Picture Annotation as a Communication Method for Non-Speaking People

Hine N.A, Wilkinson D, Gordon I.A.S. & Arnott J.L.

MicroCentre, University of Dundee, Dundee, UK.

Abstract. The task of the IPSNI II project has been to investigate the accesibility of multimedia telecommunications services for people with disabilities. Where problems were found, the project was to develop and demonstrate examples of hardware, software and service adaptations solutions that show how accessibility could be improved. Within this project, the University of Dundee has concentrated on seeking to enable non-speaking people to use a videophone service. Work has focussed primarily on the use of text by these people. Given that future multimedia systems could have picture display and annotation functionality, it was decided to investigate this possibility as an alternative communication medium for non-speaking people. It was found that under the right circumstances, for users with poor language skills, picture annotation could be more than 4 times faster than typing the equivalent text.

1. Introduction

Previous investigations by DUMC in the IPSNI II project of adaptations to a videophone for people who have difficulty speaking has concentrated on using text as the medium for exchanging information. It has become evident, however, that many non-speaking people have additional motor impairments and many non-speaking adults do not have good language skills. For this reason, an alternative medium to text was sought that could be used to convey some useful information during a videophone conversation. The alternative medium that was chosen was that of still picture display. This was added to an adapted videophone emulated service by adding a picture display service component. In addition, a set of simple picture annotation tools were added to the service to allow comments to be made over the top of the picture.

2. Study

2.1 Hypothesis & Goal

The study is intended to investigate the value of adding a picture display and annotation service component to a videophone service to allow people who have difficulty speaking to pass information to a communicating partner at the other end of a telecommunications link. In view of the fact that many adults with speaking impairments who have additional motor impairmentsand as a consequence have poor language skills through lack of practice in the use of language, the ability to display and annotate a picture is seen as offering an alternative information channel. In addition, the time taken to compose text information may be so great for some people with severe motor impairments that they could benefit from being able to highlight elements from a picture with simple annotation tools.

The hypothesis that was to be tested was that for users with impaired speech and severe motor impairments, it is quicker to annotate significant elements on a picture than it is to type the same information in words.

2.1 Method

In order to test this hypothesis, a multimedia communications interface was constructed that allowed information to be exchanged as text, speech and audio, real time full motion video, and still picture with annotation. The interface was constructed using the SuperCard rapid prototype programming environment. In addition, the interface could be adapted by adding a scanning on-screen keyboard & mouse emulator, a direct selection on-screen keyboard or text prediction, depending on the specific requirements of each user. Once the screen layout had been configured for each user, it could be saved and recalled whenever

that user used the system.

An example of an adapted interface with on-screen keyboard and text prediction is shown in figure 1 below.

A scripted scenario was constructed where each subject was required to try to arrange a holiday in New York State by discussing the holiday with a travel operator at the other end of a two way telecommunications link. As they did so they would not only type some contributions to the conversation, but would indicate to the travel operator the location of places of interest by marking a map of the state.

Figure 1: An adapted multimedia telecommunications interface

As the experiment was conducted and the script was followed, three measurements were taken:

1) the typing speed of the subjects was logged by recording the time each character selection was made. From this, an average Characters per Second (CPS) typing speed was calculated.

2) the time taken to draw a circle around three sites of interest, that is New York City, Niagara Falls and Toronto.

3) The time taken to draw two arrows on the picture to mark the location of motels at Albany and Syracuse.

Each object that was created in the interface, including text and picture annotation objects was recorded with an object description and the time that the object was created. In addition, a video of the session was recorded showing both the subject and the screen.

The interpretation of the annotation depends on a number of factors, including the comments from both parties that preceded an annotation act, the context of the conversation and the relevance of the picture to it, the type of information being conveyed. Because of this, it is not possible to suggest exactly the text that the annotation may be replacing in a conversation. For this reason, each annotation act was compared with two possible pieces of text. These are described below.

1) The annotation of New York City, Niagara Falls and Toronto may have followed a question such as "Where do you want to visit?", in which case the annotation replaced the act of typing "New York City, Niagara Falls and Toronto", a text block of 40 Characters. Alternatively, the subject may have had to be more explicit with a text block saying something like "I would like to visit New York City, Niagara Falls and Toronto", a text block of 61 characters. Using the CPS value calculated for each person, the time taken to draw the circles was compared with the time that it would have taken the person to type both these sentences. The sentences are referred to as the short sentence and the long sentence respectively.

