

Design of a Rich Multimodal Interface for Mobile Spoken Route Guidance

Markku Turunen, Jaakko Hakulinen, Anssi Kainulainen, Aleksi Melto, Topi Hurtig

Department of Computer Sciences, University of Tampere, Finland

mturunen@cs.uta.fi, jaakko.hakulinen@cs.uta.fi, anssi.kainulainen@cs.uta.fi,
aleksi.melto@cs.uta.fi, topi.hurtig@uta.fi

Abstract

We present a design of a rich multimodal interface for mobile route guidance. The application provides public transport information in Finland, including support for pedestrian guidance when the user is changing between the means of transportation. The range of input and output modalities include speech synthesis, speech recognition, a fisheye GUI, haptics, contextual text input, physical browsing, physical gestures, non-speech audio, and global positioning information. Together, these modalities provide an interface that is accessible for a wide range of users including persons with various levels of visual impairment. In this paper we describe the functional aspects and the design of the interface of our publicly available prototype system.

Index Terms: speech interfaces, multimodal interfaces, mobile applications, accessibility

1. Introduction

Mobile phones have enabled ubiquitous spoken telecommunication between humans. Recently, mobile devices have become rather powerful and well connected, and even regular mobile phones include a possibility to run custom applications (e.g., MIDlets in Java enabled phones). This has enabled multimodal and distributed mobile applications that take advantage of speech, graphics, non-speech audio, haptics, positioning information, etc. At the same time, the mobile context has brought along new application domains for speech-based and multimodal systems, such as mobile public transport navigation assistants considered here.

Multimodality can have many benefits compared to unimodal interaction. It may bring more bandwidth to the communication and provide alternative modalities for the same tasks, for example in the case of disabled users to provide speech-based and haptic alternatives for graphical elements. Unfortunately, multimodal systems are sometimes designed based on one main modality, with the other modalities simply added on top. Multimodality can also offer disadvantages, as handling several modalities together may result in cognitive overload and reduced usability, especially in demanding usage situations that arise in mobile use.

Here we present a mobile multimodal route guidance interface that uses several modalities as alternatives that support each other. In this way, the interface supports different users and usage conditions. Mobile phone keys and physical gestures can be used for navigation in two-dimensional menu structures. Speech inputs and context-sensitive, domain-optimized text input can be used for entering names and addresses of departure and destination locations. Global positioning information and physical browsing can be used as alternatives for entering the place of departure. All graphical elements, including text input, are combined with speech outputs. Non-speech audio is used for navigation and route awareness information. Finally, a two-dimensional fisheye

interface supports people with low vision, and increases usability for other users groups as well.

Together, the modalities mentioned offer always at least two alternative ways to perform a single functionality, and the design aims at maximizing the overall efficiency of the interaction among these modalities. In this paper, we first present the design of the mobile multimodal public transport guidance application TravelMan. We first present the domain and give an overview of the application functionality. Then, we describe interface design issues. Finally, the prototype implementation and future work are presented.

2. Mobile Public Transport Guidance

Transport information services, such as bus and train timetables, have a long history in spoken dialogue system research. For example, in our previous projects we have developed three different bus timetable systems [8], of which one has been publicly available for more than three years [6]. Similar systems have been developed by other researchers [4]. Typically, these kind of server-based systems offer timetables and travel plans using a telephony-based speech interface.

Here we are focusing on mobile systems that offer guidance along the whole journey, including the changes between means of transportation, (e.g., for changes from a bus to a tram). In order to construct such systems, there are two technical enablers. First, we need publicly available local and national services that offer full route descriptions for public transport. One good example is <http://journey.fi> (matka.fi). It combines several local and long-distance transport information services together, making it possible to query transport information between any two places or addresses in Finland.

Another technical enabler is positioning technology. In particular, GPS (Global Positioning System) has become widely available in car navigation. In addition, there are applications that support pedestrian guidance in various settings. Public transportation guidance combined with pedestrian guidance could be of great importance. In particular, special user groups, such as visually impaired users, could benefit from it in their everyday life.

