Flexible Multi-modal Interaction Technologies and User Interface Specially Designed for Chinese Car Infotainment System

Chen Yang, Nan Chen, Peng-fei Zhang, and Zhen Jiao

Corporate Technology, Siemens Ltd. China, 7 Wangjing Zhonghuan Nanlu, Beijing 100102
P.R. China
{cy.yang,nan.chen,pengfei.zhang,zhen.jiao}@siemens.com

Abstract. In this paper, we present a car infotainment prototype system which aims to develop an advanced concept for intuitive use-centered human machine interface especially designed for Chinese users. In technology aspect, we apply several innovative interaction technologies (most of which are Chinese language specific) to make interaction easier, more convenient and effective. Speech interaction design is especially elaborated in this aspect. While in user interface design aspect, we systematically conducted user investigation to give enlightening clue for better designing logic flow of the system and aesthetic design. Under user-centered design principle and with deep understanding of different interaction technologies, our prototype system makes transition from different interaction modalities quite flexible. Preliminary performance evaluation shows that our system attains high user acceptance.

Keywords: Car Infotainment, Chinese ASR, Chinese TTS, Chinese NLU, Chinese Finger Stroke Recognition, Melody Recognition, User-centered design.

1 Introduction

With the rapid development of China economy, the number of automobiles possessed by common Chinese household is increasing dramatically. The platform which includes navigation, mobile communication, entertainment, and information and services is usually called car infotainment system. With car infotainment system, car nowadays becomes not only a simple transportation vehicle, but more like an integrated information space with multiple functionalities. According to Strategy Analytics, a market analysis company, the China market of car infotainment alone will be worth 4 billion US dollars by the year of 2013. It is thus evident that car information system will be a hot competition focus that can differentiate different vehicles and consequently influencing customers' purchase decision.

As car infotainment system now becomes a complex information space with multiple functionalities. How to design a human-machine interface which can make the operation more effective and intuitive and user-friendly while keeping safety as the top priority is quite important. As we know, car environment is a hands-busy and eyes-busy scenario where drivers should concentrate on the road condition from time

J. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2007, LNCS 4552, pp. 243–252, 2007.

and time, with hands holding on the steering wheel. Safety is always the major concern under such a scenario. However, usually there is only a limited number of hardware buttons in a car panel. In order to use a function from car infotainment system, a series of buttons need to be pressed. It seems a touch-screen would be better than buttons alone under such circumstances. Even though buttons and touch screen have the highest interaction accuracy, they are quite dangerous to use because they may keep drivers' eyes from the road and drivers' hands away from the steering wheel. Speech interaction provides a better way to communicate with the machine in such a case since it is hand-and-eye free interaction approach.

Chinese character is ideographic, which is quite different from western alphabet system. Chinese input is a difficult issue even if there is a keyboard. Chinese speech recognition and handwriting are known as two simpler ways to input Chinese especially under such a keyboardless scenario. However, each technology has its own pro-and-cons. At present, speech is not robustness enough to noise, and its recognition accuracy is not as high as that of handwriting recognition. However, it is quite suitable for a hands-busy and eyes-busy scenario. Handwriting is robust to environment noise and has a higher accuracy. However, it is not convenient to use when users' hands and eyes are busy. Therefore smooth transition across different interaction modes need to be carefully designed to make the system easier and more effective to use.

Research work has been conducted to study the user interface design related to car infotainment system [1],[2],[3]. However, there is no systematic work done especially for Chinese user groups. In this paper, we report a car infotainment prototype especially designed for Chinese users. This paper is organized as follows: First we describe the functionalities in our prototype system and the underlying concepts. Then, in section 3, interaction technologies background is introduced. Detailed human machine interface design, including speech interaction design is elaborated in section 4, followed by a performance evaluation in Section 5.

2 Function Descriptions and Underlying Concept

Fig. 1 shows our car infotainment prototype system. As to manual interaction, there are a total of 16 hardware buttons, four of which are used as shortcut keys. Touch screen is used in our prototype system.

