A Speech Centric Mobile Multimodal Service useful for Dyslectics and Aphasics

Knut Kvale and Narada Warakagoda

Telenor R&D, N-1331, Fornebu, Norway
{knut.kvale}{narada-dilp.warakagoda}@telenor.com

Abstract

Multimodality has the potential of benefiting non-disabled as well as disabled people. We have developed a speech-centric composite multimodal interface to a map-based information service on a mobile terminal. This interface to the service has proven useful for a severe dyslectic and an aphasic. These persons did not manage to use the ordinary public information service, neither on the web (text only) nor by calling a manual operator phone (speech only). But they fairly easily employed our multimodal interface by pointing at the map on the touch screen while uttering short commands or phrases. Although this is a limited qualitative evaluation it indicates that multimodal interfaces to information services is a step in the right direction for achieving the goal of inclusive design or design for all (DfA).

1. Introduction

Today accessibility to web based information services is limited for many people with sensory impairments. A main obstacle is that the input and output channels of the services support one modality only. It is claimed that the missing access to environments, services and adequate training contributes more to the social exclusion of disabled people than their living in institutions [1].

There are two different approaches to solving this problem. One is to develop special assistive technology devices which compensate for or relieve the different disabilities. Another solution is to design services and products to be usable by everybody, to the greatest extent possible, without the need for specialized adaptation; so-called *design for all* (DfA) *or inclusive design*. An example of applying the DfA-principle is to equip electronic services and applications with intelligent modality adaptive interfaces that let people choose their preferred interaction style depending on the actual task to be accomplished, the context, and their own preferences and abilities.

Lately some efforts have been made to explore the usability of multimodal interfaces for disabled people. For instance in [2] a multimodal communication aid was tested with a global aphasia patient.

To test the hypothesis that multimodal inputs and outputs really are useful for disabled people, we have developed a flexible multimodal interface to a public web-based bus-route information service for the Oslo area. The original public service on the web, which has both HTTP and WAP interfaces, is text based (i.e. unimodal only). The users have to write the names of the arrival and departure stations to get the route information, which in turn is presented as text. Our multimodal interface for small mobile terminals converts the web service to a map-based multimodal service supporting speech, graphic/text and pointing modalities as inputs. Thus the users can choose whether to use speech or point on the map, or even use pointing and talking simultaneously (so-called composite multimodality) to specify the arrival and departure stations. The response from the system is presented as both speech and text. We believe that this multimodal interface gives a freedom of choice in interaction pattern for all users. For normal able-bodied users this implies enhanced user-friendliness and flexibility in the use of the services, whereas for the disabled users this is a means by which they can compensate for their not-well-functioning communication mode.

In a previous study five test persons with different impairments tested out our multimodal service [3,4]. In this study persons with muscular atrophy combined with some minor speaking problems had great benefit from speaking short commands or phrases while pointing on the maps, i.e. the composite multimodality. This evaluation motivated us to let a dyslectic and an aphasic person test the service. Our evaluations were performed in cooperation with Bretvedt Resource Centre [5].

This paper first briefly describes the multimodal system architecture and how the bus information system works. Then the user evaluations by the dyslectic and aphasic test user are described and discussed.

2. System and Service

2.1. System architecture

Our multimodal bus information system has a client server architecture based on the Galaxy communicator [6]. The server part of the system consists of six separate modules and a facilitator hub module. All the server side modules run on a PC, while the client runs on a PDA, here a Compaq iPAQ. The client consists of two main components handling voice and graphical (GUI) modalities. It communicates with the server over a wireless local area network (WLAN) based on the IEEE 802.11b protocol, hence making the service mobile. The server communicates with the "Trafikanten" web service (http://www.trafikanten.no) through the Internet to get the necessary bus-route information. All computationally heavy components including the speech recogniser (Scansoft SpeechPearl 2000 for Norwegian) and the speech synthesizer (Telenor Talsmann) run on the server. More details of the system architecture can be found in [7, 8, 9].