2.1. TravelMan

TravelMan is a multimodal mobile application for serving public transport information, as discussed above. The application is based on the research on mobile spoken and multimodal transport information systems carried out in a Finnish research project on speech technology [7]. It combines experiences primarily gathered from a distributed multimodal spoken dialogue system for bus timetables [5], a multimodal route navigation application [2], and a speech-based route guidance application for visually impaired [3]. Figure 1 presents two screenshots from the application.



Figure 1: TravelMan interface screenshots (translated to English, the actual system is in Finnish).

TravelMan provides route guidance information for public transport, such as metro, tram, and bus traffic in Finnish cities. In addition, information for long-distance traffic is included. There are two main functions: (1) planning a journey and (2) interactive guidance during the journey. In the journey planning phase a user gives the departure and destination addresses or places using one of the several available methods, as discussed in the next section. The user can set preferences to exclude and include certain means of transportation, to specify walking speed, and the number of search results. Routes can be also stored and exchanged between the users.

When the user has given the addresses, the system searches suitable travel plans or routes. Then, the system allows the user to browse the suggested routes. Each route description consists of a number of sub-routes with varying means of transportation. The lengths of sub-routes are given as temporal distances (durations) between locations, instead of spatial distances. Each sub-route contains detailed information, such as bus or metro stops or streets between the start and end points. After selecting a suitable route, the user can simply listen how the journey progresses, or navigate in the route description interactively. Next, we present the design of the multimodal interface for the application.

3. Multimodal Interface Design

The main design principle here is to maximize the overall efficiency of the user interface with solutions that work equally well with different modalities and support multiple simultaneous or alternative modalities. Most importantly, the main output modalities, speech and graphics, support each other. Still, it is possible to use the application with either of these modalities. This way the same interface is suitable for users ranging from those with normal vision to blind users. In particular, the combination of speech and a graphical fisheye interface supports users with limited vision. Furthermore, this is helpful to all users, since in the mobile context we all have sometimes limited vision (e.g., due to need to monitor our surroundings).

Besides speech and graphics, there are several other modalities to make the interaction richer and provide alternatives. For system outputs, the interface uses (1) synthesized speech, (2) graphical elements optimized for small displays using fisheye techniques, (3) two types of non-speech sounds,

and (4) haptics. For user inputs, the interface supports (5) telephone keys with contextual text input, (6) speech recognition, (7) physical browsing, (8) physical gestures, and (9) positioning information. Next, we describe the interface metaphor and the different modalities.

3.1. Speaking fisheye GUI

The use of TravelMan is based on multidimensional menus that are operated with directional keys of the telephone. As illustrated in Figure 1, menus are presented to users with a reel metaphor: items in a menu are on top of each other and the user can roll the reel to select menu items. The user can move from the last item to the first and the other way around. The currently selected node is enlarged. In this way, it is easier for people with low vision to see the information on the small display. As the adjacent items (both horizontally and vertically) are visible, the user has a context for the current selection. In the case of route descriptions, the previous and next sub-route of the journey can be very helpful in reviewing the route. In particular, the initial route screen shows the first, the second and the last sub-route of the journey, thus allowing a quick overview of the route (Figure 1, right). The interface is inspired by fisheye techniques such as Fisheye Menu [1]. Our approach, however, is two-dimensional, uses a reel instead of menus, and is tightly integrated with speech.

By using the reel metaphor for the interface, we can combine the local context on the visual interface and the efficiency of spoken outputs. When navigating on the route, the main information such as bus line numbers, bus stop names and departure times for each sub-route are presented. More detailed information (e.g., bus or metro stops in the route) can be browsed by selecting the sub-routes.

The content of each item in a reel is read out loud by the speech synthesizer when the item is activated, as illustrated in Figure 1. However, the spoken content is not necessarily the same as the content presented on the display since speech and text have different strengths and weaknesses. On the small mobile phone display, it is important to use very few characters for the textual presentation, so the route descriptions are carefully constructed to include only the necessary amount of characters. This makes the zooming as effective as possible.

