptions. The accuracy rates of participant's description of the action were 62.50% in the original video condition and the light blur condition, 12.50% in the hard blur condition, and 43.75% in the feature-based condition. A likelihood ratio chi-square test was performed to determine whether the accuracy of facial expression identification in different conditions were equal. The accuracies were not equal across conditions, $\chi^2(3) = 11.83$, $p < .01$. While feature-based visualization performed understandably poorer than unprocessed video, it has advantages over hard-blur filtering for providing an awareness of facial expression.

4.4.3 Ranking of Perceived Difficulty to Identify Individuals

Figure 6 shows how participants ranked the four conditions according to the perceived difficulty of identifying particular actors. A likelihood ratio chi-square test was performed to determine whether the four conditions were ranked equally. Ranking for the four conditions were not equally distributed, $\chi^2(9) = 349.12$, $p < .0001$. For the hard blur condition, 46.88% of participants ranked it in first place and 51.56% ranked it as the second. While for feature-based visualization, 51.56% ranked it in first place, and 23.44% ranked it second. The weighted average rank for hard blur filtering (expected value of rank = 1.55) is still better than feature-based visualization (rank = 1.79). Nevertheless, feature-based visualization outperformed original video (rank = 3.91) and light blur filtering (rank = 2.77).

4.4.4 Ranking of Perceived Ability to Distinguish Facial Expression

Figure 7 shows how participants ranked the four conditions in terms of the easiness to distinguish facial expressions. A likelihood ratio chi-square test was performed to determine whether the four conditions were ranked equally. Ranking for the four conditions were not equally distributed, $\chi^2(9) = 347.74$, $p < .0001$. Not surprisingly, 82.81% of participants ranked unprocessed video in first place. The weighted rank (i.e., expected value) for video was 1.17. For the feature-based condition, 17.19% of participants ranked it in first place and 48.44% in second. The weighed rank for feature-based visualization was 2.30, which is worse than video, but still better than light blur (2.67) and hard blur (3.86).

4.5 Discussion

In this study, we confirmed that we could preserve the awareness of facial expression without unveiling users' face images and identities. With feature-based visualization, participants failed to identify the performing actor from the set of four candidates. They also ranked feature-based visualization to provide better de-identification protection than light blur and original video. H1 is thus supported.

On the other hand, in terms of awareness of facial expression, although the accuracy rate of facial expression identification in feature-based visualization (43.75%) was not as good as the rates of video and light blur conditions (62.5% in both), it was still much better than the accuracy rate of hard blur condition (12.50%). In the ranking-based analysis, we also obtained a similar result where video provided the greatest awareness of facial expression, followed by feature-based visualization as the second best condition. Image-based filtering conditions (light blur and hard blur) were clearly the least effective for providing such awareness. H2 was also supported by our results. However, it is worth noting that the amount of awareness that the image blurring technique can provide would depend on the degree of blurring (i.e., radius of the Gaussian blur filter).

Based on our results of the video-rating study, overall, to deal with the tension between privacy and awareness of facial information, the technique of feature-based visualization of facial expression shows its superiority over the other methods. As we expected, the parsimonious set of facial features that we chose to visualize can effectively de-identify the identity of individuals. Also, unlike image-based techniques where all visual information is filtered unanimously, feature-based visualization also has the function to highlight important visual information necessary for conveying non-verbal communicative cues. The result confirmed the expected value of the visualization technique for communicating facial expressions.

Figure 8: The interface of KinChat. The chat window is on the left. The red and larger face on the right represents the partner's facial expression and the green face represents oneself.

Figure 9: The experimental setting for second study, which resembles the common use scenario of computer-mediated chat.

5 STUDY 2: REAL-TIME TEXT COMMUNICATION

The results from study 1 indicated that feature-based visualization can de-identify user facial images and still allow other people to perceive and recognize users' facial expressions. Now we extend our focus to examine the effects of using feature-based visualization of facial expression in real-time, interactive communication.

