Speech centric multimodal interfaces for disabled users

Knut Kvale* and Narada Dilp Warakagoda
Telenor R&I, N-1331, Fornebu, Norway

Abstract. This paper explores how multimodal interfaces make it easier for people with sensory impairments to interact with mobile terminals such as PDAs and 3rd generation mobile phones (3G/UMTS).

We have developed a flexible speech centric composite multimodal interface to a map-based information service on a mobile terminal. This user interface has proven useful for different types of disabilities, from persons with muscular atrophy combined with some minor speaking problems to a severe dyslectic and an aphasic. Some of the test 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 development of speech centric multimodal interfaces to information services is a step in the right direction for achieving the goal of design for all.

Keywords: Speech centric multimodality, mobile interface design, design for all, disabled users

1. Introduction

1.1. Why multimodal interfaces?

Today accessibility to web based information services is still 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 [10].

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) [9,10]. An example of applying the principles of DfA is to equip mobile terminals 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.

Multimodal human-computer user interfaces are able to combine different input signals, extract the combined meaning from them, find requested information and present the response in the most appropriate format. Hence, a multimodal human-computer interface offers the users an opportunity to choose the most natural interaction pattern. If the preferred mode fails in a certain context or task, users may switch to a more appropriate mode or they can combine modalities.

We believe that multimodal interfaces offer 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 (see e.g. [21,30]), whereas for the disabled users this is a means by which they can compensate for their not-well-functioning communication mode.

1.2. From computers to mobile terminals

In the last two decades there has been a huge amount of research within multimodal user interfaces. In 1980,

*Corresponding author. Tel.: +47 91664177; E-mail: knut.kvale @telenor.com.

ISSN 1055-4181/08/$17.00 © 2008 – IOS Press and the authors. All rights reserved

Bolt [5] presented the “Put That There” concept demonstrator, which processed speech in parallel with manual pointing during object manipulation. Since then major advances have been made in speech recognition algorithms and natural language processing, in handwriting and gesture recognition, as well as in speed, processing power and memory capacity of the computers. Today’s multimodal systems are capable of recognizing and combining a wide variety of signals such as speech, touch, manual gestures, gaze tracking, facial expressions, head and body movements. The response can be presented by e.g. facial animation in the form of human-like presentation agents on the screen in a multimedia system. These advanced systems need various sensors and cameras and a lot of processing power and memory. They are therefore best suited for interaction with computers and in kiosk applications, as demonstrated in e.g. [4,12,18,26,28,31,34,36].

For telecommunication services on small mobile terminals, with limited size and processing power, the multimodal functionality is usually restricted to *two* input modes: speech (audio) and touch, and *two* output modes: audio and vision. This type of multimodality, sometimes called tap and talk, is essentially *speech centric*, and will be explored further in this paper.

1.3. Testing by disabled

Although several multimodal systems have been developed, little effort has been spent on exploring the usability of multimodal interfaces for disabled people. One exception is [32] where a multimodal communication aid was tested with a global aphasia patient. Another example is [18] where speech was combined with head tracking for hands free work with computers.

Hence, to test the hypothesis that multimodal inputs and outputs really are useful for disabled people, we have developed a flexible speech centric multimodal interface on mobile terminals to a public web-based bus-route information service for the Oslo area. In this paper we focus on the user experiments where people with various disabilities and sensory impairments have applied this service.

1.4. Outline of this paper

This paper first discusses speech centric multimodal human-computer interfaces for mobile terminals. Section 3 describes the system architecture of our multimodal interface to the public web-based bus-route information service for the Oslo area. Section 4 describes the user evaluations of the system by five test persons with different impairments, as well as a dyslectic and an aphasic test user. Conclusions are provided in Section 5.

2. Speech centric multimodal interfaces on mobile terminals

In most speech centric multimodal interfaces on mobile terminals, the input combines and interprets spoken utterances and pen gestures (pointing, circling and strokes). The output information is either speech (synthetic or pre-recorded) or text and graphics.

2.1. Pen and speech are complementary

Speech centric multimodality utilises the fact that the pen/screen and speech are complementary: The advantage of pen is typically the weakness of speech and vice versa. With speech it is natural to ask one question containing several key words, but it may be tedious to listen to all information read aloud because speech is inherently sequential. With pen/graphics only, it may be hard to enter queries, but it is easy to get a quick overview of the information on the screen, as summarised in Table 1.

