

Bilingual environment

From human to computer



Automatic Recognition of mixed Ukrainian-Russian Speech

Valeriy Pylypenko, Tetyana Lyudovyk

Abstract

This work presents an approach to recognition of conversational speech with code-switching which is widespread in Ukraine now. Both inter-sentential and intra-sentential Ukrainian/Russian code-switching is handled. The approach takes into account closely related Russian and Ukrainian phonetic systems. A cross-lingual ASR system is developed. The acoustic model and pronunciation lexicon are based on Ukrainian phone set. Experiments with different types of code-switching speech (Parliamentary, TV broadcast) were conducted and results are presented. The approach is suitable especially in cases of intra-sentential code-switching where language identification is problematic.

Introduction

Code-switching means alternation of different languages within or between sentences. In countries where more than one language is spoken by a significant number of people the code-switching speech is a common phenomenon.

In Ukraine the majority of population is bilingual. Code-switching between Ukrainian and Russian languages is a frequent event penetrating the broadcast media, meetings, parliamentary debates, court hearings etc. This complicates the automatic speech recognition.

The following issues should be accounted for:

- Ukrainian and Russian are closely related languages;
- Russian language spoken in Ukraine differs from Russian language spoken in Russia;
- code-switching speech is often speech with an accent.

Related Work

The problem of code-switching is relevant for Hong Kong, Singapore (English/Chinese); Taiwan (Mandarin/Taiwanese); India (English/Hindi) and other regions.

For code-switching automatic speech recognition a multi-pass and a single-pass approaches have been proposed.

The multi-pass approach consists in finding boundaries of monolingual fragments in multilingual speech, providing language identification (LID), and using an appropriate language monolingual system to recognize monolingual fragments. This approach is highly dependent on the accuracy of finding the boundaries of monolingual fragments.

In case of closely related languages (Mandarin/Taiwanese, Ukrainian/Russian) with similar phonetics LID becomes difficult. Within a single-pass approach, a bilingual recognition system is used with one bilingual acoustic model, one bilingual language model and one bilingual pronunciation lexicon. This approach does not imply the LID.

Peculiarities of Ukrainian and Russian Phonetics

Ukrainian and Russian languages are closely related: both are members of the East Slavic group, and both of them use the Cyrillic alphabet.

All symbols of Russian consonants are present in the Ukrainian phoneme set. On the contrary, 6 of Ukrainian consonants used for speech recognition are absent in Russian.

Devoicing of consonants in some positions is present in Russian and absent in Ukrainian. In fact, in real speech this devoicing is common for both languages.

As for vowels, there are two vowel reduction levels in Russian and no vowel reduction in Ukrainian. When Ukrainians speak Russian, they do reduce unstressed vowels, but not as markedly as Russians speaking their mother language.

Our main hypothesis is that people speaking Ukrainian and Russian in Ukraine use the same set of phonemes. As the Ukrainian phonetics includes all Russian phonemes it is possible to use an acoustic model based only on Ukrainian phone set to recognize Russian and code-switching speech.

Ukrainian/Russian Speech Corpora

Vocabulary:

613K words

Language modeling:

3632 millions words from Internet and open data lang.org.ua

Acoustic modeling:

	Ukrainian	Russian	Ukr+Rus
Parliament	236h	59h	295h
AKUEM	113h	141h	254h
Broadcast	17h	13h	32h
Total	366h	213h	579h

Methods

- The automatic speech recognition system is developed using open source Kaldi toolkit.
- For creating Russian and Ukrainian lexicons, automatic grapheme-to-phoneme (g2p) rule-based converters are used. Peculiarities of conversational speech are taken into account.
- For common words, numerals, proper and geographic names multiple phonetic transcriptions are used.
- For acoustic and language models Deep Neural Learning is used.

Automatic Speech Recognition Results

	Broadcast	AKUEM	AKUEM	Parliament
Language	Ukrainian/ Russian mix	Ukrainian	Russian	Ukrainian
Total number of word-forms	53789	88507	109815	22799
Word error rate, %	23.20	14.70	14.16	10.41



Department of Speech Recognition and Synthesis,
International Research/Training Center for
Information Technologies and Systems NASU,
Kyiv, Ukraine

valeriy.pylypenko@gmail.com
tetyana.lyudovyk@gmail.com