

SUBJECTIVE EVALUATIONS OF AN EXPERIMENTAL GESTUREPHONE

Mohd Nazri Ramliy¹, Nor Azhar Mohd Arif² and Ryoichi Komiya³

^{1,3}Faculty of Information Technology, ²Faculty of Engineering
Multimedia University, Malaysia

ABSTRACT

This paper presents the findings of our subjective evaluations on the integration of gestures in telecommunication. The experimental setup for tracking and imitating the human arm gesture are described. Our research investigates the possibility of transferring this often overlooked communication medium in our daily communication, for its application in telecommunication using robotics. Based on the subjective evaluation, a maximum allowable delay for an imperceptible gesture reconstruction in the lateral setup is suggested.

1. INTRODUCTION

This research attempts to incorporate human gestures into telecommunication, focusing on developing a sensor to track human arm motions through image processing. The intention is for enhancing telecommunication as detailed in [1]. The motivation behind this research is explained below.

In face-to-face communication, we utter words and listen with our ears – utilizing, at the minimum, our auditory sense to understand the information. We also use some hand and body gestures to effectively express our thoughts, together with the verbal expression.

Casual talking with someone is best done face to face. When distance is the barrier, we usually resort to the telephone. What would happen if we had a device, which allows us to see and feel the presence of the other person in front of us, and talk to them naturally as if that person were in front of us? A device that would make two of us feel as if we were actually talking to each other, face to face. How would we feel using that device? How would we react to it? How would it affect the way we communicate? How would it change the way we work? What are its impacts on us – physically and emotionally? These are questions that we would like to know the answer to if there is such a device.

Fig. 1 illustrates the concept of such device. The device would have the components to operate: it would have the appropriate sensors (microphone, camera, etc)

and actuators (speaker, robotics) for sensing and actuating the information in order to ‘materialize’ the other party to whom we are talking to in front of us. We need to decide on what information to be sensed for transmission, and find out the minimum set of information needed to make it feel like that person is in front of us.

2. PHYSICAL COMMUNICATION COMPONENT

There were researches done on the inclusion of physical components that exist in human communication into telecommunication.

Brave, Ishii and Dahley [3] attempted the use of tangible use interfaces (TUIs) for remote collaborations. They introduced the PsyBench, which is a board with sensors for detecting the objects placed on it. Magnetic objects that are on the board can be moved using an electromagnet placed on 2-axis below the board. Two units of the system can be used in bilateral configuration to imitate any movement of one object on one side to the other. There was also work on inTouch for haptic interpersonal communication where they use two sets of three rollers that synchronize their roll angles. When one user rolls his or her palm on the roller on one side, the rollers on the other side roll too and the other user can feel the motion under his or her palm.

Oakley and O’Modhrain [4] explores the design space of the haptic communication, focusing on the idea of an asynchronous haptic instant message, similar like the way people use instant messaging service nowadays.

Ouchi and Hashimoto [5] developed a handshake telephone system where through the use of a pair of robotic hand, the users can remotely handshake with each other while talking over the telephone. Sensors built into the robotic hand measure the pressure of the user’s grip and also the user’s shaking torque. The signals are transmitted over the network. Based on the pressure and motion detected at the one side of the connection, the robotic hand at the other side reproduces the same grip and movement so that the two users feel each other’s motions and handgrip. They found that at maximum a delay of 100ms should be maintained for mutual feeling of handshake.

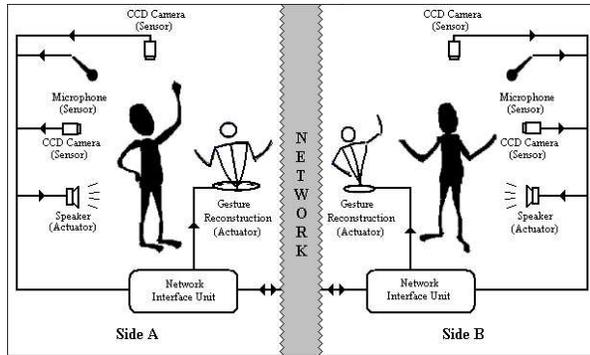


Fig. 1. The conceptual illustration of the proposed telecommunication device [2].

All these researches show that there is an interest in the implementation of physical body motion in telecommunication. The next section describes the system developed in this research for imitating the human arm motion for its application in telecommunication.

3. EXPERIMENTAL SETUP

To investigate the issues involved regarding the integration of gestures in telecommunication, we developed a system for imitating the human arm gesture and reconstructing it using robotics. The robotic arm has five degrees of freedom that approximate the freedom of the human arm motions from the shoulder to the hand. Fig. 2 shows the robotic arm.

We use image processing for tracking the user arm motion. This is done by continuously tracking the location of the shoulder, elbow and hand from two perpendicular views (front and side) and calculating the angle joints. Two CCD cameras are used for this purpose. Due to the complexity of obtaining the hand rotation from the two images, θ_5 is not calculated. The angles calculated (θ_1 , θ_2 , θ_3 and θ_4) are sent to the robotic arm for the reconstruction of the gesture through an I/O card.

