TR-IT-0203

The Effects of Experience and Interface on Users' Behavior in EMMI

Laurel Fais

February, 1997

We describe the motivation for and the design of a new interface for EMMI, including a "talking head" and greater initiative on the part of the system. We report the results of experiments in that new setting, comparing the behavior of experienced *vs.* inexperienced users, and comparing this new interface with the previous interface. The experience of the user had no effect on his/her behavior. Subjects in the new interface exchanged more information, in a more efficient manner, and required less oral prompting to use the typing options.

Table of Contents

1. Background

Spontaneous speech is the current black sheep in the language processing field, replacing speech recognition systems as the component everybody loves to hate. It's messy; it's unpredictable; it's ill-formed. In short, it's difficult to deal with given the current limitations of the field.

But it's also the final frontier, and semantic-based parsers and more robust recognition systems are taking up the challenge of making spontaneous speech make sense to machines.

An assault is being mounted from the other side as well. Recognizing that humans may be more flexible than their machine interlocutors, or perhaps simply feeling that the machines shouldn't have to do all the work, some researchers have attempted either directly or indirectly to manipulate the input to language processing systems so that it is less unruly. Direct manipulations include requiring users to type input, adhere to a specified list of utterances, speak to a form, push-to-talk, and the like. Indirect manipulations include configuring the interaction environment so that users, while apparently unconstrained, are encouraged to speak as little like a relaxed human being, and as much like text, as possible.

The work that we report here falls into this last group. For the last five years, we have been developing a multimodal interface for integration with speech translation. Initial research examined the spontaneous, task-oriented dialogues of humans speaking the same language (Fais and Loken-Kim, 1994), humans speaking different languages and interpreted by another human (Park *et al.*, 1994), and humans speaking different languages and interpreted by a "machine" translator (in a Wizard of Oz style experiment: Fais *et al.*, 1995). The subjects in each of these experiments carried out their conversations in both telephone-only and multimedia (MM) environments so that we could compare not only differences due to various interpretation settings, but also those due to diverse media environments.

The results were rather low-key (Fais *et al.*, 1996). Interested in designing a configuration of media to improve the quality of speech input for speech translation (including speech recognition), we examined such measures as disfluency rate, predictability of lexical items due to accommodation between conversational partners, quality of self-correction, and the number of words used to convey information. The results for disfluency rates and accommodation were encouraging for speech processing, if not for the use of non-speech media: subjects exhibited the lowest disfluency rates of all conditions of all three experiments in the machine-translated experiment, though there was no difference between the telephone and the MM conditions. Subjects accommodated to the lexical items used by the machine interface (making those items more predictable) more in the machine translated setting than in the human-human setting, though not as much as in the human-interpreted setting. Again, there were no significant modal differences. Finally, subjects were sometimes asked to repeat their utterances in the "machine" translated setting; these corrections also tended to be less disfluent, shorter, and more predictable.

There was one major difference between subjects' behavior in the telephone and MM conditions: subjects in the MM condition of all three settings conveyed and acquired more information during their conversational task than subjects in the telephone-only condition. On the other hand, they used more words to do it (Fais and Loken-Kim, 1996).

Communication sciences seems caught up in a "the more media the merrier" mentality. If there is any validity to the intuitions that lie behind the idea that the use of a variety of media will somehow "improve" communication, why are we having such a hard time extending that notion to interfaces with machine translation?

There are (at least) two possible answers. First, perhaps we aren't looking in the right place. And, in fact, when we look at user satisfaction, for example, or enjoyment, or comfort (by means of post-experiment questionnaires and interviews), instead of looking at disfluency rates and numbers of words, we find great advantages attributed to the availability of different communication media.

The second possible answer is that the configuration of media that we are employing is not optimal. This is a huge question. Multimedia environments, by definition, integrate a wide variety of decisions about media use into one complex, interacting system. Even with the most careful planning, it is quite possible to design a less-than-optimal system.

Below we report on our efforts to corral spontaneous speech with a multimedia user interface.

2. Interface design

With initial experience in hand, we went back to the multimedia board and attempted to identify those areas of the system that might be responsible for our lackluster results, keeping in mind the necessity of retaining those aspects of the system that made subjects feel more comfortable and more amenable to exchanging greater amounts of information. We first looked at the counter-intuitive results concerning the number of words subjects used to achieve information exchange. Our multimedia environment, dubbed EMMI (Environment for MultiModal Interactions), allows subjects to speak to one another, exchange drawings on a mutual map, type messages, and view both still pictures and a video image of their conversational partner's face. Despite the presence of all these visual, non-speech options, we found that subjects in the MM condition still used *more* words to convey information than those in the telephone setting. Closer inspection of the transcripts made clear one reason for this outcome: subjects used a lot of words to talk about the handling of the media itself, whether they should draw something, for instance, or to request their partners to type. However, even taking this "meta-media conversation" out of the calculations, subjects used no fewer words in the MM condition than on the telephone.

Though puzzled by the last result, we set it aside and attempted to address the problem of meta-media speech. The original interface in EMMI did not contain any built-in instructions or take any initiative in the use of the available media. It seems that, despite ample explanation and practice time before the experiment then, subjects were occasionally uncertain as to how to proceed and had to ask their partners about media use. For this reason, we designed the second version of the interface to include instructions about media use and to take a more aggressive part in directing the conversation.

Our second concern involved the slightly depressed rate of lexical accommodation in the machine translation condition. Noting that the highest rate had occurred between the subject and the (human) interpreter in the human-interpreted condition, we decided to include an iconic "persona" for the translator in the machine translated condition to see if we could foster a similar high rate of accommodation in that condition.

The addition of a "face" representing the machine "translator" also allowed us to solve another troublesome problem. In the initial machine-interpreted condition, subjects could hear all of the speech of all of the participants. They were sometimes confused as to what it was they were hearing, and this confusion, coupled with the time lag introduced to simulate machine translation, led to a high number of turn-taking missteps and false starts. Clearly we needed some sort of tracking mechanism by which participants could tell what was going on in the conversation.

We re-designed the interface to include not only the video image of the partner's face, but also the "face" of the machine "translator" as well as the face of the subject. When the machine was "listening" to the subject, the face looked toward the video image of the subject and had its mouth closed. When it was translating something to the subject, it faced the subject with its mouth open (and similarly for the other participant). In this

way, there was a visual representation of the status of the conversational turns. This visual representation was accompanied as well by a written description such as "The translator is listening to you," etc.

One further change was made to the system, motivated by developmental concerns rather than experience. In an attempt to bring the system one step closer to the reality of a full working configuration, we also incorporated the speech synthesis system, CHATR, designed at the Interpreting Telecommunications Research Lab at ATR. Our first Wizards were instructed to make their speech as monotonic and syllable-timed as possible, in order to simulate the laymen's impression of synthetic speech. While they succeeded admirably in sounding like cartoon robots, and in fact, unquestionably convinced all the subjects in that experiment, their speech bore little resemblance to actual speech synthesis. Incorporating real speech synthesis was an attempt to make the system at least slightly more "real."

Of course, there were aspects to the original system that we wanted to leave unaffected. We had, recall, observed lower disfluency rates and greater information exchange in the "machine" translated system and didn't want to sacrifice those aspects of the original interface. While we felt that the latter probably would not be affected by the new system, the former issue could be problematic. Speakers tend to "clean up" their speech for machines; if the "face" of the "machine translator" made users forget that they were talking to a machine, or at least revamp their notion of "machine" in some possibly counterproductive way, it might be that the subject would use more rather than less casual speech, with the result that disfluency would increase.

These, then, are the issues that lay behind the experimentation that we report below. We first give a more detailed description of the revised interface used in the experiment; then we outline the details of the experiment itself, and the results collected. We compare these results to those collected in the experiments with our previous media configuration. We examine the data from two perspectives: how did the nature of the interface affect the amount of language produced by the subjects? And how did it affect the nature of the language used? Finally, we draw some conclusions about future generations of multimedia language processing systems.

2.1 The ASysT interface

The set of screens as seen by the subjects in the experiment are given in Appendix A. Subjects are seated in front of a NeXT computer equipped with a touchscreen, and have a microphone and speakers next to them.

The initial screen introduces the subject to the "ATR Automatic System for Translation,"[1] and asks the subject to select either English, German or Korean by pushing the appropriate button. Once they select English (since the other buttons are not functional and all subjects were native English speakers), they are asked to select the appropriate office. When they choose the office they were instructed to call (see below), the ASysTant (i.e., the machine's "face") appears and introduces the system. The text of the introduction is printed on the screen below the ASysTant and is "spoken" by CHATR as well. This, and subsequent instructions which are also handled with both spoken and printed language, allows us to "train" the subjects to the peculiarities of CHATR while they have the opportunity to check their understanding against a written version.

In the next screen (which appears automatically), the ASysTant is again displayed, this time in between the video images of the two participants in the conversation. The ASysTant introduces the subject to the other participant and invites the subject to broach his/her reasons for "calling."

To this point, the CHATR output is automatic; once the subject and the other participant begin their conversation, the Wizard is responsible for sending appropriate output. He can send pre-set utterances (these utterances are listed in Appendix B; having conducted three previous experiments with the same task, it was possible to predict fairly

[1]Not its real name. However, it does make a nice acronym, "ASysT," and even provides us with the name for the machine "persona:" the ASysTant. We used this name in order to lend an air of reality to the system so that subjects would think that it was an actual working system.

accurately a high percentage of the utterances necessary), or type in utterances that fall outside the predictable course of the conversation, which are then synthesized by CHATR. The Wizard also controls the position of the ASysTant's face; he turns it away from the subject when it is listening to or translating for the other participant. The face automatically turns *toward* the subject, in the "translating" expression, when output is sent from CHATR and changes to the "listening" expression when CHATR finishes its output. The Wizard, an experienced translator and interpreter, also translates from the subjects' English to Japanese for the other participant. This is done *sotto voce* so that it is not audible to the subject. Neither is the Japanese of the other participant.

Once the subject begins the task, there are two possible screens which the other participant can use to aid their efforts. Usually the first to be called up is the map. The first time that the map appears, written and spoken instructions concerning its use appear in boxes over it. The same is true when the second screen, a form, is called up. These instructions disappear after they are "spoken" by CHATR.

Both participants can draw on the map (touchscreen) with their fingers. In order that the timing of the drawing be appropriate for the opposite partner, it is first sent to the Wizard. The Wizard then sends the drawing to its intended recipient along with the translation.[2] On the other hand, the information which both participants can type on the form appears immediately on the form of the opposite participant; generally, this information does not accompany speech and there is no need to delay its transference while a translation is being made.

The final option for use by both participants in the conversation is to type messages in a typing window. Again, the first time the typing window is opened by either participant, instruction boxes appear and the instructions are also "spoken" by CHATR. Typed messages, like spoken utterances, are also relayed through the Wizard. If translation is necessary, the Wizard can type a translated message and send it to the appropriate recipient.

At the conclusion of the conversation, the subject "hangs up" by pushing the appropriate button and the face of the ASysTant appears to thank the subject for using ASysT.

3. Method

Twenty-seven paid subjects, all native speakers of North American English, participated in the experiment. They were told to imagine that they had just arrived in Kyoto and needed to get directions for a conference they were planning to attend; relevant information about the conference was provided on a conference "flyer." In addition, they needed to make hotel reservations for their time in Kyoto during the conference. They were encouraged to play the role of "client" as completely and naturally as possible, asking whatever questions they felt were necessary in order to accomplish those two tasks. Each subject participated in just one conversation.

We conjectured that one of the reasons subjects had used so much meta-media speech in earlier experiments was because they were unfamiliar, and perhaps uncomfortable, with the system. We could not test this idea directly, since we did not have a pool of subjects who were experienced with and thus comfortable with the system. So, we expanded on the idea, proposing that users who were naive about computer use in general might in fact behave differently from those were very experienced in a computer environment. For this reason, we recruited subjects of two types: those who had little experience with computers (these subjects did not use computers on a daily basis and when they did use computers, used them at most for word processing tasks), and those with extensive computer experience (data analysis, programming, system design and the like). Fourteen subjects were considered "inexperienced;" thirteen were considered "experienced."

The other participants in the experiment were the Wizard, experienced not only in interpretation but also in Wizardry, as he had served as Wizard in a previous experiment

[2]The most recent version of the system incorporates an automatic speech/visual gesture coordinator (Loken-Kim *et al.*, 1995).

as well, and a trained, native speaker of Japanese who acted as the "agent" at the Conference Office. A schematic of the experimental configuration is shown in Figure 1. Video recordings were made of the Clients' and the Interpreters' screens; audio recordings of all speech were made on a DAT recorder, and later transcribed for filled pauses, false starts, simultaneous speech and non-speech noises as well as "normal" utterances.

Before the subjects began their conversations, the workings of the system were explained to them and they were given some time to practice on the relevant and similar components from the previous system. They also filled out an attitude survey concerning their feelings about computers. After their conversation was finished, subjects filled out a similar attitude scale (with items randomly re-ordered) concerning their feelings about ASysT, took a "quiz" to ascertain what sort of information they had gotten out of their conversation, and completed a questionnaire covering various aspects of the experience. (In addition to using the quiz to measure the information they had gathered, we also hoped that the knowledge that they would complete a quiz at the end of their conversation would induce greater concentration.) After the completion of the experiment, subjects were sent a letter informing them of the true nature of the system (with apologies to all the computer scientists among them who had been excited to think that such a system was a reality).

4. Measures

4.1. Amount of language and information

4.1.1. Words and turns. Words, including those in false starts, and turns, including feedback turns, were counted.

4.1.2. Words per information unit. "Information unit" is a loose term which we apply to a piece of information, often the size of one turn, typically exchanged between Agent and subject. The amount of the bus fare, the length of the taxi ride and the location of the train station exit are all examples. Using a standard list of these units, compiled after examining all the transcripts from previous experiments involving the same tasks, we counted the number of information units that appeared in the transcripts of the conversation. The list of information units included an "other" category so that we could give credit for all information which subjects ascertained in the experiment. In addition to a straight count of information units, we also calculated a word-per-information-unit rate, to determine, in effect, the efficiency of the information exchange.

4.2. Media use

We will not attempt a full scale description of the use of the non-speech media in these experiments here. However, one of the main motivations for incorporating a multi-media interface with automatic speech translation was the intuition that users might reduce the amount of language they used, replacing it with the use of visual media. Thus, it is appropriate to explore the relationship between the use of these options and the use of speech.

4.2.1. Map. Each occasion of drawing on the map was transcribed and classified according to the task to which it pertained, the type of drawing it was, the meaning it conveyed and its relationship to the accompanying linguistic expression (if any).

Task. Since there were two tasks required of the Clients, i.e., getting directions and making a hotel reservation, we labeled each instance of media use depending upon whether it occurred during the direction-finding or during the hotel reservation task. Some Clients also pursued topics that fell outside these two areas; media use in these areas was labeled "other."

Drawing type. An examination of the types of drawings made by the subjects showed that they tended to be one of three types: subjects drew circles and lines, and

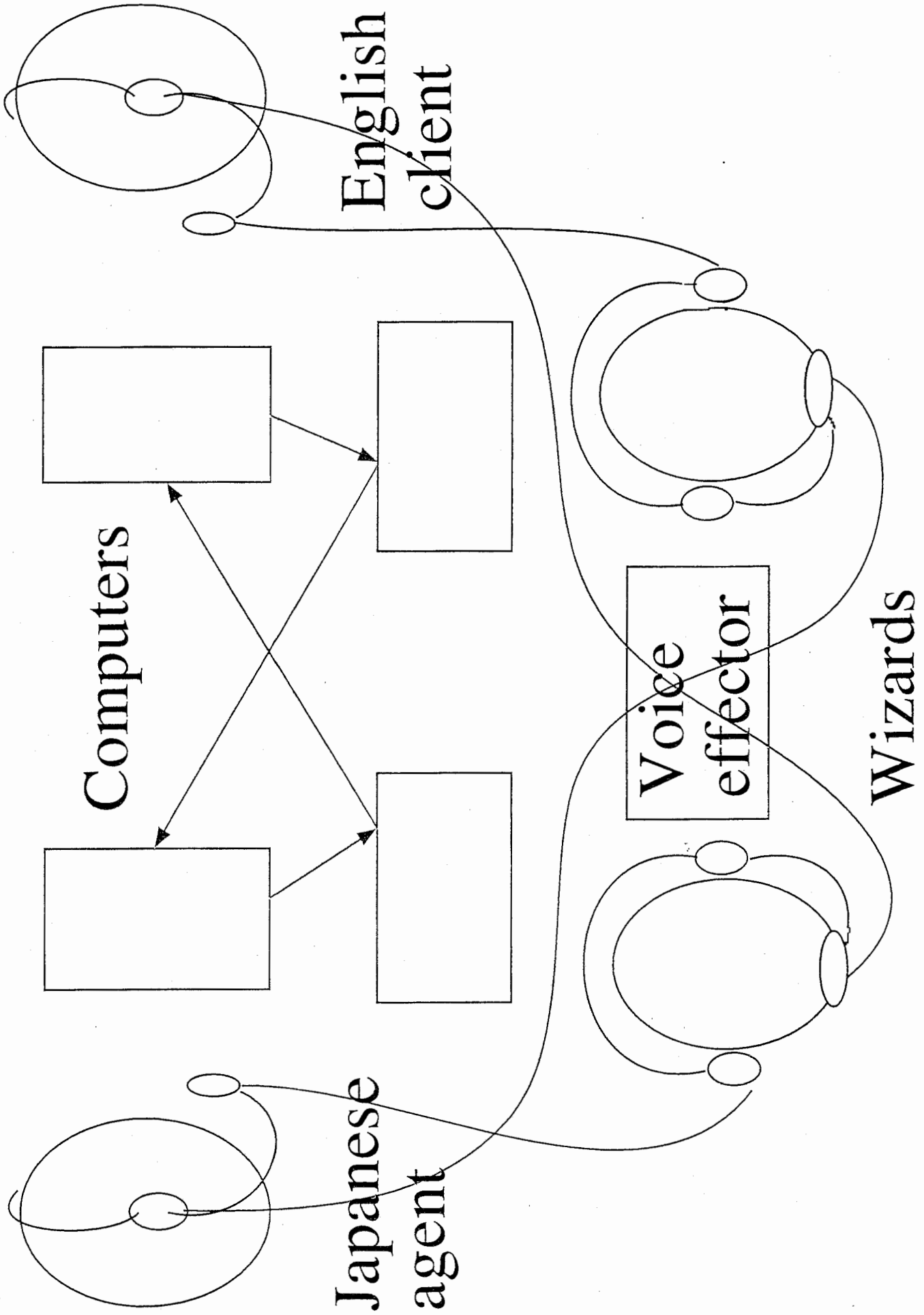


Figure 1. The configuration for the ASysT interface experiment (also for the Faceless interface experiment).

made small marks by pointing.[3] We counted the number of each of these drawing types for each subject.

