TR-IT-0173

Multimedia Route Descriptions: Experimental Results

Laurel Fais

June, 1996

Abstract

The results of an experiment to determine the optimum configuration of drawing and verbal description for the description of routes are discussed. Six groups of subjects were given the same route via different combinations of drawing and language description on a map presented on a computer screen. Subjects wrote the most accurate descriptions of the route after hearing a detailed language description; whether or not subjects also saw a drawing of the route did not affect their performance either on the written description or on the post-experiment route-drawing task. Results from these tests and from user comments indicate that the most effective means of presenting route information is with a combination of visual and detailed language descriptions. Implications for multimedia communication systems are discussed.

Table of Contents

Introduction[1]

In this paper, we discuss the results of an experiment designed to determine the most effective combination of speech and drawing for the presentation of a route through a town. The significance of this kind of information lies in the fact that giving directions is an integral part of the multimedia machine translation project conducted within the setting of the Environment for Multimodal Interaction (EMMI) in ITL, Department 4. Some background concerning this project will also make clearer the results of the experiment reported here.

Background

EMMI is part of an effort to build a system in which various communication media are integrated with machine translation. Within EMMI, users can speak, draw on a map, type to a form, or type unrestricted messages in order to perform a direction-finding task and a hotel reservation task (Loken-Kim *et al.*, 1993). EMMI can accommodate same-language interaction or bilingual interpreted interaction with either human or "machine" interpretation.

Three major experiments have been conducted in the context of EMMI in order to gather data concerning the speech and media behavior of naive users of such a system. In the first of the three experiments, subjects acting as "clients" got directions to a conference site by engaging in a cooperative dialogue with the "conference agents." In this first experiment, both "clients" and "agents" were native speakers of American English, and their interaction was human-human (mediated by various technologies; see below). In two further experiments, native American English-speaking "clients" interacted with Japanese-speaking "agents." In one of these experiments, speech was interpreted by human translators; in the other, by a simulated automatic machine translation system. In all three experiments, subjects interacted in two different communication settings: via a standard telephone, and via a computer-based, multimedia (MM) environment in which subjects could freely interact by voice, by typewritten text, by drawing on a visual image (a map), and by typing to a form (Fais, 1994; Fais *et al.*, 1995; Park *et al.*, 1994). The acoustic data for all three experiments were recorded on DAT tapes and transcribed; the visual data (drawing or typing by both agent and client) were recorded directly from the computer screens and marked on the speech transcriptions (Park *et al.*, 1995).

There are two notable features of these experiments. The first is that they were human-to-human (though mediated in the second and third experiments) and speech-based. These were not human-machine interactions, and the users were not limited to typewritten input, but could interact via speech as in usual forms of conversation. The second feature is that

[1]I would like to express my thanks to Tsuyoshi Morimoto for supporting this work. I am grateful as well to Kazuhiko Kurihara for the production of the stimuli used in this experiment and to Karen Berner for her invaluable organizational assistance.

the users were completely unconstrained; at no time were they instructed to use any particular wording or, in fact, to modify their speech in any way.

As a result, these experiments yielded important data concerning the nature of spontaneous expression in these communication environments (MM and telephone; human-human, human-interpreted, and machine-interpreted). Looking at these results, we are able to see how users naturally interact via a number of different media, what choices they use in presenting information, and how their speech behavior changes in each setting.

Once we understand the basic nature of interaction in these environments, we can identify speech or media behaviors that we would like to encourage or discourage and attempt to design our system towards those ends. For example, we found that users employed a disproportionately high number of words to convey information in the MM setting because they often had to request help concerning the use of the system itself (Fais and Loken-Kim, 1995). In our subsequent revision of the interface, we attempted to give the system more initiative in instructing and guiding the user so that the user would not have to ask how to employ the MM options available (Fais *et al.*, 1996).

In the experiments, we did not instruct the agent about how to guide the interaction or how to present information. However, this situation is not really true-to-life; in a scenario such as ours, the conference agent could very easily be trained to give information in a certain way so that it could be most readily grasped by the client. With this in mind, we designed an experiment in order to discover what is the most effective combination of drawing and speech for presenting a route such as that encountered in the conference-site-locating task of the three initial experiments. Our procedure was to present subjects with directions given in various media combinations and then test the subjects to see how much of the route they could reproduce. The media combination for the group which was most successful in reproducing the route would be considered the most effective.

As we mentioned above, the subjects in the initial experiments were free to use whatever media they chose to convey information. Examining the results of these experiments, we were able to see what the preferred patterns of media use were. Where we might have expected to see subjects *replace* speech with the use of visual media for a task such as giving directions, we found instead that they did that in only about half of the occasions on which they used visual media (Fais and Loken-Kim, 1995). The other half of the time, subjects drew on the map and gave a complete language description at the same time. Although it was clear that subjects preferred to use the media to present information in this way, we wondered if this use of media options was in fact the best for *receiving* information. This was another motivation for conducting the experiment described here.

Methods

Subjects were presented with a route through a fictitious town and then asked to reproduce that route in both written form and by drawing the route on a map. The map is shown in Figure 1. An attempt was made to design the map and the route so that they were complicated enough to be somewhat difficult to remember. This allowed presentation differences to show up better.

There were six groups of ten subjects each; each group was instructed in the route via a different combination of media. These combinations are described in Table I; the text of each of the three language descriptions appears in the Appendix.