3.1 Hardware Configuration of Entire System

Fig. 3 shows the detailed experimental system. The two CCD cameras are connected to the frame grabber, feeding the two video input for the image processing module to extract the four angles of the user's arm, which is then sent to the robotic arm in analogue voltage for the gesture reconstruction. The joint angles that form the arm pose is calculated from the two views obtained and sent to the IO card for reconstruction. Fig. 4 depicts the relation between the CCD cameras and the image processing modules in order to obtain the angles. The five angles shown in this

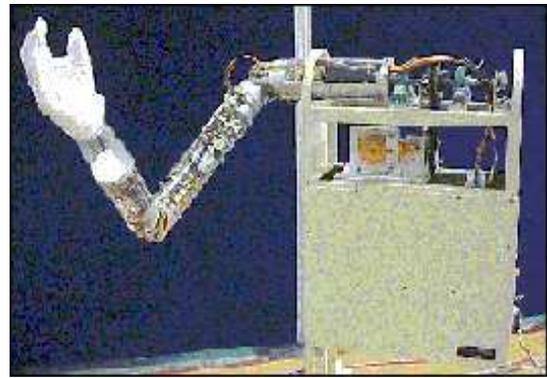


Fig. 2. The actuator — a robotic arm resembling the human arm, with 5 degrees of freedom.

figure represent the five degrees of freedom that the robotic arm has.

3.2. Sensor

3.2.1. Sensor Hardware

Two monochrome CCD cameras, a frame grabber and a PC are used as the sensor hardware. The CCD cameras are positioned perpendicularly to each other. One camera takes the image sequence of the front view of the right arm, and the other takes the image sequence of the right side of the right arm. The two video streams are fed to the frame grabber. Image processing algorithm implemented in the PC is used to track the arm motions and calculate the angles of the joints that produce the gestures. These angles are then sent to the actuator for reconstruction of the arm gestures.

3.2.2. Motion Tracking Algorithm

The tracking of the arm gestures is done by tracking the positions of the shoulder, elbow and hand in the two video streams using the grayscale correlation method. This method is chosen because of its direct and simple implementation.

4. SUBJECTIVE EVALUATION

Using the system, a subjective evaluation involving 10 assessors is conducted in order to investigate the following issues involving the imitation of human arm gestures for telecommunication:

a) *Delay* - How long is the delay that can be present in the imitation system before the user notices that there is a delay. This delay can be divided in two – one is the processing delay and the other is the network delay. The

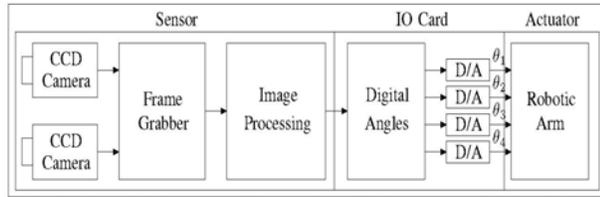


Fig. 3. Block diagram of the arm gesture transmission.

processing delay refers to the delay in the capturing and calculating the joint angles, while network delay is the delay that is inherent to the network. In the subjective evaluation, the maximum unnoticeable network delay is measured for each assessor.

- b) *Appropriateness* - Whether it is appropriate to have gesture transmission using robotics for telecommunication.
- c) *Effectiveness* - How effective was the system in imitating the user's arm motion.
- d) *Detail* - Whether it is necessary to imitate the finger motions.
- e) *Acceptance* - Whether the user can accept the idea of using robotics for telecommunication.
- f) *Size* - Whether the size of the robotics have a significant impact or not.

Some duration of network delay is added to the system in order to evaluate the maximum unnoticeable delay for transferring gestures. To produce the effect of network delay, the system was set to imitate the user's arm motion only after a certain amount of milliseconds of delay after reading the current frame and calculating the current angle joints, that is, after the processing delay. First, the delay was initially set to one second and the user was asked to perform some gestures and at the same time look at the robotic arm, to see how it feels like when there is a delay in the imitation (a network delay of one second in the imitation is very noticeable). Then, the delay was gradually reduced (500, 400, 350, 300 ... milliseconds) until the user noted that the delay is no longer noticeable. This delay is recorded for each assessor.

5. RESULTS

A total of 10 assessors took part in the subjective evaluation. For evaluating the delay, each assessor are asked to indicate when is it that they no longer notice any delay in the reconstruction of their gestures as the delay is gradually reduced from the initial one second delay. The assessors are then given questions asking them to reflect their opinion on using the system. The evaluation criteria

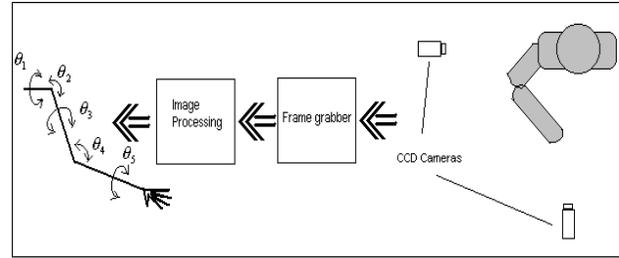


Fig. 4. Extraction of the angles from the two video sequences.

were divided into delay, effectiveness, appropriateness, detail, acceptance, and size. The following sections discuss the results regarding each of these criteria:

5.1. Delay

Table I shows the results for each of the assessors. The value shown is the maximum delay in milliseconds where they no longer notice any delay in the imitation of their gestures. Note that this delay includes the processing delay for the image processing part. Most of the assessors noted that the delay started to be imperceptible when it was down to 420ms, at which point they have to move their arm slowly in order concentrate on the reproduced gesture motion to find out whether they can still detect the delay.