We developed an enhanced text-chat system called KinChat. The system was built with a server-client architecture. The client was built as a Windows application, which has access to a Kinect motion sensor. It can track the positions of facial feature points and update the information through a mediation server. Then the sever transfers these data to other connected client programs. When the other clients receive the positional information of facial features, the system renders the visualization of the facial expression on the screen, as shown in the right hand side of Figure 8. People express their emotion through facial expression rapidly. It is rare that facial expression of emotion can last as long as five or ten seconds except when someone's emotion is very intense. Most facial expressions show up only for a few seconds [12]. Hence, the real-time mechanism of KinChat could avoid people missing an important facial message. We expect that the extra visualization can add social cues to text chat, enriching the communication medium in a beneficial means.

In the second study, we examine the effects of feature-based facial visualization on communication by comparing three conditions, text-only chat, video-added chat, and KinChat (i.e., visualization-added chat). We examine the effects of different conditions on users' communication anxiety, social attraction, and level of understanding of the communication content.

5.1 Social Anxiety in Communication

One potential benefit of using de-identified visualization rather than video to convey facial expressions is to ameliorate one's feeling of communication anxiety by reducing the amount of social cues, and thus the sense of social presence of other people.

Previous research suggested that communication over lean media (i.e., media with fewer social cues) may reduce the feeling of anxiety during communication. Studies showed that individuals are more willing to reveal personal information in text-mediated communication [15]. Similarly, individuals tend to disclose more to their partners in voice-only meetings than video conferencing [1]. Individuals tend to feel more comfortable, more confident, and safer in online communication settings [7].

5.2 Hypotheses

When applying KinChat to computer-mediated text-based communication, there are a number of possible benefits.

First, the availability of facial expressions may serve as an additional non-verbal channel, providing extra resources for individuals to use for conveying messages and grounding understanding [8].

Second, because facial expressions are de-identified by KinChat, compared to unprocessed video the richness of social cues is lessened, and thus people may be less anxious and more comfortable engaging in interpersonal conversations. As a result, the quality of communication and thus comprehension of messages may also improve. More specifically, we posit the following hypotheses:

5.2.1 H3: Level of Understanding

We expect that participants will have better understanding during the conversation when feature-based visualization of facial expression is available. The level of understanding in KinChat condition will be greater than the level in text-only communication.

5.2.2 H4: Communication Anxiety

As described earlier in the design of feature-based visualization, KinChat will not show extra information except the key facial features it traces. Therefore, private and personal information such as one's look and the surrounding context of the space one is located in will not be revealed either. Therefore, compared to video communication, using KinChat will result in lower anxiety.

5.3 Interface Design

Similar to the design of many video-chat tools, KinChat presents both one's own and partners' feature-based visualizations of faces at the same time on the same screen, so that people know how they appear to their partners. One's own face is smaller than the partner's and is presented at the corner of KinChat's interface (see Figure 8). A light indicator is used to indicate the sensing status of the Kinect sensor. The indicator lights green when the sensor detects the user, otherwise it stays red.

5.4 Experimental Study

We conducted a laboratory study to evaluate the impact of KinChat on communication with three conditions: Video (without sounds), IM chat (text only), and KinChat. To allow a fair comparison across conditions, the video condition is muted so the participants can not speak or hear any sounds from their partners. Participants still use text messages to communicate, but they can see each other in the video. There are two main windows shown on the computer screen: one for text messaging (using Skype's chat function), and another for KinChat or Video depending on the experimental condition. Participants can also freely use built-in emoticons in any condition.