Fig. 1. System appearance

Fig. 2. Main menu

Our car infotainment system has six main functions. It includes GPS navigation, communication (telephone and short message service), entertainment (media player function), personal assistance, car configuration, and network. The main menu of the system is shown in Fig. 2.

With the highly-demanded trend of seamless connection with customer electronics, in communication module, mobile phone can be connected into the platform to place or receive a phone call and receive short message. In entertainment module, there is a function called "My MP3", where external mp3 devices can be plugged into this system and use all the provided interaction technologies to select the songs stored in the mp3 device.

Our system supports multiple interaction modes: all menu items can be speech controlled. In navigation module, Chinese ASR can be used to input the target address by just speaking where to go. As another option, the destination can also be input by our Chinese Finger Stroke recognition engine. In communication module, name dialing is at users' choice when placing a phone call. In the meanwhile, when a short message is received, Chinese TTS engine will read the content of the short message aloud. In entertainment module, users can either directly speak the song's name or hum a little beats of a song to select their favorite songs

3 Multimodal Interaction Technologies

Multimodal interaction technologies are used in our car infotainment prototype system, including buttons, touch-screen, speech, handwriting, and melody recognition. This section will give some introduction on the interaction technologies used in our prototype system.

3.1 Speech Technology

Automatic Speech Recognition

Automatic Speech Recognition (short for ASR) is a technology which can translate the input human sound into a text string [4]. Fig.3 shows the system diagram of an ASR system.

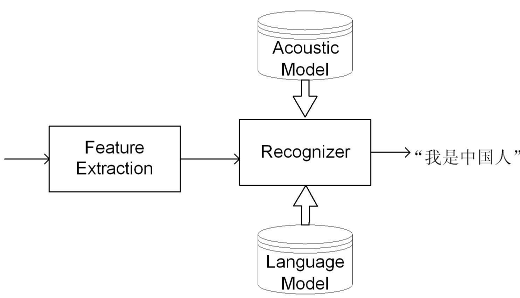

Fig. 3. System diagram of ASR

Given unknown speech, the recognizer can decode the input wave stream into text string with the help of pre-trained acoustic model and language model. Chinese spoken language is monosyllabic and tonal, which is quite different from western languages. Special care should be taken in order to get an acceptable recognition result. While acoustic model part has no much difference from that of other languages, language model part is quite language specific.

Text-to-Speech
Text-to-Speech (short for TTS) is a technology that can generate human-like speech sound according the input text string [4]. Fig. 4 shows the system diagram of a TTS system. There are three modules in a TTS system, that is: text analysis, prosody prediction, and waveform generation. Significant differences of Chinese TTS from other language TTS lie in the first two modules, i.e., TTS preprocessing.

Chinese Natural Language Understanding
Natural Language Understanding (short for NLU) is a technology that can parse the text string and get the meaning or intention of the text string [4],[5]. Speech recognition can only transcribe the input wave stream into text string. With NLU, machine can understand the intention of the text, and corresponding action will be operated. The core of understanding is the representation of semantics. Therefore, this technology is highly language dependent.

3.2 Chinese Finger Stroke Recognition

Most Chinese handwriting recognition is based on character structure [6],[7],[8]. Siemens Limited. China User Interface Design group proposed a new finger stroke recognition system which breaks out of such restriction since the character is now recognized by the stroke orders [9]. In this way, strokes are allowed to be overlapped to each other, which is quite useful when the input pad is too small to input the whole structure of a Chinese character. Fig. 5 shows a simple system diagram of finger stroke recognition engine used in our prototype system.

