Internal Use Only

002

TR-IT-0018

EMMI - ATR Environment for Multi-Modal Interactions

Kyung-ho Loken-Kim Fumihiro Yato
Kazuhiko Kurihara Laurel Fais Ryo Furukawa

September 30, 1993

ABSTRACT

This report describes ATR's Environment for Multi-Modal Interactions (EMMI) developed to collect speech and language data used in multi-media, multi-lingual, multi-party interpreting telecommunication settings. The primary task that EMMI supports is international conference registration and its subtasks: directions, reservations, and negotiations. Further details on the system's hardware and software configurations are presented here.

1 INTRODUCTION

Recent developments of application specific ICs (ASICS) [1] open up the possibilities of realizing personal communicators that integrate voice, data, handwriting, fax, electronic-mail, still images, and full-motion video within a decade. No doubt such multi-media systems will have a profound impact on the world of communications, and they will change the form of human communications forever. It is not well understood, however, how these technologies should be optimally amalgamated to induce maximum efficiency in human-machine-human communications, especially in multi-media, multi-lingual, multi-party interpreting telecommunication settings. The optimal multi-media configuration for an application, such as multi-media interpreting telecommunications, can not be obtained in an ad hoc fashion. It rather requires a series of empirical studies conducted in settings simulating those in which the intended uses are most likely to take place.

ATR's Environment for Multi-Modal Interactions (EMMI) is a simulation tool that supports a variety of realistic environments for multi-media, multi-lingual, multi-party interpreting telecommunications. EMMI is neutral in the sense that it does not contain any sort of intelligence; that is, it is simply a man-machine-man interface.

EMMI has been created specifically for collecting data about the speech and language people might use in multi-media interpreting telecommunications. We have selected an international conference registration task, and it is divided into three sub-tasks: directions, reservations, and negotiations (see section 3 for further details).

Collecting mono-lingual multi-media speech and language data requires a minimum of two participants: an agent and a client. For multi-lingual data, a third party, the translator should be added. In EMMI, participants can communicate with each other using a variety of input/output modalities: speech, text, and video image. For example, the agent, who is acting as the conference secretariat, can give directions verbally to the client over the telephone while showing and writing on the map displayed on both the agent's and the client's screens.

EMMI is equipped with an array of multi-media data collection equipment. Currently, three video cameras, two used to transmit the full-motion video images of the participants, and the third used to record the client's interaction with the system, are available. Three telephones and a Digital Audio Tape deck have been installed to collect speech-only dialogues and ensure the high quality recording of all verbal transactions.

In this report, the authors describe the hardware and the software configurations

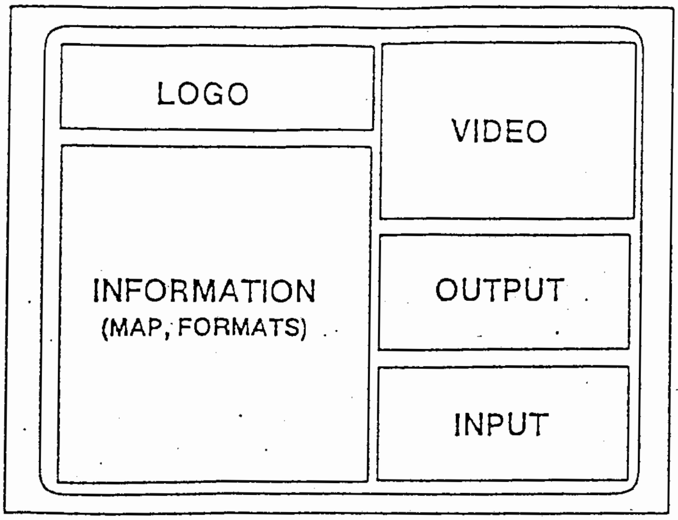


Figure 1: User Interface

of EMMI. Readers are encouraged to use the system, collect desired data, and share the data. Any comments regarding EMMI would be greatly appreciated.

2 USER INTERFACE

Participants interact through the multi-media window illustrated in Figure 1. This window is divided into four sub-windows: information, video, input/output, and logo. The INFORMATION window is used for displaying maps and reservation forms, as well as for marking and writing. For example, participants, while engaged in a dialogue, can mark and write necessary information on the map by pressing the left button of the mouse and dragging the cursor. In order to distinguish the client's and the agent's marks and writings, different colors are used. When a reservation form appears on the screen, both the agent and the client can fill out the form simply by typing.

The VIDEO window is used for displaying the full-motion video images of the client and the agent. Participants of course can turn off the video camera if they choose. The limitations of the video camera positions nevertheless do not allow direct eye contact between the participants (see section 7).

The INPUT and the OUTPUT windows are provided to aid verbal communication by allowing participants to exchange information in text. Japanese proper nouns, for example, are more easily described in text than in verbal descriptions.

