Yochina: Mobile Multimedia and Multimodal Crosslingual Dialog System

Feiyu Xu, Sven Schmeier, Renlong Ai and Hans Uszkoreit

Abstract Yochina is a mobile application for crosslingual and crosscultural understanding. The core of the demonstrated app supports dialogues between English and Chinese or German and Chinese. The dialogue facility is connected with interactive language guides, culture guides and country guides. The app is based on a generic framework enabling such novel combinations of interactive assistance and reference for any language pair, travel region and culture. The framework integrates template-based translation, speech-synthesis, finite-state models of crosslingual dialogues and multimedia sentence generation. Furthermore, it allows the interlinking between crosslingual communication and tourism-relevant content. A semantic search provides easy access to words, phrases, translations and information.

1 Introduction

The language barriers between Eastern and Western societies constitute big challenges for economic and cultural exchange. The dream of todays technophile travelers, educated by visionary science fiction, is to own a mobile speech-to-speech translation system functioning as a personal interpreter, allowing them to talk in their own language and to be understood in the language of the partner thanks to a combination automatic translation as well as speech recognition and synthesis. Recent technological breakthroughs incorporated into Apple's Siri and Google's Translate lend additional support to these expectations. However, as we know from experience and literature, both speech recognition and automatic translation are still far from being reliable [2]. Furthermore, the most reliable systems still suffer from slow re-

Feiyu Xu
Yocoy Technologies GmbH and DFKI LT Lab
e-mail: Feiyu.Xu@yocoy.com, feiyu@dfki.de

Sven Schmeier
Yocoy Technologies GmbH and DFKI LT Lab
e-mail: Sven.Schmeier@yocoy.com, schmeier@dfki.de

Renlong Ai
DFKI LT-Lab
e-mail: renlong.ai@dfki.de

Hans Uszkoreit
DFKI LT-Lab
e-mail: uszkoreit@dfki.de

sponse times depending on input length, complexity and internet access. Roaming costs still prevent travelers to enjoy the benefits of online translation and speech recognition services in most foreign countries such as China.

The Yochina crosslingual dialogue framework developed by Yocoy Technologies GmbH (Yocoy)[1] provides a realistic solution that helps foreigners to overcome language and communication barriers in countries such as China without depending on internet connection. The pragmatic approach guarantees correct translations. Yochina incorporates various language technologies such as speech synthesis, template-based translation, dialogue and semantic search [3, 4]. The framework provides the following functions:

- template-based and situation-based translation
- crosslingual dialogue
- multimodal dictionary
- semantic search
- interlinking of language and information
- spoken output

Fig. 1 Yochina: language, travel and culture guide for China

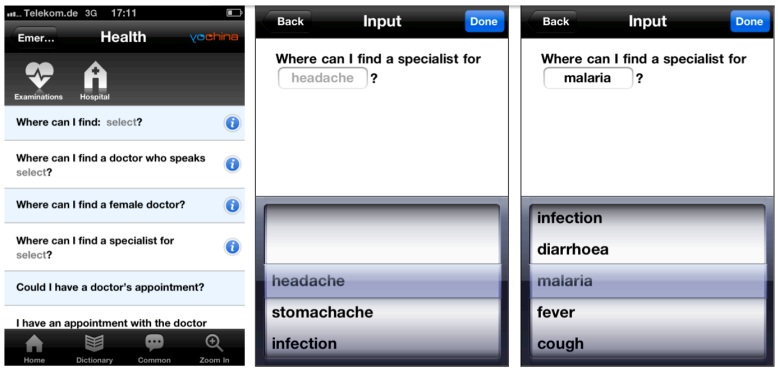
Yochina is implemented as a mobile application available for two language pairs (English to Chinese and German to Chinese) at the Apple app store. Yochina contains three major components: language guide, country- and travel guide and semantic search, as depicted in Figure 1. The remainder of the paper is organized as follows. Section 2 describes the Yochina crosslingual dialogue system. Section 3 explains a novel strategy of linking provided knowledge with covered communication situations. Section 4 shows the search function and the visualization of the search results.

[1] http://www.yocoy.com

2 Crosslingual Spoken Dialogue System

The Yochina language guide aims to help foreigners to formulate their wishes, requests and questions in their own languages by providing phrases, phrase templates and multimedia phrases. The phrase templates provide slots for filling in words or phrases expressing numbers, currencies, time points, countries, languages, body parts, medical symptoms etc. Figure 2 and Figure 3 show the entire workflow of a crosslingual dialogue containing five steps depicted by five screens. Users can choose their preferred phrases from screen (1). If they choose a template phrase such as the one on screen (2), they can fill the slots either from a pre-selected list or by free input as shown on screen (3). Screen (4) displays the translation of the sentence from screen (3) together with options for responses to be shown to the Chinese conversation partner for selection. Screen (5) shows the translation of a selected answer into the language of the user.

Figure 2 shows an example in which a user asks for medical specialists in particular areas. Figure 3 exhibits the translation of the completed request given in screen (4) together with options for responses by the Chinese conversation partners. If the users touches the speech bubble, the system will read the Chinese sentence in the yellow area. The next screen displays the translation of the selected Chinese response into the language of the user.



(1) (2) (3)

Fig. 2 Phrase templates with slots

Since most smart phones have a camera and can store pictures, we developed a new communication method allowing users to integrate pictures instead of referential phrases in their utterances. For example, the following phrase in Figure 4 refers to the object in the picture which users want to buy. Users can insert a picture from their iPhone photo album into the slot just like filling a slot with a text snippet. This new function comes in handy when a picture can spare the user from describing a complex object.

