Mobile Intelligent Virtual Agent with Translation Functionality

Inguna Skadiņa, Inese Vīra, Jānis Teseļskis, and Raivis Skadiņš

Tilde, Vienibas gatve 75a, Riga, Latvia
{inguna.skadina,inese.vira,janis.teselskis,
raivis.skadins}@tilde.lv

Abstract. Virtual agent is a powerful means for human-computer interaction. In this demo paper, we describe a new scenario for mobile virtual agent that, in addition to general social intelligence, can perform translation tasks. We present the design and development of the intelligent virtual agent that translates phrases and sentences from English into French, Russian, and Spanish. Initial evaluation results show that the possibility to translate phrases and short utterances is useful and interesting for the user.

Keywords: virtual agent, multilingual information systems, mobile applications, automated translation.

1 Introduction

Nowadays, as mobile devices become ever more powerful, communication through virtual agents seems to be the most appropriate way to interact with a computer or mobile device. Several commercial applications, e.g., *Apple Siri* and *Google Now,* are the first important steps towards multimodal communication between a user and a computer.

Human-computer interaction has been actively researched for many decades. Different aspects, like dialog management, interactivity, reactive behavior, and others, are studied. Based on these studies, different applications are proposed – assistants, tutors, simple chatbots, etc. Although various virtual agents have been developed, translation tasks were not their primary focus. Some mobile phone applications provide multilingual translation services for a particular domain [1] without a multimodal component. Others have been developed as language teachers [2].

In this paper, we present an attractive 3D multimodal virtual agent Laura[1] for *Android* devices. Laura is a freely talking 3D head with natural mimics and synchronous lip movements, as well as emotions. The agent can answer questions and handle simple dialog based on AIML, *Artificial Intelligence Mark-up Language,* [3]. Besides answering a variety of questions, the agent can also translate words, phrases, and sentences from English into Spanish, French, and Russian.

[1] First version of Laura is available from https://play.google.com/store/apps/details?id=com.tilde.laura&hl=en

E. Métais, M. Roche, and M. Teisseire (Eds.): NLDB 2014, LNCS 8455, pp. 258–261, 2014.

2 Overview of the Virtual Agent

Our work is directed by the growing need for multilingual communication. Taking into account the wide use of mobile devices and user preference for speech input, we developed the virtual agent as a user friendly, socially intelligent, and freely talking 3D head. Although Laura can communicate about a wide range of themes, its novel function is a voice translation (Fig. 1.).

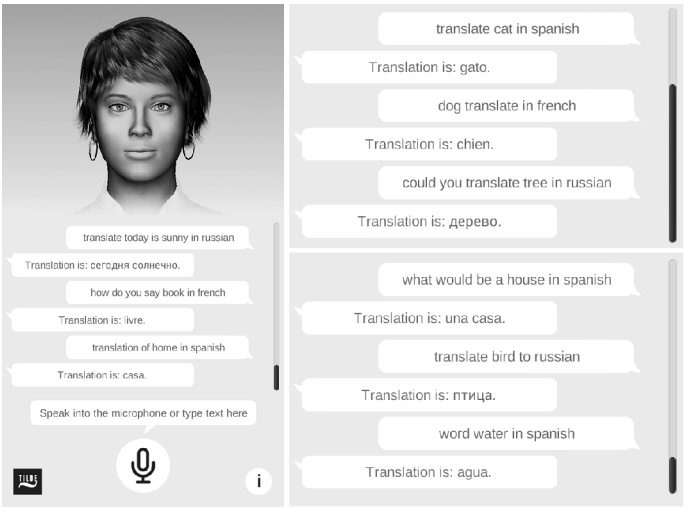

Fig. 1. Visual interface of the Laura application

The general architecture for the translation agent is shown in Fig. 2. It consists of five main constituents:

- Mobile application (implemented using the *Unity game engine* [4])
- Virtual agent web service (based on AIML language) – responsible for intelligent conversation between the agent and the user
- Automatic speech recognition (ASR) service – used for voice input
- Text translation service – used for translation of the user's queries
- Speech synthesis service (TTS) – used for speech output

Dialogue with the virtual agent can be managed through the voice command or typed in by using the keyboard. In the case of voice interaction, the user's request is recorded using the Unity engine microphone class to capture sound from the mobile device's built-in microphone. After being captured, audio is sent to the ASR service and received as a recognized text string. If the user prefers text input, text is sent directly to the Virtual agent web service that checks input patterns and gives appropriate answers accordingly.