Meaning. We made an attempt to determine what sort of meaning subjects were expressing by their use of a drawing. This admittedly required a subjective analysis. Our primary method was to consider the linguistic expression which (usually) accompanied the drawing. In typical cases such as "walk this way" accompanied by a line drawing, or "the hotel is here" accompanied by a circle, we feel confident in assigning the meanings *direction*, and *location*, respectively. However, some other cases were not so straightforward. For example, "walk down these stairs here," accompanied by a line drawing showing the way leading to the stairs as well as the route down the stairs themselves, seemed to indicate both *direction* and *location*. These sorts of examples were so labeled.

Deictic/Redundant. Our final classification involved determining whether the meaning conveyed in the drawing was redundant to that contained in the linguistic expression or whether it accompanied a deictic expression and was an integral part of the semantic content of the utterance. Thus, the circle drawn around the hotel on the map accompanying the utterance "Is it possible to get a better view of Kyoto City from the Kyoto Park Hotel" is considered redundant. The circle around the bus stop accompanying the utterance "The bus stop is here" is considered deictic. Some drawings occurred in the absence of linguistic expression; these were labeled "alone."

4.2.2. Typing. Instances in which subjects used the option to type were transcribed and classified according to the following parameters.[4]

Prompt. We examined what kind of prompt motivated the typing, i.e., whether it was in response to an oral request for information, to a visual prompt such as a slot on a form where information can be typed, or if the typing was unprompted.

Form completion interaction patterns. Judgments concerning the categorization of patterns of typing in response to the reservation form questions are extremely subjective. However, some generalizations can be made.

We examined the interaction between Agent and Client which lead up to the typing done for each slot on the reservation form. These interactions seemed to be of three main types, illustrated by the examples below.

Agent asks question; Client answers; Agent types (A?; CA; AT)

[1] Agent: Touchakuwa, itsu desuka
Wizard: What time will you arrive?
Client: two p.m.
Wizard: gogo niji
[Agent types 2p.m.]

Agent asks question; Client types answer (A?; CT)

[2] Agent: Nanjini touchaku shimasuka
Wizard: What time will you arrive?
[Client types in arrival time]

Client sees reservation form; Client types information (CT)

---

[3]Two subjects also drew lines with arrows one time each; these are not included here. In previous work (Park *et al.*, 1995), we identified "spiral" as a type of drawing as well. While there were a small number of "spirals" in this data, we simply grouped them with circles, considering them to be less "spirals' than iterated circles.

[4]They were also classified according to which task they fell into; since this typing of classification does not concern us here, and since the vast majority of the instances of typing feel into the hotel reservation task, we will not discuss this further.

[3] Wizard: Please look at this form. You can type on the form. Touch the slot first, then type.[5]

[Client types name, phone number, arrival day, departure day, check in time, check out time, number of adults, single room]

Of course, not all interaction followed only one pattern; some exhibited a mixture of the above styles. We examined each conversation and categorized the pattern of reservation form completion using the above patterns as a guide. Conversations which followed primarily one type of pattern were given that classification. Conversations which combined two or more approaches without a clear preference for one approach were classified as "mixed."

4.3. Nature of language

4.3.1. Disfluency rate. This was calculated from the total number of disfluencies, which was the sum of the filled pauses and false starts. This sum was then divided by the number of words and multiplied by 100 to give the average number of disfluencies per 100 words.

4.3.2. Lexical accommodation. One of our major interests in previous experiments was the degree to which Client and Interpreter[6] use the lexical items spoken by the other. This is what we call "lexical accommodation," and we measure the lexical accommodation rate for each conversation in the following way. First, we determine the number of different words used by both the Client and the Interpreter ("words-in-common"). We then divide that number by the total number of different words used in the conversation (by *either* Client or Interpreter); this gives us the accommodation rate. In addition, we are also interested in determining, if possible, the direction of accommodation. We postulate that the participant who does *not* use a lexical item first is accommodating;[7] in light of that definition, we counted the number of words-in-common used first by the Client and the number used first by the Interpreter.

5. Results

5.1. Experienced and inexperienced subjects

It is quite straightforward to sum up the results pertaining to the experienced and inexperienced computer users in this experiment: there were none. In terms of the amount of language used, experienced and inexperienced computer users uttered equivalent numbers of words and turns, exchanged equivalent numbers of information units and used equivalent numbers of words to do so, and used equivalent numbers of meta-media words. There were also no differences in the use of non-speech media, either drawing or typing. In terms of the nature of the language used, subjects had equivalent disfluency rates, and had equivalent lexical accommodation rates.

On the other hand, there were some differences between results for subjects using the current, revised system ("ASysT") and subjects using the previous system ("Faceless") (all subjects participated in identical tasks under identical conditions with the exception of the configuration of the interfaces as described above; a detailed description of the experiment with the Faceless interface is given in Fais *et al.*, 1995). These differences are described below.

---

[5]This is the standard instruction given when the reservation form first appears onthe screen. In this case, the client simply took the initiative to fill in the slots herself without waiting to be asked.

[6]The justification this pairing is discussed extensively in (Fais, submitted).

[7]Arbitrary as this may sound, it is based on fairly carefully thought-out previous research (Fais and Loken-Kim, 1995).

5.2. Amount of language and information

Clients in the two experiments did *not* differ in the number of words used, the number of turns taken, or per cent of meta-media words used. They did differ, however, in measures involving information (Figures 2 and 3) and in the use of non-speech media.

Figure 2. Average number of information units in the direction finding task per conversation in the two experiments.

Clients using the ASysT interface exchanged significantly more information ($p<0.04$) in the direction finding task than did subjects in the Faceless interface.

Furthermore, Clients in the ASysT interface reduced the number of words used per information unit approximately 24% for both the direction-finding task and the hotel reservation task.

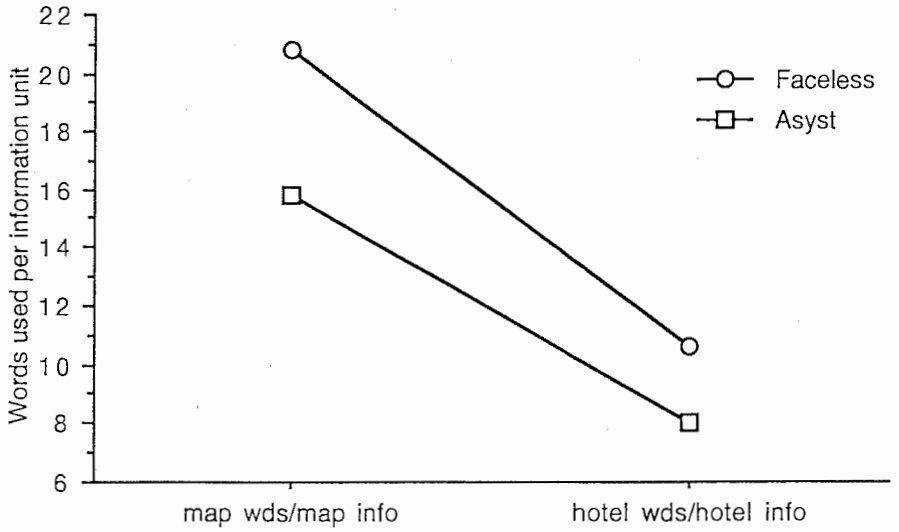


Figure 3. Number of words used by Clients per information unit for each task in each interface.

5.3. Media use

Subjects using both interfaces had virtually identical non-speech media options available to them,[8] but used these options in different ways.

5.3.1. Map drawing

Tasks. Agents in both settings drew on the map more often than their respective Clients. Further, both sets of Agents and the Client in the Faceless interface drew far more in the Directions task of the experiment. On the other hand, the Client in the ASysT interface drew as often in the Hotel reservation task as in the Direction task.

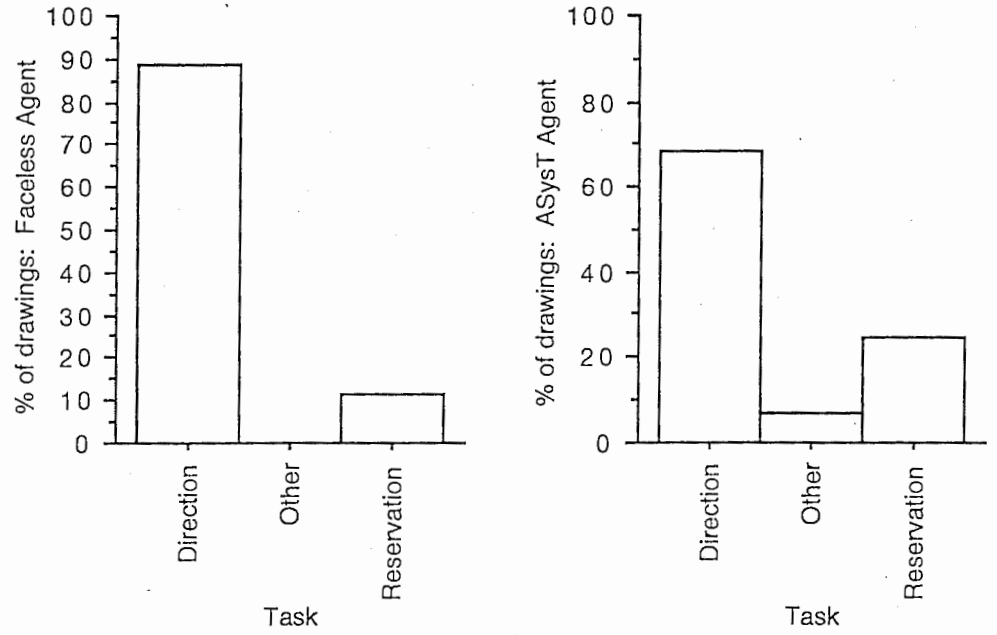


Figure 4. Distribution of drawings over tasks in each experiment, for Agents.

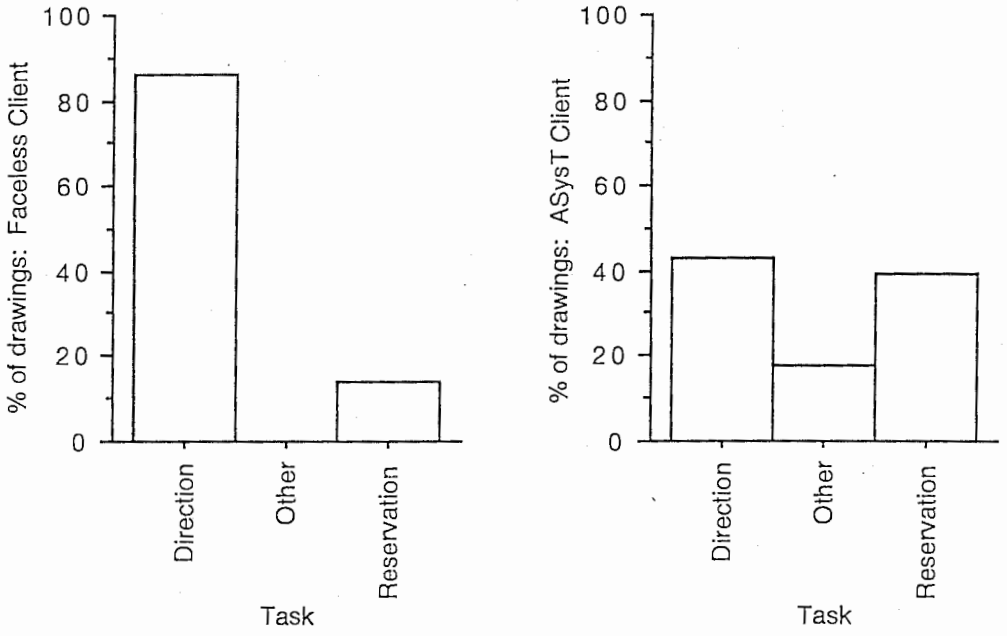


Figure 5. Distribution of drawings over tasks in each experiment, for Clients.

[8]The differences involved presentation. In the previous system, subjects received instructions before the experiment (and during, where necessary); in the current system, subjects received instructions online, in the course of the experiment.

Drawing types. Subjects were free to draw on the map in any way they chose. Clients in the Faceless interface chose to point (i.e., simply touch the screen, making a small dot) most often. However, since the resulting image was very small, part of our revisions to the Faceless interface was to render those small marks as somewhat larger circles. Agents in the Faceless interface used circles and lines equally; Agents in the ASysT interface used pointing more than either circles or lines. We conjecture that, because pointing resulted in a (small) circle, pointing may have replaced some of the use of circles by the Agent. Clients in the ASysT interface preferred circles.

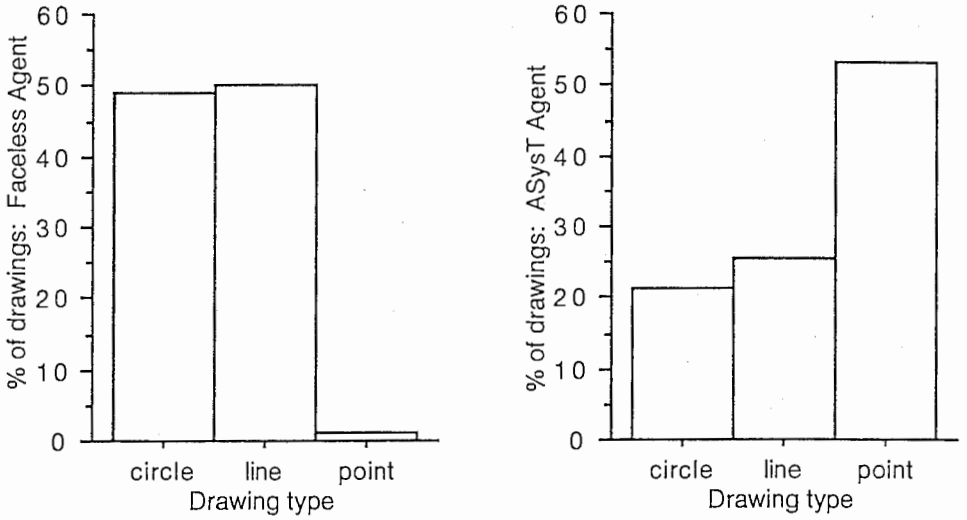


Figure 6. Distribution of drawings over drawing types in each experiment, for Agents.

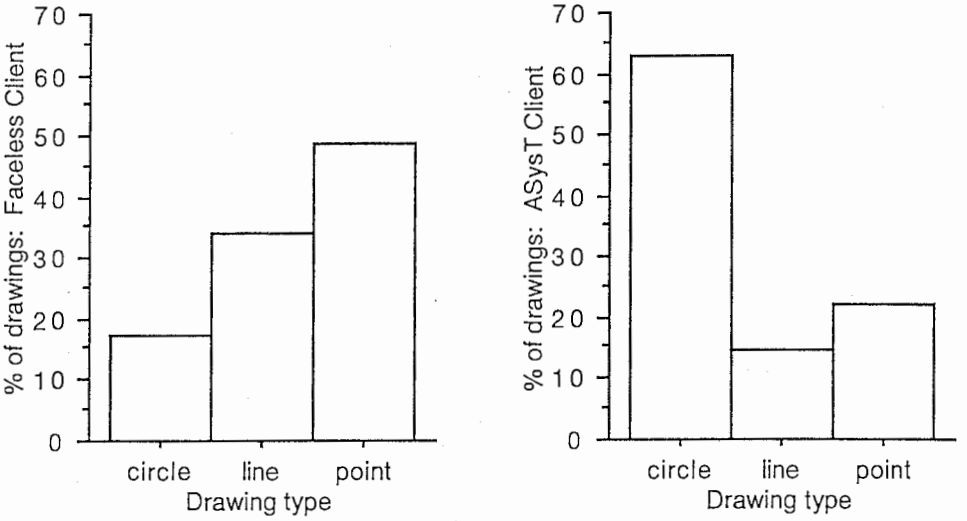


Figure 7. Distribution of drawings over drawing types in each experiment, for Clients.

Meanings. Agents in both interfaces used drawings to accompany expressions of *location* and *direction* most often (though the Agent in the ASysT interface used drawing far more often with expressions of *location*, while the Agent in the Faceless interface used it equally for *location* and *direction*). Clients, on the other hand, used drawing to accompany expressions of *identity* and *location* most frequently, to approximately equal

degrees. Another striking difference in the expression of meaning in drawings is that in the Faceless interface, subjects used drawing with a wider range of meanings, including some that were difficult to categorize. Hence in Figures 8 and 9, the labels "direction/location," "location/identity" and the like.

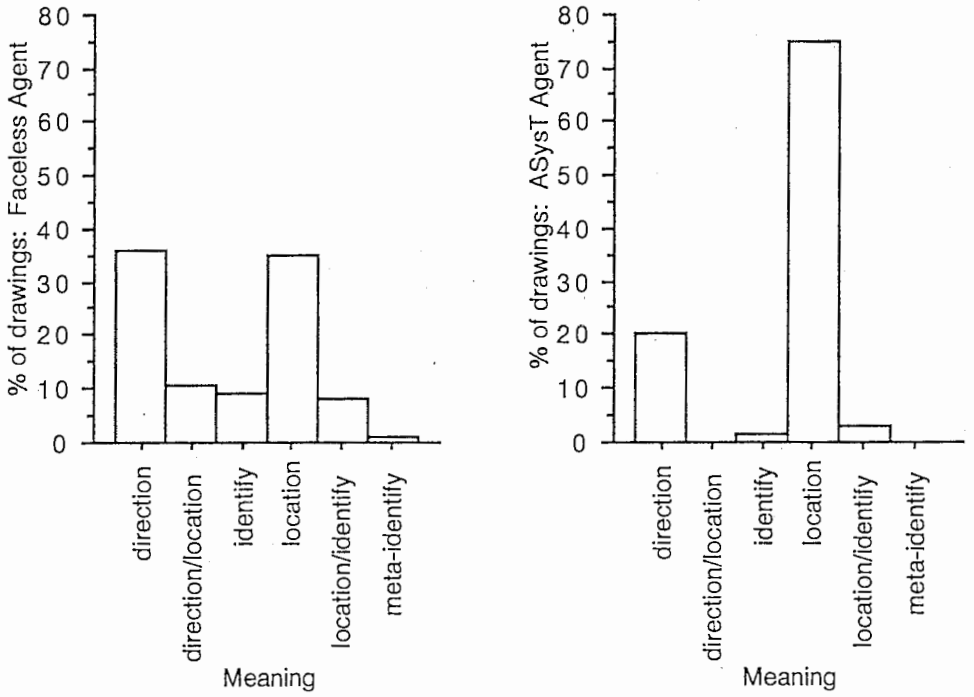


Figure 8. Distribution of drawings over meanings in each experiment, for Agents.

