COMPASS2008: Multimodal, multilingual and crosslingual interaction for mobile tourist guide applications

Ilhan Aslan[1], Feiyu Xu[1], Hans Uszkoreit[1], Antonio Krüger[1,2], and Jörg Steffen[1]

[1] DFKI GmbH, Germany
[2] University of Münster, Germany

Abstract. COMPASS2008 is a general service platform developed to be utilized as the tourist and city explorers assistant within the information services for the Beijing Olympic Games 2008. The main goals of COMPASS2008 are to help foreigners to overcome language barriers in Beijing and assist them in finding information anywhere and anytime they need it. Novel strategies have been developed to exploit the interaction of multimodality, multilinguality and cross-linguality for intelligent information service access and information presentation via mobile devices.

1 Introduction

More and more people rely on Smartphones and PDAs as their companions helping them to organize their daily activities. Having initially just focused on traditional PIM applications (such as telephone and date book), the last generation of devices equipped with GPS, enhanced connectivity (Wireless Lan and UMTS) as well as powerful processors is well prepared to support a variety of location based services. Mobile services for tourists are one of the promising domains to profit from this new development. Especially tourists in a foreign country who are often unfamiliar with the local language and culture could extremely benefit from well designed mobile services and interfaces. Tourists need help in a variety of situations, e.g. when ordering food, using public transportation or booking museum tickets and hotel rooms. However, the complex task of designing adequate mobile user interfaces for these services still remains a challenging problem. The mobile interface has to reflect the individual user's interests and background as well as the specific usage situation in order to support smooth interaction. Often, relevant content is only available in the local language and therefore not accessable by the majority of tourists. Finally, mainly due to their small screen estate and reduced interaction capabilities (i.e. the lack of a keyboard), the interaction with mobile devices is more cumbersome than with regular scale computer systems. Users experience difficulties, for instance, when posing a query or accessing a service if this includes browsing through menus with a deep hierarchy.

In the past, different lines of research have addressed these two problems. On the one hand, multimodal interfaces have been investigated as one possible solution to improving the interaction with mobile devices (see [10]). By using speech and gesture, users are able to compensate for smaller screens and a missing keyboard or mouse. On the other hand, research on multilingual and crosslingual information access has advanced the possibilities of users to exploit information sources in languages other than their own (see [7], [8] and [11]).

In this paper we would like to bring together both lines of research by introducing translation techniques, multilinguality and cross-linguality into the design of mobile multimodal interfaces. We will show that such an interface will not only combine the benefits of both approaches but will also provide a new interaction style which allows to combine modalities in different languages. This will allow tourists to communicate much more effectively with automated services and local people by using the mobile device as a mutual communication platform in their own and in the foreign language.

We will demonstrate our concepts in the context of the Compass2008 (COMprehensive Public informAtion Services) project for the Olympic Games 2008 in Beijing[3]. Compass2008 is a Sino-German cooperation aiming at integrating advanced technologies to create a high-tech information system that helps visitors to access information services during the 2008 Olympic Games in English and Chinese and a few other languages.

The next section will provide an overview of the services and the scenarios we had in mind while designing Compass2008. Section 3 elaborates on the interaction concepts and translation services and section 4 discusses the overall architecture of the system. The paper closes with a brief review of related work and draws a few conclusions.

2 Multimodal, Multilingual Service Access and Presentation

COMPASS2008 is built on top of the monolingual service platform FLAME2008 [4] that provides users services according to their personalized demands and allow service providers to register their services. However, the service access option in FLAME2008 is restricted to category-based navigation and no intelligent user interface technologies are applied. The current focus of COMPASS2008 is Olympic Games related information services. A service taxonomy is defined to describe the COMPASS2008 service scope, having taken some existing ontology and taxonomies into account, namely, the service ontology designed in FLAME2008, the tourism ontology resulted from the multilingual tourism information system MIETTA (see [11]) and the Olympic Game information classification available on the Athens 2004 Olympic Game web site. The service ontology serves for service classification and service content structuring and is very important for the service adaptive search and result presentation design.