On the other hand, speech, while rather slow as output, uses full sentences to keep the message easily comprehensible. Also the nature of the sub-route (first, last, in between) can be expressed in speech output, e.g., “Depart on bus fifteen-A from Tullaajankuja at thirteen-twenty-five” or “Final stop at Ahoniityntie at thirteen-o-five”. Since speech is quite slow output medium, the duration of spoken outputs varies considerably depending on the addresses and complexity of the routes. However, several techniques, such as tapering, are used to shorten them. The temporal nature of speech also makes it possible to skip the ends of spoken prompts by quickly moving to the next item similarly to spearcons [9]. Because of this, presenting most important information first can speed up the usage especially for blind users.

The reel metaphor works efficiently with speech also when it is used to make selections. The user can select between multiple items by moving left and right. Since the active item is always read out loud, there is no need to add any additional speech, such as listing the possible alternatives, as with more traditional GUI elements. Instead, users can review the options by browsing through them, and they do not need to use special button for selecting, since the active item is always the selected one. The only additional speech output in navigation with the reel menus is menu title and the current

functionality of the two softkeys. These are presented once after a new menu has been opened. Other than that, the speech output is consistently just the active menu item (Figure 1, left). If the user needs more guidance, context sensitive help messages can be activated using navigation keys.

The reel metaphor is designed to support multiple input and output modalities. These will be described next.

3.2. Physical gestures

The reel menus are designed to be controlled also by tilting gestures, with speech synthesis, non-speech audio and haptics giving instant feedback. The reel items can be browsed by tilting the device forward, backward, left, and right, and selections can be made using buttons or tapping the device. As in any novel interaction strategies, the factor of user-dependency must be taken into account, as different people perform certain tasks in different ways. Even though moderate results have been achieved even with 3D gestures, to eliminate or at least shorten the learning phase, this kind of a simple 2D approach is at the moment most viable. This interaction strategy can be extremely useful especially to visually impaired people, not just due to auditory feedback, but also because there is no need to handle several keypad buttons.

3.3. Non-speech audio and haptics

There are numerous studies on how to support navigation (on mobile devices) with non-speech audio [10] and haptics [11]. In order to support navigation and users' awareness of the route, we designed a non-speech auditory component for TravelMan where auditory feedback is combined with haptics. They provide generic feedback on user actions (e.g., gestures), the state of the interface (e.g., position in the reel), and the state of the application (e.g., routes have been obtained from the server). In addition, they provide application specific information, such as information on means of transportation. Non-speech audio can support users as a less intrusive, awareness supporting information source. Such support is especially valuable for people with disabilities, specifically, the visually impaired.

Navigation in the reel is supported by auditory icons and haptics. One step in the menu makes a subtle audible cue which resembles physical interaction with a reel-like object, and a synchronized nudge can be given as haptic feedback. Fast movement as multiple steps in the menu, or jumping over from the last menu item to the first, gives out a sound which resembles turning a wheel or dial, or flipping through a deck of cards or a rolodex. The cues are played as chords or melodies of two sounds. The pitch of the first sound is in relation to the position of the menu item, so advancing in the menu raises the pitch. The first sound has two tones: one for representing moving forward (e.g. wheel moving forward), and one for going backward. The second sound in the cue represents the end pitch, so that the absolute position in the menu can also be determined. These cues support the structure and behaviour of the menu interaction, and they provide feedback to the user on his or her actions. This is important when we are dealing with recognition-based technologies, such as gesture recognition.

Auditory icons are used to describe journey-related information, such as means of transport and temporal information. The auditory icons represent the four modes of local public transportation in Finland: the sounds of walking, metro trains, trams and buses. For each mode of transportation, an accelerating, constant speed and decelerating sound was chosen. The tempo of each sound was doubled to make them

more iconic and discernable from actual traffic sounds. The auditory icons can be used in two ways. First, they are part of the navigation interface, offering background information on the route. For example, when the user navigates from a tram node to a bus node corresponding auditory elements are played in the background (i.e., a sound of a tram slowing down fading to a sound of a bus starting up). Second, when the journey is in progress, the auditory interface is synchronized with the progress of the journey, providing subtle cues on forthcoming elements.