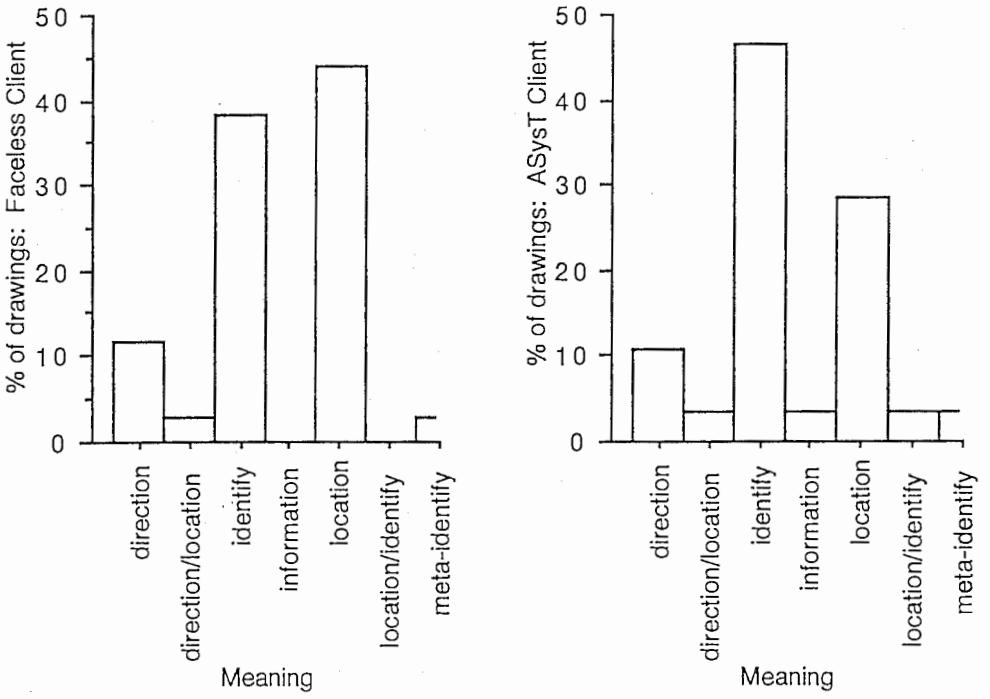


Figure 9. Distribution of drawings over meanings in each experiment, for Clients.

Deictic/Redundant. Clients in both interfaces used nearly equivalent proportions of deictic and redundant drawings: about twice as many deictic drawings as redundant ones.

Agents, on the other hand, varied widely. The Agent in the Faceless interface used, in fact, slightly *more* redundant drawings than deictic ones; the Agent in the ASysT interface used extremely few redundant drawings. Both Agent and Client in the Faceless interface also used drawings without any accompanying linguistic expression to a small degree.

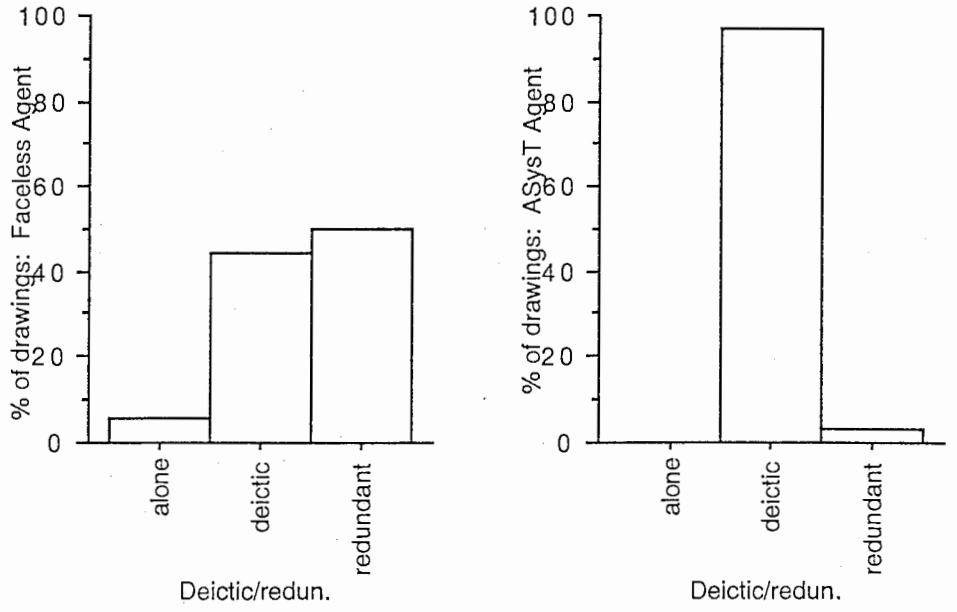


Figure 10. Distribution of drawings over deictic/redundant categories in each experiment, for Agents.

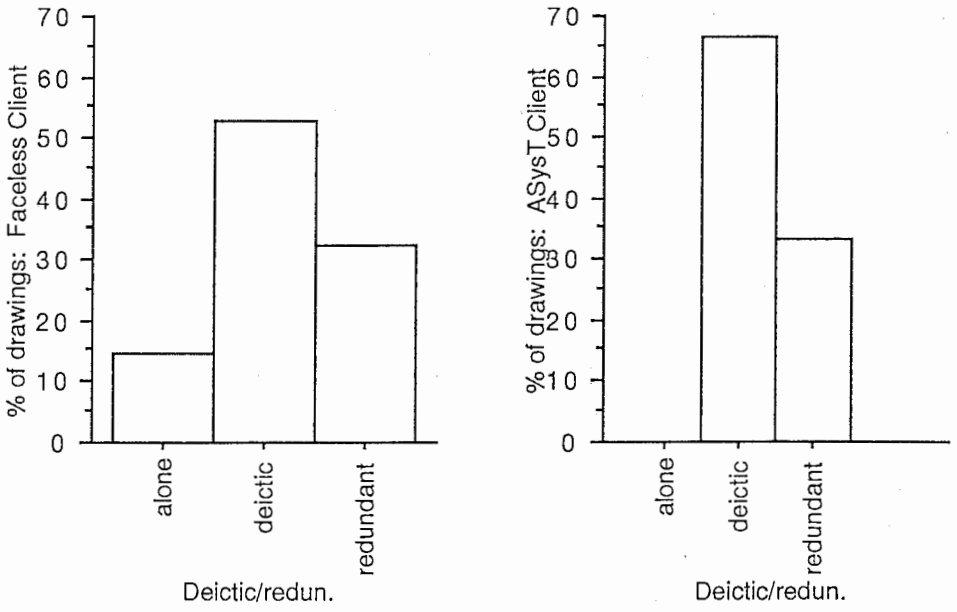


Figure 11. Distribution of drawings over deictic/redundant categories in each experiment, for Clients.

The interplay of these four factors across interface conditions and Agent/Client roles is, of course, fairly complex, but some clear correlations are apparent[9]. The use of

[9]The discussion below summarizes the main points of the data shown in Appendix C.

drawing by the Agent in the ASysT interface is the most fixed, so we will examine it first. Recall that many of the sentences used by the Agent in the ASysT interface were pre-determined. The vast majority of those sentences with which drawings could be appropriately drawn included deictics. This accounts for the high percentage of use of deictic drawings in this case. As the Agent used these sentences with different Client subjects, he established a personal pattern for the use of accompanying drawings; thus his overall performance was fairly regular. In the Directions task, the ASysT Agent associated circle and point drawing types with the meaning *location*; he associated lines with *direction*. The primary meaning conveyed with drawing in the Reservation task was *location*, accompanied primarily by point drawings. The pointing drawings were the only ones which were redundant for the ASysT Agent; most of these redundant drawings came in the Hotel task.

The Agent in the Faceless interface showed similar trends with slightly greater variety. Like the ASysT Agent, in the Directions task, he associated circles with *location* and lines with *direction*. However, the majority of this Agent's line drawings were redundant, and some circles were as well, yielding a much higher percentage of redundant drawings in this task. In the Reservation task, the Agent in the Faceless interface used circles instead of pointing to accompany *location*, but also to *identify*. Circles and *identify* were both associated with redundant drawings, which increased the percentage of those drawings in this task as well.

The Client in the ASysT interface showed drawing behavior which was essentially a subset of that of the ASysT Agent. In the Directions task, the Client primarily used lines, showing *direction*, which were deictic. In the Reservations task, the Client's role seemed to be to *identify* items which the Agent *located*. Thus, the Client used circles to *identify*; since circles could be redundant drawings for the ASysT Client, this accounts for some of the use of redundant drawings by the ASysT Client.

The Client in the Faceless condition exhibited by far the greatest variety of drawing/task/meaning/redundancy combinations. The Client used pointing in the Directions task, primarily to *locate*, but also to *identify*. He used lines primarily to *identify*, but also to show *direction*. Since these lines were frequently redundant, the Client in the Faceless condition, like the Agent, showed a certain amount of redundant drawing. In the Reservations task, the Client used circles to *locate* and *identify*, and lines to *identify* and show *direction*. Again the presence of the latter accounts for a certain percentage of redundant drawings.

5.3.2. Typing

Prompt. Oral prompts took the form of questions or requests for information, usually spoken by the Agent, or answers, usually given by the Client. A typical example of each kind of typing to an oral prompt is found in the following:

[4] Wizard: Please state phone number.
Client: OK. **[Client types phone number on the form]**
Wizard: hai
Agent: arigatou gozaimasu
Wizard: Thank you
Agent: go touchakubiwa, itsu deshoka?
Wizard: What day will you arrive?
Client: I will arrive October 24th
Wizard: juugatsu, nijuuyokkani touchaku shimasu
[Agent types arrival date on the form]

Figure 12 shows the differences between results in the ASysT interface and those in the Faceless interface.

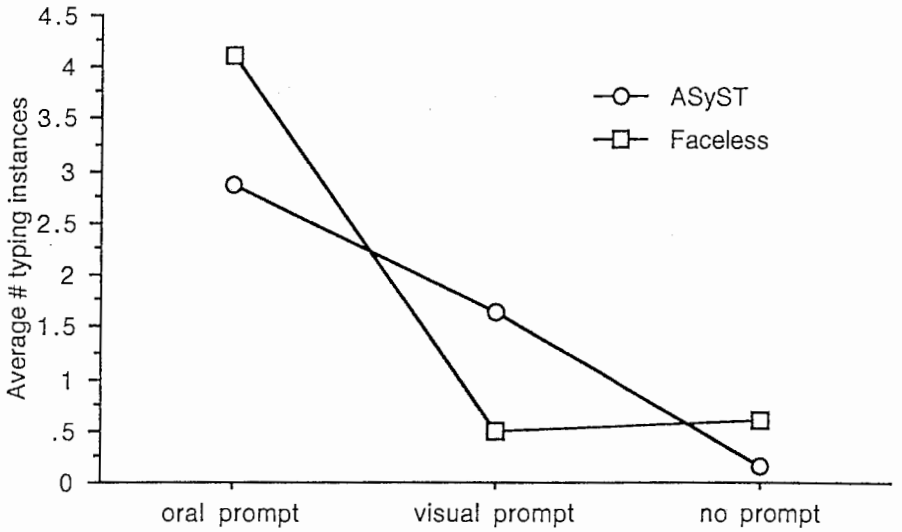


Figure 12. Number of typing instances in response to oral, visual or no prompts in each interface.

In both interfaces, there were more typing responses to oral prompts than to visual or no prompts. In the Faceless interface, typing depended much more heavily upon oral prompting than on visual prompting. In the ASysT interface, typing was motivated more frequently by oral prompts, but fairly frequently by visual prompts as well; the dependence upon oral prompts was not nearly as great as in the Faceless interface.

The results for oral prompts did not differ for Client and Agent; however, those for visual prompts and no prompt did. Figures 13 and 14 illustrate those differences.

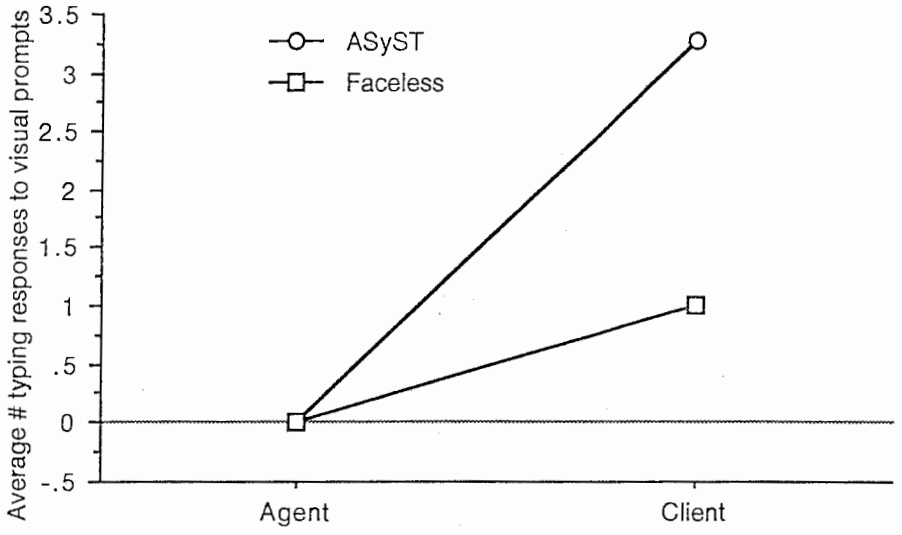


Figure 13. Number of typing responses to visual prompts in each interface, for Agent and Client.

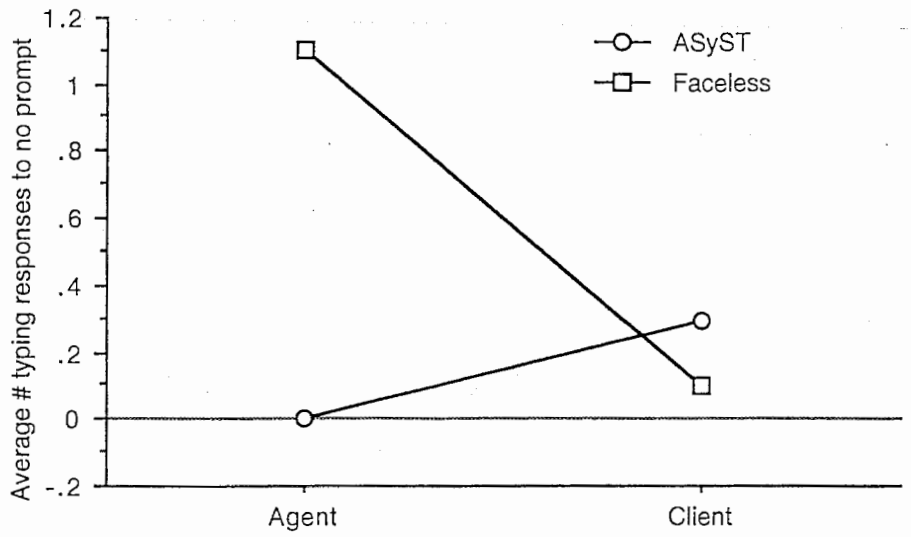


Figure 14. Number of typing responses to no prompt in each interface, for Agent and Client.

No Agents in either interface ever typed to a visual prompt, so the results shown in Figure 12 for visual prompts reflect the behavior of the Clients only. Further, the ASysT Agent never typed unprompted, though the Faceless interface Agent did, to a significantly great extent ($p < 0.006$); while neither Client typed unprompted very much. So, the results shown in Figure 12 for no prompt reflect primarily the behavior of the Agents.

Typing patterns. Table 1 lists the distributions of reservation form competition interactions according to the patterns described in section xx: Agent asks question; Client answers; Agent types (A?; CA; AT); Agent asks question; Client types answer (A?; CT); and Client sees reservation form; Client types information (CT).

Pattern	Faceless	ASysT
A?; CA; AT	50% (5)	22% (6)
A?; CT	20% (2)	22% (6)
CT	10% (1)	37% (10)
mixed	20% (2)	19% (5)

Table 1. Frequency of reservation form completion patterns for each interface.

Note that the two interfaces showed about the same percentages of subjects using mixed patterns and the type of interaction in which the Agent asked the question and the Client typed. Clients in the Faceless interface were more inclined to give their answers verbally for the Agents to type than were Clients in ASysT interface, while Clients in the ASysT interface were more inclined to take the initiative to type in information on the forms themselves than were Clients in the Faceless interface.[10]

[10]Like Oviatt and Olsen (1994), we found a small percentage of clients and agents who typed and spoke the same outut simultaneously. Oviatt and Olsen consider this sort of behavior to be the exception, but we consider it similar to the simultaneous use of drawing and lingusitic description so prevalent in the behavior of the Agent in the Faceless interface and Clients in both interfaces.

5.4. Nature of language

Subjects in the two interfaces did not differ in disfluency rate or rate of accommodation.

6. Discussion

6.1. Amount of language

A very basic contribution which an interface can make to the problem of handling spontaneous speech automatically is simply to ensure that there is less of it. Recall that, despite intuitions to the contrary, subjects in the MM condition in the Faceless experiment used *more* words to convey information than did subjects in the telephone condition.

Clients in the ASysT interface used a number of words equivalent to that used by Clients in the MM condition of the Faceless interface. Similarly, they used an equivalent percentage of meta-mode words (about 10%). Thus, the revised interface did not yield a lower number of words.

However, and more important from the point of view of communication, Clients in the ASysT interface did exchange more information than did those in the Faceless interface, and they achieved this using a lower number of words per information unit. This latter result was true for both tasks. Note that this comparison is made simply of the speech of the Client; the speech of the Wizard in the ASysT interface was so fixed that it was not a good measure of natural interaction. However, the words-per-information-unit of the ASysT Wizard was much lower than that of the Wizard in the Faceless interface; thus, had that measure been included in the calculations above (Figure 3), the difference would have been even more dramatic.

6.2. Media use

Could the reduction in the words-per-information unit in the map task have been due to greater use of visual information channels by the subjects in the ASysT interface? It seems not. Both the Client and the Wizard in the ASysT interface in fact used fewer drawings in the map task and fewer drawings per turn than in the Faceless interface. Thus it would seem that *higher* numbers of words per information unit are associated with greater use of drawing. Or, to put it another way, heavier use of drawing results in less efficient information exchange.