COMPASS2008 services are classified into three groups: (a) **Information service**: weather info, eating and drinking, city info, olympic info, etc. (b)

[3] Funded by German Federal Ministry of Education and Research, with grant no. 01IMD02A and selected for cofunding by the Chinese Ministry for Science and Technology.

Transaction service: translation services and e-commerce service (c) **Composed service**: services that integrate various services to deal with a complex situation (e.g. taxi dining service that makes use of Taxi Dialog Assistance and Smart Dining service)

We follow a two-level strategy for the service access and presentation: first level is the *general service retrieval* and second level is the *service specific information retrieval*. The general service retrieval options contain ontology-based service retrieval and keywords search. Users can click, handwrite or speak out the service category names, (e.g. *city info* and/or enter keywords such as *forbidden city*). Services belonging to the chosen category and containing information about *forbidden city* will be returned to the users in their preferred modalities.

Figure 1 depicts the navigation in the service taxonomy, stepping from the general category *eating and drinking* to its subcategory *smart dining*.

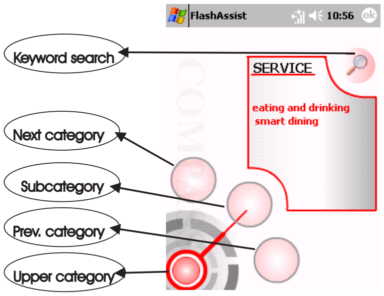

Fig. 1. Navigation in the service taxonomy

The second level strategy allows users to enter more precise and specific queries when they have selected a specific service category. A specific query template is designed for each service category. For example, users can ask for *temperature* and *air pressure* of a specific region, when they are interested in *weather information*. In the following sections, we will give a detailed description of the taxi dialog assistance and dining services. Both services provide multimodal, multilingual and crosslingual interaction concepts and help foreign tourists and Olympic Game participants to overcome the language barriers in the everyday life in Beijing.

2.1 Taxi Dialog Assistance

One of the biggest problems for foreigners in Beijing is how to make the taxi driver understand their destination requests, because most taxi drivers in Beijing can only speak and understand Chinese. Our Taxi Dialog Assistance acts as a mediator between taxi drivers and foreign visitors. It translates the destination request into Chinese and speaks to the taxi driver and asks him to enter the estimated price and distance information on the mobile device. If the location-based service is available, COMPASS2008 system will also provide interactive map and route information which eases communication and understanding between the taxi driver and the clients.

If a COMPASS2008 user chooses the service *Taxi Dialog Assistance*, a specified service query page will appear for entering the destination name. The current location of the user will be treated as the default starting point, appearing on the screen parallel to the destination calculated by the location-based service. The user is asked to specify the category of the destination, because it can be a restaurant name, a building name, an organization or institution name. Same name belonging to different categories can have different addresses. Experiences tell us that the category information is an important resource for destination disambiguation. Furthermore, we allow users to enter the addresses of the destination as additional information. Given the destination and its category, the translation engine will look up in the bilingual dictionary. In comparison to the traditional bilingual dictionary, our dictionary contains not only the translations between terms in different languages, but also tourism relevant categories, to which they belong, such as hotel, restaurant, hospital, shop, school, company, etc. These names are related to the address and location information available in the location-based services.