Figure 1 shows typical screen sequences for a user with reduced speaking ability who wants to go from “Fornebu” to “Jernbanetorget”.

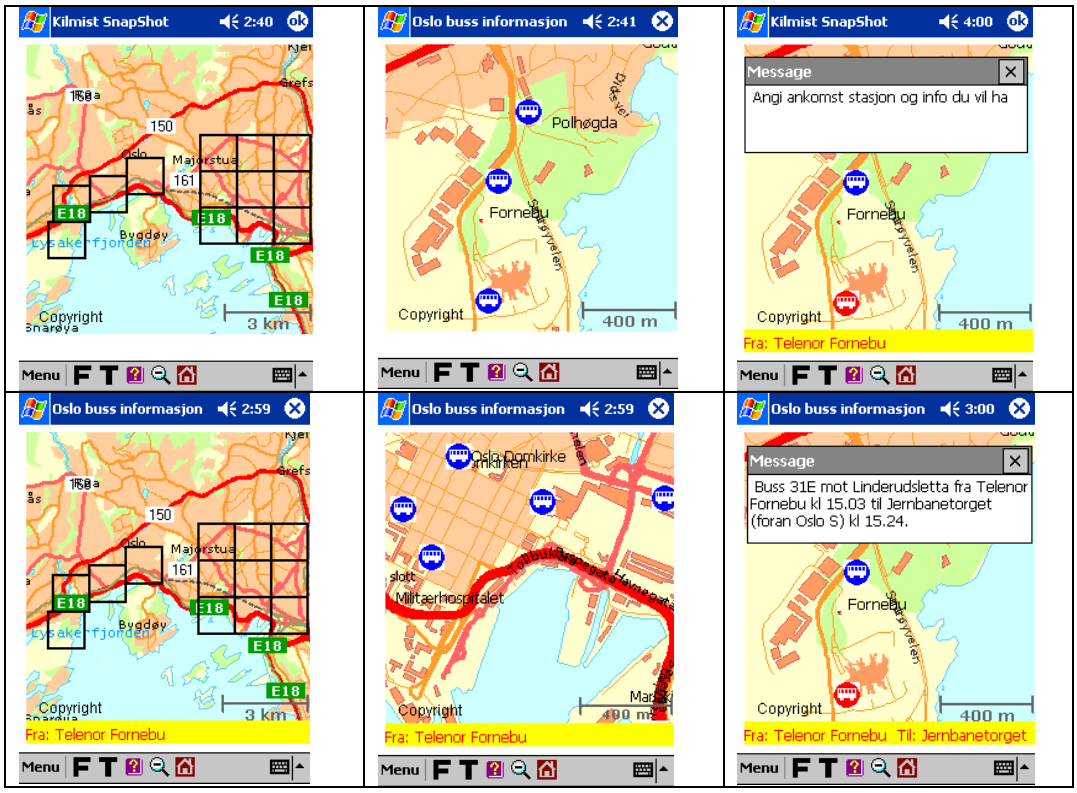

Figure 1 A typical screen sequence for a user with reduced speaking ability. 1) Overview map: The user taps on the submap (the square) for Fornebu 2) The user says "next bus here Jernbanetorget" and taps on bus station Telenor 3) The system does not recognize the arrival station. Therefore the user selects it by using pen. But first the user taps on the zoom-out button to open the overview map 4) The user taps on the submap, where bus station Jernbanetorget lies 5) The user taps on the bus station Jernbanetorget 6) The user can read the bus information

2.2. Using the service

The interface of our multimodal service is provided by the client application running on a mobile terminal. When the client is started, and connected to the server, the main page of the server is presented to the user. This is an overview map of the Oslo area where different sub-areas can be zoomed into, as shown in figure 1. Once zoomed, it is possible to get the bus stations in the area displayed. The user has to select a departure station and an arrival station to get the bus route information. The users are not strictly required to follow the steps sequentially. They can e.g. combine several of them, whenever it makes sense to do so.

