*Human Language Technologies – The Baltic Perspective*
*A. Utka et al. (Eds.)*

*doi:10.3233/978-1-61499-442-8-46*

# The Development of Conversational Agent Based Interface

Inese VĪRA[1], Andrejs VASIĻJEVS[2]
*Tilde, Latvia*

**Abstract.** In this paper we present two prototypes of 3D based virtual agents: one chatbot which in addition to the ability to hold a conversation can perform translation from English into Spanish, Russian, and French; and another which supplies currency conversion (lats to euro and euro to lats) in the Latvian language. Both chatbots are voice controlled, with natural mimicry and representations of human-like emotions. We describe the motivation, development process, design and architecture of these mobile applications. The evaluation of both applications and their usage in selected scenarios is also presented.

**Keywords.** Conversational agent, speech interfaces, TTS, ASR, chatbot, question answering, multilingual information systems, AIML, Latvian

## Introduction and Motivation

In this paper we present our work on implementing the concept of conversational agent based interfaces. According to (Karat et al., 2002) the goal of conversational technologies is to close the gap between human-computer interaction and human-human interaction by leveraging expertise in human-to-human conversational interaction. In the conversational interface command based and mouse-click based interactions are replaced by interactions in naturally formed sentences and touch or keyboard strokes, and textual messages are replaced by spoken input and speech output.

Conversational interfaces are particularly useful on mobile devices that have smaller screens and do not have a physical keyboard and mouse. Examples of a broad scale implementation of conversational interface in consumer devices are *Apple Siri*[3], *Microsoft`s Cortana*[4] and *Google Conversational Search*[5]. These virtual assistants provide speech based interfaces to perform various tasks on the smartphone – from making a phone call to creating an appointment in the Calendar and finding

[1] Tilde, Vienibas gatve 75[a], Riga, LV1004, Latvia; e-mail: inese.vira@tilde.lv
[2] e-mail: andrejs@tilde.com
[3] http://www.apple.com/ios/siri/
[4] http://www.windowsphone.com/en-us/features-8-1
[5] http://www.cnet.com/news/googles-conversational-search-arrives-with-new-chrome/

information on the Web. The virtual assistant *Nina*[6] by Nuance Communications is targeted at multichannel customer service applications via a human-like conversational interface.

In the research field there is a growing interest to explore opportunities beyond the basic functionality – to detect, recognize and react to nonverbal behavioral cues, such as laughter [1] or gestures [2], or even "sense" the smell of objects it sees by recognizing the object`s form, size or structure, for instance, pizza [3].

Our goal is to research and test the feasibility of further extending the notion of conversational interface by embodying it in a human-like virtual character. Such a character should add visual components to interactions – lip movements for speech and facial animation for displaying human-like emotions. Our assumption is that a human-like agent will make conversation more natural and appealing, complementing the cognitive aspects of conversation with an emotional dimension.

Some "embodied" virtual agent applications have already been published recently for both desktops (*Denise*[7] virtual assistant by Guile3D Studio) and mobile devices (*Talking Pocket Bot Lucy*[8], *Pocket Blonde Smart Assistant*[9] for Android devices). Although they can assist in finding information, making calls or writing text messages, they are mostly focused on entertainment and emphasize the feminine characteristics of the female agent. Other recent entrants in the field of virtual agents use a knowledge base to make conversation smarter, but are implemented as fully fictitious humanoid avatars (e.g. *Talking and Answering Jenny*[10]).

Our broader interest is to research how conversational agents can be applied beyond the field of pure entertainment: how they can be used to meet the needs for social communication; how they can make finding necessary information more natural; what role do emotional expressions play in communication with virtual agents, etc.

In addition we are interested to research multilingual aspects of conversational interface. As the research and development in this field is mostly focused on English speaking users, we are researching opportunities for extending the use of virtual agents to other languages, in particular, the applicability of the current speech technologies for Latvian.