(4) (5)

Fig. 3 Crosslingual dialogue with bidirectional translations

Fig. 4 Multimedia phrase

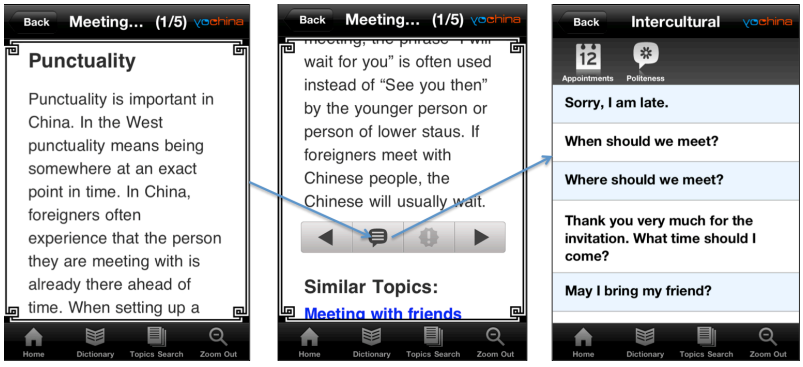


Fig. 5 Linking between communication and content

3 Interlinking of Communication and Knowledge

Traditionally language guides and travel guides are separate products. Thus there is no linking between descriptions of points of interests, historical events, cultural

information or inter-cultural hints on the one side and and useful related phrases on the other. However, in our real world, communication and information is tightly connected in many ways. During our conversations, we search for information and knowledge for understanding what we hear and better explaining what we mean. On the other hand, new information or content inspires and supports communication. In Yochina, we annotate content on country and culture with phrases which are useful in the context of the information. Figure 5 depicts one example of such linking, here between the intercultural issue of ”punctuality” and related phrases. The linking is signaled by the language bubble symbol on the second screen.

4 Semantic Search and Expansion

In Yochina, all words, phrases and content are indexed for free text search. Given a word as search query, users can find 1) the exact match and the related words with their translations and speech output, 2) phrases semantically related to the word, and 3) travel information mentioning the keyword. Figure 6 shows an example with the query *internet*. Given this query, the users are shown internet-related words and phrases or sentences around internet access and costs. Furthermore, Yochina explains the internet usage situation in China.

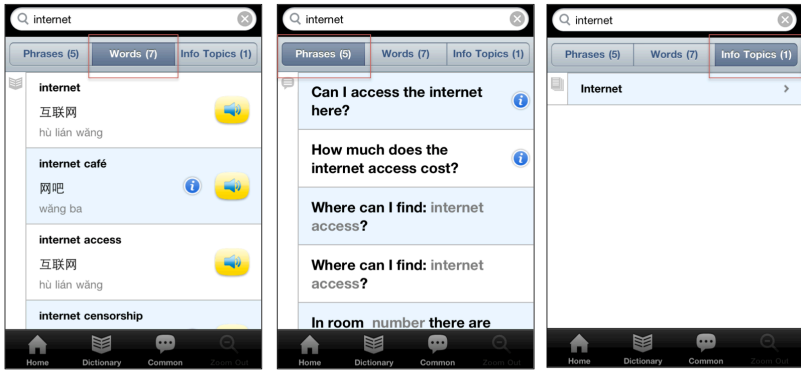


Fig. 6 Search

5 Robust offline application of embedded ASR

Although the current Yochina applications available on the market do not feature speech recognition technologies, extensive research has been conducted with the aim to integrate embedded automatic speech recognition (ASR) technology. ASR is used to activate phrases and phrase templates available in the Yochina dialogue

grammars. In [1], Yochina phrases and grsammars have been adapted to various grammar formats of the corresponding ASR tools. Special methods have been developed to convert the template phrase slot features into features allowed by the corresponding grammar formats. An n-best recognition strategy has been applied to ensure the targeted robustness. Three ASR tools have been tested for this specific task in our experiments. Fonix performs a little better than SVOX and Nuance (overall recognition rate: Fonix 87.4%, SVOX 85.9% and Nuance 84.0%). But Nuance exhibits the best recognition for non-native speakers and in noisy open air situations, while SVOX's result is best with female testers and native speakers. Although the recognition performance would not suffice for a sufficiently reliable speech-to-speech translation, it turned out that by restricting the vocabulary to the words needed for semantic access to the situation-relevant phrases a satisfactory recognition performance can be accomplished.

The upshot of the experiments were therefore that a restricted utilization of speech recognition for fast access will circumvent the problems of free speech input while still freeing the user from typing in the entire sentences and phrases to be translated.

6 Conclusion and Future Work

The demonstrated app exemplifies a thoughtful combination of various language technologies into a successful real-life application. We have argued for a creative novel pragmatic combination of matured technologies into a reliable product that avoids the pitfalls of imperfect leading-edge techniques such as completely free automatic translation and free speech recognition. Because of the modular design of the application framework, the modules for translation and speech input processing can be substituted at any time by improved MT and ASR technologies as soon as they reach the needed level of reliability. The continuous testing and gradual incorporation of maturing technologies into a modular application framework has proven an appropriate approach for securing maximal user benefits and technical competitiveness.

References

1. Ai, R.: Fuzzy and intelligent match between speech input and textual database. Master's thesis, Technical University of Berlin, Berlin, Germany (2010)
2. Feng, J., Ramabhadran, B., Hansen, J.H.L., Williams, J.D.: Trends in speech and language processing [in the spotlight]. IEEE Signal Process. Mag. **29**(1), 177–179 (2012)
3. Uszkoreit, H., Xu, F., Liu, W.: Challenges and solutions of multilingual and translingual information service systems (invited paper) (2007)
4. Uszkoreit, H., Xu, F., Liu, W., Steffen, J., Aslan, I., Liu, J., Müller, C., Holtkamp, B., Wojciechowski, M.: A successful field test of a mobile and multilingual information service system compass2008 (2007)