NL	no language	Subjects only saw the route being drawn on the map.
D	deictic	Subjects saw the route being drawn and heard a description of the route phrased in deictic expressions.
LNDr	language, no drawing	Subjects heard a natural language description of the route but saw no drawing.
LDr	language, drawing	Subjects heard a natural language description of the route and saw it being drawn on the map at the same time.
XLNDr	"extra" language, no drawing	Subjects heard a detailed language description of the route but saw no drawing.
XLDr	"extra" language, drawing	Subjects heard a detailed language description of the route and saw it being drawn on the map at the same time.

Table I. Description of each experimental condition.

The subjects who saw the drawing sat in front of a computer screen on which was displayed the map in Figure 1, and watched as a line was drawn (automatically) to trace the route. The same line drawing was used for every group which saw the drawing. The speech descriptions were taped and played for the subjects, synchronized with the line drawing. Since the line drawing had to take a long enough time to accommodate the longest, most detailed description, there were some lengthy pauses in the deictic description, especially. All subjects, even those who did not see the drawing, watched the map while the language description was being played.

The subjects were all native speakers of English; the text was checked before the experiment for any dialectal peculiarities that may have favored one variety of English over another. There was some concern over the use of the names of US Presidents for street names; as it turned out, there were no significant differences in use of street names in the written descriptions of the subjects due to variety of English spoken by the subject.

After the route was presented to them, subjects were asked to reproduce the route. Half of the subjects in each condition was asked to write down as much of the route as they could remember in words first; the other half was given a copy of the map they had seen on the computer screen and asked to trace the route first. Subjects who had written the description first, then drew the route on the map; subjects who had drawn the route first, then wrote the description of the route in words. In all cases, the subjects' first description (whether on the map or in words) was taken away before they began the second description. Thus, subjects did not have access to a copy of the map while they were writing their descriptions, or *vice versa.*

Subjects were also asked to note their suggestions and impressions of the task after they had produced both descriptions.

Measures

The descriptions were scored by hand. In the written descriptions, the numbers of both correct and incorrect assertions were counted. Subjects produced a maximum of 37 correct assertions and a minimum of three, with a mean of 16.75. Subjects produced a maximum of 10 incorrect assertions and a minimum of none, with a mean of 2.9.
A total of 25 points were possible on the route drawing task; since the drawing of the route seemed to be an easier task, there is a smaller spread in scores. Twenty-three subjects received perfect scores on the route drawing; the minimum score was 11 and the mean score was 21.15.

Results

Route description

Figure 2 shows the distribution of correct answers over each condition. The difference between the extra language condition and the other conditions is a strong trend, but not significant. (In this and all other figures below, the X-axis lists the language conditions as in Table I; "Dr" stands for "drawing" and NDr" stands for "no-drawing.")

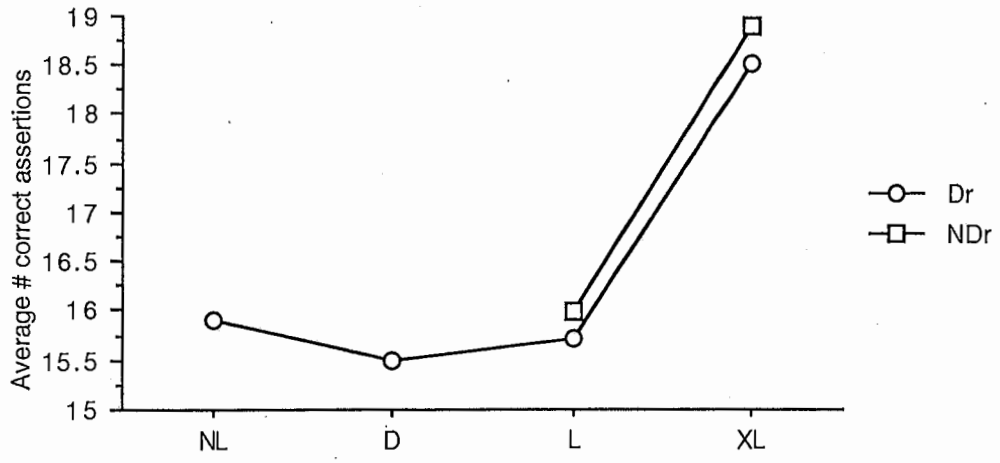

Figure 2. The average number of correct assertions in the written descriptions for each condition.

The distribution of incorrect answers over each condition is shown in Figure 3. The differences are not significant.