Hence, systems combining the pen and speech input may lead to a more efficient human-computer dialogue:

- The users can express their intentions using fewer words and select the input mode they judge to be less prone to error, or switch modes after system errors and thus facilitate error recovery.
- The system offers better error avoidance, error correction and error recovery.

2.2. Applications

Speech centric multimodal interfaces for mobile terminals can be utilised in many different applications. In e.g. [39], the complementary merits of speech and pen are utilised for entering long sentences into mobile terminals. With this interface, a user speaks while writing or writes while speaking, where the two modes complement one another to improve the recognition performance. However, the two most promising mobile applications with speech centric multimodality are form-filling and map/location-based systems.

Table 1
Comparison of the two complementary user interfaces: Only pen input and screen (visual) output versus a pure speech based input/output interface

Only pen input, screen output	Pure speech input/output
Hands and eyes busy – difficult to perform other tasks	Hands and eyes free to perform other tasks
Simple actions	Complex actions
Visual feedback	Oral feedback
No reference ambiguity	Reference ambiguity
Refers only to items on screen	Natural to refer also to invisible items
No problem with background noise	Recognition rate degrades in noisy environments

2.2.1. Form-filling

One example of form filling applications is described in [13,37], where the users interact with a Personal Information Manager (PIM) on a PDA, by tapping and touching a field and then uttering appropriate content to it.

In [23,24], two form filling applications on a PDA are described: a public train table information retrieval service and a public "yellow pages" service. Here we experienced that users rather easily corrected ASR-errors in these form-filling services. If some of the information on the screen was wrong, the user corrected it by clicking on the field with mis-recognised words and then either saying the correct word once more or tapping on the correct word from the N-best list, which is shown at the right hand side of the field. This is in contrast to automatic telephony applications where correcting speech recognition errors in speech only mode (no visual feedback) often can be very difficult and reduces the user satisfaction.

The actions and benefits of speech centric multimodality in the form filling applications are summarized in Table 2.

2.2.2. Map-based applications

Combining speech and pen gestures as inputs to mobile terminals have proven particularly useful for navigation tasks in maps. Typically, this kind of speech centric multimodal mobile applications provide easy access to useful city information, for instance restaurant and subway information for New York City [16, 17], a tourist guide for Paris [1,2,21], bus information system for the Oslo-area [22,25,38], navigational inquiries in the Beijing area [14], trip planning and guidance while walking or driving car [3], various map-tasks with "QuickSet" [29] and services aimed at public transportation commuters [15].

Our bus information system for the Oslo-area fits into these kinds of applications, and will be discussed further in Section 3.

3. A bus information system

To test the hypothesis that multimodal inputs and outputs really are useful for disabled people, we have developed a flexible speech centric multimodal interface on mobile terminals 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). The users have to write the names of the arrival and departure bus stops 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 bus stops. The response from the system is presented as both speech and text.

3.1. System architecture

Our multimodal bus information system consists of a server and a thin client (i.e. the Mobile Terminal). The client server architecture is based on the Galaxy communicator [11]. The server side comprises five main autonomous modules which inter-communicate via a central facilitator module (HUB) as shown in Fig. 1. All the server side modules run on a PC, while the client runs on a PDA, in this case 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 a web service called "Trafikanten" (http://www.trafikanten.no) through the Internet to get the necessary bus-route information. All computationally heavy components including the automatic speech

Table 2
Benefits with speech centric multimodality in form filling applications

User actions	Benefits of multimodality
Natural language input, asking for several different pieces of information in one sentence.	Speech is most natural for asking this type of questions. Speech is much faster than typing and faster than selecting in a hierarchical menu.
Reads information shown on the screen.	The user gets a quick overview – much quicker than with aural feedback reading sentence by sentence.
Taps in the field where the ASR-error occurs, and taps at the correct alternative in the N-best list.	Much easier to correct ASR-errors or understanding rejections than with complicated speech-only dialogues. Better error control and disambiguation strategies (e.g. when there are multiple matching listings for the user query).