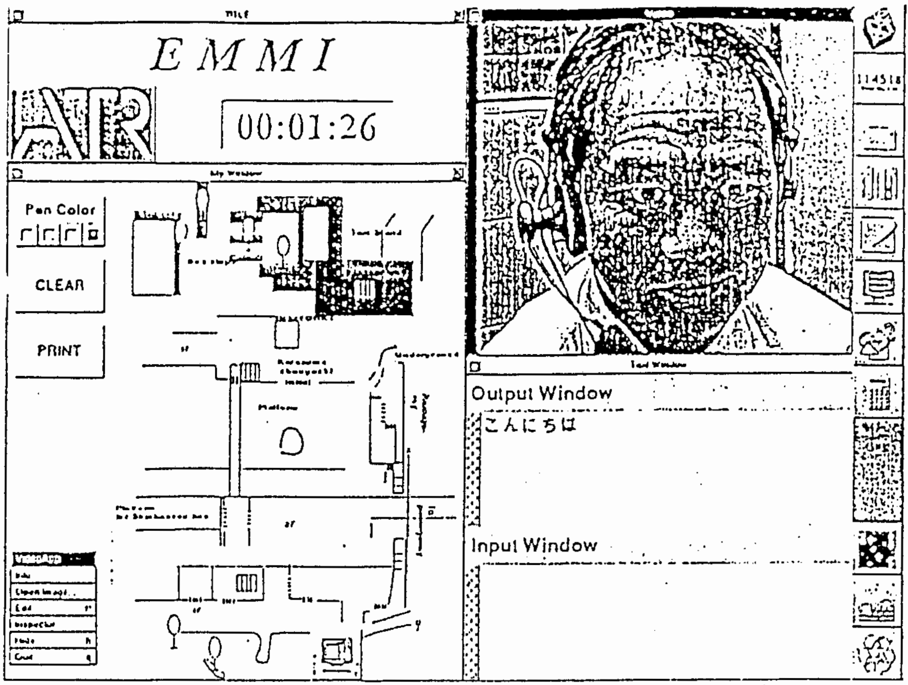


Figure 2: Client's Screen

3 TASK

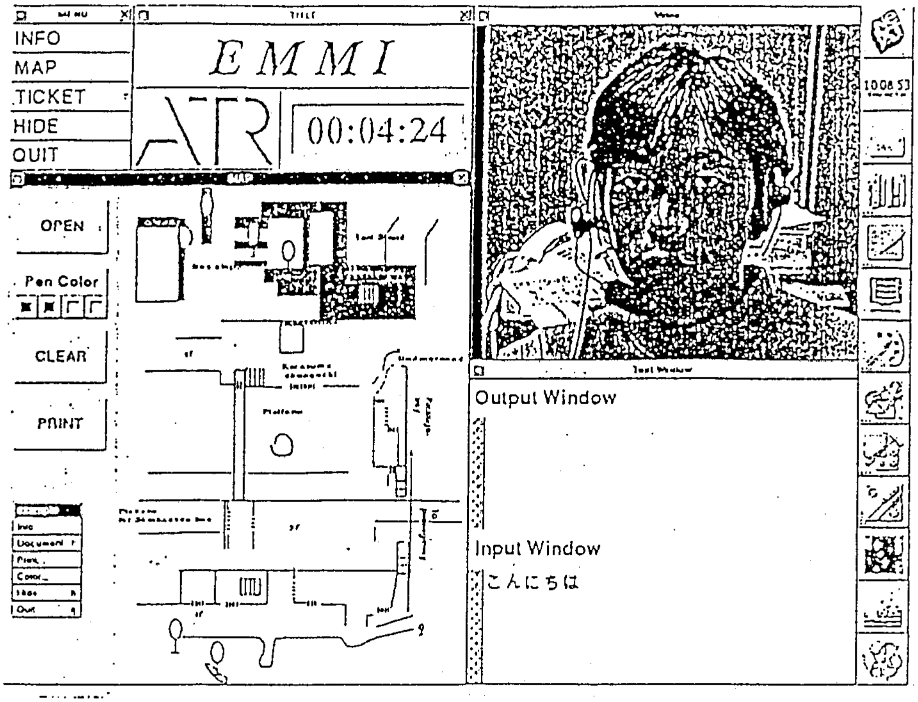
The primary task that EMMI supports is international conference registration and its sub-tasks. Specifically, the following tasks are supported.

1) DIRECTIONS TASK

For this task, a client asks the conference secretriat for directions, for example, from Kyoto Station to the Kyoto International Conference Center. The agent gives the directions by displaying one of three maps of the areas surrounding Kyoto Station, the International Conference Center, and Kyoto Park Hotel. As previously explained, the maps are displayed on both of the agent's and client's screens, and the agent and the client can engage in a dialogue while marking and writing relevant information on the maps using a mouse. Figure 2 and Figure 3 illustrate mid-session screens of the client and the agent respectively.

2) RESERVATIONS TASK

In this task, a client needs to make a reservation, such as a hotel reservation. The agent displays a reservation form on the screen and makes the reservation by filling out the form using the keyboard and the mouse. The client may also fill out the form. Currently four different reservation forms are available: train, airline (Figure 4), hotel, and package tour.