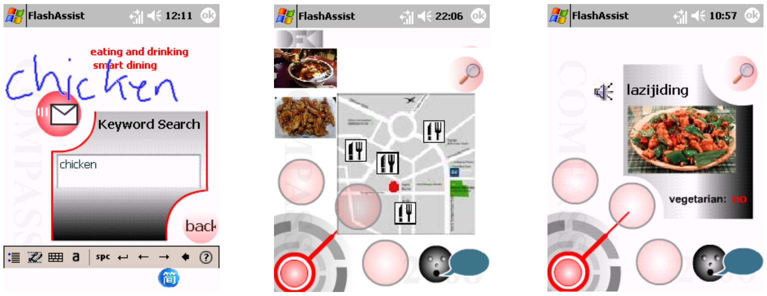

Fig. 2. Keyword search, location based results and menu navigation

2.2 Smart Dining Translation Assistance

The smart dining translation assistance is a service that helps foreigners in Beijing to find the right restaurant or food, according to their taste and preferences by providing a vivid and attractive user interface, using multimedia data. We have designed a very fine-grained multilingual database for food and restaurant information, containing e.g., Chinese food names and their audio records, name translations, food/restaurant images, taste descriptions, restaurant addresses, and even related video information. As described in Figure 1 and 2, a COMPASS2008 user can perform a general keyword search, for example, by entering *chicken* and then chooses from the matched restaurants, highlighted in the map. Furthermore, we also allow multimodal interactions in the queries, for example, (1) speech: *"show me only vegetarian food"* (2) speech: *"show me the ingredients of this dish"* + gesture: tap on image of dish from a list of dishes (3) speech: *"compare this dish"* + gesture + speech:*"to this dish"* + gesture (4) speech: *"translate this Chinese writing"* + handwriting (or gesture)

For ordering the food by the restaurant staff, the mobile device can speak out the food name in Chinese and the preferences of the user such as "vegetarian" or "no garlic" in Chinese.

These are example services, which showcase the synergetic usage of inteligent multilingual, crosslingual and multimodal interaction. Multilingual writing, multilingual voice in/output and gestures in combination with multimedia data presentation gives us the flexible means for the convenient and service adatpive multilingual and crosslingual information access and presentation via mobile devices.

3 Interaction concepts

3.1 Mobile Multimodal Interaction

The vast variety of interaction modalities available on mobile devices result from the fact, that mobile and handheld devices are used in everyday life, in different situations and context. Therefore, it seems to be difficult to provide a static interaction modality that is always suitable. For example the usage of speech input within a crowded stadium seems to be difficult, because of the background noise. The usage of a stylus to tap or write on the handheld device will only be possible if the user has both hands free. The interaction modalities we support in COMPASS2008 are based on the results of a user study describe in [10]. This study investigated preferences of users, which had no or little experience with handheld PCs, considering multimodal interaction with a mobile and multimodal interactable shopping assistant in a public environment. The subjects preferred (in addition to unimodal interaction modalities; that is, interaction with only one modality) to interact with speech in combination with gesture performed on the display of the mobile device (e.g. with the stylus). In COMPASS2008 we also support combinations of speech[4] and gesture combined with handwriting. We believe that writing is a modality that is important in a tourist and multilingual environment.

Multimodal interaction is realized in COMPASS2008 with the use of *data container pairs*. In this work we refer to *data container* as a data containing file in XML format. A data container pair consists of one XML file that encodes information of all single target objects and another XML file that encodes multimodal grammar definitions that can be used to access information about the target objects.

3.2 Translation Services

COMPASS2008 is the first service platform that combines traditional general machine translation services with some specified translation services, in order to

[4] We currently use IBM ViaVoice for speech input and Scansoft TTS for speech output, Chinese speech recognition and synthesis modules developed in the Chinese partner project MISS2 are also under consideration.

cover different demands of the foreign visitors and tourists in Beijing. We call our translation services "COMPASS2008 Translation Center". Figure 3 depicts the four major translation services in COMPASS2008. As highlighted in the above figure, different service requests need different translation services. We provide two kinds of services in the general machine translation functionality, namely, *hybrid machine translation service* and *the name translation service*. In the *hybrid machine translation service*, we integrate on the one hand available online machine translations, such as *Google* or *Babelfish*, and on the other hand, offer a more restrictive but also more reliable translation service through an extended digital tourism phrase book. We expect that current progress in statistical MT especially for the language pair Chinese-English as well as enhanced selection and voting methods will strongly improve the reliability of the full text translation service. For the time being, reliability is mainly achieved through the digital phrase book. Its purpose is to provide foreigners with essential key words, phrases and sentences they need in the most urgent situations and for communication in hotels, airports, train stations, restaurants, Olympic stadiums and other travel sites. The hybrid machine translation service will use the results of the Digital COMPASS Tourism Phrase Book if a translation can be found there. Otherwise it will call the online free machine translation services. The *name translation service* is especially designed for the mobile device. It helps foreigners to recognize Chinese characters and translate them into the preferred languages. In addition to the regular Chinese character input methods, foreigners have two options to enter expressions made up of one or several Chinese characters: (a) **handwriting**: drawing Chinese characters via stylus, (b) **photo capturing**: capturing the Chinese characters via digital camera