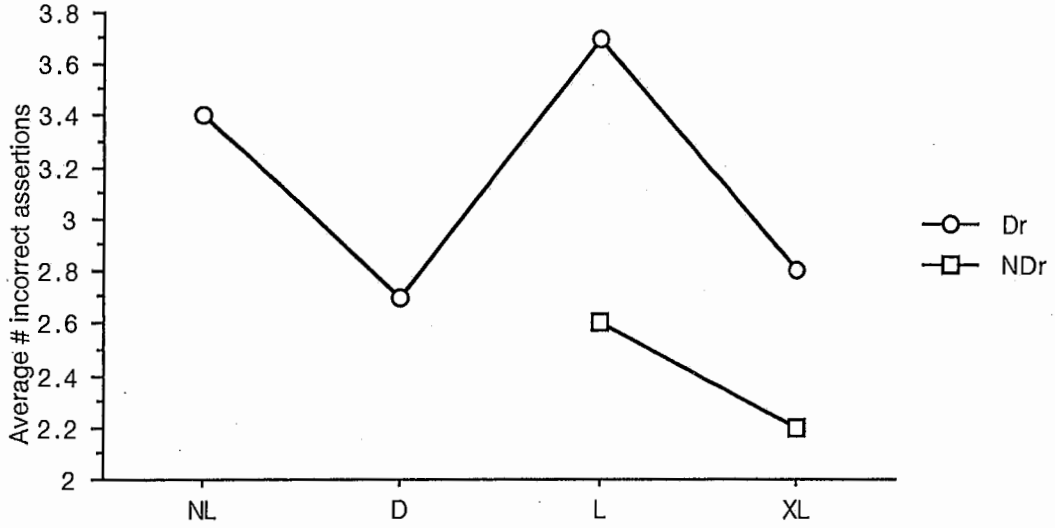

Figure 3. The average number of incorrect assertions in the written descriptions for each condition.

Figure 4 shows the distribution of scores for the route drawings for each condition. Similar to the situation for the correct assertions in the written descriptions, the differences shown below between the no-language condition and the other conditions

represent a trend; they are not significant differences. The differences between the conditions with drawing (Dr) and those without (NDr) likewise represent a trend.

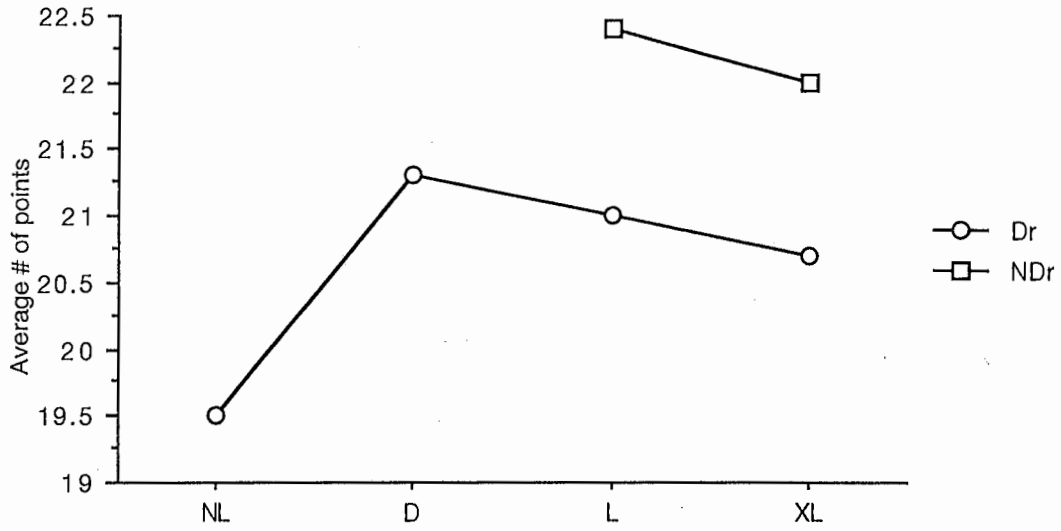

Figure 4. Average score for the route drawing for each condition.

The order in which the post-experiment tests were presented had a significant effect on the ability of the subjects to reproduce the information. In all conditions, subjects produced significantly more correct assertions and received higher scores on the route drawing when the route drawing was the first task and the written description was second. Because this was true for all conditions, we ignore it here and simply average the scores from both groups for the post-experiment tests.

However, the situation for incorrect assertions was somewhat different (Figure 5). Subjects' responses in the no-language, deictic, and language conditions followed the same trend: more accurate responses (in this case, fewer wrong answers) in the route-drawing-first group. However, the opposite was true for the extra-language condition; subjects made fewer mistakes when they wrote the route description first.

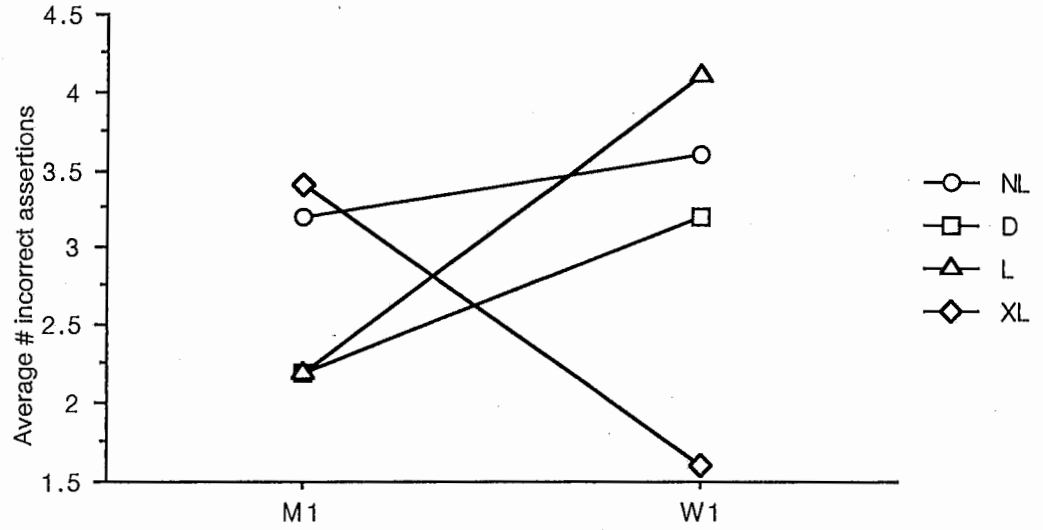

Figure 5. Average number of incorrect assertions by condition for route drawing first (M1) and written response first (W1).