Our service provides both *simultaneous inputs* (i.e. the speech and pointing inputs are interpreted one after the other in the order that they are received) and *composite inputs* (i.e. the speech and pointing inputs at the "same time" are treated as a single, integrated compound input by downstream processes), as defined by W3C [10]. Users may also communicate with our service monomodally, i.e. by merely pointing at the touch sensitive screen or by speech only. The multimodal inputs can be combined in several ways, for instance:

- The user says the name of the arrival bus station and points at another bus station on the map, e.g.: "I want to go from Jernbanetorget to here"
- The user points at two places on the screen while saying: "When does the next bus go from here to here"

In both scenarios above the users point at a bus station within the same time window as they utter the underlined word, "here". In order to handle two pointings within the same utterance, we defined an asymmetric time window within which speech and pointing are treated as a composite input if:

- Speech is detected within 3 seconds after a pointing
- Pointing is detected 0.85 second before the speech signal ends

Both pointing and speech can be used in all operations including navigation and selecting bus stations. Thus the user scenarios can embrace all the possible combinations of pointing and speech input. The received bus route information is presented to the user as text in a textbox and this text is also read aloud by synthetic speech.

Thus we expect that the multimodal service may prove useful for many different types of disabled users, such as:

- Persons with hearing defects and speaking problems may prefer the pointing interaction.
- Blind persons may only use the pure speech-based interface
- Users with reduced speaking ability may use a reduced vocabulary while pointing at the screen.

3. User evaluations

Since disabled users may have low self confidence we tried to create a relaxed atmosphere and we spent some time having an informal conversation before the persons tried out the multimodal service. In these evaluations the test persons brought relatives with them. The dyslectic user had his parents with him, while the aphasic was accompanied by his wife. The evaluation situation may still have been perceived as stressful for them since two evaluators and two speech therapists were watching. This stress factor was especially noticeable for the young dyslectic.

The multimodal interaction pattern was new to the users and it was necessary to explain this functionality for them. It has been shown [11] that different introduction formats (video versus text) have impact on how new users actually use a multimodal service. For this evaluation we applied so-called *model based learning*, where a trusted supervisor first showed how he used the service and carefully explained the functionality.

3.1. In-depth evaluation of a severe dyslectic test user

Dyslexia causes difficulties in learning to read, write and spell. Short-term memory, concentration, personal organisation and sequencing may be affected. About 10% of the population may have some form of dyslexia, and about 4% are regarded as severely dyslexic [12].

Our dyslectic test person was fifteen years old and had severe dyslexia. He could, for instance, not read the names of the buses. Therefore he was very uncertain and had low self-confidence. He was not familiar with the Oslo area. Thus we spent more than an hour discussing, explaining and playing with the multimodal system. The dyslectic sat beside his trusted supervisor/speech therapist who showed him how to ask by speech only for bus information from "Telenor" to "Jernbanetorget". The speech therapist repeated and rephrased the query: "Bus from Telenor to Jernbanetorget" at least five times, and the dyslectic was attentive.

However, when we asked the dyslectic test person to utter the same query, he did not remember what to ask for. Therefore we told him the two bus station names he could ask for: "From Telenor to Jernbanetorget". He had however huge problems with remembering and pronouncing these names, especially "Jernbanetorget" because it is a long word. Hence we changed the task to asking for the bus route information: "From Telenor to Tøyen", which were easier for him. But he still had to practise a couple of times to manage to remember and pronounce these two bus stations.

Then he learned to operate the PDA and service with pointing only. After some training, he had no problem using this modality. He quickly learned to navigate between the maps by pointing at the "zoom"-button. The buttons marked **F** and **T** were intuitively recognised as **F**rom station and **T**o station respectively.

Now we told him that it is not necessary to formulate full sentences when talking to the system, one word or a short phrase is enough to trigger the dialogue system. He then hesitatingly said "Telenor". The system responded with "Is Telenor your from station?", and he answered "yes". In situations where the system did not understand his confirmation input, "yes", he immediately switched to pointing at the "yes" alternative on the screen (he had no problem with reading short words). If the bus station has a long name he could find it on the map and select it by pen instead of trying to use speech.