In order to understand this result, let us examine the results concerning redundant and deictic drawing. Recall that many of the sentences used by the Agent in the ASysT interface were pre-determined. The vast majority of those sentences with which drawings could be appropriately drawn included deictics. This accounts for the high percentage of use of deictic drawings in this case. Because the Agent's drawing was constrained by the pre-set sentences at his disposal, we will not consider this behavior to be typical of natural drawing behavior in this setting. This is confirmed by examining the frequency of use of redundant drawing by the other subjects in the two experiments. This frequency for both sets of Clients was about 33%; that for the Agent in the Faceless interface was over half. These results indicate that the use of drawings that are redundant to the message in the speech stream is a fairly natural behavior in unconstrained situations. Thus, more frequent drawing increases the likelihood that some of the drawings will be redundant, which accounts at least for why the numbers of words used is not reduced.

But why should it be increased? In previous work, we examined the role played by meta-media speech, utterances such as "I will type it for you now," or "Can you see what I just drew?" (Fais and Loken-Kim, 1996) We concluded that meta-media speech accounted for the greater number of words used in the MM condition, but did *not* account for the counter-intuitive results that (excluding meta-media speech), the efficiency rates, i.e., words per information unit) were the *same* for telephone and MM conditions.

The ASysT interface showed levels of meta-media speech similar to those for the Faceless interface, about 10%. As discussed above, subjects in the ASysT interface had lower words-per-information-unit rates than did subjects in the Faceless interface. Yet,

as in (Fais and Loken-Kim, 1996), subjects in the ASysT interface had words-per-information rates *equivalent* to those for subjects using the telephone, even when discounting the meta-media speech. Thus, despite better instruction built into the interface and despite more initiative on the part of the Agent/system, subjects still are not as efficient exchanging information using a multimedia interface as they are over the telephone.

The results for the hotel reservation task seem more straightforward and in accordance with our intuitions. Recall that subjects in the ASysT interface had fewer words per information unit than did those in the Faceless interface in this task as well. However, the reason for that is easier to see. There was less oral prompting in the ASysT interface (Figure 12) and many more instances of the form filling pattern in which the Client simply typed his/her response with no prompt. Both of these behaviors then, resulted in fewer words used per information unit exchanged in the hotel reservation task.

6.3. Nature of language

Another purpose of the revisions in the interface was to see if further reductions in disfluency rates and in the number of words used could be achieved. Given the results above (no significant differences in these areas) and the trends illustrated in the figures below, we conjecture that our results may have "bottomed out" for speakers in the context of a human-machine-human, multimedia interfaced automatic translation interaction. It may be difficult to improve these measures significantly beyond this point.

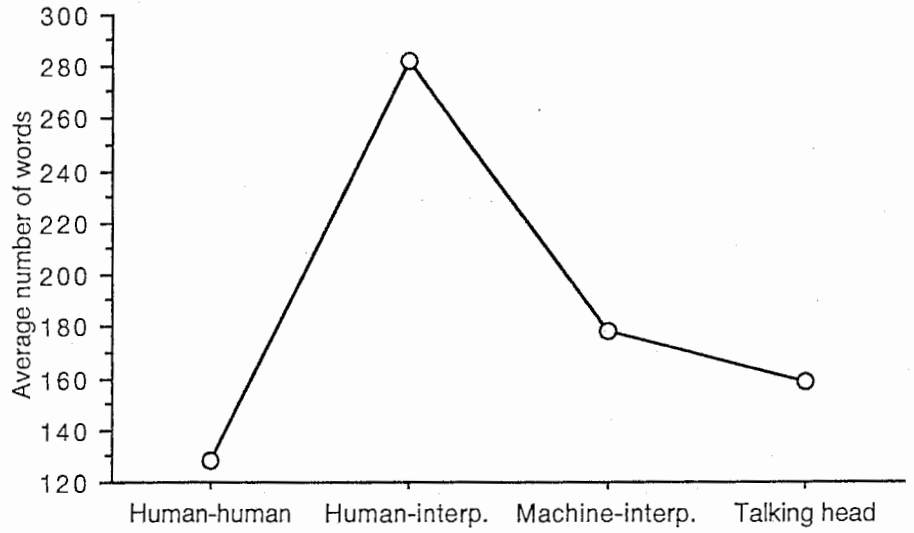


Figure 15. Average number of words per conversation for each experimental configuration.

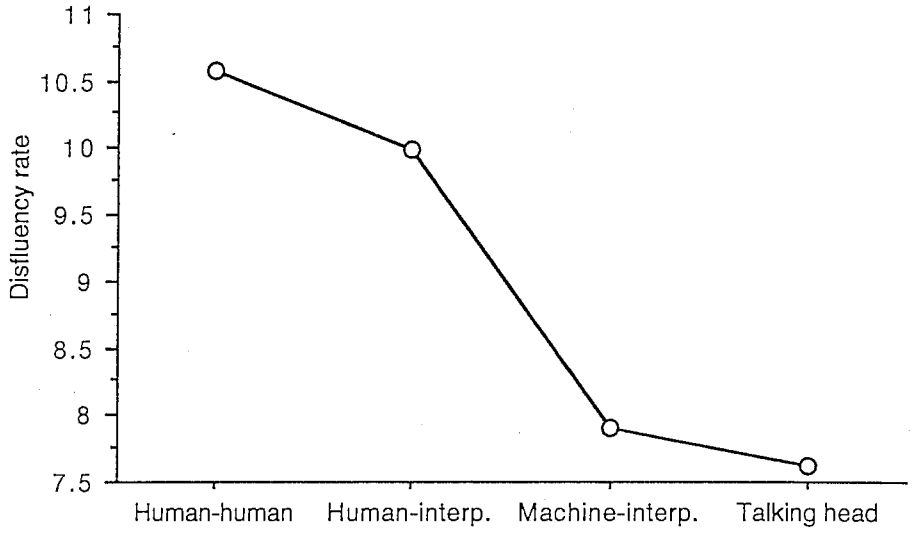


Figure 16. Disfluency rates for each experimental configuration.

However, this is not the case for accommodation. There was, in fact, less accommodation in the ASysT interface than in the Faceless interface (Figure 17). Why there should be less accommodation to an interface that includes a personification in the form of a facial image is a question that is currently under further investigation.

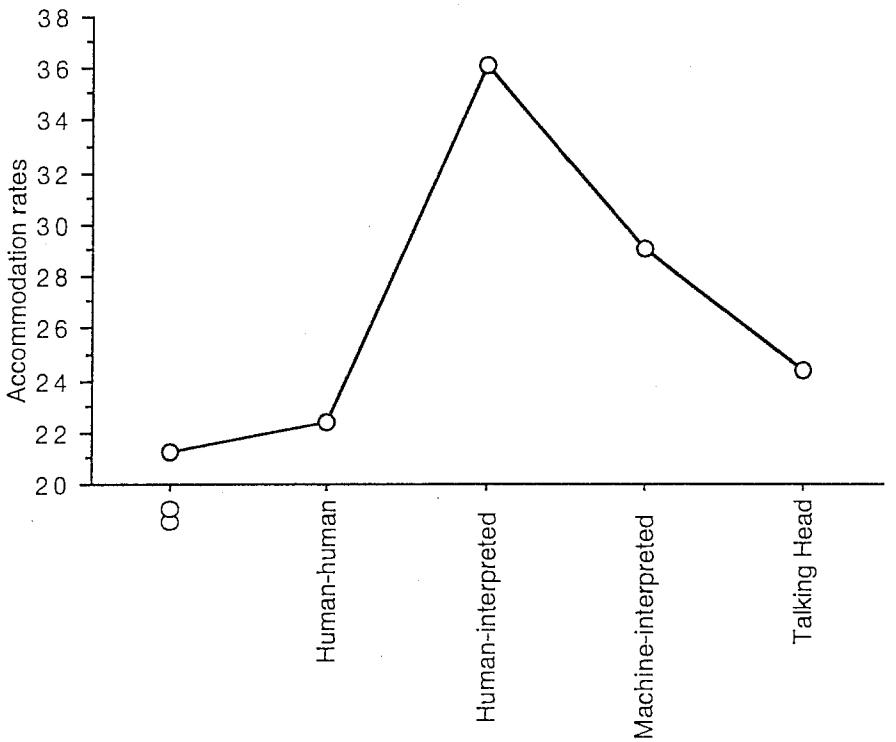


Figure 17. Accommodation rates for each experimental configuration. "CO" is the level for coincidental overlap (Fais and Loken-Kim, 1995).

7. Summary

We have seen three major results in this work. Subjects in the ASysT interface:

- subjects in the ASysT interface exchanged more information in the direction-finding task
- subjects in the ASysT interface were more efficient about exchanging information in both tasks
- subjects in the ASysT interface needed less oral prompting in order to use the typing options.

These results suggest that the revisions made to the interface made the subjects feel more confident about their ability to deal with the tasks via the multi-media interface, both because they themselves took more initiative to type and because they were able to get more directions more efficiently.

We can confirm these conjectures by examining the responses subjects made to the post-experiment questionnaire, in which they were asked their reactions to all of the aspects of the experiment (complete summaries of the questionnaire results are found in Appendix D). A majority of the subjects in the ASysT interface reported feeling "comfortable" with the system and many cited the confidence they had in receiving the information as one of the "best parts of the experiment." The multimedia aspects of the system, i.e., the map, the form, the typing option, the touchscreen, etc., were overwhelmingly rated as "very useful."

8. Bibliography

Fais, Laurel, and Kyung-ho Loken-Kim, 1994. Effects of Communicative Mode on Spontaneous English Speech. Technical Report of the Institute of Electronics, Information, and Communication Engineers, NLC94-22 (1994-10).

Fais, Laurel, and Kyung-ho Loken-Kim, 1995. Lexical Accommodation in Human-interpreted and Machine-interpreted Dual Language Interactions. Proceedings, ESCA Workshop on Spoken Dialogue Systems, Vigso, Denmark.

Fais, Laurel, and Kyung-ho Loken-Kim, 1996. How many words is a picture really worth? Proceedings, ICSLP 96 International Conference on Spoken Language Processing, Philadelphia, USA, pp. 2179-2181.

Fais, Laurel, Kyung-ho Loken-Kim, and Young-Duk Park, 1995. Speakers' Responses to Requests for Repetition in a Multimedia Language Processing Environment. Proceedings, International Conference on Cooperative Multimodal Communication, CMC/95, Eindhoven, The Netherlands.

Fais, Laurel, Suguru Mizunashi, Kyung-ho Loken-Kim, and Kazuhiko Kurihara, 1996. EMMI progress report: An evaluation of research done with the first EMMI interface. ATR Technical Report TR-IT-0172. Kyoto, Japan: ATR Interpreting Telecommunications Research Laboratories.

Fais, Laurel, submitted. Lexical Accommodation in Human- and Machine-Interpreted Dialogues. *International Journal of Human-Computer Studies.*

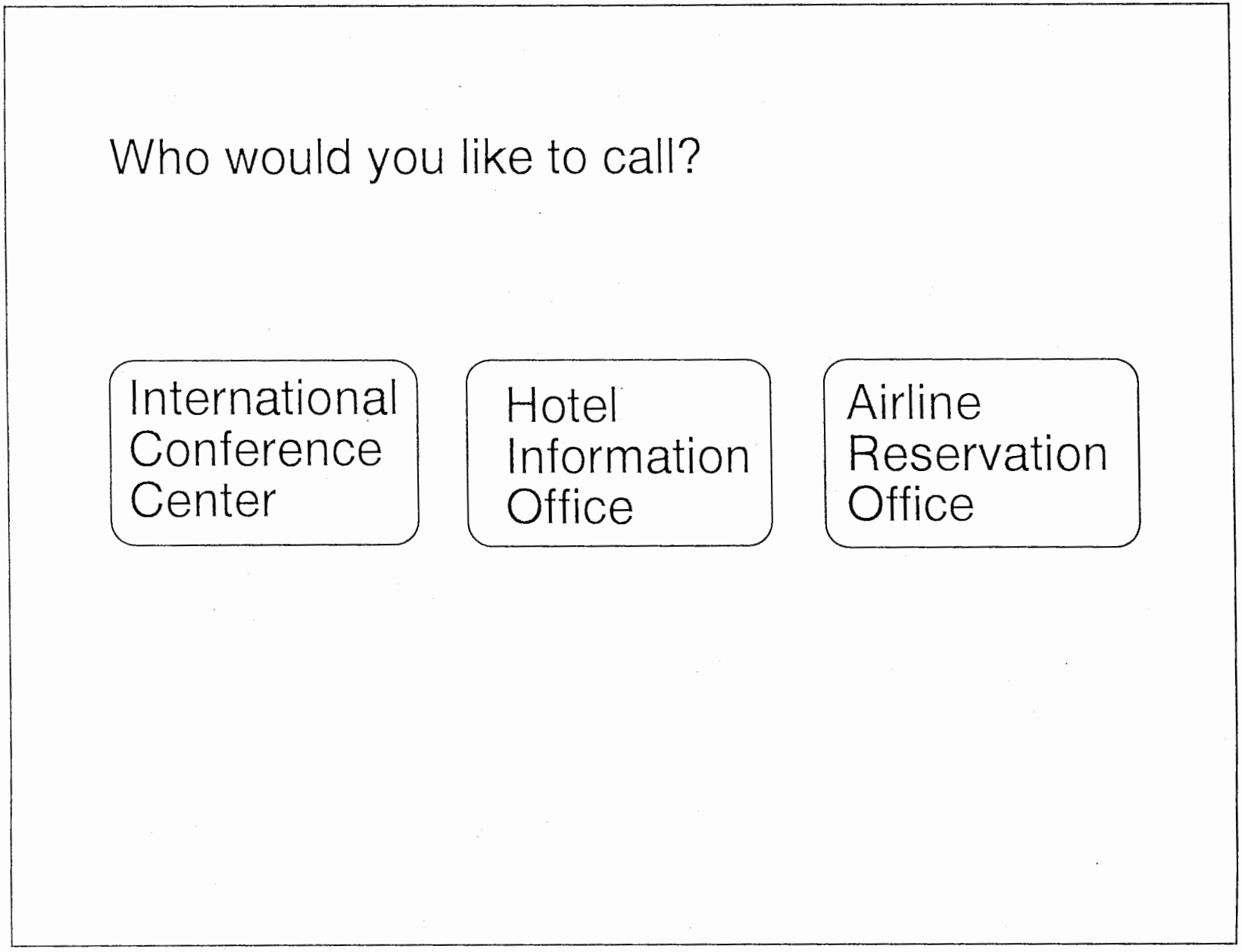
Loken-Kim, Kyung-ho, Suguru Mizunashi, Mutsuko Tomokiyo, Laurel Fais and Tsuyoshi Morimoto, 1995. Analysis and integration of multimodal inputs in interpreting telecommunications. IPSJ Workshop.

Park, Young-Duk, Kyung-ho Loken-Kim, and Laurel Fais, 1994. An experiment for telephone versus multimedia multimodal interpretation: Methods and subject behavior. Technical Report of the Interpreting Telecommunications Research Laboratory, ATR, TR-IT-0087. Kyoto, Japan.

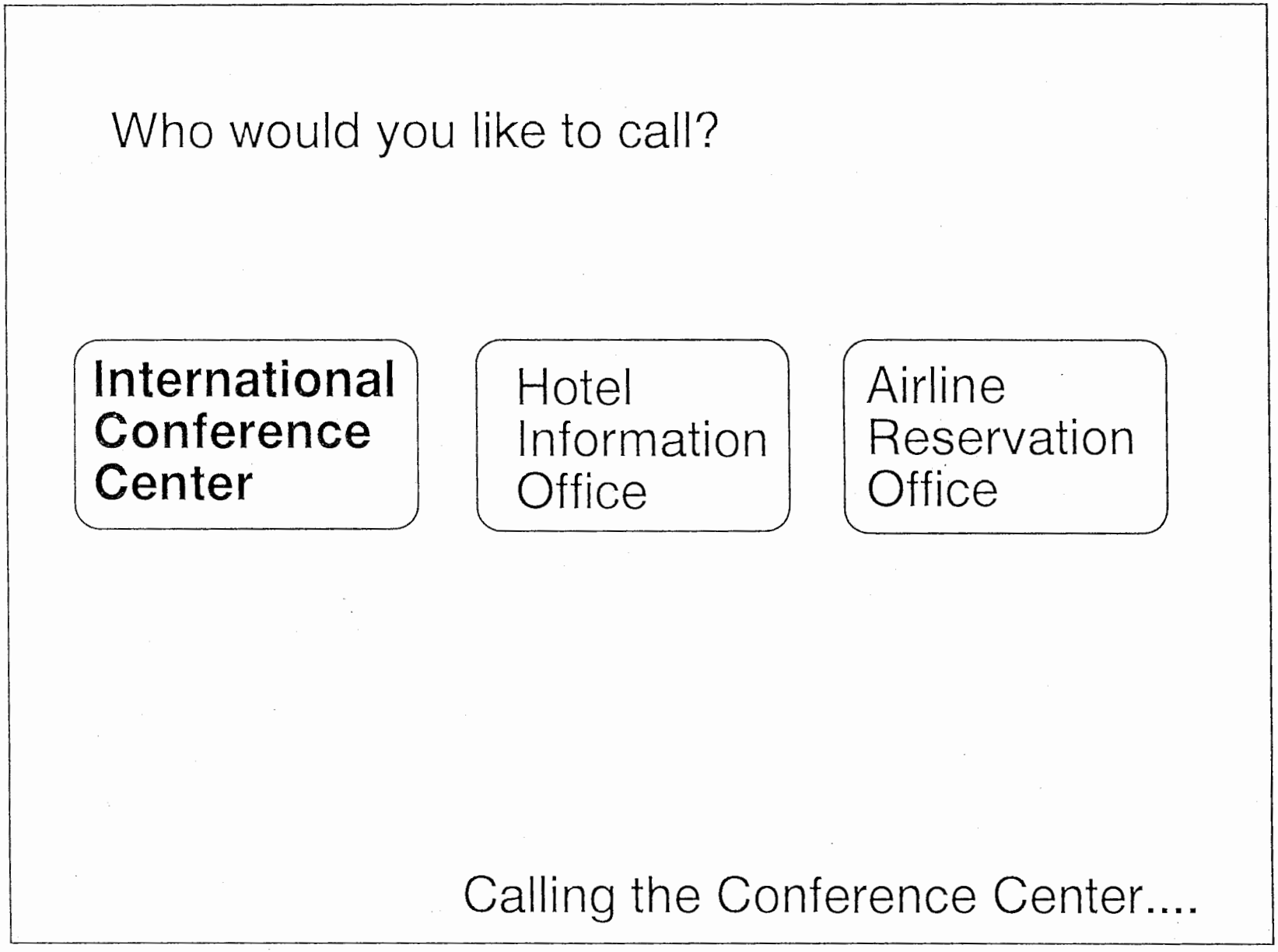
Park, Young-Duk, Kyung-ho Loken-Kim, Laurel Fais, and Suguru Mizunashi, 1995. Analysis of gestures behavior in a multimedia interpreting environment. ATR Technical Report TR-IT-0091. Kyoto, Japan: ATR Interpreting Telecommunications Research Laboratories.