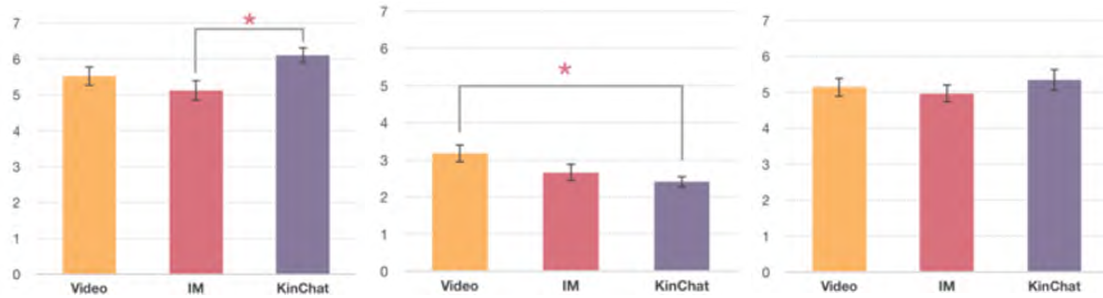

Figure 10: Left: Level of understanding by type of medium used for communication. Middle: Communication Anxiety by type of medium used for communication. Right: Social Attraction by type of medium used for communication.

5.5 Procedure

We recruited a total of 45 participants to attend a simulated, one-on-one conversational interview. We told participants that we will randomly assign them to play either the role of interviewer or interviewer. We also told them that a participant who arrived earlier and was assigned to play the role of interviewer will ask a number of questions, and interviewee will need to answer those questions. However, all of the participants were recruited as interviewees and the interviewer was a trained confederate arranged by the researchers. At the end of the experiment, we debriefed the participants with explanations of the experimental design. We didn't notice adverse reactions from the participants.

The confederate interviewed each of the participants for 15 minutes in a semi-structured fashion by using one of the three chat conditions. Each condition had 15 participants. All of them were students recruited from two major universities in Taiwan. See Figure 9 for the physical setting of the study.

The topic of conversation was controlled to focus on college and campus life. The conversation started by asking participants "What major are you studying?" followed by questions such as "What courses had you taken?" and "Why you chose to study such subjects?" After the session, participants were asked to complete a questionnaire. We also invited four participants to attend a semi-structured usability interview concerning their experience of using the chat tool and their perception of its usefulness. The usability interview consisted of questions concerning how the participant felt about this system, such as "according to your use of the chat tool, what's the pros and cons of the system?" and "in your opinion, what can be improved in this system?" The usability interview lasted about 15 minutes for each participant.

5.6 Measures

To measure participants' experience and social perceptions in communication, participants were asked to take a post-experimental survey for assessing their communication anxiety, social attraction, and level of understanding in the conversation.

5.6.1 Level of Understanding

Perceived level of understanding in communication was assessed with three items used in a previous CMC study [14]. Sample questions include "I had to think harder to understand the ambiguous information in his (her) ideas," and "I believed that he (she) could understand my ideas clearly." Cronbach's alpha reliability for the three items reached .79.

5.6.2 Communication Anxiety

Six items from a common communication apprehension scale were used to assess communication anxiety [5]. Sample items include "I prefer to listen rather than to talk in the conversation," and "My body feels tense when I talk with him or her." Cronbach's alpha reliability was .80.

5.6.3 Social Attraction

Five items from a validated scale for interpersonal attraction in interpersonal communication were used [17]. Sample items include "I think s/he could be a friend of mine," and "We could never establish a personal friendship with each other." Cronbach's alpha reliability for the sampled items was .59.

All questions have been translated to Mandarin Chinese to match the language proficiency of our participants. An independent English-Chinese bilingual speaker checked and confirmed that the translations were semantically consistent with the English versions.

5.7 Results

5.7.1 Level of Understanding

There was a significant main effect of communication media on level of understanding between conditions, $F(2, 42) = 3.48$, $p = .04$, $\eta_p^2 = .14$. The mean of participants' rating of how much they understood their partners was 5.51 (SD = 1.05) in video condition, 5.15 (SD = 1.06) in IM condition and 6.08 (SD = 0.89) in KinChat condition (see Figure 10 left). Post-hoc comparisons using Tukey HSD indicated that participants had significantly higher perceived understanding in KinChat than in regular text-only IM ($p = .03$). H3 was supported.