3.4. Speech inputs and contextual text input

For entering addresses and names of departure and destinations places, TravelMan has two primary options: speech inputs and contextual text inputs. In the first option the user can speak the full address including street names and numbers, places of interests, and city names. If there are multiple alternatives, the system presents a clarification dialogue. There are two sources for alternative choices: n-best speech recognition results and multiple entities with the same name (e.g., if the city name is missing, there would be numerous "shopping streets" in the search results). Naturally, the reel metaphor is used also in the clarification menu.

Technology-wise, speech recognition uses a distributed architecture to be described in the next section. Since the recognition takes place on the server, and as there are better alternatives for menu navigation, it is not efficient in the current interface to use speech recognition for other purposes than giving addresses.

The second way to enter addresses is contextual text input optimized for the set domain. Most importantly, text input is designed to be fully accessible for visually impaired people. When the user types characters the system performs a lookup for addresses and speaks out loud the most likely character sequences according to the current domain. When there is a reasonable amount of results, the system informs the user of available choices. This is performed in the background so the user can continue typing, or select the correct address using the reel interface.

3.5. Positioning and physical browsing

The application supports GPS devices for two purposes; as an alternative method for giving a departure address based on the current location of the user. (i.e., coordinates are read from the GPS device and used to set the departure place), and for contextual real-time guidance during the journey. For example, when the user has boarded a bus TravelMan will inform him/her about the progress of the journey based on the GPS information. Without GPS, traveling time is used to estimate the current location, so the application works either as a dynamic guidance system with GPS support, or as a referential route description guide without such support.

A simple form of physical browsing can be used to give destination addresses. In physical browsing, the user points at physical objects in the environment to receive information. Here, the mobile phone's camera is used to read data matrixes embedded in selected bus stops in the Helsinki metropolitan area. For a given bus stop, its data matrix contains the address of the stop with possible additional information. Similarly, it would be possible to use RFID technology for the same functionality.

4. Prototype and User Evaluations

The rich multimodal user interface, as presented in the previous section, requires many novel technical solutions. The application runs on MIDP 2.0 compatible Series 60 mobile phones (e.g., Nokia 6600, N72, and E61), and uses a locally installed Finnish speech synthesizer (<http://www.bitlips.fi/>). A separate Bluetooth GPS device is used for positioning information. The application supports state-of-the-art technology (<http://www.upcode.fi/>) to recognize two dimensional data matrixes needed in the physical browsing interface. For that purpose, a phone with an internal camera is required. For physical gestures, there are two main techniques for capturing physical gestures. The physical gesture interface will first be developed for the Nokia 5500 smartphone, which is equipped with 3D accelerometers. In addition, a version where device movement is detected from camera data will also be implemented. There should be no difference in the usage of these two versions, but the movement detection of the camera version might be less robust.

For speech recognition, the application uses a distributed system architecture running a server-based Finnish speech recognizer (<http://www.lingsoft.fi/>) [5]. Since the vocabulary is rather large, a server-based recognition is needed to make the recognition robust and fast enough. It is noteworthy that at the same time when the recorded speech is streamed to the server, the application retrieves route related information as well. In this way, connections to the server are kept at minimum, and speech recognition itself is not the only reason for a distributed system.

We have developed several prototypes focusing on specific techniques (e.g., physical gestures, non-speech audio), and several integrated versions with varying capabilities and technical requirements will be publicly released in conjunction with the 57th UITP (International Association of Public Transport) world congress. With a public pilot use, an extensive amount of feedback can be gathered from real users. We have previously had good experiences in using this technique [6]. In addition, more focused usability studies will be carried out for the advanced functionalities of the system. For example, we will carry out focused user studies on the usability of speech inputs compared to contextual text inputs, on the usability of physical gestures for menu navigation, and on the usefulness of non-speech audio combined with haptic feedback.