Finally we introduced the composite multimodal input functionality. We demonstrated queries as: "from here to here" simultaneously tapping the touch screen and saying "here". The dyslectic then said "from here" and pointed at a bus station shortly afterwards. Then he touched the 'zoom out' button and changed map. In this map he pointed at a bus station and then said: "to here". This request was correctly interpreted by the system which responded with the bus route information. Both the speech therapists and the parents were really surprised by how well the young severe dyslectic boy managed to use and navigate this system. His father concluded: "When my son learned to use this navigation system so quickly - it must be really simple!".

3.2. In-depth evaluation of an aphasic test user

Aphasia refers to a disorder of language following acquired brain damage, for example, a stroke. Aphasia denotes a communication problem, which means that people with aphasia have difficulty in expressing thoughts and understanding spoken words, and they may also have trouble reading, writing, using numbers or making appropriate gestures.

About one million Americans struggle with aphasia [13]. There is no official statistics of the number of aphasic persons in Norway. Approximately 12000 suffer stroke each year and it is estimated that about one third results in aphasia. In addition, accidents, tumours and inflammations may lead to aphasia, giving a total of about 4000-5000 new aphasia patients each year in Norway.

Our test person suffered a stroke five years ago. Subsequently he could only speak a few words and had paresis in his right arm and leg. During the first two years he had the diagnosis global aphasia, which is the most severe form of aphasia. Usually this term applies to persons who can only say a few recognizable words and understand little or no spoken language. Our test person is no longer a typical global aphasic. He has made great progress, and now he speaks with a clear pronunciation and prosody. However, his vocabulary and sentence structure are still restricted, and he often misses the meaningful words - particularly numbers, important verbs and nouns, such as names of places and persons. He compensates for this problem by a creative use of body language and by writing numbers. He sometimes writes the first letter(s) of the missing word and lets the listener guess what he wants to express. This strategy worked well in our communication. He understands speech well, but may have problems interpreting composite instructions. He is much better at reading and comprehending text than at expressing what he has read.

Because of his disfluent speech, characterized by short phrases, simplified syntactic structure, and word finding problems, he might be classified as a *Broca's aphasic*, although his clear articulation does not fit completely into this classification.

He is interested in technology and has used a text-scanner with text-to-speech synthesis for a while. He knew Oslo well and was used to reading maps. He very easily learned to navigate with the pen pointing. He also managed to read the bus information appearing in the text box on the screen, but

he thought that the text-to-speech reading of the text helped his comprehension.

His first task in the evaluation was to get bus information for the next bus from "Telenor" to "Tøyen" by speaking to the service. These stations are on different maps and the route implies changing buses. Therefore, for a normal user, it is much more efficient to ask the question than pointing through many maps and zooming in and out. But he did not manage to remember and pronounce these words one after the other.

However, when demonstrated, he found the composite multimodal functionality of the service appealing. He started to point at the from-station while saying "this". Then he continued to point while saying "and this" each time he pointed - not only at the bus stations but also at function buttons such as "zoom in" and when shifting maps. It was obviously natural for him to talk and tap simultaneously. Notice that this interaction pattern may not be classified as a composite multimodal input as defined by W3C, because he provided exactly the same information with speech and pointing. We believe, however, that if we had spent more time in explaining the composite multimodal functionality he would have taken advantage of it.

He also tried to use the public bus information service on the web. He was asked to go from "Telenor" to "Tøyen". He tried, but did not manage to write the names of the bus stations. He claimed that he might have managed to find the names in a list of alternatives, but he would probably not be able to use this service anyway due to all the problems with reading and writing. The telephone service was not an alternative for him at all because he was not able to pronounce the station names. But he liked the multimodal tap and talk interface very much and characterised it spontaneously as "Best!", i.e. the best alternative for him to get the information needed.