5.7.2 Communication Anxiety

There was a significant difference on communication anxiety between conditions, $F(2, 42) = 3.28$, $p = .048$, $\eta_p^2 = .135$. Participants' level of anxiety was 3.17 (SD = 0.94) in video, 2.4 (SD = 0.89) in KinChat, and 2.65 (SD = 0.62) in text-only IM (see Figure 10 middle). A smaller score indicates that participants perceived less tension when interacting with the confederate. Post-hoc comparisons using Tukey HSD indicated that participants had lower anxiety when they communicated over KinChat than video ($p = .04$). H4 was supported.

5.7.3 Social Attraction

In terms of how socially attractive participants felt about their partners, the levels of perceived social attraction in video (M = 5.13, SD = 1.04), KinChat (M = 5.3, SD = 1.16), and IM (M =

4.96, SD = 0.95) were not significantly different, $F(2, 42) = 0.47$, $p = .63$, $\eta_p^2 = .02$. (see also Figure 10 right).

5.7.4 Usability Interview

We transcribed the post-experimental usability interviews and identified salient topics touched on during the interview by scanning through the transcripts repeatedly. The interviews appear to support our findings from the experiment. Participants considered it difficult to identify interlocutor's emotions in text-only chat. They also reflected their everyday experience and reported that they tend to avoid using video when it is not convenient, for example, after a bath or with strangers as reported in the interviews.

6 Discussion

According to the results from Study 1, the feature-based visualization technique can successfully de-identify users' face images. Although the technique sacrificed some awareness of facial expression, it is still overall a better choice than image-based filtering techniques.

By applying the technique to support interactive text communication, we see significant improvements in both the social and cognitive aspects of communication. Results show that adding real-time visualization of facial expression to text chat may increase understanding and decrease communication anxiety. People better understood text messages when they saw others' facial features and head movement with KinChat.

As a nonverbal cue, facial expressions may supplement verbal communication. The results we obtained provide initial support that using simple feature-based visualization to mediate facial expression can also effectively mediate the nonverbal cues necessary for supporting understanding in communication. Meanwhile, participants felt less anxious because of the de-identification of faces, and the reduction in anxiety may also be beneficial to overall quality of communication as people will be able to better engage in the interaction.

6.1 Implications to CMC Design and Research

The paper has contributions and implications to the technical and social aspects of CMC design. First, in the technical aspects, we present a simple, easy-to-implement yet effective interaction technique for adding desirable nonverbal cues to text-based communication without disclosing interlocutors' identities. Unlike video or image-based filtering, the technique delivers facial cues selectively. This attribute makes the technique an ideal solution for realizing *adaptive awareness* in CMC. For example, if a user wants to increase (or reduce) the awareness of non-verbal cues, they may choose to deliver facial features selectively without dramatically altering the affordances of the communication channel (e.g., switching between text chat and video conferencing).

As discussed at the beginning, model-based techniques such as the use of avatars serve as another option for visualizing facial expressions. However, we noted that the complexity and the building cost associated with avatars may outweigh the benefits they can reap. One major concern is that choosing which avatar to use is not straightforward and may introduce unexpected impact on communication due to visual attributes of the models, such as race, gender, and even cuteness. Also, avatars normally do not allow selective visualization of facial features (try picturing an avatar without a nose!), making user-controlled adaptability less feasible.

Second, regarding the social aspect of CMC design, KinChat opens up a new way to augment text chat with extra nonverbal cues, or to appropriately adjust the richness of social cues in video communication to an extent less intimidating to users. KinChat provides an opportunity to re-consider and even integrate the two competing theoretical models in CMC, cue-filtering or media richness [9] versus hyperpersonal or social information processing [26]. The uniqueness of KinChat is that it enables *selectively highlighting* (rather than filtering) nonverbal cues. This property goes beyond the scenarios that either theoretical model has considered before. For example, it is now possible for users to engage in strategic hyperpersonal communication by choosing what nonverbal cues to convey without worrying about the disclosure of unintended information.