In this paper we describe the architecture, key components, and the approach to visualization for our generic virtual agent technologies. Based on these technologies we have created and evaluated two experimental agents – Laura and Eriks – for three usage scenarios. Laura has been created in a version for conversational dialog (chatbot scenario) and in a version for a translation scenario. Eriks has been developed as a task specific agent to perform currency conversions.

## 1. Architecture, Components and Visualization of Virtual Agents

Conversation interface software should integrate a number of key technologies to enable dialog with a virtual agent: voice activation (if the agent is operated fully hands-

6 http://www.nuance.com/landing-pages/products/nina/

7 https://guile3d.com/en/

8 https://play.google.com/store/apps/details?id=com.kauf.botv3.talkingpocketbotlucy&hl=lv

9 https://play.google.com/store/apps/details?id=com.brainyfriends.voice.game.pocket.blonde

10 https://play.google.com/store/apps/details?id=com.kauf.talking.account1.TalkingandAnsweringJenna Bot

free), speech recognition and transcription, intent and meaning analysis, data search and query, generation of response, and text to speech for voice based response [4]. In our solution the activation of user interaction is manual by pressing the microphone icon or typing a text in the dialog box.

Figure 1 represents the general architecture and the main components of our solution for generic virtual agent technologies. It uses a web-service approach with three-tiers: (1) a mobile application performing input and output functions and a graphical rendering of all animations; (2) online web-services for dialog processing tasks: intent and meaning analysis, data query, generation of response, user profile management; (3) external web-services such as a translation service, Web search, querying of external knowledge bases etc.

Text-to-speech (TTS) and automatic speech recognition (ASR) also rely on external web-services or on the device's built-in features (e.g. Android voice input feature). The calculator module for the Eriks application is to only locally perform data processing for the currency conversion. The overall dialogue is based on Artificial Intelligence Mark-up Language (AIML) [5].

This architecture enables extensions with additional services to support multilingual and multimodal communication and sources of knowledge for intelligent dialog.

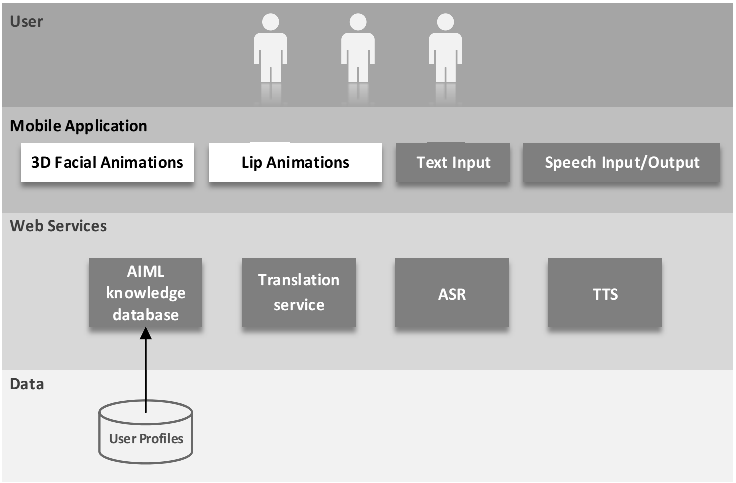


**Figure 1.** Components of the conversional agent application.

We needed to take into account the specifics of mobile devices such as performance, the screen and text/voice controls, as well as the specifics of the touch interface. To achieve optimal performance and user experience we elected to use a 3D head as the avatar not the whole body. As it is the close-up of the face, we have had to make high quality visuals with human-like mimicry, lip animation, and emotions.

After a comparative study of different technologies we selected *Autodesk Maya*[11] and the *Unity* game engine[12] as development and running environments. We used *Autodesk Maya* for creating the 3D avatar, lip and facial animations. The *Unity* game engine came handy when organizing all assets, including the user interface, creating 3D heads with textures and animations, lip-syncing, as well as connecting with the web

[11] http://www.autodesk.com/products/autodesk-maya/overview
[12] https://unity3d.com/unity

services. An important task was finding the best solution for natural-looking lip animation to correspond to the spoken phonemes and user friendly and smooth visual expressions represented by 3D virtual avatars, male and female. After each important update a user study was performed with the aim of assessing the new improved features.