From Table I, the average network delay that the assessors no longer notice in their imitated gesture is 392.5ms. The maximum unnoticeable network delay recorded was 410ms for assessor number 8, and the minimum unnoticeable network delay recorded is 380ms for assessor number 1. From this result, we can consider a maximum acceptable network delay of 392.5ms for the context of gesture reconstruction as good enough for having an almost realtime gesture reconstruction with an imperceptible delay. Note that the experiment was done in a unilateral way. For bilateral configuration, some more experimentation is needed to confirm whether this finding would hold true or not.

For the case of tele-handshaking mentioned in Section 2, Ouchi and Hashimoto [5] proposed a delay of 100ms to maintain a mutual feeling of handshake. The difference between the 392.5ms suggested delay found in this research and that 100ms can be attributed to the fact that our touching sense is more sensitive than our seeing sense -we can easily sense that something is moving by touching it, whereby the eye needs more displacement before we notice that it is moving. For example it is hard for us to tell whether a car engine is running just by looking at it, but it is easy to determine that the engine is running if we touch the hood using the palm of our hands, where we can feel the hood vibrating because of the

TABLE I
Maximum Unnoticeable Network Delay

Assessor	Maximum unnoticeable network delay (ms)
1	380
2	385
3	390
4	385
5	400
6	390
7	400
8	410
9	400
10	385
Maximum	410
Minimum	380
Average	392.5

running engine. So in the case of handshake it is the touch sense that is mostly involved for experiencing the gesture, hence the difference in the delay measured.

5.2. Effectiveness

Eight assessors noted that the effectiveness of the system in imitating their arm gesture was satisfactory, while the other two rated it as excellent.

5.3. Appropriateness

Eight of the assessors agree that it is appropriate for gestures to be reconstructed using robotics for telecommunication. One assessor mentioned that he was not sure about the appropriateness and another one has the opinion that perhaps virtual reality (videophone) is a better technology for this purpose.

5.4. Detail

Seven of the assessors felt that the finger motion imitation is not necessary for bigger hand/body gestures. Two felt that it is necessary, and one was not sure.

5.5. Acceptance

All ten of the assessors accept the idea of using robotics for telecommunication, with one of them noted that he would accept it to some extent.

5.6. Size

On the size of the robotics, eight of the assessors agreed that the size of the robotics can be smaller than the size of

the human arm if it can clearly show the gestures. One felt that the size has to be the same. While another one felt that the size must be of the average human arm size. Asked on the case where the whole human body is imitated, eight of the assessors felt that the size does not have to be the same as the human body, while the other two felt that the size has to be the same.

6. CONCLUSIONS

For satisfactory gestures reconstruction, the robotics system must be able to match the human arm motion, both in term of speed and agility so that it can successfully reconstruct the original gesture without its system delay. There is a potential for the inclusion of gesture in telecommunication. The subjective evaluation conducted has shown that the idea of transferring the normal gestures are made during conversation for telecommunication use is quite acceptable and appropriate. The authors believe that robotics can be an attractive mode of telecommunication where gestures, along with video, audio and data are all transferred together. The physical presence of the gestures would surely give the interacting user a new kind of feeling, sensation and perception during telecommunication. This will greatly enhance the telecommunication experience. The size of the actuator for reconstructing the gestures does not have to be the same as the corresponding user's body size, as long as it can clearly show the gesture. For example, using a miniature robotics that resembles the human body can be a possible implementation which results in a smaller system.

7. REFERENCES

- [1] R. Komiya, *Virtual Reality Telecommunication System – A Person-to-Person Multimedia Communication System*, IEEE Globecom 2000 Conference Record, Vol.3 of 3, pp1354-1359, November 2000.
- [2] Algirdas Pakstas, Ryoichi Komiya, *Virtual Reality Technologies for Future Telecommunications Systems*, John Wiley & Sons, LTD., ISBN 0-470-84886-3, 2002, pp 127.
- [3] S. Brave, H. Ishii and A. Dahley, *Tangible Interfaces for Remote Collaboration and Communication*, Proceedings of CSCW'98.
- [4] I. Oakley, and S. O'Modhrain, *Contact IM: Exploring Asynchronous Touch Over Distance*, Proceedings of CSCW 2002, New Orleans.
- [5] K. Ouchi and S. Hashimoto, *Handshake Telephone System to Communicate with Voice and Force*, IEEE International Workshop on Robot and Human Communication 1997.