We also investigated how long the written descriptions were. Figure 6 summarizes this information. The difference between the extra language group and the no language group on one hand, and the deictic and language groups on the other, is significant ($p<0.05$).

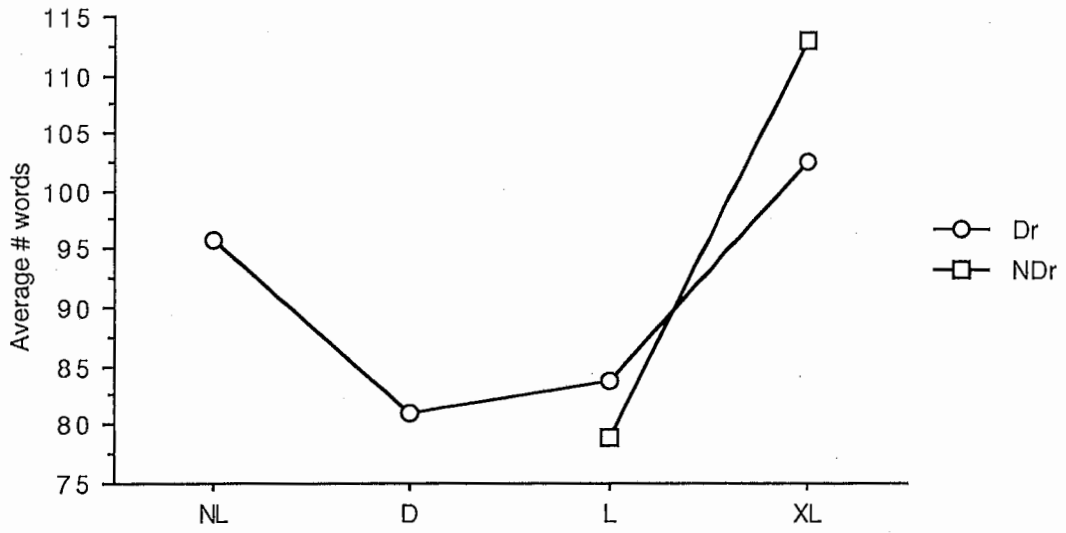

Figure 6. Average number of words in the written descriptions of subjects in all conditions.

Discussion

The presence or absence of the drawing had no effect upon the number of correct assertions made by subjects in the written descriptions. The crucial aspect in this regard was the presence of a detailed language description. The same level of correct information was conveyed in the no-language, deictic and language conditions; only when a very detailed description was provided were subjects able to reproduce greater amounts of information. Recall that the writing-first group of the extra-language condition produced markedly fewer incorrect answers than all the other groups. This is contrary to the trend shown by all the other groups: fewer incorrect assertions in the route-drawing-first group. However, it *is* consistent with the situation for correct answers: if it is in fact the presence of detailed language which encourages a high number of right answers, then it may not be surprising that fewer incorrect answers are made when that detailed description is "freshest," i.e., when the written response was made immediately after listening to the description. However, this does not explain why the same reasoning doesn't hold for correct assertions; despite the fact that the language description was freshest immediately after viewing the route, subjects still produced more correct assertions when they wrote their descriptions after drawing the route on the map.

The explanation probably lies in the nature of the tasks. Reproducing a spatial description is easier to do in spatial terms, i.e., by drawing on a map. Once subjects had "practiced" the route by drawing it on the map first, they were better able to describe the route in words. Describing the route in words first (before drawing the route on the map) requires that both the difficult task of recall and that of reproducing a spatial description in linguistic terms be done at the same time. Subjects were able to perform better on both tasks when they were separated: recall the route first (by drawing it), and then describe it in words. Furthermore, subjects knew that they were expected to do both the writing and the drawing tasks. There is the very real possibility that some subjects in the drawing-first group used the map to rehearse the route in order to prepare for the written description. Note, too, that the avoidance of incorrect answers (better in the writing-first group) and the production of correct ones are different kinds of tasks. It may be easier to avoid mistakes when the linguistic description is "freshest;" it may be easier to produce

more correct answers when recall has been aided by doing the task in the more appropriate (i.e., the spatial) mode first.

In summary, then, subjects produced more correct assertions after they had drawn the description first, and only in the case in which they had heard a detailed description of the route. Whether or not they had also seen the drawing of the route did not affect their ability to describe the route in words.

If we look now at the number of points received on the route drawings, we see a slightly different picture. Subjects who heard no language description did the worst; subjects who heard *any* sort of language description (including merely deictics) did better. Thus, it seems that some sort of language description aids spatial recall. Some of the subjects themselves were aware of this. After they had completed their written and drawn responses, subjects were asked simply to write down any comments or suggestions they had about the experiment. Some noted that "it was useful to have the voice in addition to the visuals; it served to focus my attention for the duration of the route." However, note the situation with respect to the presence or absence of drawing. Subjects actually did better at drawing the route themselves when they did *not* see a drawing of the route on the screen. Recall, however, that none of these differences are significant; thus they are only suggestive of a possible trend.