6.2 Limitations and Future Work

In this paper, we present KinChat as a proof-of-concept prototype. We are aware of numerous opportunities to improve the technical, behavioral, and theoretical aspects of the work. The emergence of these opportunities strengthens, but not weakens, the value of the current work.

The interface can be further improved in its usability. In the current design, text messages and face animation are displayed in two different windows, which may make it difficult for users to notice facial features when they pay attention to the chat. The studies were limited in terms of the number of participants involved, reliability of measurement instruments (e.g., alpha reliability is only around .6-.8 for the instruments used in Study 2), and the experimental setting per se. Replicating the studies with varieties of stimuli, tasks, samples of participants, and conditions will help consolidate the findings.

In terms of motion sensing, it is valuable to track not only facial features but also the movements of hands and other body parts to provide more visualizations as non-verbal cues. With more non-verbal cues, we may also add the function of feature management and selection for designers and users to choose what features to convey and highlight.

The visualization of facial expression can also have applications other than augmenting online text chat. One possibility is that the visualization can help diminish discrimination in an interview by blocking cues related to one's look, age, gender or race. Another application is on supporting intercultural collaboration. By selectively highlighting individuals' visible features, it can prevent possible prejudice and help people communicate better.

7 Conclusion

KinChat is a technique for achieving de-identified visualization of facial expression in computer-mediated communication (CMC) by integrating the technologies of motion sensing and 2D graphical visualization. The approach allows designers and users to highlight and deliver key facial features as nonverbal cues to supplement text chat. The technique is conceptually more parsimonious and technically more straightforward than other de-identification techniques, such as image-based blurring and modeling-based avatars.

Through two interconnected studies, we show that the proposed technique appeared to preserve both awareness of facial cues and de-identification of the interlocutors as we expected. In real-time communication, using KinChat to augment text chat resulted in better understanding of messages and reduced communication anxiety, confirming its utility.

We note that the artifact of KinChat represents a new breed of communication channel with properties and affordances beyond what traditional CMC theoretical models can cover. KinChat enables strategic uses of nonverbal facial cues in task-oriented and relational communication, and it brings together both the benefits of rich- and lean modes of communication. The design and social spaces of KinChat deserve further investigation.

ACKNOWLEDGEMENTS

The work is supported in part by Ministry of Science and Technology, Taiwan (102-2221-E-007-073-MY3; 104-3115-E-007-001; 105-2628-E-007-004-MY2), Microsoft Research Asia and Google Faculty Research Award. We thank anonymous reviewers and the associate chair for the constructive feedback, which help us improve the work vastly.

REFERENCES

[1] J. N. Bailenson, N Yee., D. Merget, & R. Schroeder (2006). The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. Presence: *Teleoperators and Virtual Environments, 15*, 359–372.

[2] J. N. Bassili (1979). Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face. *Journal of personality and social psychology 37*, 2049-2058.

[3] A. L. Baylor (2009). Promoting motivation with virtual agents and avatars: role of visual presence and appearance. *Philosophical Transactions of The Royal Society B 364*, 1525, 3559-3565.

[4] D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, & S. K. Nayar (2008). Face swapping: automatically replacing faces in photographs. *Proceedings of the SIGGRAPH 2008 Conference on Computer Graphics and Interactive Techniques*.

[5] S. Booth-Butterfield & M. Gould (1986). The communication anxiety inventory: validation of state-and-context communication apprehension. *Communication Quarterly, 34*, 194-205.

[6] M. Boyle, C. Edwards & S. Greenberg. (2000). The effects of filtered video on awareness and privacy. *Proceedings of the CSCW 2000 Conference on Computer Supported Cooperative Work*.

[7] S. E. Caplan (2005). A social skill account of problematic Internet use. *Journal of Communication, 55*, 721-736.

[8] H. H. Clark, & S. E. Brennan (1991). Grounding in communication. In L. B. Resnick & J. M. Levine (Eds.), *Perspectives on Socially Shared Cognition* (pp. 127-149), Washington DC: APA.