Appendix A: Monitor screens for the Client interface in ASysT.

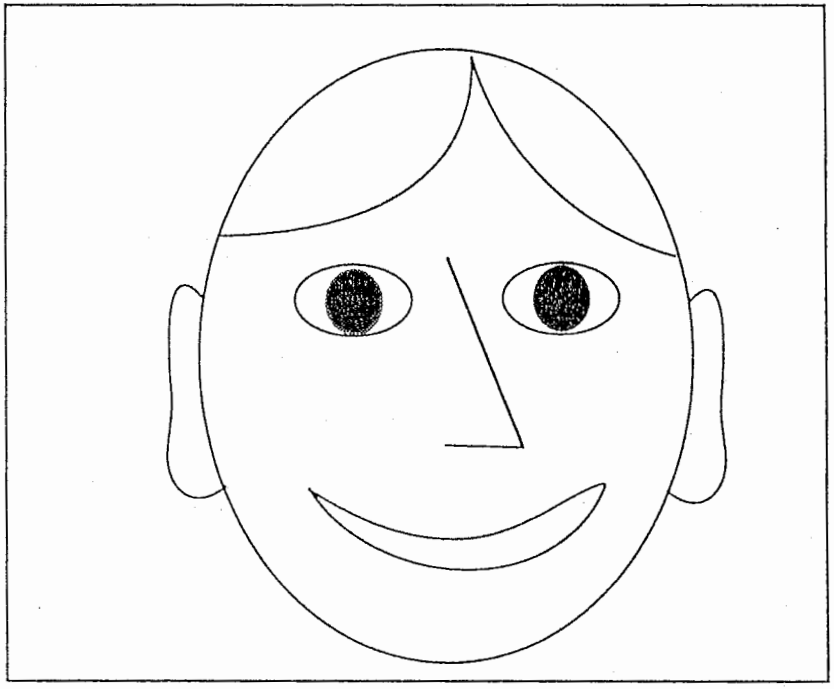
Screen 1. This screen is what the Client sees first.



Screen 2. After pushing "English" on the previous screen, this screen appears.



Screen 3. After the Client pushes the "International Conference Center" button.



Hello, this is the Conference Center. I am your computer ASysTant®.
I will translate your conversation. Please speak slowly and clearly.

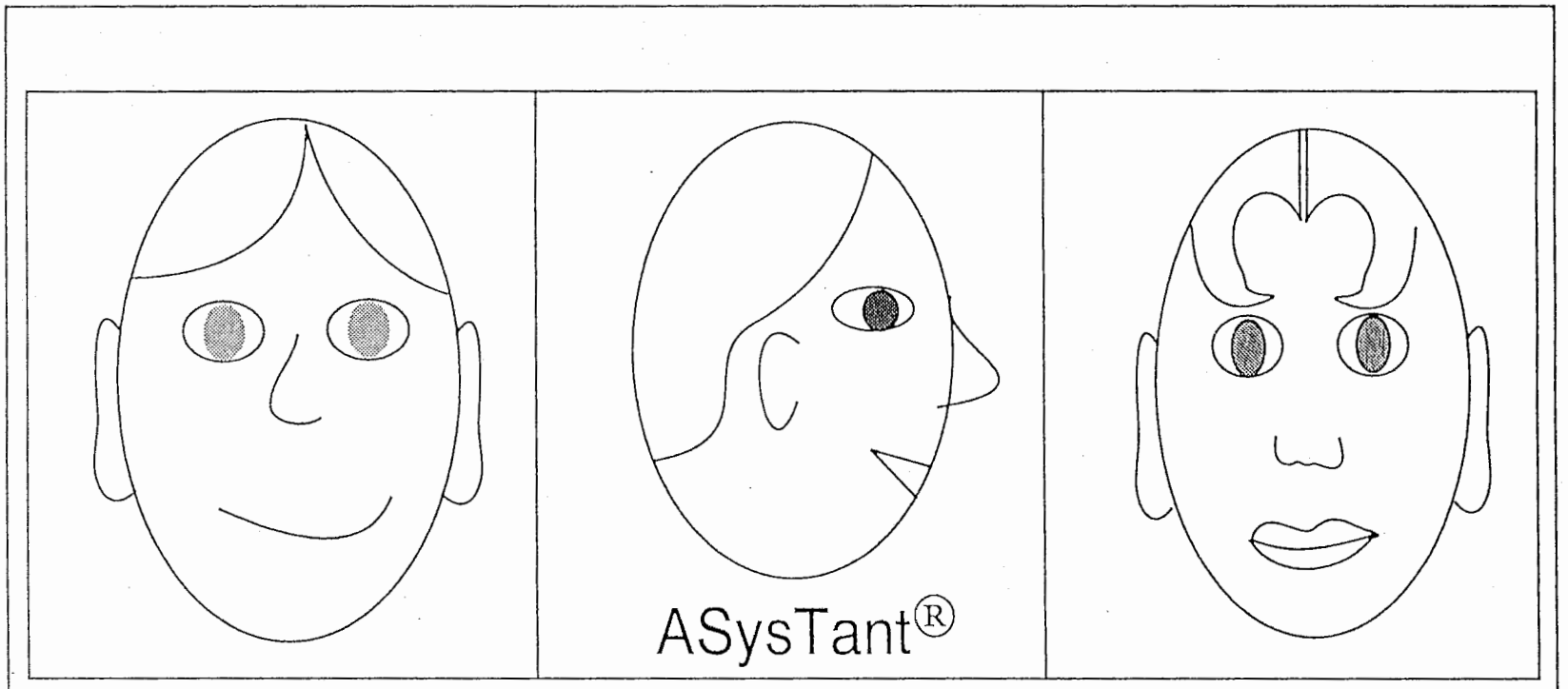
If you want me to stop translating a sentence, please push Interrupt

If you want me to repeat a translation, please push Repeat

If you want me to rephrase a translation, please push Rephrase

Now, here is the conference agent...

Screen 4. This is also "spoken" by CHATR.



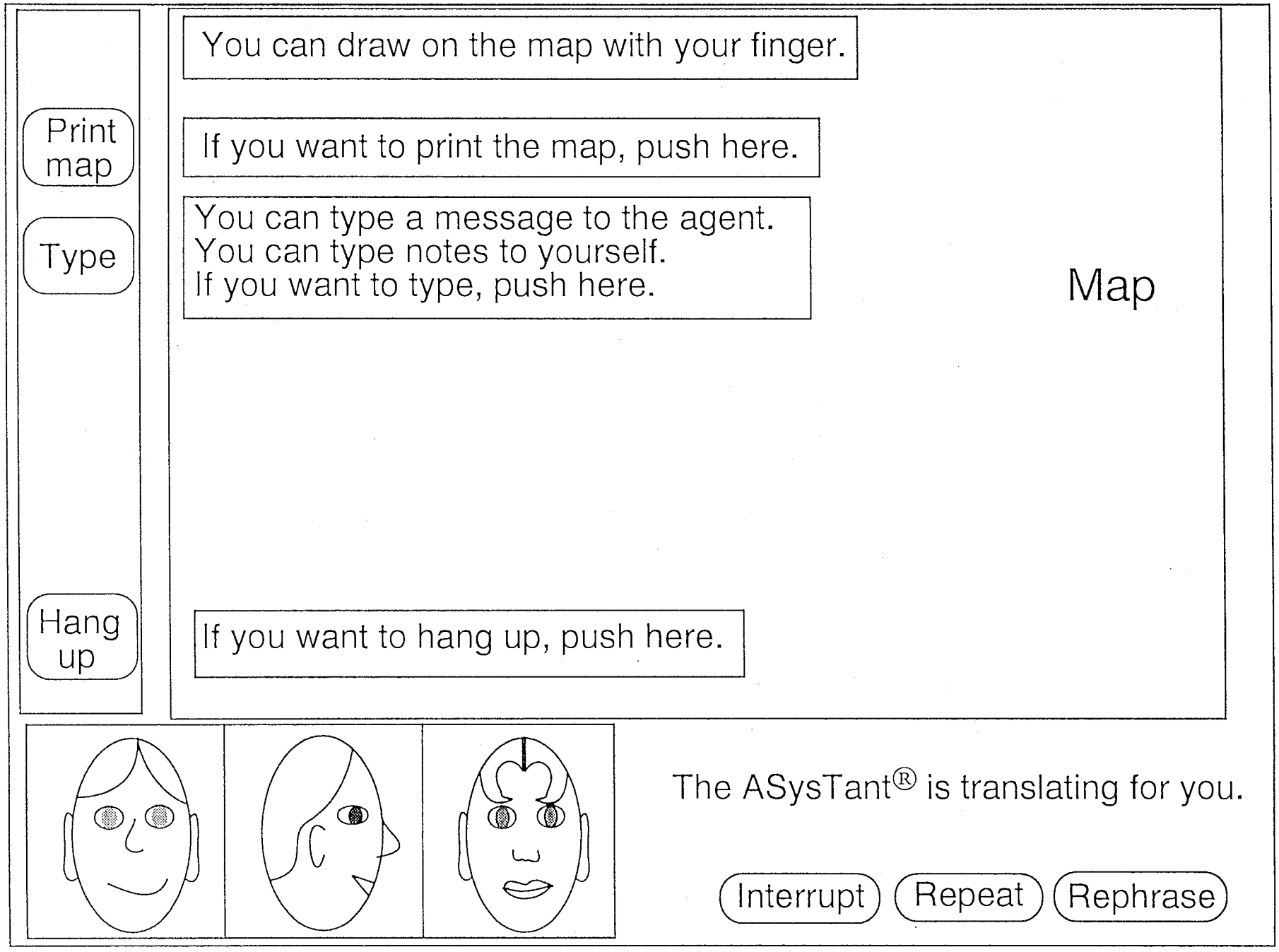
This is the
conference agent.

How can we
help you?

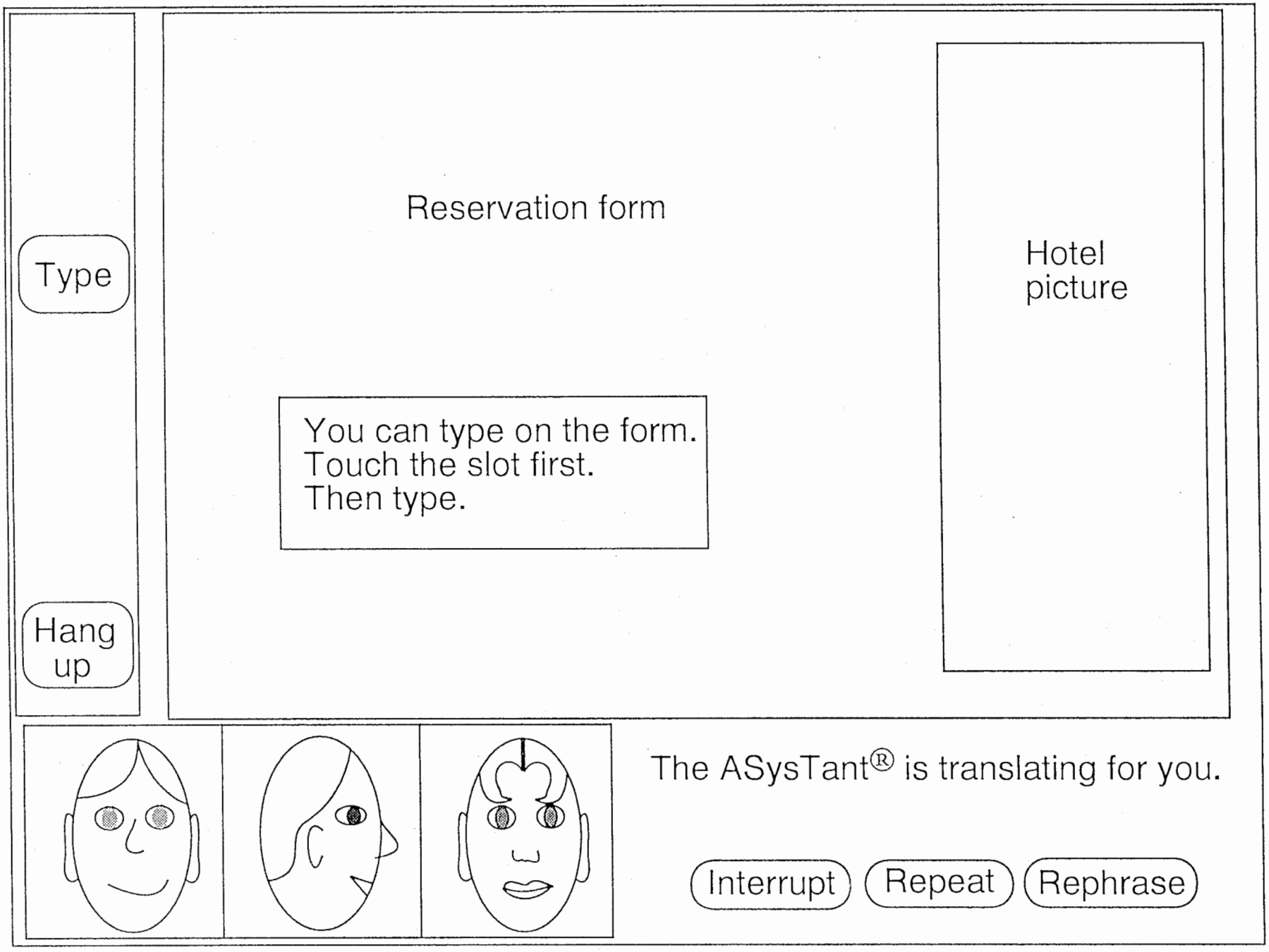
The ASysTant® is translating for you.

Interrupt Repeat Rephrase

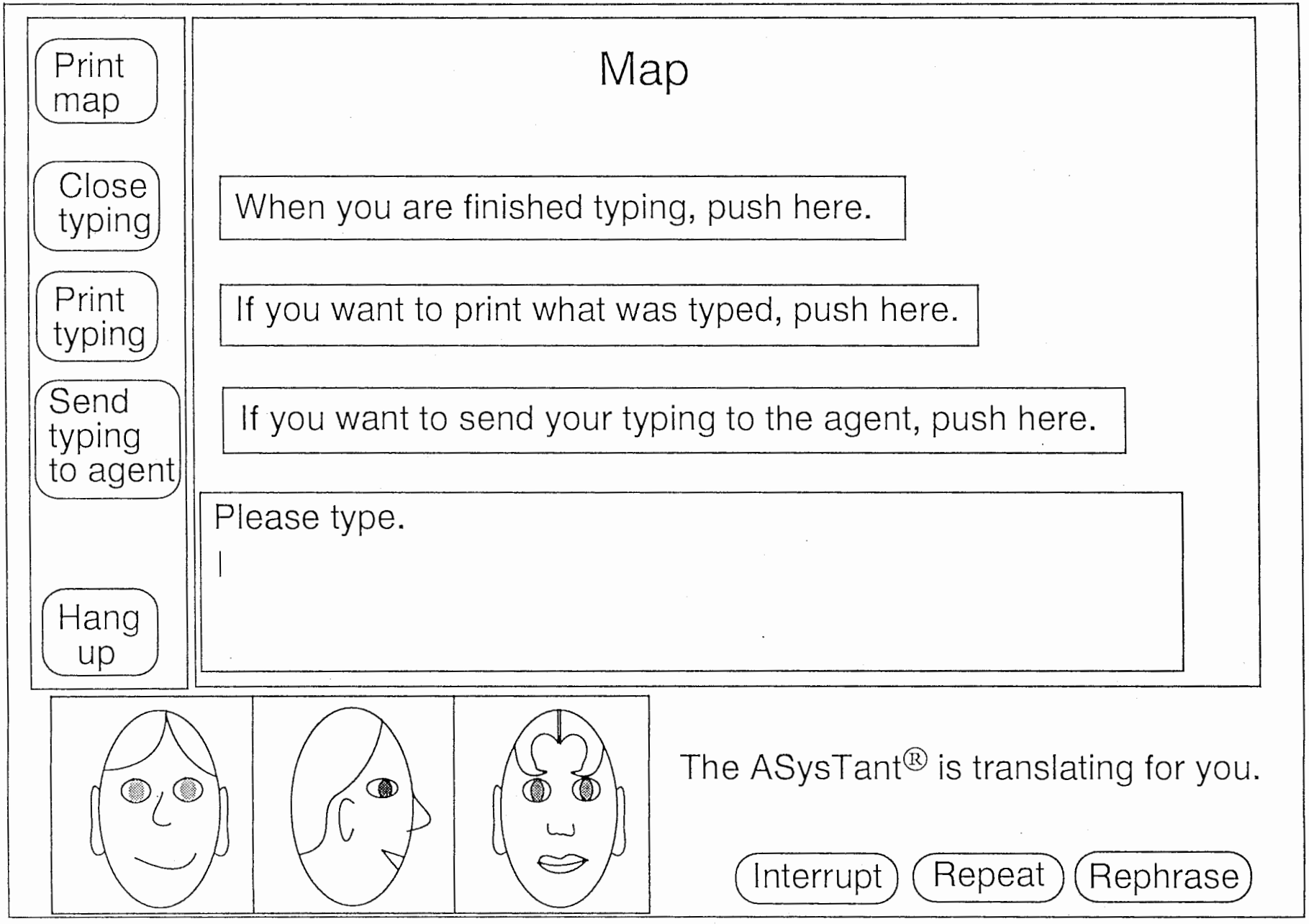
Screen 5. The video image on the right is the Client.