However, it is interesting to speculate on why it might be the case that subjects made better scores on the route drawing in the *absence* of having seen a drawing. That accompanying language made the drawing easier to reproduce is not a surprise; language serves both as a way to focus attention on the route and to "codify" the route for later recollection. However, that an accompanying *drawing* made the drawing *harder* to reproduce is somewhat surprising. It may be the case that only listening to a language description forced subjects to construct the visual routes themselves as they listened to the description and watched the map; these self-constructed routes would be easier to reproduce later than routes that subjects simply watched passively. In fact, one subject in a no-drawing condition described exactly this process: "my sense of direction is visual so I actively traced the route on map [while I listened]." A related possibility is that the drawing focused the subjects' attention *too* narrowly. That is, in watching the drawing, subjects were prevented from taking in other visual information that might have helped them retain and reproduce the route better. One subject who did not see the drawing but only heard a language description voiced exactly this explanation: "At first I thought maybe a small dot or something could be used to follow on the screen what the voice said but later I'm glad it wasn't because I think my eyes and mind (hence, concentration) would have been distracted. As it was, I could listen to the voice and scan the entire map at the same time."

Another possible explanation is that the presence of language somehow interfered with subjects' ability to retain the spatial information presented in the drawing. However, if this were the case, subjects in the no-language condition should have performed better than subjects in the language-and-drawing conditions. In fact, they performed worse. Thus, we conjecture that the presence of language aided the retention of the spatial information concerning the route, and that the absence of a drawing on the screen might have forced subjects to more actively focus their attention on the spatial aspects of the information being presented.

This kind of result is mirrored in a way by the results for the number of words subjects used in their descriptions. Ignoring for the moment the extra-language condition, we see that subjects used more words when they had not listened to a language description. This corresponds to the situation above in which subjects more accurately represented the route drawing when they hadn't seen the drawing on the screen. Subjects who listened passively to a linguistic description did not write as many words as subjects who did not listen to a description but instead had to construct a description for themselves as the drawing was presented. In effect, the linguistic description focused their attention on

only those words used in that description (as the drawing might have focused their attention on only that portion of the map) and this restricted their later written output. The presentation of a detailed description (the extra-language condition), on the other hand, may have suggested to the subjects aspects of the description they would not have constructed or noticed themselves, so that their written responses following a detailed description were longer than they would have been otherwise.

Thus it seems that the most effective environment in this experiment for conveying information is the extra-language condition: this condition produced both the most accurate written descriptions and some of the most accurate route drawings. The effects of the presence or absence of the drawing are inconclusive; no real difference was found either for the written descriptions or for the route drawing..

However, even though the results of the route drawing task were inconclusive, the preference expressed in subjects' comments is overwhelming. Fully a quarter of the subjects, from all conditions, commented that they felt that following a visual image was easier than just listening to a verbal description. Some went so far as to say that they "didn't really listen to the instructions" but relied on the drawing. Three subjects in no-drawing conditions said they felt that a drawing of the route would have helped them.

Often when subjects commented on the benefits of having a visual image, it was in the context of comparison to a strictly verbal description; they clearly preferred a visual presentation to a verbal-only presentation. Some subjects in the no-drawing conditions claimed that listening only once, and listening only, was not enough. On the other hand, some subjects in language-and-drawing conditions commented positively about the oral instructions. They felt, for example, that " the audio explanation enhanced the drawing."

In addition, we should also take into account other factors involved in conveying information, such as confidence in the information. The clear preference of subjects for visual information in this task is a strong indication that they have more confidence in their ability to retain information presented in this way. This is a factor which cannot be ignored in making recommendations concerning the use of multimedia.

Furthermore, we saw in previous work, as discussed above, that when subjects are free to choose which media they will use to present visual information, they use both verbal and visual means. This corresponds exactly to our findings in this experiment: subjects were able to reproduce the most correct assertions and draw the most accurate maps when given a detailed verbal description, and they felt more confident about their information when it had been presented to them visually as well.

Finally, this experiment was conducted in the theoretical setting of a non-working system. That is, the task is a hypothetical one, and is part of a constructed scenario. The criteria for effectiveness, then, are hypothetical. If, for example, actual users need to retain a verbal description of the route, then a detailed language description with accompanying drawing is the best mode of presentation, since that produces the best results and is consonant with subjects' feelings of confidence in the presentation. If, for some reason, users will need to draw the route, then the description alone may be most effective. In short, the actual needs of the users of the system will be critical in deciding the optimum media configuration.

Language notes

Before we conclude the discussion of this experiment it may be interesting to note some further comments made by the subjects and to compare those comments to some of the features of the linguistic descriptions they wrote. Subjects expressed some clear preferences concerning the presentation of directions; these are borne out in their written descriptions in interesting ways.

Street names. Subjects overwhelmingly commented that it was very difficult to remember the street names in the directions. Several suggested that (too many) street names should not be used. Figure 7 shows the average use of street names in the written descriptions for each condition.