こんにちは

Figure 3: Agent's Screen

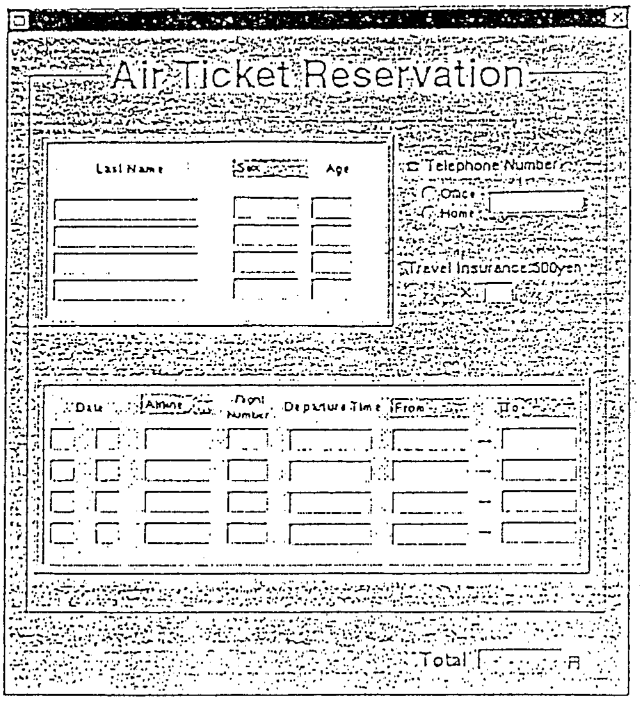
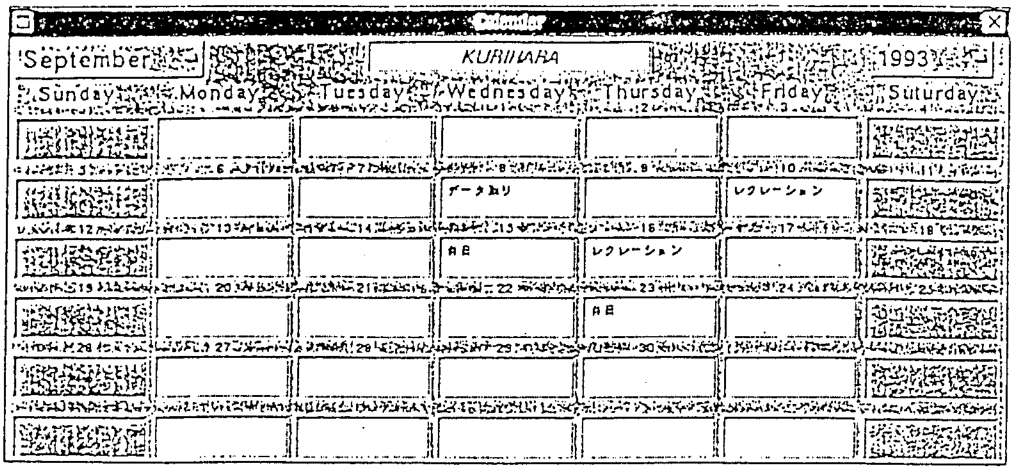


Figure 4: Airline Reservation Form



September

KURIHARA

1993

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
				2	3	4
5	6	7	8 データ取り	9	10 レクレーション	11
12	13	14	15 AB	16 レクレーション	17	18
19	20	21	22	23 AB	24	25
26	27	28	29	30		

Figure 5: Calendar

3) NEGOTIATIONS TASK
In this task, the agent negotiates with the client on the client's possible paper presentation date and time using the calendar displayed on the screen (Figure 5). It is assumed that the conference schedule has not been determined yet.

4 LAYOUT

Figure 6 illustrates the current layout of the laboratory where EMMI is located. Each workstation is physically separated from the others, and sound-absorbing partitions have been placed in between to ensure minimum transmission of the participants' voices over the partitions.

5 HARDWARE CONFIGURATIONS AND SCHEMATICS

Table 1 is a list of equipment used for EMMI. There are two NeXT computers: one for the agent, and the other for the client. A SUN Sparc Station has been allocated for the translator. All three computers are equipped with a keyboard and a mouse. A touch screen for each computer will be added later.

One Digital Audio Tape deck, microphone amplifiers, and headphones have been installed to obtain high quality speech transmissions and recordings.

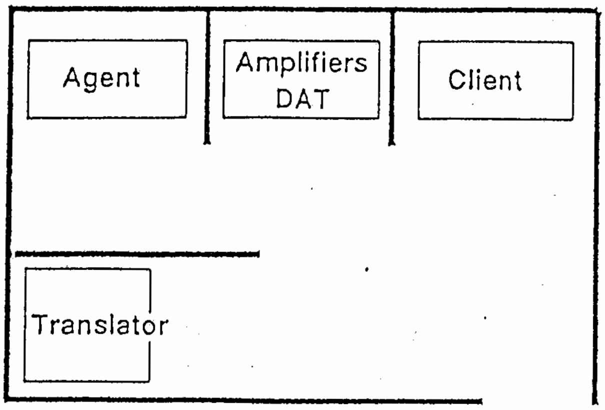


Figure 6: Laboratory Layout

Two video cameras, connected to the Video Monitor interface of the NeXt Cube, are used to transmit video motion images of the agent and the client. A third video camera has been installed to capture the client's interactions with EMMI.

Three telephones have been installed to collect telephone-to-telephone speech-only dialogues. Presently, two telephone lines are used; one for the client, and the other for the agent. The line to the agent is interconnected with another telephone for the translator.

A scanner is attached to the SUN workstation to scan the maps and other graphic images.