4. Conclusions

Our composite multimodal interface to a map-based information service on a mobile terminal has proven useful for a severe dyslectic and an aphasic. The severe dyslectic and aphasic could neither use the public service by speaking and taking notes in the telephone-based service nor by writing names in the text-based web service. But they could easily point at a map while uttering simple commands. Thus, the multimodal interface is the only alternative for these users to get web information.

These qualitative evaluations of how users with reduced ability interacted with the multimodal interface are by no means statistically significant. We are aware that there is big variation among aphasics, and even the performance of the same person may vary from one day to the next. Still, it seems reasonable to generalise our observations and claim that for severe dyslectic and certain groups of aphasics a multimodal interface may be the only useful interface to public information services such as bus timetables. Since most aphasics have severe speaking problems they probably will prefer to use the pointing option, but our experiment indicates that they may also benefit from the composite multimodality since they can point at the screen while saying simple supplementary words.

Our speech-centric multimodal service allowing all combinations of speech and pointing has therefore the potential of benefiting non-disabled as well as disabled people, and thereby achieving the goal a common of design for all.

5. Acknowledgements

We would like to express our thanks to Tone Finne, Eli Qvenild and Bjørgulv Høigaard at Bredtvet Resource Centre for helping us with the evaluation and for valuable and fruitful discussions and cooperation.

We are grateful to Arne Kjell Foldvik and Magne Hallstein Johnsen at the Norwegian University of Science and Technology (NTNU) for inspiration and help with this paper.

This work has been financed by the BRAGE-project of the research program "Knowledge development for Norwegian language technology" (KUNSTI) of the Norwegian Research Council..

6. References

[1] ETSI, ``Human Factors (HF); Multimodal interaction, communication and navigation guidelines'', ETSI EG 202 191 v1.1.1., 2003.

[2] Pedersen, J.S., Dalsgaard, P., Lindberg, B., ``A Multimodal Communication Aid for Global Aphasia Patients''. Proc. Interspeech 2004 – ICSLP, Jeju Island, Korea. 2004

[3] Kristiansen, M., *Evaluering og tilpasning av et multimodalt system på en mobil enhet*, Master thesis NTNU (in Norwegian) 2004.

[4] Kvale, K., Warakagoda, N. D, Kristiansen, M., ``Evaluation of a mobile multimodal service for disabled users'', Proc. the 2nd Nordic Conference on Multimodal Communication Gothenburg, Sweden, 2005.

[5] Bredtvet Resource Centre. http://www.statped.no/bredtvet

[6] Galaxy communicator. http://fofoca.mitre.org/

[7] Kvale, K., Warakagoda, N.D. and Knudsen, J.E., ``Speech centric multimodal interfaces for mobile communication systems'', Telektronikk nr.2, pp. 104-117, 2003

[8] Kvale, K., Knudsen, J.E., and Rugelbak, J., ``A Multimodal Corpus Collection System for Mobile Applications'', Proc. Multimodal Corpora - Models of Human Behaviour for the Specification and Evaluation of Multimodal Input and Output Interfaces, Lisbon, pp. 9-12, 2004.

[9] Warakagoda, N. D., Lium, A. S. and Knudsen J. E., ``Implementation of simultaneous co-ordinated multimodality for mobile terminals'', The 1st Nordic Symposium on Multimodal Communication, Copenhagen, Denmark, 2003.

[10] W3C "http://www.w3.org/TR/2003/NOTE-mmi-reqs-20030108/" presence verified 04/5/05.

[11] Kvale, K., Rugelbak, J., Amdal, I.,. ``How do non-expert users exploit simultaneous inputs in multimodal interaction?'', Proc. International Symposium on Human Factors in Telecommunication, Berlin, pp.169-176, 2003.

[12] Dyslexia Institute. http://www.dyslexia-inst.org.uk/ presence verified 04/5/05.

[13] Brody, J.E., ``When brain damage disrupts speech'', In The New York Times Health Section, p. C13, June 10, 1992.