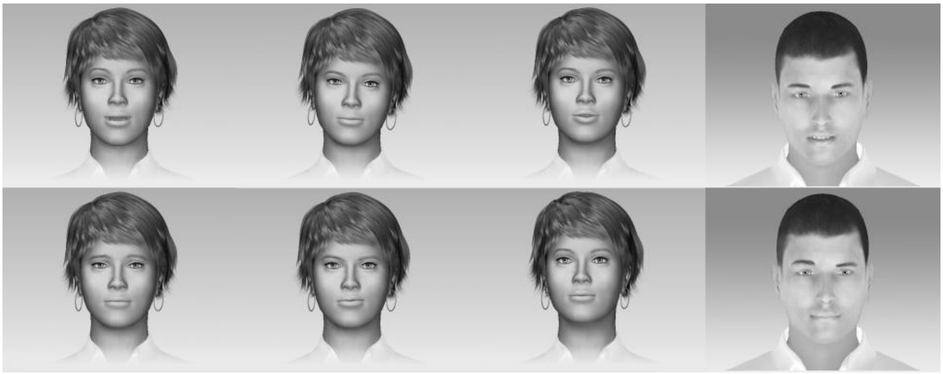

**Figure 2.** Head animations and emotions of virtual agents Laura and Eriks.

For the further research we selected three scenarios for which we created virtual agents: virtual conversation (chatbot) scenario, translation scenario and, a currency conversion scenario.

## 2. Virtual Conversational Agent Laura

The objective of the first scenario was to create an engaging and entertaining English speaking female chatbot. Its main task is to keep up a conversation mimicking natural human-human interaction without deeper intelligence or subject specific knowledge. We applied AIML technology from ALICE (Artificial Linguistic Internet Computer Entity) [5] chatbot inspired by Weizenbaum's ELIZA [6]. AIML (Artificial Intelligence Markup Language) is an XML dialect where user input patterns are provided and responses to these patterns are defined. There can be more than one answer to the question or sentence which is usually defined by a random function.

```
<category>
 <pattern>
 TRANSLATE TO FRENCH LANGUAGE *
 </pattern>
 <template>
 <translate lang="fr"><star/></translate>
 </template>
</category>

<category>
   <pattern>
   TRANSLATION OF WORD * IN SPANISH
   </pattern>
  <template>
   <translate lang="es"><star/>
   </translate>
  </template>
</category>
```


**Figure 3.** Sample of translation patterns in AIML language (left) and as displayed to the user (rights).

We also experimented with the emotional state of the chatbot regarding the spoken text (Figure 4).

User: How are you today?
Laura: I am feeling fine. Thank you.

User: How are you today?
Laura: Not so well today, but thank you for asking.

User: How are you today?
Laura: Good. Really good. How are you?

**Figure 4.** Mimicking emotional states.

We introduced an additional tag <emotion> for expressing emotions like smile, sadness, and happiness. The AIML sample code for the emotional dialog is shown in Figure 5.

```
<category>
<pattern>HOW ARE YOU TODAY</pattern>
<template>
  <random>
    <li>I am feeling fine <emotion>smile</emotion>.Thank you. </li>
    <li>Not so well today <emotion>sad</emotion>, but thank you for asking.</li>
    <li>Good. <emotion>happy</emotion> Really good. How are you? </li>
  </random>
</template>
</category>
```