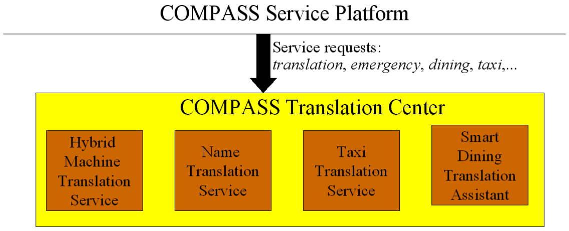
COMPASS Service Platform
Service requests:
translation, emergency, dining, taxi,...
COMPASS Translation Center
Hybrid Machine Translation Service
Name Translation Service
Taxi Translation Service
Smart Dining Translation Assistant

Fig. 3. COMPASS Translation Center

The first option is supported by a software for Chinese handwriting on the Pocket PC such as the CE-Star Suite for Pocket PC 2003 2.5. Users can draw the Chinese characters, the system will then suggest the most similar characters and let the users choose the right one. For the second option, the corresponding Chinese OCR software still needs to be selected from a range of available options.

The Taxi Dialog Assistance and the Smart Dinning services described above use the *Taxi Translation Service* and the *Smart Dining Translation Assistance* for the translation task. Both specialized translation services reach a higher accuracy by employing fine-grained dictionaries modelled for the specific domains.

3.3 Multilingual and Crosslingual Interaction

In addition to the number of modalities, the COMPASS system has to deal with multiple languages. Some modalities are connected to a language (e.g. speech and handwriting). Others are language independent (e.g. gesture) but can nevertheless be used to facilitate or complement language communication.

The COMPASS platform allows the combination of multimodal techniques and communication on several languages. The system is multilingual in the sense that three languages are currently supported with the options of adding more. Since the completed information system for the Olympic Games will not support more than five to six languages, we still have to expect a large number of users for whom none of these supported languages will be their mother tongue. For these users who have a certain command of e.g. English or German but no knowledge of Chinese, multimodal presentation techniques can greatly facilitate communication.

In addition to this multilingual setup, the system also offers crosslingual functionalities that help to overcome language barriers. These functionalities are important for the communication between people who do not speak a mutual language. They can also help tourists to find their way in an environment where signs, menues, instructions are expressed in a language they do not master.

However, some of the target users of the COMPASS2008 system may be bilingual to a certain degree (e.g. will be able to speak English and some Chinese). These users are also allowed to input their queries and commands in mixed-language modalities; for example, a bilingual user may ask a question in English and relate it to a Chinese writing (e.g. they will ask "How do I pronounce this" and write down in Chinese characters the name of a location or object). To achieve cross-lingual modalities, the data container pairs that are used for multimodal interaction need to be changed. There are two approaches to do this, either a second data container pair that describes the data in the second language can be used or the existing data container pair can be extended.

4 Compass2008 Architecture

The COMPASS2008 service platform has two main parts: the frontend system and the backend system. The backend system is mainly responsible for preparation of data resources for service retrieval, multilingual services, multimodal interaction and location-based services. In this paper, we will focus on the frontend system architecture. The frontend system allows users to register their profiles and retrieval and access the useful services. In this context, we will only describe the central system architecture and concentrate on the multilingual and multimodal service retrieval and access. We distinguish the server architecture from the client architecture. The server architecture includes computing and storage intensive processes. The relevant processes for multimodal interaction are mainly embedded in the client architecture. The right hand side of Fig. 4 depicts the main components in the COMPASS2008 server. The COMPASS2008 frontend