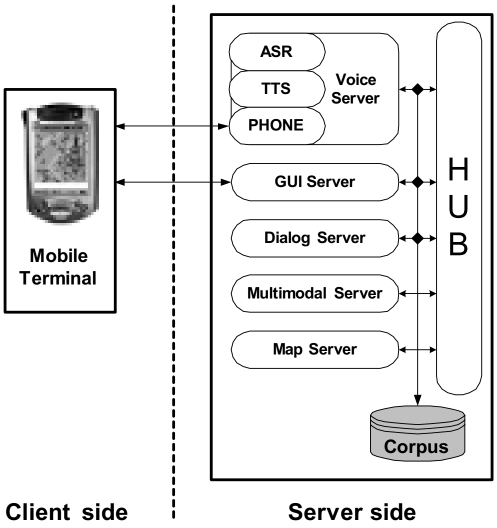


Fig. 1. The overall architecture of our multimodal test platform.

recogniser (SpeechPearl 2000 for Norwegian) and the speech synthesizer (Telenor Talsmann) run on the server. More details about the system architecture can be found in [22,23,38].

Recently, a more mobile version of the system was developed, where the client-server communication takes place over the 3G/UMTS network instead of WLAN [33]. In this version, the client runs on a Qtek 9000 3G mobile terminal, whereas the server side has only small modifications.

3.2. 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 Fig. 2.

Once zoomed, it is possible to get the bus stops in the area displayed. The user has to select a departure bus stop and an arrival bus stop to get the bus route information. The users are not strictly required to follow the steps sequentially. They can for example 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 [35]. Users can 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:

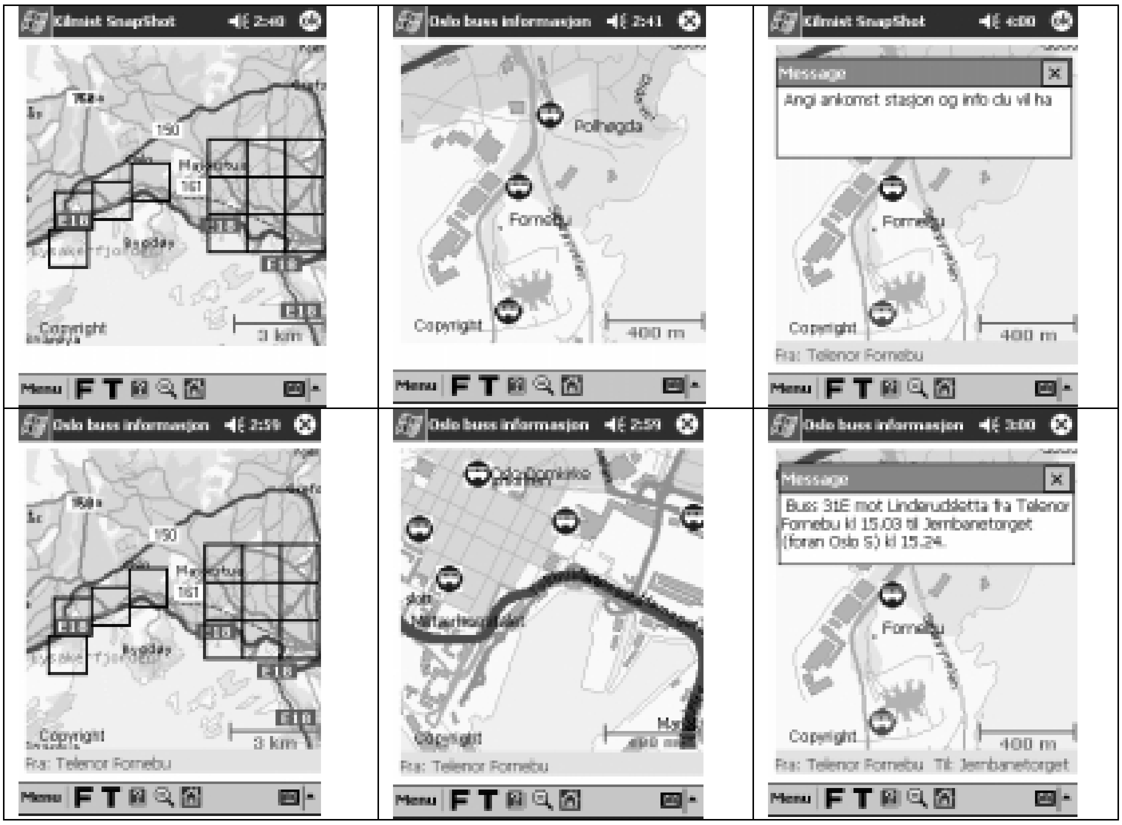
– The user says the name of the arrival bus stop and points at another bus stop on the map, e.g.: "I want to go from Jernbanetorget to *here*"
– The user points at two places on the screen while saying: "When does the next bus leave from *here* to *here*".