Figure 7 is the hardware schematic for the current equipment. The section within the dotted line has not been completely installed.

6 SOFTWARE CONFIGURATIONS

EMMI has been developed on a NeXT computer using Interface Builder and Objective C [2]. The software construction of the client's side is almost the mirror image of that on the agent's side. As illustrated in Figure 8, the two top processes, i.e., start.new.csh on the agent's and the client's sides activate EMMI by sending messages to corresponding TEL processes. Upon receiving the messages, the agent TEL process displays the telephone window illustrated in Figure 9 while the client TEL process displays a slightly different one (Figure 10). This is the initial ideling state of EMMI waiting for a client.

Table 1: Equipment

Computers	Two NeXT Cube workstations, SUN Sparc Station-3 (Xerox 6401)
Audio Equipment	Two SONY Digital Audio Tape Decks DTC-77ES
Amplifiers	Two ONKYO Integrated Stereo Amplifier A-812EX, Two Technics Mic Mixing Amplifiers SH-3026
Headphones	Four SONY MDR CD850 Headphones Three Sennheiser HMD 410 Headphones with microphones
Video Equipment	SONY Video Hi-8 Handycam PRO 3DD, Two SONY Video Hi-8 Handycams
Video Recorder	Sony Video Television Recorder - SONY WALKMAN - GV-500
Telephone	Three telephones
Scanner	Sony Color Video Scanner UY-T55V

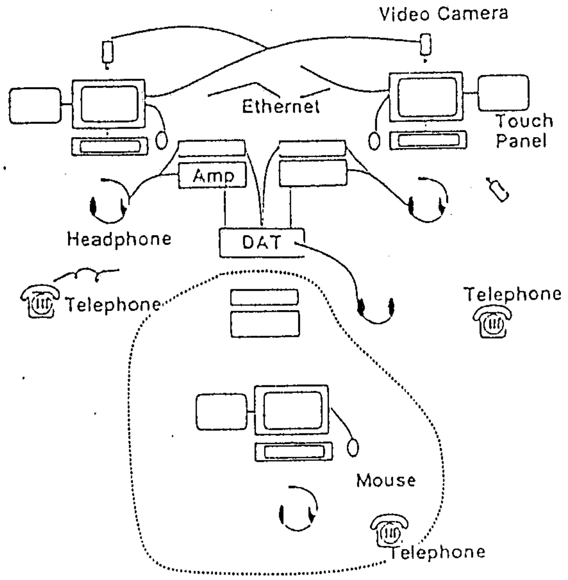


Figure 7: Hardware Schematic

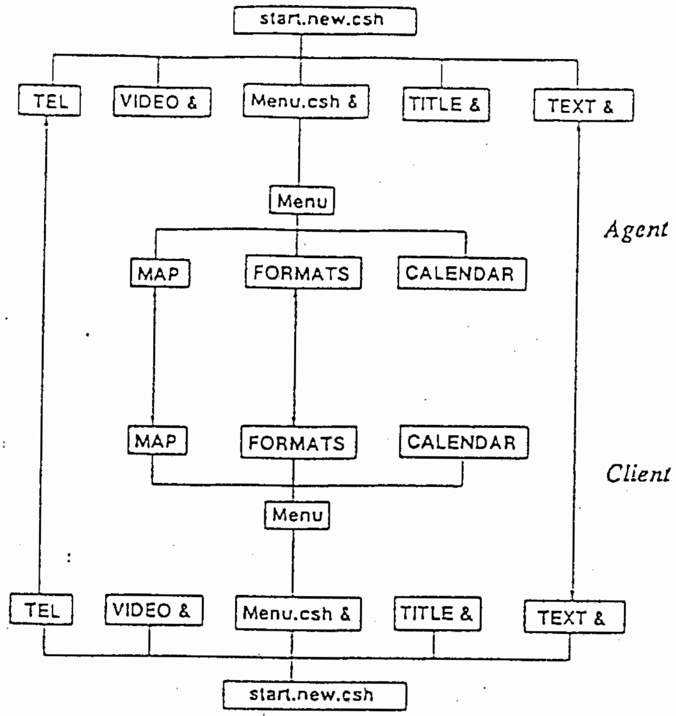
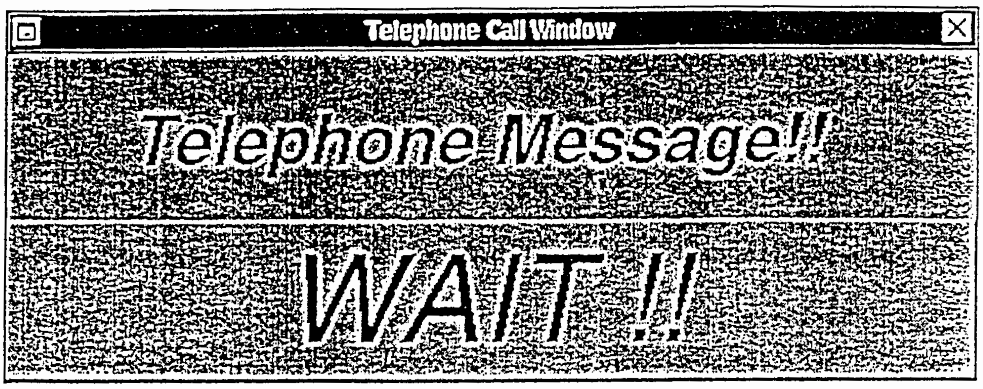


Figure 8: Software

When a client types in the telephone number of the conference center, the number is transmitted to the agent TEL process. The process, then, generates the sound of a telephone ring, and switches the window to Figure 11.