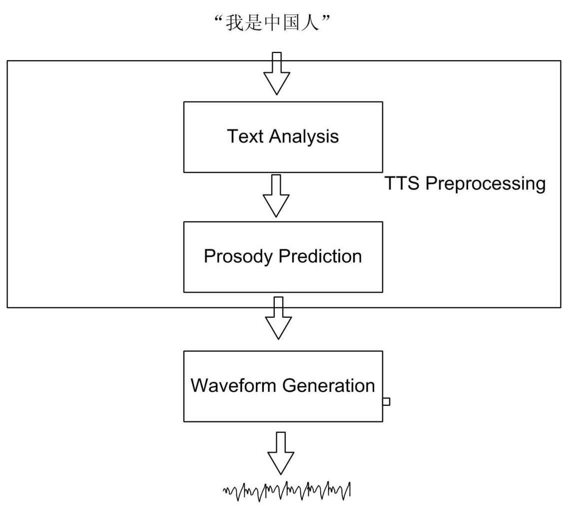

Fig. 4. System diagram of TTS

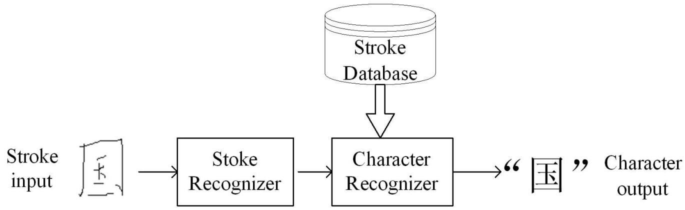

Fig. 5. System diagram of Chinese Finger Stroke Recognition

3.3 Melody Recognition

Music retrieval using the name of a song or a singer is quite familiar to us. However, quite often, we may often encounter such a scenario: you are quite familiar some rhythm and beats of the song, but you can not remember exactly what the song's name is. With the technology named melody recognition, you can easily perform such a task. Typical melody recognition steps include [10],[11].

1. Feature extraction: pitch contour of the music segment is extracted and transcribed as string symbols to be used in following search engine.
2. Search engine: String match is used to find the best alignment between the transcribed music segment and the pre-stored melody database.

4 Human Machine Interface Design

4.1 User Interface Design

In this section, we follow User-centered Design (UCD) procedure to study the Chinese users' typical requirements on the infotainment system, especially the navigation system. The UCD lifecycle starts with users' requirements collection and analysis. The analysis results are used for functional modeling and design. After the prototype is developed, usability evaluation is conducted to collect the users' feedback for the iterative redesign [12].

Chinese drivers have typical driving context, i.e., traffic jam, rules-breaking conditions, overwhelmed traffic indications and signs etc. The specific context may result in typical driving behaviors and requirements from Chinese users. Three main methodologies are used for user investigation: interview, field observation, and focus group. First, thirty drivers were invited to a preliminary short interview to get an overview of the typical driving tasks and their characteristics. Every interview took about thirty minutes. After the interviews, three representative routes in three representative driving contexts (weekday, weekend, and holiday) are selected for the tasks in the field observation, in which twenty filed observations were schemed to discover the habitual interaction language and timing of Chinese users. Then, eight focus groups were held to define function features of the infotainment system and to organize the navigation process.

Fig. 6. Tab label in destination input interface

After conducting systematic user requirement investigation and qualitative and quantitative analysis, we have following interesting findings from Chinese users:

1. Most of the participants mentioned a bad user experience in searching a function in a complicated user interface. They prefer a simple interaction flow
2. Most Chinese users prefer bright color and large icon
3. Most participants think speech is a convenient interaction technology especially in car environment. However they quite concern about the recognition rate. They hope there will be appropriate way to correct the errors when recognition mistakes occur.

Based on the findings, following design strategies are adopted:

1. To make the interaction flow simpler,
1) A tab label is used in destination input interface, where different search approaches, for a instance, all addresses, point of interest (POI), archived addresses are opened by pressing on the their corresponding labels. In this way, traditional vertical interaction depth is smartly transformed to horizontal one. The interface is shown in Fig. 6.
2) Four frequently used functions are put in shortcut key button positions as shown in Fig. 2. This is to ensure to quickly transition from one function to another function
3) Four software buttons are placed in each interface. The text changes according to the interaction result without adding extra interaction levels. For example, in real-time navigation interfaces, the text contents of certain software buttons change at different navigation status, such as at the beginning, in the middle, and at the end of the navigation.