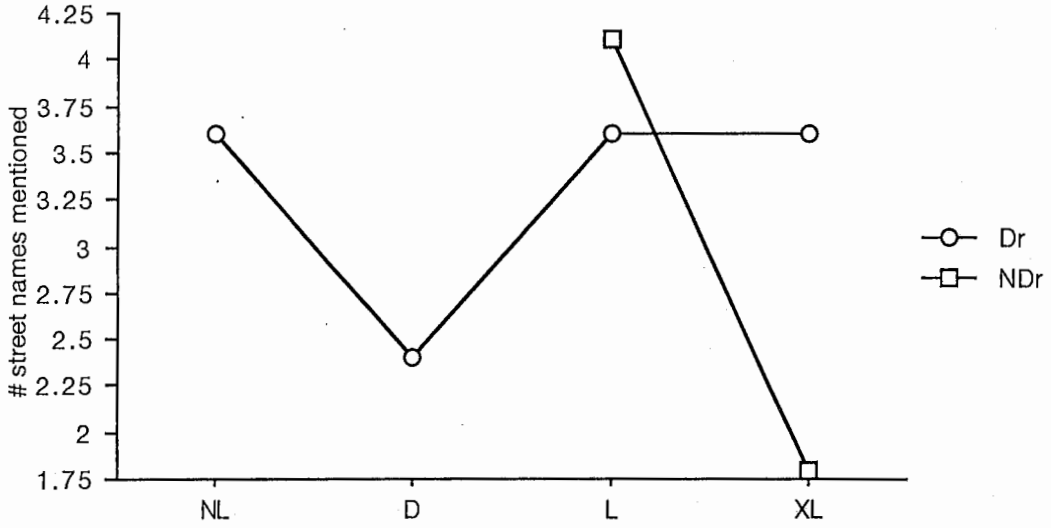

Figure 7. The average number of times street names occurred in written descriptions for each condition.

Focusing for a moment on the drawing conditions, we see the same sort of "repressing" effect that we saw above. That is, in conditions where there is no language at all or in which street names were mentioned, there is the same level of use of street names in the written descriptions. But in the deictic condition, in which there is a language description, but it does not include street names, subjects tended to use fewer street names in their descriptions. The use of street names seems to be one "natural" strategy for describing a route (hence the high level of the use of street names in the no-language condition), but when subjects heard a description in which that strategy was not employed, they also did not use it as much.

The behavior in the extra-language, no-drawing condition is puzzling. It is not clear why subjects should have abandoned the use of street names in this particular condition. Perhaps the lack of a drawing which would focus subjects' attention on specific streets and thus their names, allowed subjects to pay more attention to the landmarks mentioned in the extra-language description. They then reproduced these landmarks rather than street names in their written descriptions.

As for subjects' claims that it was extremely difficult to remember street names, they seem to have been accurate. The language description mentions 12 different street names; the extra-language description mentions 13. Subjects were able to reproduce less than a third of these names in their written descriptions, and 30% of these were incorrect. It is not clear whether this is due to the nature of the street names (since they were often similar, such as "Birch" and "Beech," they may have been difficult to remember), to the high density of street names on the map, or to something intrinsic in the nature of street names. Whatever the reason, street names were not an effective means by which our subjects could remember the route.

Landmarks. Subjects also commented quite frequently that they preferred to be given landmarks. To examine this assertion, we looked at the use of the word "brick" in the written descriptions of the subjects as a way of ascertaining how much use subjects actually made of landmark characteristics in their own descriptions. "Brick" does not occur in the deictic description; it occurs once in the language description and five times in

the extra language condition. Figure 8 shows the frequency with which it occurs in the written descriptions of the subjects.

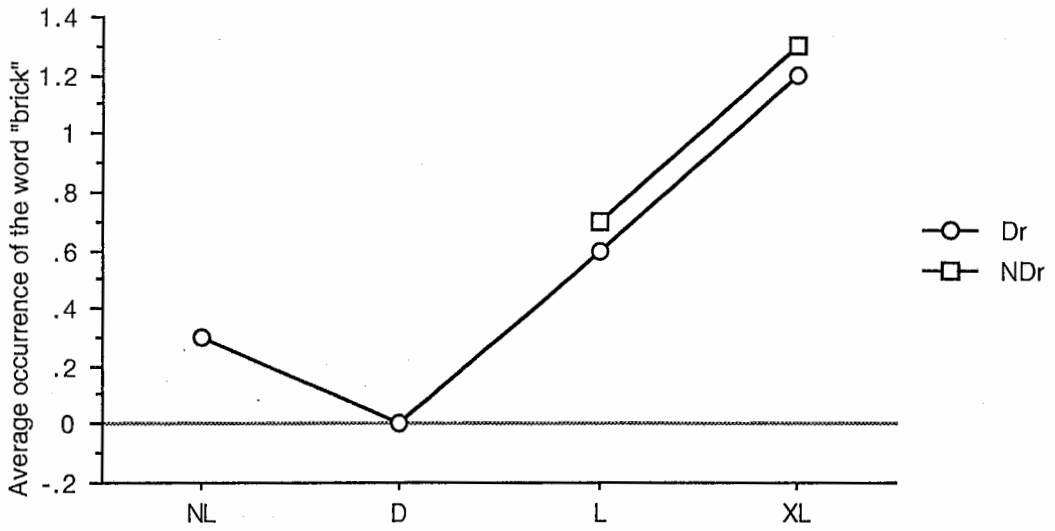

Figure 8. Average number of time the word "brick" occurs in the written descriptions for each condition.

In this case, the presence or absence of a drawing did not affect the results, which are similar to those noted above. Subjects not constrained by having heard a linguistic description (i.e., the no-language condition) used "brick" occasionally. This is consistent with subject comments that they prefer the use of landmarks in giving directions. Subjects who heard the word (the language and extra-language conditions) used it in proportion to the number of times they heard it in the description. Subjects who heard a language description which did not use the word "brick" (the deictic condition), did not use it in their descriptions.