The agent can answer the call by pushing the return key which automatically executes four background processes running in parallel: TITLE, TEXT, VIDEO, and Menu.csh. The TITLE process displays the EMMI logo window and the clock using the control object (Figure 12). The TEXT process displays input and output windows through which participants communicate using keyboards (Figure 13). The VIDEO process displays a video window on which full-motion video facial images of the participants are projected. The Menu.sch displays a small menu window at the top left corner of the agent's screen (Figure 14), so the agent can select items on the menu. This menu window, however, is not displayed on the client's screen.

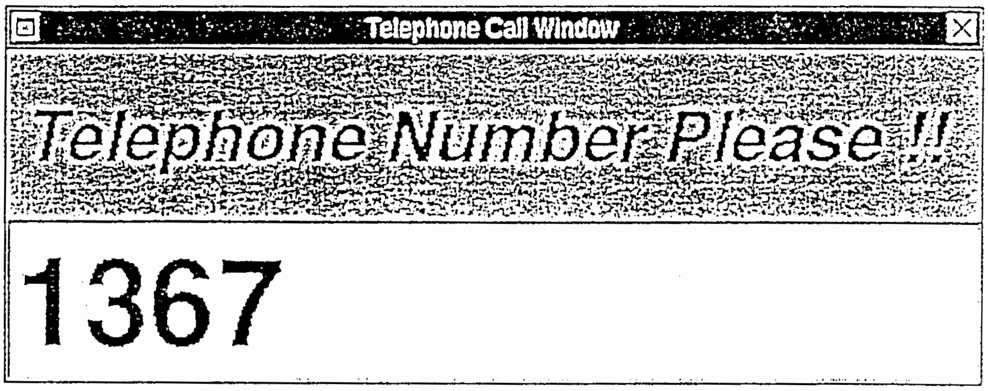
For clients who read only English, there is an English version start.new.csh.eng process (Figure 15).



Telephone Message!!

WAIT !!

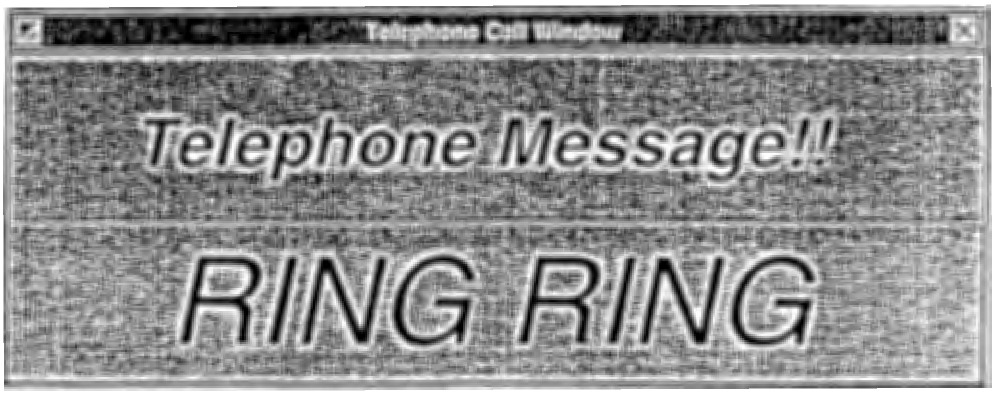
Figure 9: Agent's Initial Window



Telephone Number Please !!

1367

Figure 10: Client's Initial Window

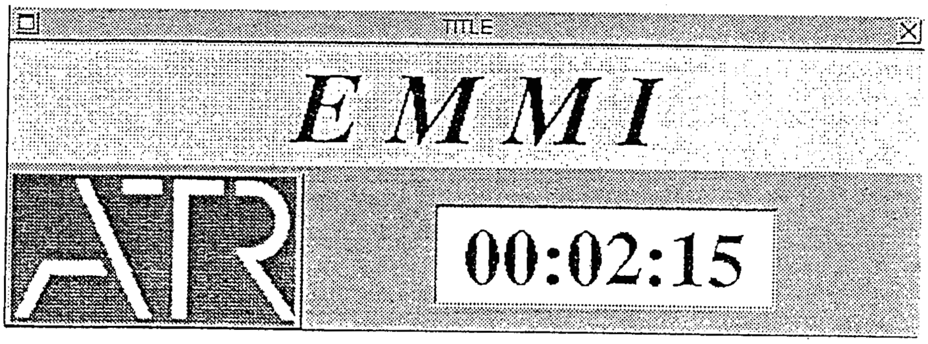


Telephone Call Window

Telephone Message!!