[9] R. L. Daft, & R. H. Lengel (1986). Organizational information requirements, media richness and structural design. *Management Science, 32*, 554-571.

[10] D. Derks, A. E. R. Bos, & J. v. Grumbkow. (2007). Emoticons and social interaction on the Internet: the importance of social context. *Computers in Human Behavior*, 23, 842-849.

[11] C. Shahani-Denning (2003). Physical attractiveness bias in hiring: what is beautiful is good. *Journal of Personality*, 14–17.

[12] P. Ekman, & W. V. Friesen (1975). *Unmasking the Face: A Guide to Recognizing Emotions from Facial Cues*. Englewood Cliffs, N.J.: Prentice-Hall.

[13] T. Erickson, & W. A. Kellogg (2000). Social translucence: an approach to designing systems that support social processes. *ACM Transactions on Computer-Human Interaction*, 7, 59–83.

[14] G. Gao, H-C. Wang, D. Cosley, & S. R. Fussell (2013). Same translation but different experience: the effects of highlighting on machine-translated conversations. *Proceedings of the CHI 2013 Conference on Human Factors in Computer Systems*.

[15] A. N. Joinson (2001). Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology*, 31, 177-192.

[16] Y. B. Kafai, D. A. Fields, & M. Cook (2007). Your second selves: avatar designs and identity play in a teen virtual world. *Proceedings of the Digital Games Research Association (DiGRA) 2007 Conference*.

[17] J. C. McCroskey, & T. A. McCain (1974). The measurement of interpersonal attraction. *Speech Monographs, 41*, 261-266.

[18] P. L. McLeod (2011). Effects of anonymity and social comparison of rewards on computer-mediated group brainstorming. *Small Group Research*, 42, 475–503.

[19] C. Neustaedter, S. Greenberg, & M. Boyle (2006). Blur filtration fails to preserve privacy for home-based video conferencing. *ACM Transactions on Computer-Human Interaction*, 13, 1-36.

[20] L. O'Brien, & J. Murnane (2009). An investigation into how avatar appearance can affect interactions in a virtual world. *International Journal of Social and Humanistic Computing 1*, 192–202.

[21] J. Peña, J. T. Hancock, & N. A. Merola (2009). The priming effects of avatars in virtual settings. *Communication Research*, 36, 838-856.

[22] R. Ratan, & B. S. Hasler (2014). Playing well with virtual classmates: relating avatar design to group satisfaction. *Proceedings of the CSCW 2014 Conference on Computer Supported Cooperative Work*.

[23] K. Rennert, & E. Karapanos (2013). Faceit: supporting reflection upon social anxiety events with lifelogging. *Proceedings of the CHI 2013 Extended Abstracts on Human Factors in Computer Systems*.

[24] L. C. Tidwell, & J. B. Walther (2002). Computer-mediated communication effects on disclosure, impressions, and interpersonal evaluations: getting to know one another a bit at a time. *Human Communication Research*, 28, 317-348.

[25] J. B. Walther, & J. K. Burgoon (1992). Relational communication in computer-mediated interaction. *Human Communication Research, 19*, 50-88.

[26] J. B. Walther (1996). Computer-mediated communication impersonal, interpersonal, and hyperpersonal interaction. *Communication Research, 23*, 3-43

[27] H-C. Wang, S. R. Fussell, & L. D. Setlock (2009). Cultural difference and adaptation of communication styles in computer-mediated group brainstorming. *Proceedings of the CHI 2009 Conference on Human Factors in Computer Systems*.

[28] N. Yee, & J. Bailenson (2007). The Proteus effect: The effect of transformed self representation on behavior. *Human Communication Research 33*, 271-290.

[29] Q. A. Zhao, & J. T. Stasko (1998). The awareness-privacy tradeoff in video supported informal awareness: A study of image-filtering based techniques. Technical Report, Georgia Institute of Technology. GIT-GVU-98-16.