At first glance, it appears that subjects were simply repeating what they heard. However, this is more than just a case of accommodation to the language description. First, subjects who did not hear the word (the no-language condition) did use the word spontaneously in their written descriptions in response to the visual stimuli of brick buildings on the map. Second, *even with those visual stimuli still present*, subjects did not use "brick" when the language description did not contain it. This is another illustration of the extent to which the language description guided the responses at least (and perhaps the perceptions) of the subjects. We have already referred to this phenomenon above in discussing the focusing function of the language descriptions.

Compass directions. The use of the word "north" is a purely linguistic example of the same sort of phenomenon. Figure 9 shows the distribution of the use of the word "north" over all conditions.

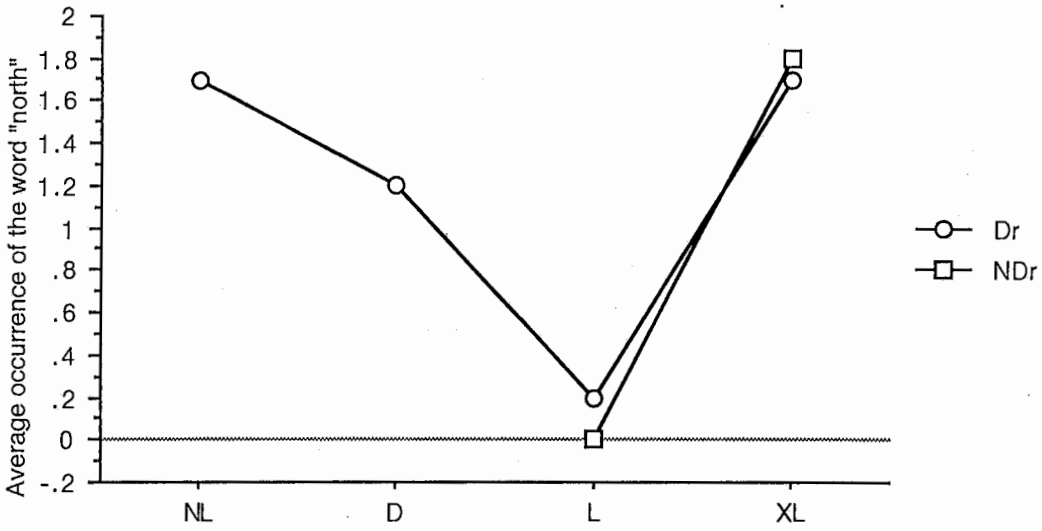

Figure 9. Average number of time the word "north" occurs in the written descriptions for each condition.

In this case, the word "north" does not occur at all in the language description, and occurs twice in the extra-language condition. The results for the no-language and the deictic conditions are consistent with the occasional comments made by subjects that they prefer using compass directions to using "right" or "left." Note that, unlike in the "brick" case, subjects hearing the deictic description did use "north." This may be because, even though there was no reference to "north" in the deictic description, there are phrases such as "turn this way." "This way" is neutral between "turn right" and "turn north;" thus, subjects were free to phrase the information in either way when writing their own descriptions. Notice that in the language condition, in which "north" did not appear, subjects did not use this word either, while they used it frequently after hearing the extra-language description in which it did occur.

In summary, then, these linguistic considerations illustrate once again the power of the language description to focus even visual attention (as in the case of "brick") in a route description setting. This effect can be manipulated in giving directions such that salient landmarks can be easily highlighted. Furthermore, we have seen the natural propensity of subjects to use compass directions and landmarks and to avoid the use of street names in reproducing directions. These are reasonable recommendations for the construction of verbal descriptions of routes.

Conclusions and future directions

We were somewhat surprised when we examined our initial data and discovered that subjects preferred to give directions using both verbal descriptions (either deictic descriptions or full language descriptions), *and* a drawing (Fais and Loken-Kim, 1995). However, the results of the experiment discussed above bear out this inclination: subjects wrote and drew more accurate representations of the route having heard a detailed language description; further, subjects far preferred to receive visual information about the route along with the verbal information. This is in fact a strong argument for the use of multimedia systems, not because one medium is more efficient than another and so will replace it, but because a number of media used in combination are the most effective way to present and receive information.

Clearly this can be only an initial conclusion. As discussed above, real users in real situations will most accurately dictate what media combinations are effective.

Furthermore, this experiment dealt with a small subset of the many uses to which multimedia systems are put; similar experimentation is necessary in other cases before more general conclusions can be reached. However, the suggestion that users prefer to present and receive information via multiple media is an important one; it indicates that we cannot think in terms of "splitting up" information to be presented in different modes, but must think about presenting the same information in supplementary ways.

Of course, there are still unanswered questions and future work. The motivation behind including an extra-language condition in this experiment was the thought that "too much" language might make it *more* difficult for subjects to reproduce the route accurately. Though this was not the case in this experiment, there are still indications that this is a real possibility. Two subjects in the extra-language condition commented on the "meaningless clutter of verbiage," and that "the additional descriptions of buildings started to seem irrelevant or distracting." It may have been the case that the reason the extra-language condition produced good results was that in that condition, relevant and useful landmarks were mentioned, whereas in the language description, street names were used, which, as we saw above, are not the optimal means for conveying directions. This is a question for future study.

Of course, there are numerous other considerations to explore. One is the effect of looking at the map even when there was no drawing. That in itself of course is visual input. What would be the effect if subjects not only did not see a drawing, but in fact did not see the map at all? We would conjecture that subjects' performance on reproducing the route would be fairly poor, but the study remains to be done.

