



Smarter Driving with IDA, the Intelligent Driving Assistant for Singapore

Andreea I. Niculescu, Ngoc Thuy Huong Thai, Chongjia Ni, Boon Pang Lim, Kheng Hui Yeo and Rafael E. Banchs

Institute for Infocomm Research (I²R), A*STAR, Singapore

{andreea-n, nthhthai, nicj, bplim, yeokh, rembanchs}@i2r.a-star.edu.sg

Abstract

In this paper we present our works towards creating a natural language platform for an intelligent driving assistant (IDA) for smart parking in Singapore. In particular, we are focusing on the challenges of designing and implementing reliable spoken dialogue components that enable drivers to communicate hands-free with the system. These components require: spoken language dialogue design, data collection, as well as training of speech recognition (ASR) and natural language understanding (NLU) modules. The main objective of IDA is to help drivers to find suitable parking, online monitor car park availability and redirect drivers when the number of free available spots drops to a critical level. As such, this speech-enabled application contributes to a more sustainable city by decreasing traffic congestion, fuel expenses and time waste for all drivers on the road.

Index Terms: multi-modal dialogue system, speech recognition, natural language understanding, smart parking application, interaction design

1. Introduction

Despite being ranked top for smooth traffic flow, efficient road network, road quality and public transportation [1] Singapore is confronted during peak-hours with heavy traffic jams and lack of clear and timely information about parking spaces. To avoid such stressful events an intelligent driving assistant (IDA) application was developed within a project collaboration between I²R, Continental Automotive Singapore and TUM-Create. The application helps drivers to take smart parking decisions, keeps monitoring the parking availability and guides drivers to the desired car park. Since the use of mobile devices while driving is strictly prohibited under Singapore law, the application interacts with the driver using spoken dialogue. Both speech and dialogue technologies are fundamental for the application's effectiveness and ease of use. However, building reliable spoken language components in a noisy environment (inside of a driving car) where drivers speak English in highly accented way (as the case in Singapore) and are merely concerned with a different activity (driving) rather than speaking with system is certainly a challenging task. Therefore, in this paper we are focusing on the design and development process of IDA's spoken dialogue components in this particular demanding context.

2. Dialogue design

The application is designed to turn on once the driver starts the car engine. At the beginning IDA greets the driver and asks about the driving direction and parking intention. Upon driver's confirmation IDA searches in an online database – provided by the local land authority (LTA) - for available car park lots and makes three car park suggestions. The car parks are by default ranked on distance proximity to the driving

destination. The driver can change the default setup configuration using speech or touch. If the driver dislikes the suggestions he/she can request for alternatives up to 3 times before IDA switches to manual mode; here the driver can search by himself for available parking once the car is standing. If the number of available lots drops to a critical level IDA informs the driver and starts searching for parking alternatives. When the driver reaches the destination the application can be set to send an SMS with the exact parking location (see fig.1).

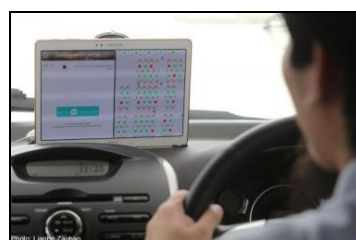


Figure 1: Test driver using IDA

The driver has the option to turn off the application at any moment in time by saying: "Dismiss". Fig. 2 shows IDA's interaction flow design.

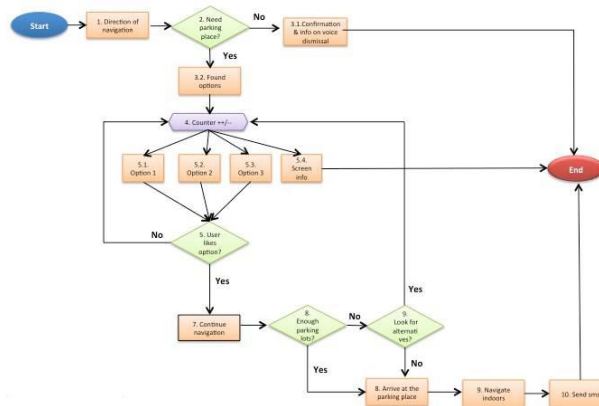


Figure 2: Interaction flow design

2.1. IDA's personality

To shorten the dialogue and enhance the user experience we designed some personality features for IDA through the use of humor. When the driver keeps rejecting IDA's parking's suggestions as being "too expensive" the system shows disapproval: "Expensive? This is cheap lah! If you want cheaper go Malaysia! Anyway, I don't have anything below this price! Are you sure you don't want this place?". The surprise effect is increased by the fact that IDA uses typical Singlish¹ words and expressions when she gets upset. IDA

¹ Singlish is a locally spoken English variety

also gives similar responses to undecided drivers who keep asking for the "next" option: "*Again next? You drive me crazy! This place is nice and cheap! Why not taking?*". IDA's dominant personality shows up each time the driver persists in rejecting her suggestions for more than 3 times in a row, i.e. just before turning into manual mode. In cases where the noise level reaches a critical level hindering the ASR, IDA asks the driver to stop talking to other passengers, to turn off the radio or to speak louder. To note that for common, native English speakers IDA's remarks might sound rude, however the expressions are common in the Singaporean slang and do not hold the usual negative connotation.