In both scenarios above the users point at a bus stop within the same time window as they utter the underlined word, "*here*". In order to handle such inputs, 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 is registered
– Pointing is registered within 0.85 second after speech is detected,

where registration of pointing is instantaneous and detection of speech is completed at the end point of the speech signal. In order to handle two touches on the screen within the same utterance, an integration algorithm that uses two such time windows was employed [38].

Both pointing and speech can be used in all operations including navigation and selecting bus stops. Thus the user scenarios can embrace all the possible



Angi ankomst stasjon og info du vil ha

Buss 31E mot Linderudsletta fra Telenor Fornebu kl 15.03 til Jernbanetorget (foran Oslo S) kl 15.24.

Fig. 2. 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 stop Telenor. 3) The system does not recognize the arrival bus stop. 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 stop Jernbanetorget lies. 5) The user taps on the bus stop Jernbanetorget. 6) The user can read the bus information.

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 will prove useful for many different types of disabled users, such as:

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

4. User evaluations

4.1. Introducing the multimodal for new users

The multimodal interaction pattern was new to the test users and it was necessary to explain this functionality to them. In a user experiment with normal able-bodied persons we discovered that different introduction formats (video versus text) had a noticeable effect on user behaviour and how new users actually interacted with the multimodal service [21]. Users who had seen a video demonstration used simultaneous pen and speech input more often than the users who had had a text only introduction even if the same information was present in both formats. In our user experiments, 9 of 14 subjects who had seen the video demo, applied simultaneous pen and speech input instantly.

We therefore applied two different strategies in the introduction for the disabled test persons:

- For the scenario-based evaluation we produced an introduction video showing the three different interaction patterns: Pointing only, speaking only, and a combination of pointing and speaking. We did not subtitle the video, so deaf people had to read the information on a text sheet.

– For the in-depth evaluation of the dyslectic user and the aphasic user we applied so-called model based learning, where a trusted supervisor first showed how he used the service and carefully explained the functionality.

Since disabled users often 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 the scenario-based evaluations only the experiment leader and the test person were present. The in-dept evaluations were performed in cooperation with Bretvedt Resource Centre [6]. In the in-dept evaluations the test persons brought relatives with them.

The dyslectic user had his parents with him, while the aphasic user 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 in the young dyslectic.

4.2. Scenario-based evaluation

A qualitative scenario-based evaluation followed by a questionnaire was carried out for five disabled users. The goal was to study the acceptance of the multimodal service by the disabled users.

The users were recruited from Telenors handicap program (HCP). They were in their twenties with an education of 12 years or more. The disabilities of the five users are:

– Muscle weaknesses in hands
– Severe hearing defect and a mild speaking disfluency
– Wheelchair user with muscular atrophy affecting the right hand and the tongue
– Low vision
– Motor control disorder and speech disfluency.

The scenario selected for this evaluation involved finding bus route information for two given bus stops. The users had to complete the task in three different manners: By using pen only, speech only and by using both pen and speech. The tests were carried out in a quiet room with one user at a time. All the test persons were able to complete the tasks in at least one manner:

– They were used to pen-based interaction with PDAs so the pen only interaction was easy to understand and the test users accomplished the task easily. Persons with muscle weaknesses in hands or with motor control disorder demanded the possibility of pointing at a bigger area around the bus stops. They also suggested that it might be more natural to select objects by drawing small circles than by making a tap (see also [27]). The person with hearing defects and speaking disfluency preferred the pen only interaction.
– The speech only interaction did not work properly, partly because of technical problems with the microphone and speech recogniser and partly due to user behaviour such as low volume and unclear articulation.
– The multimodal interaction was the last scenario in the evaluation. Hence some persons had to have this functionality explained to them again before trying to perform this task. The persons with muscular atrophy combined with some minor speaking problems had great benefit from speaking short commands or phrases while pointing at the maps.

In the subsequent interviews all users expressed a very positive attitude to the multimodal system and they recognized the advantages and the potential of such systems [19,20,25].

4.3. 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 [8].

Our dyslectic test person was fifteen years old and had severe dyslexia. He could, for instance, not read the destination names on 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 to travel 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 to just say the names of the two bus stops: “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 simplified the task to asking for the bus route information: “From Telenor to Tøyen”, which was easier for him. But he still had to practise a couple of times to manage to remember and pronounce the names of these two bus stops.

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.