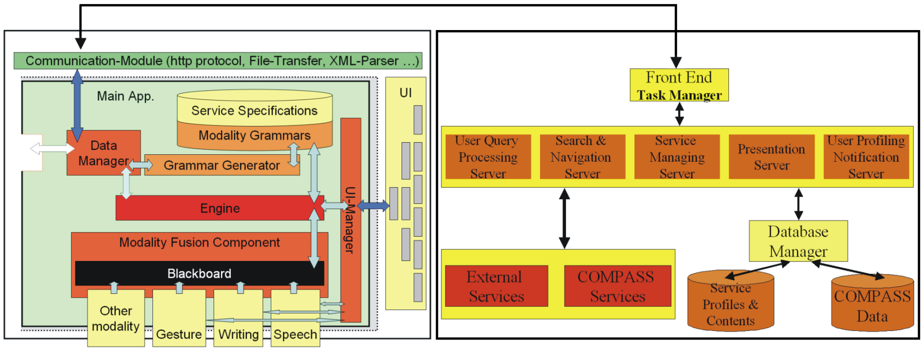

Fig. 4. Overview of client and server archtiecture

system contains two managers: the task manager and the database manager. The task manager is responsible for the workflow and the interaction between the sub-components, while the database manager takes care all database access activities. The main subcomponents are (i) *Query Processing Server*: Responsible for processing the user queries sent by the clients, including multi-modal interaction and query translation tasks. A query can be speech commands, keywords, questions, service categories, locations and filled forms etc. (ii) *Search Server*: Responsible for index and database search. It includes both general service retrieval and the service specific retrieval. (iii) *Service Manager Server*: Responsible for service-specific applications (iv) *Presentation Server*: Responsible for the generation of the presentation pages displayed in the browsers. It includes components for multilingual and multimodal presentation. (v) *User Profiling and Notification Server*: Responsible for setting and updating of user pro-files and notifying the user when new information is available or updated. In Fig. 4, on the left hand side an overview of the client architecture is given. A communication module uses the HTTP protocol to make requests to the services server. The services server presents information (e.g. data container pairs) in XML format. The XML files are parsed, the grammar is loaded and the user interface manager is informed. The main subcomponents are: (i) *Modality Fusion*: Responsible for handling input modalities. (ii) *UI-Manager*: Controls connection between Flash based (Flash is a vector animation software for the web, because they are so lightweight) UI set and C++ based control logic. (iii) *Communication*: Pipes requests to the server component. (iv) *Engine*: Coordinates threads. (v) *Data (and Grammar) Manager*: Manages data container pairs. (vi) *Flash based UI set*: Pipes user input to UI-Manager and is responsible for data presentation.

The user interface manager is connected via socket technology with a topological ordered set of Flash user interface templates. A main Flash template serves as the entry point for the flash user interface. Depending on the commands the main Flash template receives from the user interface manager, the main template pipes information to child templates and initiates the presentation of data and reports interaction with the user interface.

5 Related Work

In this section, we compare multimodal systems that have been developed in the past to assist users in tourist domains. One of the most prominent mobile spatial information systems is the GUIDE system [3], which provides tourists with information on places of interest in the city of Lancaster. The GUIDE system allows simple point gestures only and was not explicitly designed to explore multi-modal research issues. The HIPS [6] project aimed at designing a personalised electronic museum guide to provide information on objects in an exhibit. The presentations were tailored to the specific interests of a user with the help of a user model and the user's location within the rooms of a museum. The implementation platform was a notebook and touchscreen which allowed for simple gestures and speech commands, but both modalities were not fused and processed in parallel. The REAL system [2] is a navigation system that provides resource adapted information on the environment. The user can explicitly perform external gestures by pointing to landmarks in the physical world to obtain more information. REAL does not allow for speech interaction. In contrast, DEEP Map [5], another electronic tourist guide for the city of Heidelberg, combines both speech and gestures (mainly pointing) to allow users to interact more freely with the presented mapbased presentations. SmartKom [9] is one of the first systems that follow the paradigm of symmetric modality (see Section 3.3). Input to SmartKom can be provided through the combination of speech and gestures. SmartKom then provides travel assistance for the city of Heidelberg through synthesised speech and through gestures performed by a life-like character. None of the described systems allows for a multilingual and crosslingual interaction comparable to the Compass2008 system. Most related work has been designed for one primary language, the support for multiple languages in the context of multimodal systems is to our knowledge a new concept. Description of additional mobile navigation systems (without multimodal interaction capabilities) for tourist applications can be found in the survey [1].