RING RING

Figure 11: Agent's Window



TITLE

EMMI

ATR

00:02:15

Figure 12: Logo Window

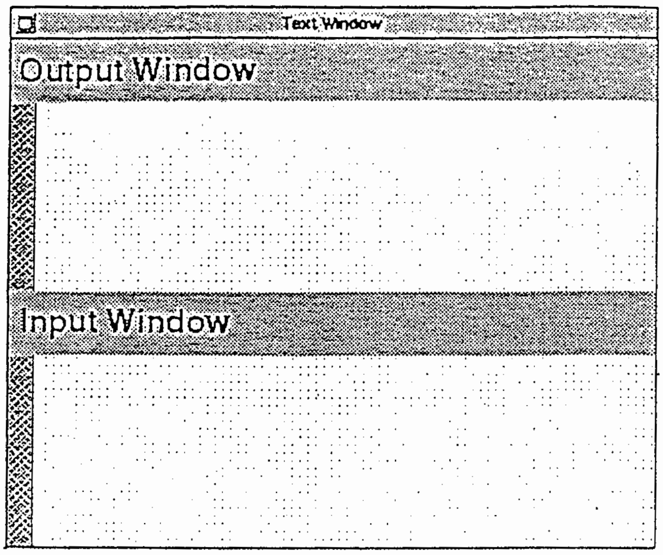


Figure 13: Input/Output Window

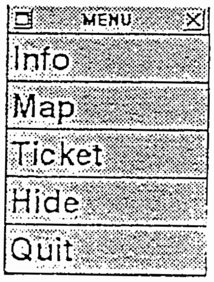


Figure 14: Menu Window

7 FURTHER IMPROVEMENTS

As with any other man-machine system, EMMI needs to go through a period of evolution and fine-tuning. Some of the current limitations can be eliminated simply by replacing the current hardware settings. Marking and writing, for example, on the CRT screen with a mouse is awkward for people who are not accoustomed to doing it. Our pilot study [3] indicates that many subjects prefer taking pen and paper notes instead of writing on the screen. Awkwardness in using a mouse may prevent them from using the mouse more actively. In this case, replacing the CRT with a flat LCD tablet with pen input capabilities may facilitate handwriting.

Another problem is that the present video camera positions do not allow direct eye contact between participants. One would have to look straight into the camera lens in order to make direct eye contact. In other words, the screen and the camera would have to be one and the same. With the current setting, it seems that one is talking to another who is looking away. However, this is not a major problem because participants look at the information window most frequently and seem to keep the video image in their peripheral vision.

Record keeping is another area that needs to be improved. The client's interactions with EMMI, for example, were initially video-taped by placing a video camera behind the right-hand of the client. It turned out to be almost useless, because the client's writings and markings on the screen were too faraway and too small to be recognized. The situation was somewhat improved when the video camera was moved to behind the left shoulder of the client, and closer to the screen. In this position, although the camera had a better view of the screen, it was often obstructed by the instruction sheet the client was waving. We are currently making an effort to record the client's interactions and gestures directly on the computer, or to download from the computer to a video camera.

参考文献

[1] Lauren BRUST, Mean-SEA Tsay. *Mixing signals voltages on chip, IEEE SPECTRUM (8/1993).*

[2] Bruce F. WEBSTER. *The NeXT Book, Eddison Wesley Publishing Company (1989).*

[3] Ryo FURUKAWA, Fumihiro YATO, Kyung-ho LOKEN-KIM, *Analysis of Telephone and Multimedia Dialogues, ATR Technical Report TR-IR-0020 (1993).*

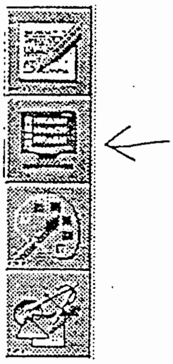
APPENDICES

A MANUAL

A.1 AGENT'S MANUAL

A: HOW TO START EMMI

1) Turn the CRT on by pushing the power botton at the right bottom of the screen.
2) Type the followings:
   User ID: ATR
   Password: EMMI
3) Using the mouse, select the following icon displayed on the right side.



4) Type the following on the screen:

```
anext-1> cd /kazuhiko/WORK
anext-1> start.new.csh : for Japanese agent and client
anext-1> start.new.csh.eng : for English agent and client
```

5) Using the mouse, select ハイド (h) on the grey window located on the top-left corner of the screen.
6) Now wait for a client to call you.
7) When the bell rings, answer it by clicking the bottom part of the window using the mouse.

A.2 CLIENT'S MANUAL

A: HOW TO START EMMI

1) Follow from step 1) to 6) of agent manual.

7) Now type 1367 and return to call the agent.

B INSTRUCTIONS

Prepared by Laurel Fais and translated into Japanese by Ryo Furukawa

Instructions

Thank you for coming today. I want to tell you a little bit about this experiment, and give you some instructions about what you are going to do.

In today's experiment we are trying to find out how people speak to one another in different contexts. We will ask you to play a role and speak to another person in two different settings. We'll tell you more about who you are supposed to be and what you should say later. Right now, let's look at the two different settings.

One should be familiar to you. It is just a simple telephone. Youwill be wearing a headset so that we can record your speech, but other than that, the telephone you will be using is exactly like other telephones that you use every day.