Then we told him that it was unnecessary to formulate full sentences when talking to the system, one word or a short phrase was enough to trigger the dialogue system. He then hesitatingly said “Telenor”. The system responded with “Is Telenor your from bus stop?”, 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 stop had 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 stop shortly afterwards. Then he touched the ‘zoom out’ button and changed map. In this map he pointed at a bus stop 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!”.

4.4. 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 [7]. There is no official statistics for the number of aphasic persons in Norway. Approximately 12000 people suffer stroke every year and it is estimated that about one third of these results in aphasia. In addition, accidents, tumours and inflammations may lead to aphasia, giving a total of about 4000–5000 new aphasia patients every 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 has 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 can be classified as a Broca’s aphasic, although his clear articulation does not completely fit 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 bus stops 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 stops 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 [35], 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 stops. 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 bus stop names. But he liked the multimodal tap and talk interface very much and spontaneously characterised it as "Best!", i.e. the best alternative for him to get the information needed.

5. Conclusions

Our speech centric composite multimodal interface to a map-based information service on a mobile terminal has proven useful for different types of disabilities, from persons with muscular atrophy combined with some minor speaking problems, to 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 wide 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 from our observations and claim that for severe dyslectic and certain groups of aphasics a multimodal interface can 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 users, and thereby achieving the goal of a common of design for all.

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.

References

[1] L. Almeida et al., Implementing and evaluating a multimodal and multilingual tourist guide, in: *Proc. International CLASS Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems,* J. van Kuppevelt et al., eds, Copenhagen, 2002, 1–7.

[2] L. Almeida et al., The MUST guide to Paris. Implementation and expert evaluation of a multimodal tourist guide to Paris, in: *Proc. ISCA* (*International Speech Communication Association*) *tutorial and research workshop on Multi-modal dialogue in Mobile environments* (*IDS2002*), Kloster Isree, Germany, 2002.

[3] D. Bühler and W. Minker, The SmartKom Mobile Car Prototype System for Flexible Human-Machine Communication, in: *Spoken Multimodal Human-Computer Dialogue in Mobile Environments,* Springer, Dordrecht (The Netherlands), 2005.

[4] J. Beskow et al., Specification and Realisation of Multimodal Output in Dialogue System, in: *Proc. 7th International Conference on Spoken Language Processing* (*ICSLP 2002*), Denver, USA, 2002, 181–184.

[5] R. Bolt, Put That There: Voice and Gesture at the Graphics Interface, *Computer Graphics* **14**(3) (1980), 262–270.

[6] Bredtvet Resource Centre. http://www.statped.no/bredtvet presence verified 24/5/07.

[7] J.E. Brody, When brain damage disrupts speech, in: *The New York Times Health Section,* June 10, 1992, C13.

[8] Dyslexia Action. http://www.dyslexiaaction.org.uk/, presence verified 24/5/07.

[9] ETSI. Human Factors (HF); Guidelines for ICT products and services; "Design for All". Sophia Antipolis, 2002. (ETSI EG 202 116).

[10] ETSI. Human Factors (HF); Multimodal interaction, communication and navigation guidelines. Sophia Antipolis, 2003. (ETSI EG 202 191).

[11] Galaxy communicator. http://communicator.sourceforge.net/, presence verified 24/5/07.

[12] J. Gustafson et al., Adapt- A Multimodal Conversational Dialogue System In An Apartment Domain, in: *Proc. 6th International Conference on Spoken Language Processing* (*ICSLP 2000*), Beijing, China. Vol. II, 2000, 134–137.

[13] X. Huang et al., MiPad: A multimodal interaction prototype, in: *Proc. ICASSP,* 2001.

[14] P.Y. Hui and H.M. Meng, Joint Interpretation of Input Speech and Pen Gestures for Multimodal Human Computer Interaction, in: *Proc. INTERSPEECH – ICSLP'2006,* Pittsburgh, USA, 2006, 1197–1200.

[15] T. Hurtig, A Mobile Multimodal Dialogue System for Public Transportation Navigation Evaluated, in: *Proc. MobileHCI'06,* Helsinki, 2006.

[16] M. Johnston et al., Multimodal language processing for mobile information access, in: *Proc. International Conference on Spoken Language Processing* (*ICSLP-2002*), 2002, 2237–2240.

[17] M. Johnston, B. Srinivas and V. Gunaranjan, *MATCH*: multimodal access to city help, in: *Automatic Speech Recognition and Understanding Workshop,* Madonna Di Campiglio, Trento, Italy, 2001.