2) The indication of the arrows to mark motel locations could be preceded by a question such as "Where would you like a motel?", in which case the annotation would simply replace the text "At Albany and Syracuse", a text block of 22 characters. Alternatively, the subject may have had to be more explicit with a text block saying something like "I would like stay in a motel at Albany and Syracuse", a text block of 51 characters. Using the CPS value calculated for each person, the time taken to draw the arrows was compared with the time that it would have taken the person to type both these sentences. The sentences are referred to as the short sentence and the long sentence respectively.

2.2 Subjects

Three non-speaking people with additional motor impairments took part in the study. Subject 1 used a roller ball to move from place to place on the screen and to move to keys on an on-screen keyboard. She moved the rollerball and selected the "mouse" button with her chin. Subject 2 had very slow movements as a result of muscular dystrophy and also used a rollerball but rolled it with a finger. Subject 3 used a switch by his head to select keys or mouse emulator controls from a scanning keyboard and mouse emulator.

All the subjects were regular users of the computer system on which the experiment was conducted, and were used to handling their chosen adaptation. A practice run was conducted with each subject in order to check that the adaptations were set up properly and to allow the subjects to become familiar with the operation of the system.

2.3 Apparatus

The system used for the experiments consisted of two Apple Macintosh computers connected by an ethernet network. Each terminal had a Flexcam camera which has a stereo microphone built in. The audio was fed to a pair of loudspeakers beside the terminal of the other communicating party, and the video was displayed on the screen using a video overlay function built into the computer. Text, pictures and annotation changes were conveyed from one terminal to the other via the ethernet network. The two terminals were situated in different rooms which were acoustically isolated from each other.

3. **Results**

From the typing conducted during the study, the subjects were found to be able to type, with relevant adaptations, at an average rate of:

Subject 1 0.045 CPS
22.4 secs per character
Subject 2 0.18 CPS
5.55 secs per character
Subject 3 0.066 CPS
15.15 secs per character

The time taken for the subjects to perform the two annotations and to type the equivalent sentences are shown in table 1 and 2 below.

Subject	Annotation Time/Secs	Short Sentence/Secs	Long Sentence/Secs
Subject 1	143	896	1388
Subject 2	240	222	345
Subject 3	577	606	939

Table 1: Time taken to communicate intention to visit New York City, Niagara Falls and Toronto

Subject	Annotation Time/Secs	Short Sentence/Secs	Long Sentence/Secs
Subject 1	Subject did not do this.	493	1142
Subject 2	76	122	283
Subject 3	335	318	773

Table 2: Time taken to communicate requirement for a motel at Albany and Syracuse

Figure 2: Comparison of annotation and typing times to communicate intention to visit New York City, Niagara Falls and Toronto	Figure 3: Comparison of annotation and typing times to communicate requirement for a motel at Albany and Syracuse

4. **Discussion & Conclusions**

In general, the results show that the addition of the picture display and annotation function was a valuable tool for all three users, particularly where the information being conveyed removed the requirement to provide explanatory or introductory text. In general, the amount of time to type the facts or to highlight them

on the picture was about the same.

The clear exception to this general conclusion was subject one, who has a very low typing speed. This is due to rather poor language skills, manifest in poor spelling and reading (including recognising words on the predictor) and poor recognition of the letters on the on-screen keyboard. This may have been compounded by the fact that the layout of the keys was according to the QWERTY keyboard layout, a layout that she would not have been very familiar with. Some improvement could be expected if the letters were arranged alphabetically.

In conclusion, it seems from this preliminary study that:

1) For those users with reasonable language skills and a speech and motor impairment, annotation of a picture takes about as much time as typing the same facts, but may save significant amounts of time if explanation is not needed in the annotation but it would be if text is given.

2) For those users with poor language skills in addition to speech and motor impairments, picture annotation can save a significant amount of time.

For this reason, picture annotation is seen as a valuable additional service component when making a videophone accessible to non-speaking people.

Future work will attempt to confrim these conclusions, and try to determine the combination of circumstances and impairments that could benefit from picture annotation in addition to text. This could form the basis of future adaptations to multimedia telecommunications services.

5. References

[1] Blattner M. M. and Dannenberg R. B., (1992) “Multimedia Interface Design”, Addison-Wesley Publishing Company

[2] Neilson I. and Lee, J. (1994) Conversations with Graphics - Implications for the design of natural-language graphics interfaces, International Journal of Human-Computer Studies, Vol 40, No. 3

[3] Schneiderman B., (1992) “Designing the User Interface”, Addison-Wesley Publishing Company

[4] Weisbecker A., Machate J. and Koller F., (1993) “Guidelines & Rules for Development of MADE Multimedia Applications”, FhG-IAO, Deliverable of Esprite Project 6307 (MADE1)