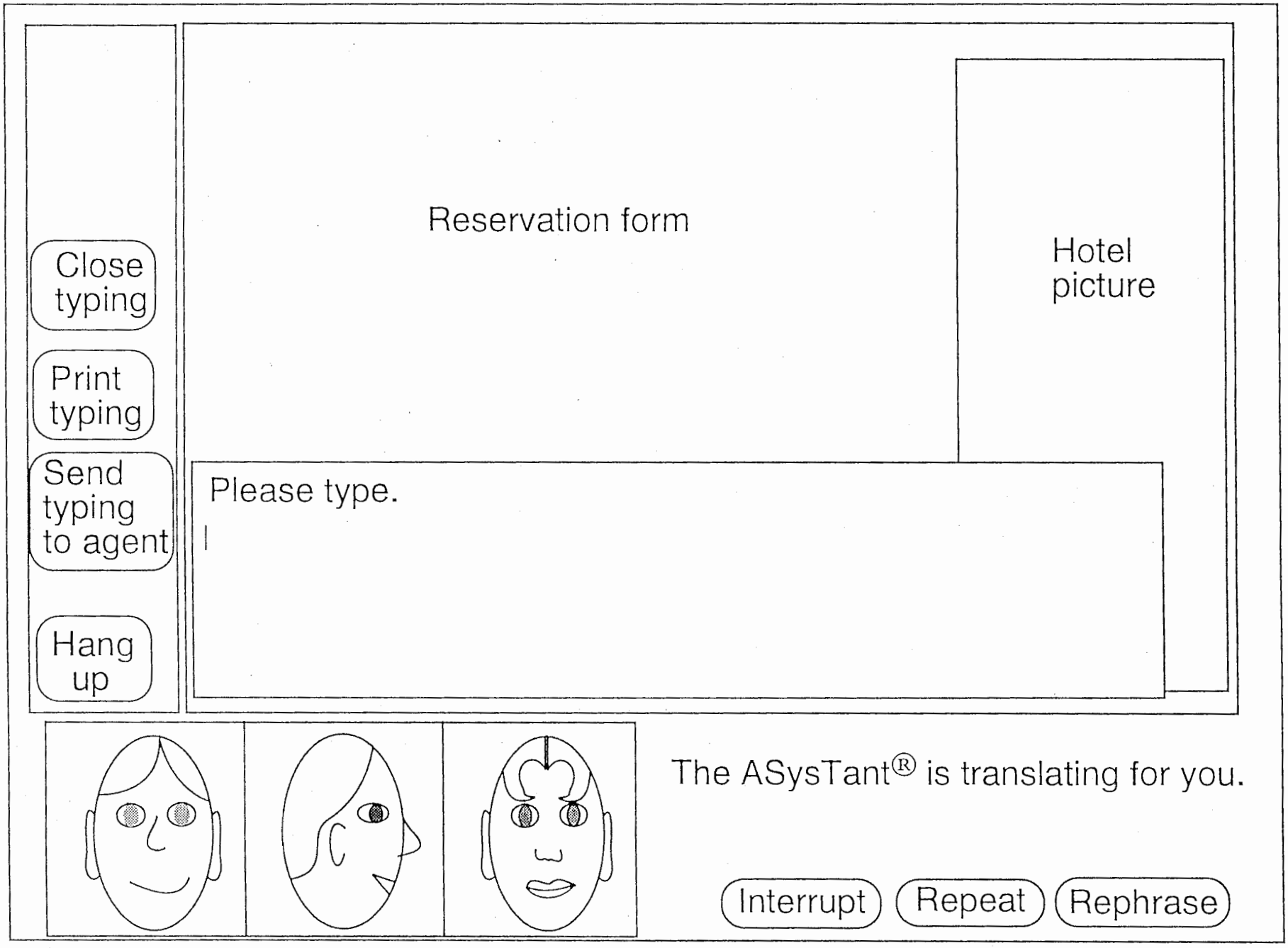
Screen 6. These instructions are shown (and "spoken" by CHATR) the first time the map appears.



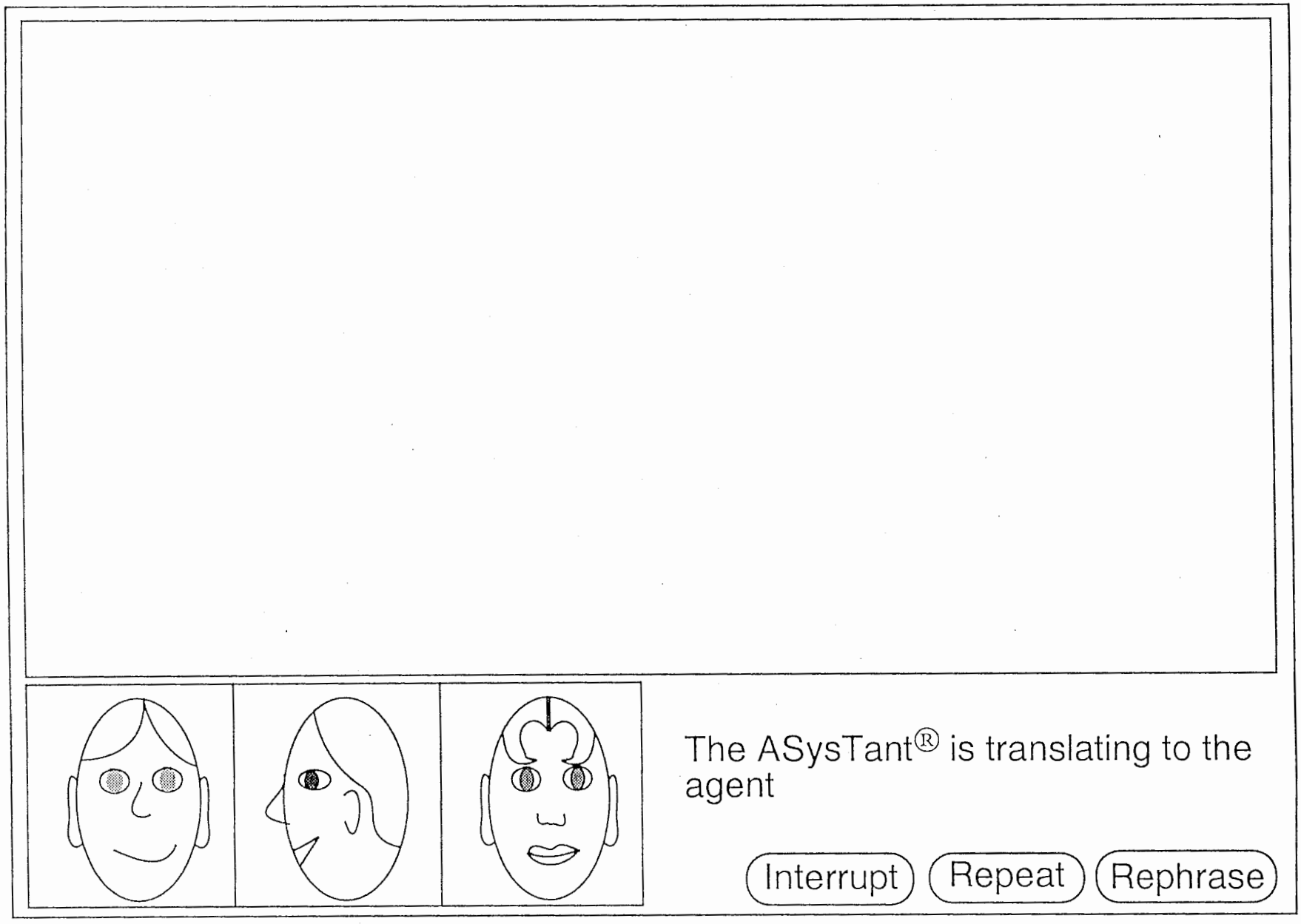
Screen 7. These instruction are shown (and "spoken" by CHATR) the first time the reservation form appears.



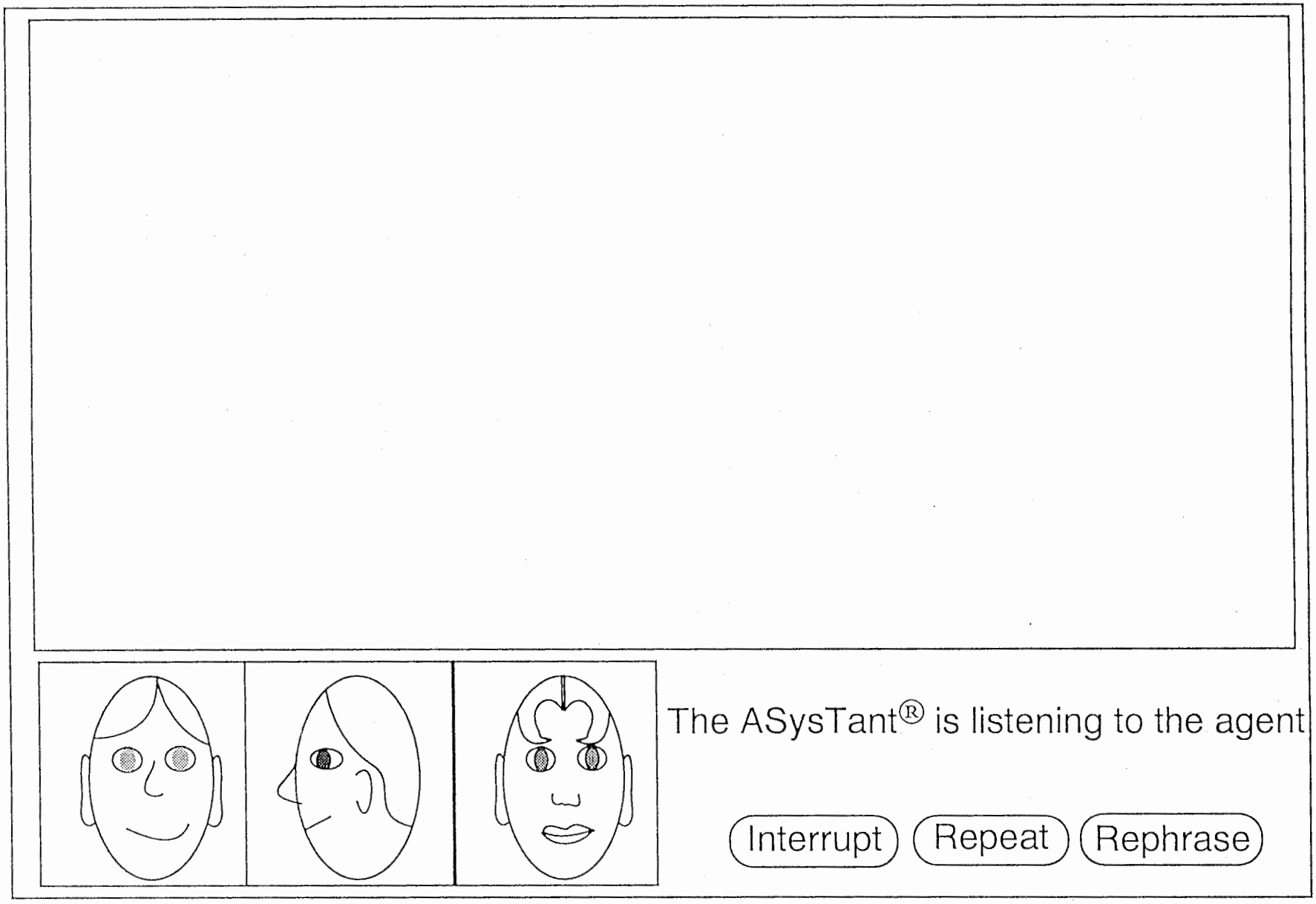
Screen 8. These instructions are shown (and "spoken" by CHATR) the first time the typing box is used.



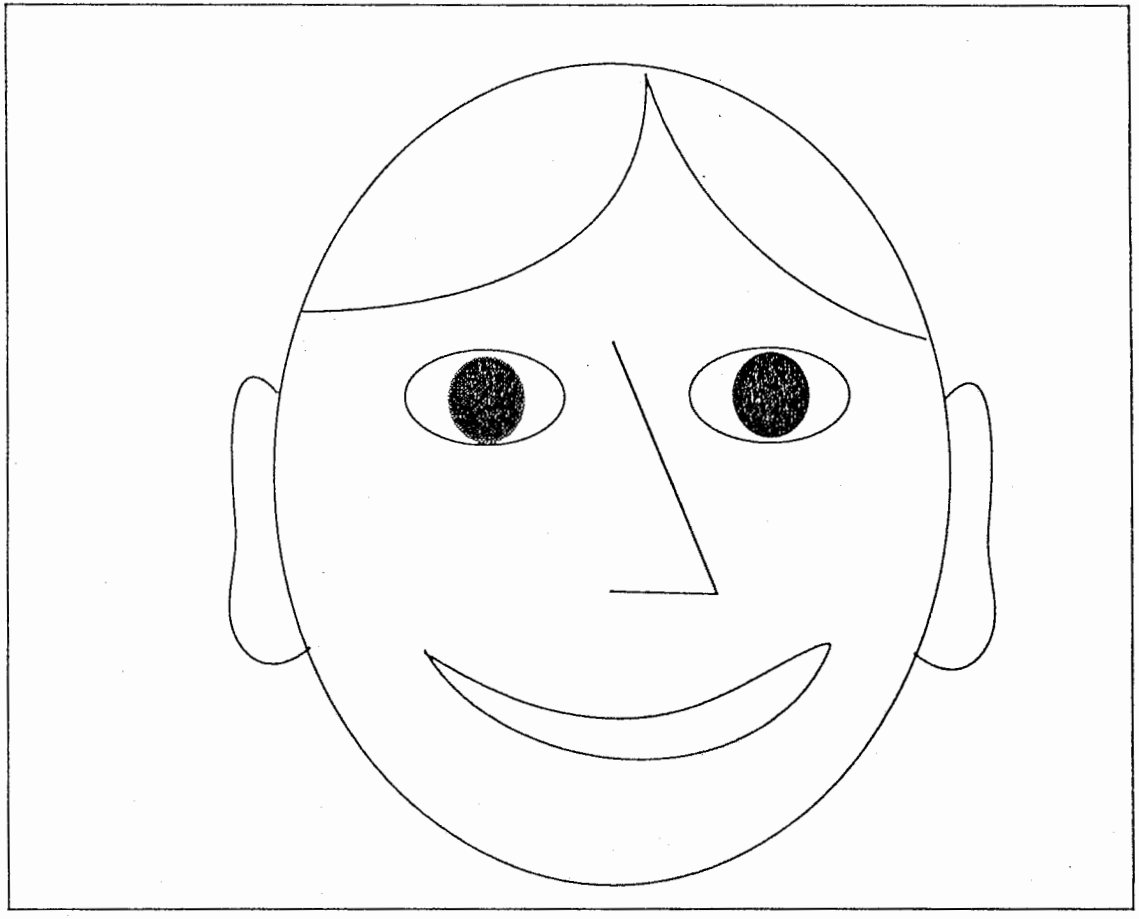
Screen 9. The typing box may be used on top of the reservation form as well.



Screen 10. When "ASysT" "translates" for the Agent, the face is turned toward the Agent with its mouth open. Likewise, when it "translates" for the Client, the face is turned toward the Client with its mouth open. (See previous screens.)



Screen 11. When "ASysT" "listens" to the Agent, the face is turned toward the Agent with its mouth closed. Likewise, when it "listens" to the Client, the face is turned toward the Client with its mouth closed.



Thank you for using ***ASysT®***

Screen 12. When the Client pushes the "Hang up" button, this screen appears.

Appendix B: Pre-set sentences for the Wizard

Hello. This is the International Conference Center. How can we help you?
You can get to the Conference Center by bus or taxi. Which would you prefer?

Bus
If you take the bus, it will cost two hundred yen.
The bus takes about thirty minutes.
From Kyoto Station, take bus number twenty three to Keage Station.
You can find the bus here.
The bus stop is here.

Taxi
If you take the taxi, it will cost about two thousand yen.
The taxi takes about fifteen minutes.
You can find the taxi here.
Please tell the taxi driver you want to go to the kokusai koryuu senta.

Map/Kyoto Station
Please look at the map.
I will show you a map.
Where are you inside Kyoto Station?

After bus/taxi
The Conference Center is here.
Please walk this way.
When you get off the train, please walk this way.
Please walk from Keage Station.
When you get off the bus, please walk this way.
The Conference Center will be a large grey building here.

Hotel
What kind of hotel would you like?
The hotel closest to the conference center is the X.
The X Hotel is here.

Reservation form
Please look at this form.
Please type your name and phone number in the slots on the form.
Please type your daytime phone number in the slot.
On what day will you begin your stay?
On what day will you leave the hotel?
When will you arrive?
When will you depart?
How many people are in your party?
How many adults and how many children are in your party?
Would you like a single room or a twin room?
A single room has one bed.
A twin room has two beds.
The price for a single room per person per night at the X Hotel is X yen.
The total cost for your stay will be X.
Would you like breakfast only or breakfast and dinner?

Meta-conversation
Please repeat.
Is that acceptable?
Is this correct?
Please wait a moment.

I will type it for you.

Ending
Is there anything else we can help you with?
Thank you for calling the International Conference Office. Have a pleasant visit.

Appendix C: Correlations among the factors of Task, Meaning, Drawing Type and Redundancy

For Agents:

Percents of Row Totals for Task, Meaning
Split By: EMMI
Cell: Asyst

	direc...	direction/loc...	iden...	loca...	location/ide...	meta-ide...	Totals
Direction	26.891	0.000	0.000	7.3E1	0.000	0.000	100
Other	16.667	0.000	0.000	8.3E1	0.000	0.000	100
Reserva...	4.651	0.000	6.977	7.7E1	11.628	0.000	100
Totals	20.690	0.000	1.724	7.5E1	2.874	0.000	100

Percents of Column Totals for Task, Meaning
Split By: EMMI
Cell: Asyst

	direc...	direction/loc...	identify	location	location/ide...	meta-ide...	Totals
Direction	88.889	•	0.000	66.923	0.000	•	68.391
Other	5.556	•	0.000	7.692	0.000	•	6.897
Reserva...	5.556	•	100	25.385	100.000	•	24.713
Totals	100	•	100	100	100.000	•	100

The total for one or more columns was zero.

Percents of Row Totals for Task, Meaning
Split By: EMMI
Cell: Faceless

	direc...	direction/loc...	iden...	loca...	location/ide...	meta-ide...	Totals
Direction	40.260	11.688	5.195	3.4E1	7.792	1.299	100
Other	•	•	•	•	•	•	•
Reserva...	0.000	0.000	40	50	10.000	0.000	100
Totals	35.632	10.345	9.195	3.6E1	8.046	1.149	100

The total for one or more rows was zero.