With all of these considerations, the interaction depth can be controlled within three levels, thus guaranteeing the simple work flow. In this way, users can access to their desired function more quickly.
2. To make the usage of the interface more easily,
1) We design the interaction flow according to the logic from user's perspective. Fore example, in navigation function, the users' main aim is to input a destination and then quickly begin the real navigation function.
2) Self-explaining name of each menu and buttons are carefully chosen to user to understand their functions more easily.

3) Same functions are put in the same position of each interface for quick learning of the system. For example, “return” buttons are always put in the last position of four soft key buttons in each interface. In this way, users can find their desired function more quickly.

In aesthetic design, we use large symbolic icons and bright colors according to the result of user investigation. In order to satisfy Chinese users’ preference of personalization interface, we have a theme-changing configuration function for users to change the style of the system outlook in our prototype system.

4.2 Speech Interaction Design

Since speech recognition and natural language understanding technologies do not have as high accuracy as keyboard or buttons, whether the interface with speech technology is effective enough is largely dependent on how well the speech interaction is designed [4]. During the design process, the strength and weakness of speech technologies should be fully aware. Following are the highlights of speech interaction design in our prototype system:

1. Speech is enabled by push-to-talk button and the activation button is put in the car steering wheel to facilitate the use of speech technology;
2. No further confirmation is needed when using the speech function in order to reduce the disturbance to drivers;
3. Speaker-independent technology is utilized to avoid the annoying training process before usage of speech function;
4. Isolated speech recognition grammar is used in command and control scenario to ensure high recognition rate;
5. Dynamic vocabulary and grammar are exploited in different interfaces to enhance the speech recognition rate of each interface;
6. Each speech-enabled icon has a text attached for user easy to know what command should utter;
7. Differentiate the speech-enabled and speech-disabled commands by using different colors;
8. Synthesized speech is used to give feedback of users’ each action and to prompt guidance information during navigation;
9. NLU technology is supported in entertainment module to make this interface more user-friendly since recognition accuracy is not crucial to user experience in this function. Instead, by using this technology, users can communicate with the machine with some degree of freedom;
10. Error handling strategy: since speech recognition is not always correct, it is important to have error handling strategy. The most direct way is to support different modes of interaction. In our prototype system, touch screen alone can perform all of the functionalities. And hardware buttons can accomplish frequently used operations. Therefore, if some recognition mistakes occur, users can easily turn back to the previous interface to make corresponding corrections.

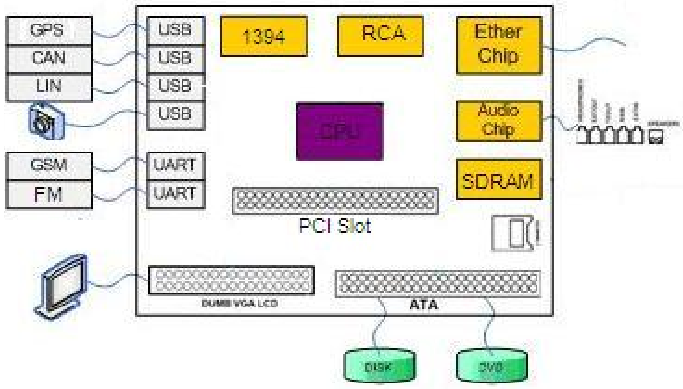

Fig. 7. Hardware layout of the car PC

4.3 Hardware Platform

Our car infotainment prototype system runs on car PC. The motherboard used is VIA EPIA M-Series. Fig. 7 shows hardware layout of the car PC. The fanless VIA Eden processors runs at speeds of up to 1GHz. The motherboard features the VIA CLE266 chipset with embedded MPEG-2 accelerator and integrated 2D/3D graphics core, which ensure smooth DVD playback and a rich overall entertainment experience. The latest high-speed connectivity is supported with IEEE 1394 and USB2.0. This car pc also supports S-Video and RCA TV-Out and 10/100 Ethernet.