[18] A. Karpov, A. Ronzhin and A. Cadiou, A Multi-Modal System ICANDO: Intellectual Computer AssistaNt for Disabled Operators. In: Proc. INTERSPEECH – ICSLP 2006, pp. 1998-2001, Pittsburgh, USA. 2006.

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

[20] K. Kvale and N.D. Warakagoda, A Speech Centric Mobile Multimodal Service useful for Dyslectics and Aphasics, in: *Proc. INTERSPEECH – EUROSPEECH'2005,* Lisbon, 2005, 461–464.

[21] K. Kvale, J. Rugelbak and I. Amdal, How do non-expert users exploit simultaneous inputs in multimodal interaction? in: *Proc. International Symposium on Human Factors in Telecommunication* (*HfT 2003*), Berlin, 2003, 169–176.

[22] K. Kvale, J.E. Knudsen and J. Rugelbak, A Multimodal Corpus Collection System for Mobile Applications, in: *Proc. Multimodal Corpora – Models of Human Behaviour for the Specification and Evaluation of Multimodal Input and Output Interfaces,* Lisbon, 2004, 9–12.

[23] K. Kvale, N.D. Warakagoda and J.E. Knudsen, Speech centric multimodal interfaces for mobile communication systems, in: *Telektronikk,* Vol. 99, No. 2-2003, 2003, 104–117.

[24] K. Kvale, N.D. Warakagoda and J.E. Knudsen, Speech-centric multimodal interaction with small mobile terminals, in: *Proc. Norwegian signal processing symposium* (*NORSIG 2001*), Trondheim, 2001, 12–17,

[25] K. Kvale, N.D. Warakagoda and M. Kristiansen, Evaluation of a mobile multimodal service for disabled users, in: *Proc. 2nd Nordic conference on multimodal communication,* Gothenburg, 2005, 242–255.

[26] S. Oviatt et al., Designing the user interface for multimodal speech and gesture applications: State-of-the-art systems and research direction, *Human Computer Interaction* **15**(4) (2000), 263–322.

[27] S. Oviatt et al., Integration and synchronization of input modes during multimodal human-computer interaction, in: *Proc. Conference on Human Factors in Computing Systems: CHI '97,* New York: ACM Press, 1997, 415–422.

[28] S. Oviatt, Multimodal interface research: A science without borders, in: *Proc. 6th International Conference on Spoken Language Processing* (*ICSLP 2000*), Beijing, Vol. III, 2000, 1–4.

[29] S. Oviatt, Multimodal system processing in mobile environment, in: *Proc. of the Thirteenth Annual ACM Symposium on User Interface Software Technology* (*UIST'2000*), ACM: New York, N.Y., 2000, 21–30.

[30] S. Oviatt, R. Coulston and R. Lunsford, When Do We Interact Multimodally? Cognitive Load and Multimodal Communication Patterns, in: *Proc. of ICMI,* 2004.

[31] S. Oviatt, Ten Myths of Multimodal Interaction, *Communications of the ACM* **42**(11) (1999), 74–81.

[32] J.S. Pedersen, P. Dalsgaard and B. Lindberg, A Multimodal Communication Aid for Global Aphasia Patients, in: *Proc. Interspeech 2004 – ICSLP,* Jeju Island, Korea. 2004.

[33] T. Schie, Mobile Multimodal Service for a 3G-terminal, M.S. thesis, Norwegian University of Science and Technology, 2006.

[34] SMARTCOM – Dialog-based Human-Technology Interaction by Coordinated Analysis and Generation of Multiple Modalities, http://www.smartkom.org/start_en.html presence verified 24/5/07.

[35] W3C, Multimodal Interaction Requirements, NOTE 8 January 2003, http://www.w3.org/TR/2003/NOTE-mmi-reqs-20030108/, presence verified 24/5/07.

[36] W. Wahlster et al., SmartKom: Multimodal Communication with a Life-Like Character, in: *Proc. EUROSPEECH-2001,* Aalborg, Denmark, 2001, 1547–1550.

[37] Ye-Yi Wang, Robust language understanding in MiPad, in: *Proc. EUROSPEECH-2001,* Aalborg, Denmark, 2001, 1555–1558.

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

[39] Y. Watanabe et al., Semi-synchronous speech and pen input, in: *Proc. ICASSP 2007,* 2007, pp. IV. 409–412.