**Figure 5.** Example of the application of the tag <emotion>.

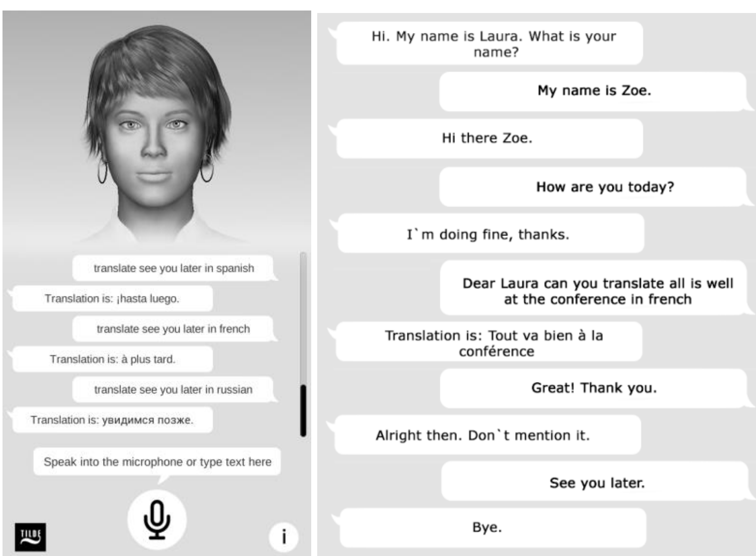


**Figure 6.** The chatbot Laura with the translation feature.

In the second scenario we extended the chatbot functionality with a translation feature. We introduced a new AIML tag <translate> with two inputs (Figure 3): the text and the language.

To define the translation patterns a user survey was performed to establish how people ask for translation. The most common phrases were “WORD translate in LANGUAGE” and “translate WORD to LANGUAGE”, in addition to such requests as “Hi, could you, please, translate WORD in LANGUAGE”. In our prototype application we implemented translation from English to the Spanish, Russian, and French languages (Figure 6).

## 3. Virtual Agent Eriks for Currency Conversion

Our third experimental application is for a currency conversion scenario. This work had two goals:

- to research the application of conversational agents for specifically defined tasks;
- to research how current spoken technologies for Latvian can be applied to a conversational interface.

This year Latvia went through a currency conversion from Latvian lats to the euro. The application allows users to say (or type) the amount in Latvian lats or euros and the application converts it into the opposite currency. It recognizes numbers (0 to 999) with or without the fractional part, and with or without the currency at the end (48 *lati*; 8,03; 170,11 *eiro*). The result is spoken by the 3D virtual male Eriks and also shown on the screen. The application is unique for its speech interface in the Latvian language – automatic speech recognition (ASR) [7] for numbers and Latvian text-to-speech synthesizer (TTS) [8] .


**Figure 7.** The user interface of the currency conversion application Eriks.

The application is the first application providing automatic speech recognition for the Latvian language.

## 4. Evaluation

There were two questionnaires conducted to assess the experimental application Laura [9]. 8 people participated in the first evaluation and 23 in the second, with a 50:50 gender breakdown and majority of participants in the 20-40 age group. Most of the participants (83%) thought that the virtual agent is friendly. Regarding the dialogue and the answers given by the virtual agent Laura, 76% of respondents were satisfied with the replies provided, and 83% of users assessed the interaction to be simple.

The second evaluation was related to the translation task and involved the evaluation of the interaction with the user and quality of the given translation answers. 70% of the users preferred to work with the 3D avatar and the voice interface, while 17% preferred communication using voice and keyboard, but 13% preferred communication through text only. To the question “Was the answer correct?” - 11% responded with “yes, always”, 72% with “yes, mostly”, 11% with “no, most of the time”, and 6% with “answer was incorrect”. However, only 33% responded positively to the question “Did Laura understand the question asked”?

To assess the translation scenario, participants were asked to evaluate the natural language processing quality for short and longer phrases and sentences separately. Assessing the quality of ASR, 63.4% of the speech recognition results were correct for short phrases or sentences, while for longer sentences only 45.2% were recognised correctly. Perhaps the rather low recognition rate was influenced by the fact that participants were not native speakers of English. 72.2% of short sentences and 53.6% of longer sentences were correctly recognized as translation requests. In evaluation of translation quality, 61.8% short sentence translations and 42.8% longer sentences translations were assessed as correct.