The second setting is a multi-media simulator that has been built in our lab. Before I show you how it works, I need to ask you a couple of questions. [answers are recorded on experiment record sheet]

First, could I ask you your age?
How much experience have you had with computers? What kinds of computers?
Do you use them daily? weekly? Have you used them more in the past than you do now?
What do you use them for?
Have you ever used a mouse before? How good do you feel you are at handling a mouse?
How good do you feel you are at using a keyboard?
Have you ever used headphones for listening? Have you used a headset for speaking into?
Do you have any speech, vision, or hearing problems that you know of?

Thank you. Now let me show you how this works. You will sit here, and wear this head set, through which you can hear the person you are going to be talking to. This microphone will pick up your voice and send it to the other person. When you are connected to the other person, and the computer is on, you will be able to use the keyboard and the mouse in various ways. Let's practice so that you are comfortable with the set-up.

First, try speaking into the microphone. Can you hear the person at the other end? [short conversation to get accustomed to the headset]

Let me show you how the mouse works. By moving the mouse like this, you move a cursor on the screen. If you want the cursor to stay in a certain place, click the button like this. [omit if person is familiar with mouse; allow practice time if unfamiliar]

If you click the mouse here, you can type in this box. What you are typing appears on the other person's screen. What the other person types on the other keyboard will appear on your screen here. [allow time to practice, including seeing text from other machine appear on screen]

You can also see a map on your screen. You can draw on your map by using the mouse. Position the cursor where you would like to start drawing; then push the button down and move the mouse. You will draw a line. When you want to end the line, let up on the button. You can clear the map by positioning the mouse cursor here and clicking the button. [allow practice time]

Now, let me tell you what you will be talking about. You should imagine that you are a person who wishes to attend a particular conference being held in Kyoto. Here is the flyer for the conference. [attached] You have no knowledge of how to travel around in Kyoto, and, in particular, you know only the name of the conference site and you do not know where it is. You have arrived in Kyoto Station and need to know how to get to the conference site. You will call the Conference Information Office to get the information you need.

You will play the role of client twice, the first time using the telephone/MM set-up [which condition is first varies with the subject], the second time using the MM set-up/telephone. When you are finished, I will ask you a couple questions about your impressions of the experiment.

(Telephone:) This is the telephone number you should call to get information from the agent at the conference office. When it is time to begin the experiment, please dial this number.

(MM:) The number you should use to get information from the agent at the conference office is listed next to "Infone" on the conference flyer. When it is time to begin the experiment, please type this number into the keyboard and hit the return key.

The person on the other end will be the agent at the conference office and will be able to answer your questions about how to get to the conference site. For the purposes of this experiment, it is best if you are as relaxed and as natural as possible, and that you play your role as completely as possible. Feel free to ask the agent any questions you need to, or to make any comments you would normally make in such a situation. Simply end your conversation in a natural way when you feel you have as much information as you need to get to the conference site. Do you have any questions?

You may begin any time you are ready.

Name: ____
Date: _______________________
Telephone _____ MM _____
Task:
Language:
Handedness:
Male Female
Age:

Pre-experiment questions

Computer experience:
kind:
frequency:
purpose:

Proficiency with
mouse:
keyboard:
headphones:
head mike:

Problems:
Speech:
Vision:
Hearing:

Comments from agent:

Post-experiment interview

(After the experiment:) Thank you very much. Would you please fill out this questionnaire before yu go?

1. How would you rate how much you enjoyed the experiment?

Telephone:

a real bore   kind of interesting   fun   had a great time

Multi-media setting:

a real bore   kind of interesting   fun   had a great time

2. How would you rate how easy it was?

Telephone:

simple   some effort   had to work at it   difficult

Multi-media set-up:

simple   some effort   had to work at it   difficult

3. How would you rate the usefulness of:

Telephone:

very useful	served some purpose	an inconvenience	worthless

Map:

very useful	served some purpose	an inconvenience	worthless

Keyboard:

very useful	served some purpose	an inconvenience	worthless

4. Do you have any impressions you'd like to note?

5. Do you have any suggestions?

Second International Symposium on Interpreting Telecommunications

September 14-16, 1993
The International Conference Center

Keynote speaker

Dr. Loken-Kim, ATR
"The Role of Multimedia Contexts in Machine-aided Translation"

Sponsored by:
The Information Processing Society of Japan
The Institute of Electronic, Information and
Communications Engineers of Japan
Japanese Society for Artificial Intelligence

For more information, contact:
Conference Information Office
2-2 Hikaridai
Seika-cho Soraku-gun
Kyoto 619-02
Phone: (07749-5-) 1368
Infone: (07749-5-) 1367

実験要領

今日は私たちの実験に御協力いただきありがとうございます。この実験についての説明と、被験者の方にしていただきたいことの説明をさせていただきます。

私たちはこの実験で、人々が異なった状況において互いにどのようにしゃべるのかを調査します。あなたはある役を与えられ、二つの異なった状況で会話をします。与えられる役と話す内容については後でお話します。まず最初に、二種類の機械を御覧下さい。

一方の機械はおなじみでしょう。ただの電話機です。あなたの会話を録音するためにヘッドフォンマイクを使用しますが、それを除いてはこの電話機はあなたがいつも使っているものと全く同じものです。

もう一つの機械は本研究室で開発されたマルチメディアシミュレーターです。この機械の使い方を説明する前に、質問をいくつかさせていただきます。

あなたのイニシャルを教えて下さい。

あなたの年齢を教えて下さい。

いままでにコンピューターをお使いになった経験はどのくらいありますか?