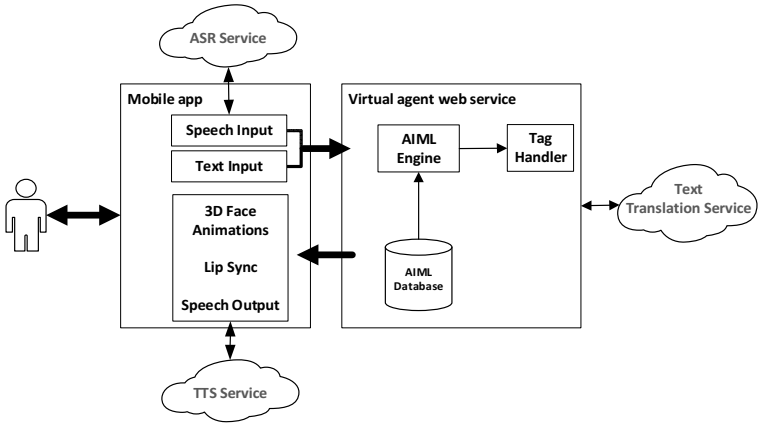

Fig. 2. Design of the system

Laura's replies are based on AIML knowledge that is extended with external web services. The text is sent to the AIML knowledge database to search for the proper answer. The basic structure of the AIML file is defined by tags [3], including beginning and ending tag, category, pattern, and template. The advanced AIML database can include tags for defining the avatar`s emotions, tasks for AIML connectivity with a mobile device, and properties to connect with external services like text translation. For text translations, we implemented our new <translate> tag with two inputs for the text and the language.

Before implementation, a user study was performed to find out how people usually ask for translation. The most common phrases were "translate WORD to LANGUAGE" and "WORD translate in LANGUAGE", in addition to "Hi, could you, please, translate WORD in LANGUAGE".

An answer generated by the Virtual agent web service is sent to the TTS web service, and the user can hear the answer. To make the 3D model talk, we used language independent real time Lip Sync engine that dynamically creates mouth movements.

3 Evaluation

For evaluation, a questionnaire was used containing three groups of questions – information about the user's experience, questions related to the general features of the agent, and questions related to the translation task. 18 people between the ages of 20 and 50 (56% aged 20-30), with a gender breakdown that was 50:50, were involved. Most of the participants (61%) had experience using mobile devices for more than 3 years, while 28% used mobile devices for less than 1 year. Most of the participants (83%) thought that the agent is user friendly and 76% of participants were satisfied with the answers provided by the agent. Voice input was used by 83% of participants.

For the translation task, we analyzed two aspects – interaction and correctness of translation. Where it concerns the simplicity of interaction, 83% said "yes". To the question "Was the answer correct?", 11% responded "yes, always", 72% – "yes, mostly", 11% - "no, most of the time", and 6% said that the "answer was incorrect". However, for the question "Did Laura understood the question?", we got 67% negative answers (probably because of speech recognition quality), therefore we plan to analyse reasons of this contradictionary answer with help of additional evaluation.

Participants were asked to write down what they said to the agent. Most of participants used simple queries like "translate WORD to LANGUAGE" and "WORD translate in LANGUAGE". Most interesting queries were: "Dear Laura, could you, please, translate WORD in LANGUAGE" and "Do translation for WORD in Spanish language" or "Laura, please translate WORD in LANGUAGE".

In regards to what improvements users would like to see in the near future, most of the participants mentioned the addition of more languages. Some advised to make the application's primary function be translation, while others suggested it work more like a dictionary.

4 Conclusion

In this paper, we presented the intelligent virtual agent that translates phrases and sentences from English into French, Russian, and Spanish. Preliminary evaluation results show that the possibility to translate phrases and short utterances is useful and interesting for the user. Following user suggestions, we plan to add more languages for the translation task as well as make conversation with our agent multilingual.

Acknowlegments. The research leading to these results has received funding from the research project "Information and Communication Technology Competence Center" of EU Structural funds, contract nr. L-KC-11-0003 signed between ICT Competence Centre and Investment and Development Agency of Latvia, research No. 2.1 "Natural language processing in mobile devices".

References

1. Paul, M., Okuma, H., Yamamoto, H., Sumita, E., Matsuda, S., Shimizu, T., Nakamura, S.: Multilingual Mobile-Phone Translation Services for World Travelers. In: COLING 2008 22nd International Conference on Computational Linguistics: Demonstration Papers, pp. 165–168. COLING Demos (2008)
2. Hazel, M., Mervyn, A.J.: Scenario-Based Spoken Interaction with Virtual Agents. In: Computer Assisted Language Learning, vol. 18, pp. 171–191. Routledge, part of the Taylor & Francis Group (2005)
3. Wallace, R.S.: The Anatomy of A.L.I.C.E. In: Epstein, R., Roberts, G., Beber, G. (eds.) Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer, pp. 181–210. Springer (2008)
4. Unity game engine, `https://unity3d.com/unity`