5. Conclusions and Future Work

We have presented a rich multimodal user interface for mobile route guidance. The paper introduced a novel interface offering several alternative and supporting modalities to make mobile applications usable for different user groups and usage situations. We have also planned several interface improvements, which will be summarized next.

Currently, auditory icons provide basic awareness information with sounds of transportation means (e.g., vehicles). Next, we will add more awareness information, such as soundmarks representing important places and objects (e.g., landmarks) on the routes. For speech inputs, we will investigate adding a local speech recognizer for selected tasks. We have also planned accentuated maps and graphical icons to be used in the fisheye interface. Accentuated maps will offer information on those routes that require walking between stops. For example, if the user needs to walk between several streets to change from a tram to a bus. Simplified, accentuated maps can be very helpful in these kind of situations, in par-

ticular for people with low vision and unfamiliar to surroundings. Physical gestures could also be extended to let the user manipulate this route map with simple tilting gestures. Several studies have proven this to be a very intuitive technique; tilting forward pans the map north, tilting backward pans the map south, and so on. Finally, graphical icons could be used as an alternative for the textual route descriptions. This could help people in certain user groups, make the graphical interface less language independent, and save valuable space on the display.

6. Acknowledgements

This work is supported by the Finnish Funding Agency for Technology and Innovation (FENIX -programme, New Methods and Applications in Speech Processing -project).

7. References

- [1] Bederson, B. B., Fisheye Menus. Proceedings of ACM Conference on User Interface Software and Technology (UIST 2000), 217-226, 2000.
- [2] Jokinen, K., and Hurtig, T., User Expectations and Real Experience on a Multimodal Interactive System. In Proceedings of Interspeech 2006.
- [3] Koskinen, S., and Virtanen, A., Public transport real time information in Personal navigation systems for special user groups. In Proceedings of 11th World Congress on ITS, 2004.
- [4] Raux, A., Langner, B., Bohus, D., Black, A. W., and Eskenazi, M., Let's Go Public! Taking a Spoken Dialogue System to the Real World. Proceedings of Interspeech 2005: 885-888, 2005.
- [5] Salonen, E. -P., Turunen, M., Hakulinen, J., Helin, L., Prusi, P., and Kainulainen, A., Distributed Dialogue Management for Smart Terminal Devices. In Proceedings of Interspeech 2005: 849-852, 2005.
- [6] Turunen, M., Hakulinen, J., and Kainulainen, A., Evaluation of a Spoken Dialogue System with Usability Tests and Long-term Pilot Studies: Similarities and Differences. In Proceedings of Interspeech 2006: 1057-1060, 2006.
- [7] Turunen, M., Hurtig, T., Hakulinen, J., Virtanen, A., and Koskinen, S., Mobile Speech-based and Multimodal Public Transport Information Services. In Proceedings of MobileHCI 2006 Workshop on Speech in Mobile and Pervasive Environments, 2006.
- [8] Turunen, M., Hakulinen, J., Salonen, E.-P., Kainulainen, A., and Helin, L., Spoken and Multimodal Bus Timetable Systems: Design, Development and Evaluation. Proceedings of 10th International Conference on Speech and Computer (SPECOM 2005): 389-392, 2005.
- [9] Walker, B. N., Nance, A., and Lindsay, J., Spearcons: Speech-based Earcons Improve Navigation Performance in Auditory Menus. In Proceedings of ICAD 2006, 2006.
- [10] Jones, M., Bradley, G., Jones, S., and Holmes, G., Navigation-by-Music for Pedestrians: an Initial Prototype and Evaluation. In Proceedings of the International Symposium on Intelligent Environments 2006: 95-101, 2006.
- [11] Van Erp, J.B.F., Van Veen, H.A.H.C., Jansen, C., and Dobbins, T., Waypoint navigation with a vibrotactile waist belt. ACM Transactions on Applied Perception (TAP), vol. 2, issue 2, pp: 106-117, 2005.