Percents of Column Totals for Task, Meaning
Split By: EMMI
Cell: Faceless

	direc...	direction/loc...	identify	location	location/ide...	meta-ide...	Totals
Direction	100	100.000	50.000	83.871	85.714	100.000	88.506
Other	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Reserva...	0.000	0.000	50.000	16.129	14.286	0.000	11.494
Totals	100	100.000	100	100	100.000	100.000	100

Percents of Row Totals for Task, Deictic/redun.
Split By: EMMI
Cell: Asyst

	alone	deictic	redundant	Totals
Direction	0.000	100.000	0.000	100.000
Other	0.000	91.667	8.333	100.000
Reservation	0.000	88.372	11.628	100.000
Totals	0.000	96.552	3.448	100.000

Percents of Column Totals for Task, Deictic/redun.
Split By: EMMI
Cell: Asyst

	alone	deictic	redundant	Totals
Direction	·	70.833	0.000	68.391
Other	·	6.548	16.667	6.897
Reservation	·	22.619	83.333	24.713
Totals	·	100.000	100.000	100.000

The total for one or more columns was zero.

Percents of Row Totals for Task, Deictic/redun.
Split By: EMMI
Cell: Faceless

	alone	deictic	redundant	Totals
Direction	6.494	46.753	46.753	100.000
Other	·	·	·	·
Reservation	0.000	30.000	70.000	100.000
Totals	5.747	44.828	49.425	100.000

The total for one or more rows was zero.

Percents of Column Totals for Task, Deictic/redun.
Split By: EMMI
Cell: Faceless

	alone	deictic	redundant	Totals
Direction	100.000	92.308	83.721	88.506
Other	0.000	0.000	0.000	0.000
Reservation	0.000	7.692	16.279	11.494
Totals	100.000	100.000	100.000	100.000

Percents of Row Totals for Drawing type, Task
Split By: EMMI
Cell: Asyst

	Direction	Other	Reservation	Totals
circle	89.189	5.405	5.405	100.000
line	90.909	4.545	4.545	100.000
line + arrow	100.000	0.000	0.000	100.000
point	48.913	8.696	42.391	100.000
Totals	68.391	6.897	24.713	100.000

Percents of Column Totals for Drawing type, Task
Split By: EMMI
Cell: Asyst

	Direction	Other	Reservation	Totals
circle	27.731	16.667	4.651	21.264
line	33.613	16.667	4.651	25.287
line + arrow	.840	0.000	0.000	.575
point	37.815	66.667	90.698	52.874
Totals	100.000	100.000	100.000	100.000

Percents of Row Totals for Drawing type, Task
Split By: EMMI
Cell: Faceless

	Direction	Other	Reservation	Totals
circle	76.190	0.000	23.810	100.000
line	100.000	0.000	0.000	100.000
line + arrow	100.000	0.000	0.000	100.000
point	100.000	0.000	0.000	100.000
Totals	88.506	0.000	11.494	100.000

Percents of Column Totals for Drawing type, Task
Split By: EMMI
Cell: Faceless

	Direction	Other	Reservation	Totals
circle	41.558	·	100.000	48.276
line	55.844	·	0.000	49.425
line + arrow	1.299	·	0.000	1.149
point	1.299	·	0.000	1.149
Totals	100.000	·	100.000	100.000

The total for one or more columns was zero.

Percents of Row Totals for Drawing type, Meaning
Split By: EMMI
Cell: Asyst

	direc...	direction/loc...	iden...	location	location/ide...	meta-ide...	Totals
circle	0.000	0.000	0.000	100	0.000	0.000	100
line	79.545	0.000	0.000	20.455	0.000	0.000	100
line + ar...	100	0.000	0.000	0.000	0.000	0.000	100
point	0.000	0.000	3.261	91.304	5.435	0.000	100
Totals	20.690	0.000	1.724	74.713	2.874	0.000	100

Percents of Column Totals for Drawing type, Meaning
Split By: EMMI
Cell: Asyst

	direc...	direction/loc...	identify	location	location/ide...	meta-ide...	Totals
circle	0.000	•	0.000	28.462	0.000	•	21.264
line	97.222	•	0.000	6.923	0.000	•	25.287
line + ar...	2.778	•	0.000	0.000	0.000	•	.575
point	0.000	•	100	64.615	100.000	•	52.874
Totals	100	•	100	100	100.000	•	100

The total for one or more columns was zero.

Percents of Row Totals for Drawing type, Meaning
Split By: EMMI
Cell: Faceless

	direc...	direction/loc...	iden...	location	location/ide...	meta-ide...	Totals
circle	0.000	0.000	1.9E1	66.667	11.905	2.381	100
line	69.767	20.930	0.000	4.651	4.651	0.000	100
line + ar...	0.000	0.000	0.000	100	0.000	0.000	100
point	100	0.000	0.000	0.000	0.000	0.000	100
Totals	35.632	10.345	9.195	35.632	8.046	1.149	100

Percents of Column Totals for Drawing type, Meaning
Split By: EMMI
Cell: Faceless

	direc...	direction/loc...	identify	location	location/ide...	meta-ide...	Totals
circle	0.000	0	100	90.323	71.429	100.000	48.276
line	96.774	100.000	0.000	6.452	28.571	0.000	49.425
line + ar...	0.000	0.000	0.000	3.226	0.000	0.000	1.149
point	3.226	0.000	0.000	0.000	0.000	0.000	1.149
Totals	100	100.000	100	100	100.000	100.000	100

Percents of Row Totals for Drawing type, Deictic/redun.
Split By: EMMI
Cell: Asyst

	alone	deictic	redundant	Totals
circle	0.000	100.000	0.000	100.000
line	0.000	100.000	0.000	100.000
line + arrow	0.000	100.000	0.000	100.000
point	0.000	93.478	6.522	100.000
Totals	0.000	96.552	3.448	100.000

Percents of Column Totals for Drawing type, Deictic/redun.
Split By: EMMI
Cell: Asyst

	alone	deictic	redundant	Totals
circle	•	22.024	0.000	21.264
line	•	26.190	0.000	25.287
line + arrow	•	.595	0.000	.575
point	•	51.190	100.000	52.874
Totals	•	100.000	100.000	100.000

The total for one or more columns was zero.

Percents of Row Totals for Drawing type, Deictic/redun.
Split By: EMMI
Cell: Faceless

	alone	deictic	redundant	Totals
circle	0.000	59.524	40.476	100.000
line	11.628	30.233	58.140	100.000
line + arrow	0.000	100.000	0.000	100.000
point	0.000	0.000	100.000	100.000
Totals	5.747	44.828	49.425	100.000

Percents of Column Totals for Drawing type, Deictic/redun.
Split By: EMMI
Cell: Faceless

	alone	deictic	redundant	Totals
circle	0.000	64.103	39.535	48.276
line	100.000	33.333	58.140	49.425
line + arrow	0.000	2.564	0.000	1.149
point	0.000	0.000	2.326	1.149
Totals	100.000	100.000	100.000	100.000

Percents of Row Totals for Meaning, Deictic/redun.
Split By: EMMI
Cell: Asyst

	alone	deictic	redundant	Totals
direction	0.000	100.000	0.000	100.000
direction/location	•	•	•	•
identify	0.000	66.667	33.333	100.000
location	0.000	99.231	.769	100.000
location/identify	0.000	20.000	80.000	100.000
meta-identify	•	•	•	•
Totals	0.000	96.552	3.448	100.000

The total for one or more rows was zero.

Percents of Column Totals for Meaning, Deictic/redun.
Split By: EMMI
Cell: Asyst

	alone	deictic	redundant	Totals
direction	•	21.429	0.000	20.690
direction/location	•	0.000	0.000	0.000
identify	•	1.190	16.667	1.724
location	•	76.786	16.667	74.713
location/identify	•	.595	66.667	2.874
meta-identify	•	0.000	0.000	0.000
Totals	•	100.000	100.000	100.000

The total for one or more columns was zero.

Percents of Row Totals for Meaning, Deictic/redun.
Split By: EMMI
Cell: Faceless

	alone	deictic	redundant	Totals
direction	16.129	19.355	64.516	100.000
direction/location	0.000	44.444	55.556	100.000
identify	0.000	0.000	100.000	100.000
location	0.000	70.968	29.032	100.000
location/identify	0.000	85.714	14.286	100.000
meta-identify	0.000	100.000	0.000	100.000
Totals	5.747	44.828	49.425	100.000

Percents of Column Totals for Meaning, Deictic/redun.
Split By: EMMI
Cell: Faceless

	alone	deictic	redundant	Totals
direction	100.000	15.385	46.512	35.632
direction/location	0.000	10.256	11.628	10.345
identify	0.000	0.000	18.605	9.195
location	0.000	56.410	20.930	35.632
location/identify	0.000	15.385	2.326	8.046
meta-identify	0.000	2.564	0.000	1.149
Totals	100.000	100.000	100.000	100.000

For Clients:

Percents of Row Totals for Task, Meaning
Split By: EMMI
Cell: Asyst

	dire...	direction/lo...	ide...	inform...	loc...	location/id...	meta-id...	Totals
Direction	1.5E1	4.878	3E1	2.439	5E1	0.000	4.878	100
Other	0.000	0.000	40	0.000	60	0.000	0.000	100
Reser...	6.250	0.000	8E1	0.000	6.25	6.250	0.000	100
Totals	11.29	3.226	4E1	1.613	4E1	1.613	3.226	100

Percents of Column Totals for Task, Meaning
Split By: EMMI
Cell: Asyst

	dire...	direction/lo...	ident...	inform...	locat...	location/id...	meta-id...	Totals
Direction	8.6E1	100.000	4.2E1	100.000	8.3E1	0.000	100.000	6.6E1
Other	0.000	0.000	7.692	0.000	1.3E1	0.000	0.000	8.065
Reserv...	1.4E1	0.000	50	0.000	4.348	100.000	0.000	2.6E1
Totals	100	100.000	100	100.000	100	100.000	100.000	100

Percents of Row Totals for Task, Meaning
Split By: EMMI
Cell: Faceless

	dire...	direction/lo...	ide...	inform...	loc...	location/id...	meta-id...	Totals
Direction	1.5E1	4.878	3E1	2.439	5E1	0.000	4.878	100
Other	0.000	0.000	40	0.000	60	0.000	0.000	100
Reser...	6.250	0.000	8E1	0.000	6.25	6.250	0.000	100
Totals	11.29	3.226	4E1	1.613	4E1	1.613	3.226	100

Percents of Column Totals for Task, Meaning
Split By: EMMI
Cell: Faceless

	dire...	direction/lo...	ident...	inform...	locat...	location/id...	meta-id...	Totals
Direction	8.6E1	100.000	4.2E1	100.000	8.3E1	0.000	100.000	6.6E1
Other	0.000	0.000	7.692	0.000	1.3E1	0.000	0.000	8.065
Reserv...	1.4E1	0.000	50	0.000	4.348	100.000	0.000	2.6E1
Totals	100	100.000	100	100.000	100	100.000	100.000	100

Percents of Row Totals for Task, Deictic/redun.
Split By: EMMI
Cell: Asyst

	alone	deictic	informa...	redun...	Totals
Direction	0000	8.2E1	0.000	18.182	100
Other	0000	60	0.000	40.000	100
Reserva...	0000	5.5E1	0.000	45.455	100
Totals	0000	6.7E1	0.000	33.333	100

Percents of Column Totals for Task, Deictic/redun.
Split By: EMMI
Cell: Asyst

	al...	deictic	informa...	redun...	Totals
Direction	•	50.000	•	22.222	40.741
Other	•	16.667	•	22.222	18.519
Reserva...	•	33.333	•	55.556	40.741
Totals	•	100	•	100.000	100

The total for one or more columns was zero.

Percents of Row Totals for Task, Deictic/redun.
Split By: EMMI
Cell: Faceless

	alone	deictic	informa...	redun...	Totals
Direction	1.7E1	4.7E1	3.333	33.333	100
Other	•	•	•	•	•
Reserva...	0.000	80	0.000	20.000	100
Totals	1.4E1	5.1E1	2.857	31.429	100

The total for one or more rows was zero.

Percents of Column Totals for Task, Deictic/redun.
Split By: EMMI
Cell: Faceless

	alone	deictic	informa...	redun...	Totals
Direction	100	77.778	100.000	90.909	85.714
Other	0.000	0.000	0.000	0.000	0.000
Reserva...	0.000	22.222	0.000	9.091	14.286
Totals	100	100	100.000	100.000	100

Percents of Row Totals for Drawing type, Task
Split By: EMMI
Cell: Asyst

	Direc...	Other	Reserva...	Totals
cir...	5.882	2.9E1	64.706	100
line	100	0.000	0.000	100
point	100	0.000	0.000	100
To...	40.741	1.9E1	40.741	100

Percents of Column Totals for Drawing type, Task
Split By: EMMI
Cell: Asyst

	Direc...	Other	Reserva...	Totals
cir...	9.091	100	100.000	62.963
line	36.364	0.000	0.000	14.815
point	54.545	0.000	0.000	22.222
To...	100	100	100.000	100

Percents of Row Totals for Drawing type, Task
Split By: EMMI
Cell: Faceless

	Direc...	Ot...	Reserva...	Totals
cir...	66.667	0000	33.333	100
line	83.333	0000	16.667	100
point	94.118	0000	5.882	100
To...	85.714	0000	14.286	100

Percents of Column Totals for Drawing type, Task
Split By: EMMI
Cell: Faceless

	Direc...	Ot...	Reserva...	Totals
cir...	13.333	•	40.000	17.143
line	33.333	•	40.000	34.286
point	53.333	•	20.000	48.571
To...	100	•	100.000	100

The total for one or more columns was zero.

Percents of Row Totals for Drawing type, Meaning
Split By: EMMI
Cell: Asyst

	dire...	direction/lo...	ide...	infor...	loc...	location/id...	meta-id...	Totals
ci...	0.000	0.000	6E1	0.000	3E1	4.348	4.348	100
line	43.75	12.500	4E1	0.000	0000	0.000	0.000	100
po...	0.000	0.000	2E1	0.000	7E1	0.000	4.545	100
T...	1.1E1	3.279	4E1	0.000	4E1	1.639	3.279	100

Percents of Column Totals for Drawing type, Meaning
Split By: EMMI
Cell: Asyst

	dire...	direction/lo...	iden...	inform...	loca...	location/id...	meta-id...	Totals
ci...	0.000	0.000	5.4E1	•	3E1	100.000	50.000	3.8E1
line	100	100.000	2.7E1	•	0.000	0.000	0.000	26.23
po...	0.000	0.000	1.9E1	•	7E1	0.000	50.000	3.6E1
T...	100	100.000	100	•	100	100.000	100.000	100

The total for one or more columns was zero.

Percents of Row Totals for Drawing type, Meaning
Split By: EMMI
Cell: Faceless

	dire...	direction/lo...	ide...	infor...	loc...	location/id...	meta-id...	Totals
ci...	0.000	0.000	6E1	0.000	3E1	4.348	4.348	100
line	43.75	12.500	4E1	0.000	0000	0.000	0.000	100
po...	0.000	0.000	2E1	0.000	7E1	0.000	4.545	100
T...	1.1E1	3.279	4E1	0.000	4E1	1.639	3.279	100

Percents of Column Totals for Drawing type, Meaning
Split By: EMMI
Cell: Faceless

	dire...	direction/lo...	iden...	inform...	loca...	location/id...	meta-id...	Totals
ci...	0.000	0.000	5.4E1	•	3E1	100.000	50.000	3.8E1
line	100	100.000	2.7E1	•	0.000	0.000	0.000	26.23
po...	0.000	0.000	1.9E1	•	7E1	0.000	50.000	3.6E1
T...	100	100.000	100	•	100	100.000	100.000	100

The total for one or more columns was zero.

Percents of Row Totals for Drawing type, Deictic/redun.
Split By: EMMI
Cell: Asyst

	alone	deictic	informa...	redun...	Totals
cir...	0000	58.824	0.000	41.176	100
line	0000	100	0.000	0.000	100
point	0000	66.667	0.000	33.333	100
To...	0000	66.667	0.000	33.333	100

Percents of Column Totals for Drawing type, Deictic/redun.
Split By: EMMI
Cell: Asyst

	al...	deictic	informa...	redun...	Totals
cir...	•	55.556	•	77.778	62.963
line	•	22.222	•	0.000	14.815
point	•	22.222	•	22.222	22.222
To...	•	100	•	100.000	100

The total for one or more columns was zero.

Percents of Row Totals for Drawing type, Deictic/redun.
Split By: EMMI
Cell: Faceless

	alone	deictic	informa...	redun...	Totals
cir...	0.000	6.7E1	0.000	33.333	100
line	0.000	6.4E1	0.000	36.364	100
point	2.9E1	4.1E1	0.000	29.412	100
To...	1.5E1	5.3E1	0.000	32.353	100

Percents of Column Totals for Drawing type, Deictic/redun.
Split By: EMMI
Cell: Faceless

	alone	deictic	informa...	redun...	Totals
cir...	0.000	22.222	•	18.182	17.647
line	0.000	38.889	•	36.364	32.353
point	100	38.889	•	45.455	50.000
To...	100	100	•	100.000	100

The total for one or more columns was zero.

Percents of Row Totals for Meaning, Deictic/redun.
Split By: EMMI
Cell: Asyst

	alone	deictic	information	redundant	Totals
direction	0.000	66.667	0.000	33.333	100.000
direction/location	0.000	100.000	0.000	0.000	100.000
identify	0.000	61.538	0.000	38.462	100.000
information	.	.	.	.	.
location	17.391	52.174	0.000	30.435	100.000
location/identify	0.000	100.000	0.000	0.000	100.000
meta-identify	0.000	50.000	0.000	50.000	100.000
Totals	6.667	60.000	0.000	33.333	100.000

The total for one or more rows was zero.

Percents of Column Totals for Meaning, Deictic/redun.
Split By: EMMI
Cell: Asyst

	alone	deictic	informat...	redundant	Totals
direction	0.000	11.111	.	10.000	10.000
direction/location	0.000	5.556	.	0.000	3.333
identify	0.000	44.444	.	50.000	43.333
information	0.000	0.000	.	0.000	0.000
location	100.000	33.333	.	35.000	38.333
location/identify	0.000	2.778	.	0.000	1.667
meta-identify	0.000	2.778	.	5.000	3.333
Totals	100.000	100.000	.	100.000	100.000

The total for one or more columns was zero.

Percents of Row Totals for Meaning, Deictic/redun.
Split By: EMMI
Cell: Faceless

	alone	deictic	information	redundant	Totals
direction	0.000	66.667	0.000	33.333	100.000
direction/location	0.000	100.000	0.000	0.000	100.000
identify	0.000	61.538	0.000	38.462	100.000
information	.	.	.	.	.
location	17.391	52.174	0.000	30.435	100.000
location/identify	0.000	100.000	0.000	0.000	100.000
meta-identify	0.000	50.000	0.000	50.000	100.000
Totals	6.667	60.000	0.000	33.333	100.000

The total for one or more rows was zero.

