INTELLIGENT VOICE SYSTEM FOR KAZAKH

Zh.Yessenbayev*[1], M.Karabalayeva[2], A.Sharipbayev[2], F.Shamayeva[3], N.Saparkhojayev[4], M.Kalibekov[4], U.Yapanel[5]

[1]NURIS, Nazarbayev University, Astana, Kazakhstan; *zhyessenbayev@nu.edu.kz; [2]L.N. Gumilev Eurasian National University, Astana, Kazakhstan; [3]Korkyt-Ata Kyzylorda State University, Kazakhstan; [4]Suleyman Demirel University, Almaty, Kazakhstan; [5]Audience Inc., 331 Fairchild Drive, Mountain View, CA 94043 USA

INTRODUCTION.

The proposed project is dedicated to developing a prototype of an intelligent voice system with an interactive dialog mode in the Kazakh language for call-centers, information desks and dispatching services. Mathematical models and software of the system were developed. This includes the development of the algorithms of speech recognition and synthesis of words and phrases in Kazakh as well as the collection and processing of speech data for training and testing the system.

METHODOLOGY.

During the project implementation, speech data was collected in a professional studio. The data includes about 30 hours with orthographic transcriptions of Kazakhwords and phrases uttered by 169 speakers chosen to represent evenly geographical, gender/age and social (education) aspects.

SPEECH RECOGNITION AND SYNTHESIS.

For speech recognition, we used context-dependent tied-state continuous HMM with 8 Gaussian mixtures per state trained on MFCC vectors extracted from the data collected. The performance was 4.1 % WER. For speech synthesis we used HMM-based synthesis trained on audio recordings from Kazcorpus (http://www.kazcorpus.kz).

SYSTEM IMPLEMENTATION.

The system was implemented in Java featuring the following functionality (Fig. 1):

- Kazakh speech recognition using Sphinx4;
- Kazakh speech synthesis using MaryTTS;
- Call processing using Asterisk IP PBX;
- Dialog management and system configuration through web GUI;
- Support for VoiceXML and VoIP

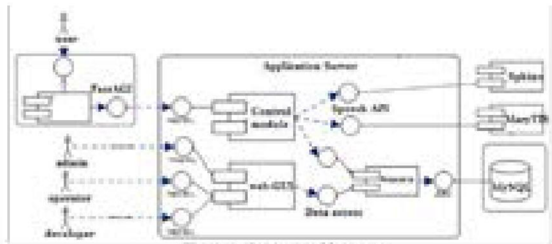

Figure 1. System architecture

CONCLUSIONS.

We implemented a prototype of the system IVoS, collected audio data, and performed speech processing experiments. We plan to improve the performance of speech recognition and synthesis modules and deploy a speech-enabled service within the university.

ACKNOWLEDGMENTS.

The work was supported by the grant of the Ministry of Education and Science of the Republic of Kazakhstan.