One subject commented that "the voice directions are somewhat confusing in the fact that they are ahead of the line drawing." In fact, the verbal directions and the drawings were timed to be fairly synchronous; however, the subject's comment brings up another interesting question. In examining our initial data, we noted that when subjects do make drawings that accompany language descriptions, the drawing and description are synchronous only about half of the time. The rest of the time, the drawing tends to come before the description (Loken-Kim et al., 1995). Perhaps the subject felt that the presentation was odd because he would have preferred the drawing to come first and the verbal description to come later. This would be another possible way to coordinate the presentation of visual and verbal information.

Other possibilities also suggest themselves. Circling landmarks instead of drawing a line would combine the subjects' preference for using landmarks with a balance between using visual information and yet allowing subjects to construct their own routes mentally. It caters to subjects' preference for visual information, yet gives them an active role in constructing the information they are receiving which, as we saw above, seems to lead to the best reproduction of that information.

Whatever configuration of media is used in future systems, the main finding of this experiment suggests that we must take the attitude, not that one medium will be responsible for one type of information and replaced by another medium for another type of information, but that information is most effectively conveyed between humans in a way that seems redundant to a machine, that is, through a variety of media in concert. Our task is to discover what combinations of media best meet the needs of users of multimedia systems.

References

Fais, L. 1994. Effects of communicative mode on spontaneous English speech. Technical Report of the IEICE, **NLC94-22,**(1994-10), Tokyo, Japan, pp. 1-8.

Fais, L., S. Mizunashi, and K-H. Loken-Kim. 1996. EMMI progress report: An evaluation of research done with the first EMMI interface. ATR Technical Report TR-IT-0172, ATR ITL, Kyoto, Japan.

Fais, L., and K-H. Loken-Kim. 1995. How many words is a picture really worth? Technical Report of the IEICE, SP95-81(1995-12) , The Institute of Electronics, Information and Communication Engineers.

Fais, L., K-H. Loken-Kim, and Y.D. Park. 1995. Speakers' Responses to Requests for Repetition in a Multimedia Cooperative Dialogue, in Pro., International Conference on Cooperative Multimodal Communication, (Eindhoven, The Netherlands, May 24-26, 1995), pp. 129-144.

Loken-Kim, K-H., F. Yato, K. Kurihara, L. Fais, and R. Furukawa. 1993. EMMI - ATR environment for multi-modal interactions. ATR Technical Report TR-IT-0018, ATR ITL, Kyoto, Japan.

Loken-Kim, K-H., S. Mizunashi, M. Tomokiyo, L. Fais, and T. Morimoto. 1995. Analysis and integration of multimodal inputs in interpreting telecommunications. IPSJ Workshop.

Park, Y., K-H. Loken-Kim and L. Fais. 1994. An experiment for telephone versus multimedia multimodal interpretation: methods and subjects' behavior. ATR Technical Report TR-IT-0087, ATR ITL, Kyoto, Japan.

Park, Y., K-H. Loken-Kim, L. Fais, and S. Mizunashi. 1995. Analysis of gesture behavior in a multimedia interpreting experiment; human vs. Wizard of Oz interpretation method. ATR Technical Report TR-IT-0091, ATR ITL, Kyoto, Japan.

Appendix: Language descriptions of the route

Deictic

Walk out of this house and turn this way. Turn like this and walk until you get to here. Then turn this way and walk to here. At this point, turn like this and walk here, go over this and continue over this. Turn this way and go over this. Cross this way and follow this to here, and then turn this way. Turn like this at this point and walk until you get here, and then turn this way. Turn this way and then like this here and once more like this, and you are there.

Language

Walk out of the house that is in front of the pool and turn left on Evergreen Street. Turn right on Oak Street and walk until you get to the brick building. Then turn left onto Willow Street and walk to the T-junction. At the T-junction, turn right onto Elm street and walk along the building, go over the overpass, and continue over the bridge. Turn left at Birch Street and go over another bridge. Cross President Way and follow Coolidge Street to the next intersection and then turn left onto Roosevelt Street. Turn right at the intersection onto Hoover Street and walk until you get to where it makes a T-junction, and then make a right onto Eisenhower Street. Make a left onto Truman Street and then the next left after that onto Adams Street and one more left into the driveway and you're at the Post Office.

Extra Language

Walk out of the house that is in front of the pool, the one opposite the striped building, and turn left on Evergreen Street so that you are walking north. Turn right on Oak Street--you will have a white building on your right and grey building on your left-- and walk until you see the large brick building in front of you. Then turn left onto Willow Street, walking along the brick building, again north, and walk to the T-junction where you'll see another brick building with a parking lot to the left. At that T-junction, turn right so that you are walking between the two brick buildings on Elm street, go over the overpass that goes over Route 151, walk around the park, and continue over the bridge that goes over the little stream from the pond in the park. Turn left at Birch Street, which is the second street after you walk over the bridge and go over another bridge over the same stream; you will also walk by two houses on your left: one grey and one brick. Cross President Way and bear left onto Coolidge Street and then follow Coolidge Street to the next intersection where you should turn left onto Roosevelt Street. Turn right at the next intersection (which is the only direction you can go; you can't make a left there); that is Hoover Street and you should walk until you get to where it makes a T-junction, and then make a right onto Eisenhower Street. Make a fairly quick left onto Truman Street and then the very next left after that onto Adams Street; at this point you should see the Post Office on your left and so if you make one more left into the driveway, you're at the Post Office.