Percents of Column Totals for Meaning, Deictic/redun.
Split By: EMMI
Cell: Faceless

	alone	deictic	information	redundant	Totals
direction	0.000	11.111	.	10.000	10.000
direction/location	0.000	5.556	.	0.000	3.333
identify	0.000	44.444	.	50.000	43.333
information	0.000	0.000	.	0.000	0.000
location	100.000	33.333	.	35.000	38.333
location/identify	0.000	2.778	.	0.000	1.667
meta-identify	0.000	2.778	.	5.000	3.333
Totals	100.000	100.000	.	100.000	100.000

The total for one or more columns was zero.

Appendix D: Results of post-experiment questionnaires for the ASysT interface experiment.

Post-Experiment Questionnaire: Inexperienced Subjects

Thank you very much for participating in the experiment. We would appreciate knowing what you thought about the various aspects of this system. Please answer the following questions. Put an "X" anywhere along the scale you feel is appropriate to rate your response to the question.

ASyST®

1. How would you rate the understandability of the voice used by ASyST®?

almost unintelligible	sometimes hard to understand	fairly clear	perfectly clear
	xxxxxx	xxx	xxxxx

2. What was your reaction to the voice synthesis? (Please list other reactions if none of the choices here are appropriate):

grating	irritating	monotonous	neutral	interesting	pleasurable
x	x x	xxxxx	xx	x xxx x	x

4. Good intonation
8. Funny

3. How did you feel about ASyST®'s "persona" (face)?

objected to it	looked odd	didn't care	looked OK	appreciated it
x x	xxxx	xxx	xx	xxx

4. A little too mechanical in appearance
8. Objected--get a wig and a smile--look like a human being
10. appreciated: It gave a personal touch though without the hair he looks a bit like an alien. I suggest hair and glasses (academic look)

4. How comfortable did you feel communicating through ASyST®?

anxious, very uncomfortable	nervous	OK	very comfortable
x x	xx	xxxxx x x	x x

2. OK--electronic machines always make me slightly uncomfortable
3. At first I was nervous but I became more and more comfortable as I went along
4. ditto
10. ditto
22. ditto

5. How would you feel about using ASyST® in a real communication situation in the future?

wouldn't use it	would prefer human interpreter	doesn't matter	would prefer ASyST®	can't wait until that's possible
xx	x	x	x x	xxxxx

4. An excellent system given situation and circumstance but not an alternative to human contact
8. Prefer human: because of speed

Experimental Setting

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

	xxxx	x xxxx	xxxx
a real bore	kind of interesting	fun	had a great time

7. How would you rate how easy it was to accomplish your goals?

	x	xxxxxxxx	x xxx
difficult	had to work at it	some effort	simple

How would you rate the usefulness of each of the following for accomplishing your goals?

8. Map:

		x xxxx	x	xxxxxxxx
worthless	an inconvenience	served some purpose		very useful

3. below some purpose: better is the whole area was shown and then the specific area blown up

9. Reservation form:

		x xxx		xxxxxxxxxx
worthless	an inconvenience	served some purpose		very useful

10. The picture of the hotel:

x		x xxxxxx		xxx
worthless	an inconvenience	served some purpose		very useful

3. didn't ask to see it
4. didn't look at it
10. some purpose: helped to ease possible anxiety over what the hotel is like
21. some purpose: helpful to see what kind of structure it was

11. Typing option:

	xx	xxxx	xxxxxxx
worthless	an inconvenience	served some purpose	very useful

21. very useful: when I couldn't understand the voice

12. Touchscreen:

		x	x	xxxxxxxxxx
worthless	an inconvenience	served some purpose		very useful

10. useful (useful for typing, too): more comfortable using keyboard and touchscreen; talking to a computer is AWKWARD
21. very useful: helped to clarify directions on the map

13. Printing option:

worthless	an inconvenience	served some purpose	very useful
x		x xx	x xxxxxxx

14. "Interrupt" button:

worthless	an inconvenience	served some purpose	very useful
x x xxxx			xxx

4. didn't use
18. didn't use
22. didn't use

15. "Repeat" button

worthless	an inconvenience	served some purpose	very useful
		x x	x x xxxxxx

16. "Rephrase" button:

worthless	an inconvenience	served some purpose	very useful
x	x x	x x	xxxx

10. inconvenience: didn't need to use it
17. questions 14, 15, 16: didn't use
22. didn't use

17. Messages like "The ASySTant is listening to you:"

worthless	an inconvenience	served some purpose	very useful
x x	xxx	xxx	xxxxxx

18. The picture of the ASySTant listening to you, etc.:

worthless	an inconvenience	served some purpose	very useful
x	x	xxxxxxx	x xxx

10. some purpose: seeing the interpreter's face made it easier to know how long to wait and made it more personal In real life I would have been more likely to joke around with the interpreter
17. very useful: makes it easier to see when a message is coming

19. The picture of the agent:

worthless	an inconvenience	served some purpose	very useful
	x	xxx	xxxxxxxxx

20. Your picture:

worthless	an inconvenience	served some purpose	very useful
xxxx	xx	xxxxx	x

10. some purpose: some purpose, yes but also distracting
18. why was I frozen?

22. What was the best part of the experiment for you?
2. Receiving information in a measured, orderly fashion.
3. Knowing that my message was received and understood.
4. Enjoyed listening to the voice translation. Two-way interaction was very good, as was the screen displays of the map.
5. Being able to get information and seeing map to confirm my location.
9. A feeling of control in filling out your own registration for a hotel.
12. Seeing the agent smile and be confused just like I was.
13. Being able to communicate with someone in a foreign language.
The chance to do reservations and get information from my home computer
18. Being able to ask about slightly unrelated things, where to eat, etc.
21. As a confused tourist, it helped me to relax, knowing that someone could understand me. Once I got used to it, my thinking became clearer; at first it was intimidating and confusing but its simplicity put me at ease rather quickly.
22. Knowing I was communicating to someone in a different language. I felt I could ask all my questions and they would be answered.

23. The worst part?
2. Occasionally the vocal function is fuzzy or slightly garbled. As well, there is a hollowness or "tinniness" that can be occasionally unpleasant.
3. When I couldn't find the delete button.
4. Uncomfortable talking into the microphone. Maybe telephone is less intimidating.
5. Trying to understand the synthesized voice.
6. The delay waiting for the translation; at times I wanted to ask a quick follow-up question but I was prevented from doing so.
8. Long wait to translate
9. Waiting! Speech is an incredibly slow form of communication. Set up a system that would let you choose your options more quickly, flow-chart style with hot keys...
10. Beginning the experiment. I wasn't sure whether to type, speak.... Also sometimes the screen was blocked so I couldn't get back to the previous screen.
12. The synthesizer said a word twice and I didn't understand it. Also in the beginning it talks at you too much.
13. The computer voice was generally clear and understandable but obviously not as clear as a human voice and thus required some repeating and typing of messages.
18. The agent was friendly, but seemed to have only a small window of specific information
22. At first I was unsure that the computer would understand my needs.

24. Do you have any impressions you'd like to note?
2. The system is easy to use and the map, picture and print functions are a particular asset. The system is a touch slow. Some delay is desirable to allow the user to acclimatize to the system, but the present delay is long enough to discourage use of the system.
4. The system is really helpful.
5. This will be very helpful in the future. 6. ditto 13. ditto
12. Too slow; the voice should be more human
17. I wasn't sure when to type.
21. The system will be useful and make visitors feel more comfortable, receiving map printouts etc. The slowness of the system seemed confusing until I got used to it.

25. Do you have any suggestions for improvement?
2. My only suggestions are in the nature of polishing. It's hard to envisage addition to the functions themselves.
Improve sound quality.
The accent of the voice is pleasant and projects an image of credibility, sophistication, etc.
The drawing function requires a broader line to better distinguish between it and the map. lines.

Instruction labels (PRINT, TYPE, etc.) are similar enough to one another to be misleading to a novice user.
The cursor too small to be noticed easily.
Overhead position of video camera gives unflattering view of user--maybe to the side is better.
3. Clearer voice
4. Voice translations of messages are not really necessary.
5. Improvement on smoothness of synthesized speech.
6. Cut out translation delay
on the maps, show pictures of destinations if possible
dual screen for map and typing
I wasn't able to type when translation was in process
when agent displayed a map/typed a message, it automatically erased the previous screen and prevented you from going back to it.
8. Some people talk to themselves or think out loud. What about this situation?
9. Print button should say Print Map.
Calling the hotel doesn't require a person on the other end. The conference information may, but it could be updated daily with any amount of information. I take it back. I forgot it was translating. This is partially due to the little reliance on speech and partially due to the overabundance of visual stimuli.
Registration form requires method of payment or credit card number slot
10. The face of the agent is one of the best ways to personalize it. Make it *clear* that the user has two options--to speak or write.
12. no keyboard, no video, happy, smiling female, Japanese face and feminine voice.
faster translation
I was afraid to touch the interrupt button and interrupt the person so I think it doesn't need an interrupt button.
[Made a drawing of screen: flashing signals for client and agent to talk; buttons to push for different hotels; split screen: map and hotel form] I don't think it is necessary to type as the agent can type what I want.
13. The English accent made the computer voice even more difficult to understand. Certain vowels sounds were not clear.
18. Direct the agent to be a bit more free with offering information...
The conversation was mostly natural, though slow except I did feel it was too focused on my task
21. When a tourist first arrives it is very confusing. If the system could ask specific questions, it would be helpful....
2. It might be useful for the system to print a map of your route, just so you have it.
It would have been nice to have been given a number to call if there was a problem.

Post-Experiment Questionnaire: Experienced Subjects

Thank you very much for participating in the experiment. We would appreciate knowing what you thought about the various aspects of this system. Please answer the following questions. Put an "X" anywhere along the scale you feel is appropriate to rate your response to the question.

ASyST®

1. How would you rate the understandability of the voice used by ASyST®?

x x xxxxxx x xx x x

almost unintelligible | sometimes hard to understand | fairly clear | perfectly clear

20. Sometimes hard: especially numbers

2. What was your reaction to the voice synthesis? (Please list other reactions if none of the choices here are appropriate):

xx x x xx xxx x xxxx

grating | irritating | monotonous | neutral | interesting | pleasurable

3. How did you feel about ASyST®'s "persona" (face)?

x x xxx xxxx x x x xxx

objected to it | looked odd | didn't care | looked OK | appreciated it

0. Not really realistic except for 3D look; too fat. A bodiless head may seem funny or strange to some people; didn't matter to me.
16. looked odd: not a guy you'd trust on instinct

4. How comfortable did you feel communicating through ASyST®?

xxxx xxxxxxx x x

anxious, very uncomfortable | nervous | OK | very comfortable

14. beyond OK: like that I could see the agent

5. How would you feel about using ASyST® in a real communication situation in the future?

xxx x x x xxx x xx

wouldn't use it | would prefer human interpreter | doesn't matter | would prefer ASyST® | can't wait until that's possible

0. would prefer IF pocket-sized, faster, clearer voice; otherwise much prefer interacting with a real person
7. Once it gets up to speed it would be quite amazing to use via Internet or something
14. prefer ASyST: if human isn't available
16. prefer ASyST: if faster
24: prefer human: for less well-defined task; prefer ASyST: to no interpreter at all

Experimental Setting

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

x xx xxxxx xxxxxx

a real bore | kind of interesting | fun | had a great time

7. How would you rate how easy it was to accomplish your goals?

	xx	xxx x x x x	x	xxxxx
difficult	had to work at it	some effort		simple

How would you rate the usefulness of each of the following for accomplishing your goals?

8. Map:

			xxxxxxxxxxxx x
worthless	an inconvenience	served some purpose	very useful

13. very useful: N marker absolutely necessary

9. Reservation form:

		x xx	x xxx	xxxxxxx
worthless	an inconvenience	served some purpose		very useful

10. The picture of the hotel:

x	x	x	xxxxxx	xx	xxx
worthless	an inconvenience	served some purpose			very useful

0. inconvenience: It made me think I was at the hotel (if I were really at Kyoto station, I wouldn't think that, probably)
13. worthless: not interested in hotel room's appearance
26. some purpose: gave me something to look at while waiting for the translation

11. Typing option:

	x	xxx	x x	xxxxxxx
worthless	an inconvenience	served some purpose		very useful

1. no choice--didn't use
7. below some purpose: for the agent to type. Don't give me the option; I was confused
11. very useful: since computer voice is sometimes hard to understand

12. Touchscreen:

		xx x	x	x xxxxxxxxxx
worthless	an inconvenience	served some purpose		very useful

0. very useful: I found it refreshing not to have to interface with a computer through a keyboard for once.
7. below some purpose--just need to learn; used to a mouse

13. Printing option:

	x	x	xx	xxxxxxxxx
worthless	an inconvenience	served some purpose		very useful

14. didn't need it but could be useful
19. didn't use
26. worthless and very useful: didn't always print what I wanted; when it did it was good

14. "Interrupt" button:

xxx xxxx xxx xx

worthless | an inconvenience | served some purpose | very useful

0. beyond inconvenient: didn't have a need. Perhaps when the system is faster it may come in handy
7. didn't use
13. didn't need or use; not sure if it's necessary
19. didn't use

15. "Repeat" button:

x xx x x xxxxxxxxx

worthless | an inconvenience | served some purpose | very useful

16. "Rephrase" button:

x x x xxx x x x xxx

worthless | an inconvenience | served some purpose | very useful

7. below very useful: didn't use but see the application
13. didn't use but probably good option
14. for question 14, 15, 16: didn't need them in this situation, but good features
20. some purpose: didn't get it to work
24. some purpose: with more practice, might help more

17. Messages like "The ASySTant is listening to you:"

x x xxxx x xxxxxxx

worthless | an inconvenience | served some purpose | very useful

7. above worthless: didn't notice
13. very useful: good to see computer prompting like this
26. very useful: once I noticed it and only because of delay

18. The picture of the ASySTant listening to you, etc.:

xx xx xxxx xx x xxx

worthless | an inconvenience | served some purpose | very useful

13. same as question 17
24. some purpose: but confusing

19. The picture of the agent:

x xxxxx x xxxxxx

worthless | an inconvenience | served some purpose | very useful

0. some purpose: it did humanize the machine interaction even though the voice and picture didn't really match
1. very useful: makes it more personal which makes the computer less daunting
13. not necessary in experiment
19. very useful: So I know I'm actually talking to a human
24. some purpose: seemed more personal

20. Your picture:

x xx x xx xxx xxxx

worthless | an inconvenience | served some purpose | very useful

0. inconvenience: served no purpose for me but perhaps it did for the agent
1. Makes it more friendly
7. inconvenience: very bad picture and it didn't change
13. not necessary in experiment
19. inconvenience: I was a still picture so I couldn't see what I was doing
24. worthless: put me off

22. What was the best part of the experiment for you?
0. quite good English from [CHATR]
11. Highly sophisticated in its ease of use.
15. Seeing how voice, touchscreen and keyboard are integrated
16. Translation was successful. All my questions were answered.
19. Being able to accomplish my goal
20. System was good for exchanging info: map w/whiteboard, forms, speech and image.
25. Drawing on the map

23. The worst part?
0. Having to hit the Repeat or Rephrase button too many times.
7. Slow and at points a bit confusing. This does not replace personal contact.
11. The ASyST face--looks similar to a Roman head on a stick.
15. Trying to understand the synthesized voice.
16. Slowness of response and voice distortion
19. Not knowing if it was my turn to talk or if I talked when the face was not looking at me, I didn't know if it would respond or not.
23. Being told I had to repeat [colloquial English] at the end of the experiment.
24. Waiting and not being able to type while waiting for the translation.
25. The time delay. Without the "listening" and "translating" messages it would have been very confusing because of the delay.
26. Not knowing when I should speak until I got used to the delay.

24. Do you have any impressions you'd like to note?
7. Every "early adopter" will want to try it.
10. Complexity and accuracy of translation amazed me...
16. With improvement in speed, I think people will like ASyST. I'd appreciate using it.
19. Turn-taking cues were difficult to pick up on.
23. The persona's most useful feature was changing its orientation to reflect what was happening.
24. I felt free to ask any complex question, but did speak more slowly and clearly than usual. Having many modes possible is great; if it had been less frustrating I would have used it more.
25. Most non-Japanese speaking people would not know how to make reservations and this would be very useful.

25. Do you have any suggestions for improvement?
0. Speed needs to be increased. Products based on this technology won't be accepted as easily. ...I've seen the often negative reactions of people when the system was too machine-like, impersonal, etc.
7. Get it faster and improve the graphic display to show only those things I need to do, i.e., a red flashing button to something that shows when it's my turn to speak, or a ping-pong, something that notifies me that it's my turn.
Also I didn't use the keyboard at all; the less a user has to deal with, the better.
I believe once it's fast enough, there'll be no need for the little face.

11. Time and speed of translation need improvement.
16. The ASyST picture could be a woman who doesn't look threatening. Voice could be more pleasing also.
The picture of me may be useful in that I know I am being seen, but it is much larger than necessary.
Picture of the agent should be centered and could be smaller.
Maybe the map of hotels should come up before I'm asked what hotel I'd like to stay in.
20. Should be more of a contained environment, fewer windows and simpler. A one-time user loses time tracking them all.
I found the Asystant to be of little help whereas the "please wait" signs were my main source of information about conversation flow. Perhaps arrows or highlighting the speaker image.
I found the typing to be puzzling. Does the agent see the client's typing without pressing the "send" button? Is the text translated?
23/ Hotel picture should be of higher resolution.
[problem with typing during typing instructions] ...Add a multiple-level repeat capability next-previous, next-next-previous, etc.
24. Make it possible to type on the form while waiting for a translation
Make it possible to move [the typing] window around so you can see the map under it.[chart conversation flow] another additional way: red light: don't talk; green light: do talk; yellow light: it's thinking.
I can imagine that people will learn to use it more efficiently. Perhaps the person on the other end can take a role in teaching them.
Some questions I dropped or never asked because of the sense that I didn't need the information enough to try again. But if I had to rely on my own Japanese I'd drop even more.
Numbers often seemed garbled.
25. Make sure the text window only comes up when it should.
The explanation pop-ups made me think I was supposed to push on the windows instead of the buttons.
The user should choose whether explanations are given.
26. While I saw the agent laugh, I didn't get the sense we were having a conversation.
The signals for when who should speak took some getting used to.
I used the Repeat button since it seemed quicker due to translation delay.
Seemed like it should have been information retrieval since anything out of the ordinary required a special translation.