3. Data collection

When we started working on the dialogue design we realized that we did not know how a typical driver would talk with IDA, i.e. what particular sentences, phrases, words etc. would use. To solve this problem we created on paper 5 use case scenarios, each one containing 6-8 questions that IDA would typically ask while interacting with the driver. Further, we asked 20 persons (5 females and 15 males) to fill in the answers. The users were local Singaporeans or permanent residence holders with a good level of English and driver license. We collected 341 sentences from a total of 100 simulated dialogues. The sentences helped us to improve the original dialogue design, as well as to enhance the vocabulary and adapt the language model (LM) to the current application domain. Additionally, 5 hours of speech data were collected from 4 different speakers. The data was collected in two different environments - office (quiet) and car (noisy). The collections served to train, test and adjust the existent acoustic models (AM) to the Singaporean accent.

4. Generating ASR and NLU components

We trained a large vocabulary continuous speech recognizer (LVCSR) with a Deep Neural Network Hidden Markov Model (DNN-HMM) architecture using the Kaldi open source toolkit [2]. The AM was trained from over 700 hours of relatively clean mobile phone speech data, with a roughly equal distribution of UK, US and Singapore English accents. A target lexicon was generated from the 341 text sentences collected as part of potential dialogue interactions, to which "must-have" words, such as local parking destinations, alternative pronunciations and driving specific terms were added. This reduced the lexicon size from over 148000 to 5300 entries. The LM was adapted from an initial LVCSR LM trained from over 1TB of sentences from the English Gigaword corpus through linear interpolation with a target LM. This target LM was trained from sentences selected from 150GB of web-discovered text applying two heuristics: the first uses cross-entropy [3], accepting only sentences above a threshold when evaluated with a domain LM trained with the target lexicon. The second accepts sentences containing at least one of the "must-have" words in the target lexicon. This biases the LM towards sentences that might involve driving and parking actions.

Our initial analysis using the unadapted AM showed a severe mismatch between the mobile phone channel and the in-car microphone used by our application. Additionally, the in-car reverberation and other usual driving sounds worsen the acoustic conditions. In spite all these obstacles, we were able to achieve roughly a 5 to 6% Word Error Rate (WER) for clean, in-office environment and 14% WER when driving.

The 1st best ASR result was then passed to the NLU module. The module uses a XML-based grammar to convert the string into a list of values. This list is used to identify the type of query and its arguments; for example, the request "*How about Takashimaya?*" would be classified as a request for parking at Takashimaya shopping mall or at a nearby location. Shorter utterances, such as "*yes*", "*no*", or "*I want this*" were also mapped to positive or negative responses using certain rules which would then advance the scenarios outlined in Figure 2.

5. System architecture

The application GUI was developed in Java for the Android platform. GPS, Google Map, TTS, SMS are built-in features of the mobile/tablet with APIs from the standard SDK. We designed a custom protocol over TCP/IP for the client application and an ASR/NLP backend server to communicate over Wi-Fi/3G/4G links. The system architecture can be illustrated through the following use case:

1. The driver presses a Bluetooth button placed on the steering wheel to start recording the voice command. The audio is streamed to the back-end server via 4G/LTE.
2. On the backend server the ASR module recognizes the speech and sends the results to the dialogue module.
3. The dialogue module processes the text and sends the response to the client.
4. The backend server retrieves parking info from LTA and returns the parking info to client.
5. The client combines the parking info and dialogue response in a single answer which is then displayed on the tablet screen and played back using TTS.

6. Conclusions and future work

In this paper we present our work towards creating a natural language platform for IDA, an intelligent driving assistant for smart city parking in Singapore. In the future, we are planning to extend the system with multi-lingual capabilities, create iOS and Windows versions and improve the current GUI interface. Additionally, two user studies are currently planned to evaluate the user experience with IDA and to measure sustainability effects achieved by daily application use.

7. Acknowledgements

IDA application was developed as part of larger collaboration project within the newly established ITS lab concerning smart cities. The ITS Lab gathers researchers from automotive supplier Continental, A*STAR Institute for Infocomm Research (I²R), and the Technical University of Munich's Create center in Singapore (TUM-Create). The authors would like to thank all those who contributed to this project.

8. References

- [1] <http://www.stb.gov.sg/>
- [2] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely, "The Kaldi Speech Recognition Toolkit" in IEEE 2011 Workshop on Automatic Speech Recognition and Understanding, 2011.
- [3] A. Rousseau, "XenC: An Open-Source Tool for Data Selection in Natural Language Processing" The Prague Bulletin of Mathematical Linguistics, 2013, pp.73-82.