6 Conclusions

We have tried to demonstrate that a systematic and conceptually thought-through combination of multimodal input and output techniques, multilingual and crosslingual communication and location-sensitive functionalities can greatly enhance tourist assistance systems. Such a combination can yield much more than an aggregation of functionalities. It enables relevant new functionality and poses a number of exciting research challenges. Especially the combination of language technology with other modalities and location sensitivity offers many novel opportunities such as: (i) Multimodal output can help the user to understand information presented in one of the supported languages even if this language is not the user's mother tongue. (ii) Multimodal presentation can help the user to interpret untranslatable expression such as certain food names and names of places. (iii)To know the location and situational context can help the translation service in disambiguation and selecting the most appropriate output. (iv) Multimodal input can help the user to enter unfamiliar script and facilitate interpretation. (v) Because of the modular architecture in which services are specified and parametrized by means of a complex ontology, new services and combinations can be added. In our ongoing and planned research and development the COMPASS2008 platform serves three purposes: (i) It is a research tool for investigating currently existing and any new forms of interaction among multimodal input, multimodal presentation, multilingual setup, crosslingual capabilities and location-sensitive functionalities. (ii) It is a tool for developing, testing and demonstrating functionalities to be offered for the information services of the 2008 Olympic Games. (iii) It is an extendable and adaptable base for developing navigation, information and assistance services for general tourism, cultural exploration and large international events.

References

1. J. Baus, K. Cheverst, and C. Kray. In *Map-based mobile services - Theories, Methods and Implementations*. Springer.
2. J. Baus, A. Krger, and W. Wahlster. A resource-adaptive mobile navigation system, 2002.
3. K. Cheverst, N. Davies, K. Mitchell, A. Friday, and C. Efstratiou. Developing a context-aware electronic tourist guide: some issues and experiences. In *CHI*, pages 17–24, 2000.
4. Gartmann, R.; Han, Y.; Holtkamp, B. FLAME 2008 - Personalized Web Services for the Olympic Games 2008 in Beijing. In *Cunningham, P.: Building the Knowledge Economy. Issues, Applications, Case Studies. Vol.1*, Amsterdam, Nederlands, 2003.
5. C. Kray. Situated interaction on spatial. In: DISKI 274, Akademische Verlagsgesellschaft Aka GmbH, 2003.
6. R. Oppermann and M. Specht. A context-sensitive nomadic exhibition guide. In *HUC '00: Proceedings of the 2nd international symposium on Handheld and Ubiquitous Computing*, pages 127–142, London, UK, 2000. Springer-Verlag.
7. H. Uszkoreit. Cross-lingual information retrieval: From naive concepts to realistic applications. In *Proc. of the14th Twente Workshop on Language Technology*, 1998.
8. H. Uszkoreit and F. Xu. Modern multilingual and crosslingual information access technologies. In *Proc. of Multilingual Information Service System for the Beijing 2008 Olympics Forum, CHITEC*, Beijing, China, 2004.
9. W. Wahlster, N. Reithinger, and A. Blocher. SmartKom: Multimodal Communication with a Life-Like Character. In *Proc. of Eurospeech 2001, pp.1547-1550.*, 2001.
10. R. Wasinger, A. Krüger, and O. Jacobs. Integrating intra and extra gestures into a mobile and multimodal shopping assistant. In *Pervasive*, pages 297–314, 2005.
11. F. Xu. *Multilingual WWW — Modern Multilingual and Crosslingual Information Access Techologies*, chapter 9 in Knowledge-Based Information Retrieval and Filtering from the Web, pages 165–184. Kluwer Academic Publishers, 2003.