またどのようなコンピューターをお使いになりましたか?

毎日お使いですか? 週に一度? 昔は今よりも頻繁にお使いでしたか?

何にお使いですか?

マウスを使ったことはおありですか?

マウスを使うとどのくらい使い易いですか?

キーボードを使うとどのくらい使い易いですか?

何か聞くのにヘッドフォンを使ったことがおありですか?

何かしゃべるのにヘッドフォンマイクを使ったことがおありですか?

あなたの知る範囲で、喋ったり、見たり、聞いたりするのになにか障害をお持ちですか?

どうもありがとうございます。ではマルチメディアシミュレーターの使い方を説明させていただきます。あなたはここに座って、ヘッドフォンマイクを装着します。するとそこから、あなたがこれから会話する人の声が聞こえます。ま

たあなたの声はマイクが拾って、相手側に送られます。あなたの側のコンピューターと相手側のコンピューターとが動いていて互いに接続されていれば、あなたはキーボードやマウスをいろいろな目的に利用できます。この機械に慣れるために練習をしてみましょう。

まず、マイクに向かって話してみて下さい。相手の人の声が聞こえますか?
(ヘッドフォンマイクに慣れるために短い会話をする)

次にマウスの使い方の説明をします。マウスをこのように動かすと、画面上のカーソルを動かすことができます。もしカーソルをある場所に止めておきたければ、ボタンをこのようにクリックして下さい。
(被験者がマウスに慣れていれば省略する。慣れていなければ練習時間をとる)

ここでマウスをクリックすると、この四角の中に文字をタイプできます。あなたがタイプしている文字は、相手側の画面に現れます。相手の人がタイプしている文字は、あなたの側の画面に現れます。
(練習時間をとる。このときに相手側からの文章が画面に現れるのを見せる)

また画面上には地図も出ています。あなたはマウスを使って、地図の上に書き込むことができます。線を引き始めたい場所にカーソルを持ってきて下さい。そしてボタンを押し、そのままマウスを動かして下さい。これで線を引くことができます。線を引き終わりたいときには、ボタンを放して下さい。またマウスカーソルをここに持ってきてボタンをクリックすると、地図上の線を消すことができます。
(練習時間をとる)

それでは、これからしていただく会話について説明します。あなたはこれから京都で開かれるある会議に出席しに行くところです。手元には会議についてのチラシがあります。しかしあなたは京都の地理について何も知りません。そして何よりも、会議場の名前を知っているだけでその場所は知らないのです。京都駅に着いたあなたは、会議場に行く方法を見つけなければなりません。そこで会議の事務局を呼び出すことにします。

あなたにはこれからサービスの利用者の役を二回やっていただきます。最初は電話機/マルチメディアシミュレーターで、二度目はマルチメディアシミュレーター/電話機で実験します。実験の後、実験の印象についての幾つかの質問をさせていただきます。

(電話機:)これは会議の事務室で情報サービスを受けるために利用する電話番号です。実験を始める時間になったらこの番号をダイヤルして下さい。

(マルチメディアシミュレーター:)会議の事務局の情報サービスを受けるための番号が、会議のチラシの「情報ネット」の横に書いてあります。実験を始める時間になったら、キーボードにこの番号をタイプして下さい。

相手側の人は会議の事務員で、あなたが会議場にいく方法を教えてくれます。この実験の目的のために、あなたが出来るだけリラックスして自然に振舞うこと、また自分の役になりきることが望まれます。道を尋ねる時にあなたがするような質問や説明はどんどんして下さい。会議場に行くために必要な情報が手に入ったと思ったら、自然に会話を終って下さい。

何か質問が有りますか?

では準備が出来たら始めて下さい。

実験後のインタヴュー

ありがとうございました。

お帰りになる前に次の質問にお答えいただけますようお願い致します。

1.実験はどのくらい面白かったですか?

電話:

すごく退屈　　まあまあ　　楽しかった　　夢のようだった

マルチメディアシミュレータ:

すごく退屈　　まあまあ　　楽しかった　　夢のようだった

2.扱い方の難易度はどうでしたか?

電話:

簡単　　少しまごつく　　努力が必要　　難しい

マルチメディアシミュレータ:

簡単　　少しまごつく　　努力が必要　　難しい

3.次の道具の実用性をどう思いますか?

電話:

非常に有用　　目的によっては便利　　不便　　何の価値もない

地図:

非常に有用　　目的によっては便利　　不便　　何の価値もない

キーボード:

非常に有用　　目的によっては便利　　不便　　何の価値もない

4.実験の印象などで書きたいものがあれば書いて下さい。

5.何か意見があれば書いて下さい。

第二回 翻訳電話通信 国際シンポジウム

1993年 9月 14、15日
京都国際交流センター

基調演説

「Machine-aided Translation における
マルチメディア文脈の役割について」

ローケン・キム博士、ATR

提供:
日本情報処理学会
日本電子情報通信学会
日本人工知能学会

連絡先:
〒619-02
京都市相楽郡生花町光台2-2
会議事務局

電話：
1368

情報ネット:
1367