For Laura application user feedback received by Google Play evaluation system is 3.3 stars from 235 people, with 14689 installs. Application is recommended on Google+ by 126 people.

For the currency converter application Eriks we incorporated a visual feedback option showing the good (thumbs up) and the bad (thumbs down) review of the current recognized number. In total 2896 votes were received – 1393 positive and 1503 negative. Although negative votes exceed positive by 7%, the rate of positive votes is still relatively high taking into account that the usage of the application in real-life situations is in noisy environments.

Eriks application results on App Store show 3 stars from 13 ratings in total. On Google Play this application is rated with 3.2 stars from 69 respondents, with a total installs by 1690 people. It is recommended by 41 respondents on Google+.

## 5. Conclusion

In this paper we have presented the ongoing work towards the development of mobile conversational agents Laura and Eriks. Evaluation shows that Laura’s ability to answer questions and successfully handle simple dialog in English is positively received by the majority of users. Laura’s novel translation feature also shows positive evaluation results confirming our hypothesis that users prefer human-embodied virtual agent against speech-only dialog.

For the next step, we plan to extend the functionality for dialog in several languages, to experiment with the agents' emotional responses during the dialog, and to increase the intelligence of the agent by adding web-service based knowledge sources.

## Acknowledgments

The research leading to these results has received funding from the research project "Information and Communication Technology Competence Center"[13] of EU Structural funds, contract nr. L-KC-11-0003 signed between the ICT Competence Centre and the Investment and Development Agency of Latvia, research No. 2.1 "Natural language processing in mobile devices".

We would like to thank our colleagues Inguna Skadiņa and Jānis Teseļskis for their contribution in the research and development, as well as all the participants in the evaluation activities.

## References

[1] S. Scherer et al. Spotting laughter in naturalistic multiparty conversations: a comparison of automatic online and online approaches using audiovisual data. *ACM TiiS: Special Issue on Affective Interaction in Natural Environments*, 2(1):4:1{4:31, 2012.

[2] M. Thiebaux et al. Smartbody: behavior realization for embodied conversational agents. In *Proc. of AAMAS'08*, AAMAS '08, pages 151-158, 2008.

[3] T. Balint, J. M. Allbeck, What's Going on? Multi-sense Attention for Virtual Agents. In R. Aylett, B. Krenn, C. Pelachaud, H. Shimodaira (eds.) *IVA 2013* (2013), LNCS, vol. 8108, pp. 349-357. Springer, Heidelberg.

[4] T. Mozer, Speech's Evolving Role, In *Consumer Electronics… From Toys to Mobile. In Mobile Speech and Advanced Natural Language Solutions*, pp. 23-34. Springer, New York, 2013.

[5] R. S. Wallace, The Anatomy of A.L.I.C.E. In R. Epstein, G. Roberts, G. Beber (eds.) *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer* (2009), pp. 181-210. Springer.

[6] J. Weizenbaum, ELIZA - a computer program for the study of natural language communication between man and machine. In *Communications of the ACM*, 9(1), 36-45, 1966.

[7] A. Salimbajevs, M. Pinnis, Towards Large Vocabulary Automatic Speech Recognition for Latvian, In *Proceedings of the 6th International Conference on Human Language Technologies—the Baltic Perspective*, 2014

[8] K. Goba, A. Vasiļjevs, Development of Text-To-Speech System for Latvian. In *Proceedings of NODALIDA 2007,* Tartu, Estonia, 2007.

[9] I. Vīra, J. Teseļskis, I. Skadiņa, Towards the development of the multilingual multimodal virtual agent, In *PolTAL 2014*, Lecture Notes in Computer Science, Springer, 2014.

[13] http://www.itkc.lv