5 Evaluation

Preliminary inspection was conducted to test the usability of our car infotainment system prototype. Six people, 3 males and 3 females, participated in the inspection. Their ages range from 24 to 35. The evaluation consists of five dimensions: learnability, efficiency, efficiency, error tolerance and overall satisfaction. Results are shown in Table 1.

Table 1. Performance evaluation

learnability	efficiency	efficiency	error tolerance	overall satisfaction
4.7	4.3	5.0	4.0	4.3

Performance evaluation shows that our system attains high user acceptance. Comprehensive usability test regarding functionalities, interaction and objective satisfaction will be conducted in the next step to further refine the prototype system.

6 Summary

In this paper, we have reported a car information prototype system especially designed for Chinese users. Three main functions, i.e., navigation, entertain, and communication, have been realized in our prototype system at present. With the profound understanding of different interaction technologies (their strength and weakness), innovative multi-modal interaction technologies including Chinese ASR, Chinese TTS, Chinese NLU, Melody recognition, and Chinese Finger Stroke Recognition have been organically applied in the system to make the interaction more effectively and user friendly. In user interface design, we follow the user centered design principle and design the user interface which especially caters for Chinese users' preference. Preliminary evaluation shows that our system has a better user acceptance. Systematic usability test will be conducted in the next step to further refine the prototype system.

Acknowledgement. The authors would like to thank CT IC5 and IC7 of Siemens AG for their great support for this project. The users' requirement analysis is a part of automotive cockpit research work from Ms. Barbara KNAPP. Thanks also go to Junyan CHEN, Rui YANG, Jian NI, Liang ZHANG, Yi-fei Xu, and Heng WANG who contributed a lot to this project. We would also like to thank Wei ZHOU and Ming-hui TIAN for their kind support, and Xiangang QIN and Yanghua LIU for their encouragement and fruitful discussion.

References

1. Ekholm, A.: personal and ubiquitous computing, 6, 153–154 (2002)
2. Akesson, K.P, Nilsson, A.: Designing leisure applications for mundane car-commute. personal and ubiquitous computing 6, 176–187 (2002)
3. Test, J., Fogg, B.J., Maile, M.: CommuterNews: A Prototype of Persuasive In-Car Entertainment. In: Proc. of CHI 2000, pp. 24–25 (2000)
4. Huang, X.-D., Acero, A., Hon, H.-W.: Spoken language Processing - A guide to theory, algorithm and system development. Prentice-Hall, Englewood Cliffs (2001)
5. Allen, J.: Natural Language Understanding, 2nd edn. The Benjamin/Cummings Publishing Company, Menlo Park, CA (1995)
6. Shi, D., Damper, R.I., Gunn, S.R.: Offline Handwritten Chinese Character Recognition by Radical Decomposition. ACM Transactions on Asian Language Information Processing 2(1), 27–48 (2003)
7. Liu, C.L, Jaeger, S., Nakagawa, M.: Online Recognition of Chinese Characters: The State-of-the-Art. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(2), 198–213 (2004)
8. Leung, W.N., Cheng, K.S: A Stroke-order Free Chinese Handwriting Input System Based On Relative Stroke Positions and Back-Propagation Networks. In: Proceedings of the 1996 ACM symposium on Applied Computing, pp. 22–27 (1996)
9. Cao, X.: Chinese Handwriting Recognition Based on Finger Stroke Order, China patent, patent no, 200510066546.8

10. Pardo, B., Birmingham, W.: Query by Humming: How good can it get?. In: Workshop on Music Information Retrieval, SIGIR (2003)
11. Kosui, N., Nishihara, Y., Yamamuro, M., Kushima, K.: A practical query-by-humming system for a large music database. In: Proceedings of the eighth ACM international conference on Multimedia, pp. 333–342 (2000)
12. Vredenburg, K., Isensee, S., Righi, C.: User Centered Design: An Integrated Approach. Prentice-Hall, Inc (2003)