Testing of the Speech Recognition Systems Using Russian Language Models

Konstantin Aksyonov
Ural Federal University
Ekaterinburg, Russia
bpsim.dss@gmail.com

Dmitry Antipin
Ural Federal University
Ekaterinburg, Russia
diantigers@gmail.com

Tamara Afanaseva
Ural Federal University
Ekaterinburg, Russia
t.afanaseva42@gmail.com

Igor Kalinin
LLC "UralInnovation"
Ekaterinburg, Russia
igor_kalinin@hotmail.com

Ilya Evdokimov
Ural Federal University
Ekaterinburg, Russia
psp720@mail.ru

Aleksandr Shevchuk
Ural Federal University
Ekaterinburg, Russia

Andrey Karavaev
Ural Federal University
Ekaterinburg, Russia
karav197@yandex.ru

Uriy Chiryshev
Ural Federal University
Ekaterinburg, Russia
iurii.chiryshev@mail.ru

Egor Talancev
Ural Federal University
Ekaterinburg, Russia
i.spyric@gmail.com

Abstract

This article includes the results of testing an existing speech recognition systems, using the Russian language model and dictionary. Objects of the research are question-answering systems, call-center processes, methods and systems for language processing. Several problems with Russian language speech recognition were identified and studied.

Introduction

In the research, algorithms for speech recognition are considered. These algorithmsare used by the following systems for speech recognition: Sphinx [3], Yandex.Speech Kit [7, 8], Google Cloud Speech-to-Text [1, 5-6].

CMU Sphinx [3] – is a major open-source cross-platform project for speech recognition, developed in Carnegie Mellon University. It includes series of the Sphinx systems and program for studying acoustic model, which is called Sphinx Train.

Currently, CMU Sphinx includes language models for the several languages: English, Russian, German, Chinese and others. It gives an opportunity to create acoustic models for other languages. The project uses BSD license, which allows distributing the product commercially, the system also includes tools for speech recognition (keyword definition, pronunciation evaluation). The development is still in progress.

Yandex.Speech Kit – is a complex of speech technologies, developed by Yandex Company, it includes speech recognition and synthesis. Speech Kit is used as the cloud service Speech Kit Cloud and Speech Kit Mobile [7-8].

Speech Kit Cloud SDK – is a program, which allows developers to use instruments for speech recognition made by Yandex. Infrastructure of the service's design based on probability of high loads, to provide access and trouble-free system operations.

Interaction with Speech Kit Cloud is performs HTTP API and includes different functions:

1. Interactive Voice Response

2. Automatic calls to transmit information about new services, to confirm an order or delivery, to remind about the record, the collection of meter readings.
3. Inquiries by phone without operator's participation, recording for reception and maintenance.
4. The voice interface of "smart house" systems.
5. The voice interface of robots.
6. Site management, using voice.

Speech Kit Cloud is designed to recognize small speech fragments about 30 seconds long [7].

Speech Kit Mobile SDK [8] is a program that allows to embed speech recognition and synthesis into a mobile application on iOS, Android or WindowsPhone. Speech Kit Mobile SDK is used in the following Yandex services: Yandex Search, Browser, Taxi, Maps, Navigator, Translator, Market, Music, Launcher, Auto, Keyboard.

Google Cloud Speech-to-Text is a technology, which allows developers to convert audio into text by applying neural network models, using the API, which Google Cloud Speech provides. The API supports 120 languages and options for supporting the user base. The system allows using voice commands, managing, rewriting an audio from call centers and more. It can handle streaming or pre-recorded audio in real time using Google's machine learning technology.

1 Deploying process of the speech recognition systems for testing the methods and algorithms

Deploying CMU Sphinx and Google Speech Recognition.

To deploy the Pocket Sphinx system and Google Speech Recognition, a solution from a third-party developer was used, it is represented as a Speech Recognition Python package. It can be installed using the Pip tool:

pip install SpeechRecognition

The installation requires Python 3.3 and later, Pip. The Python wrapper package for the Pocket Sphinx system must be installed as well:

pip install pocketsphinx

We used Python 3.6. Installation was performed in the Windows terminal from PyCharm (using virtual environment) (Figure 1).

Figure 1. Installing the Speech Recognition package

By default, only the model for English is installed, the model for Russian is available by reference:

https://sourceforge.net/projects/cmusphinx/files/Acoustic%20and%20Language%20Models/Russian/

Files of one of the downloaded models should be placed in the pocket sphinx-data folder in the folder of the speech_recognition package (Figure 2).

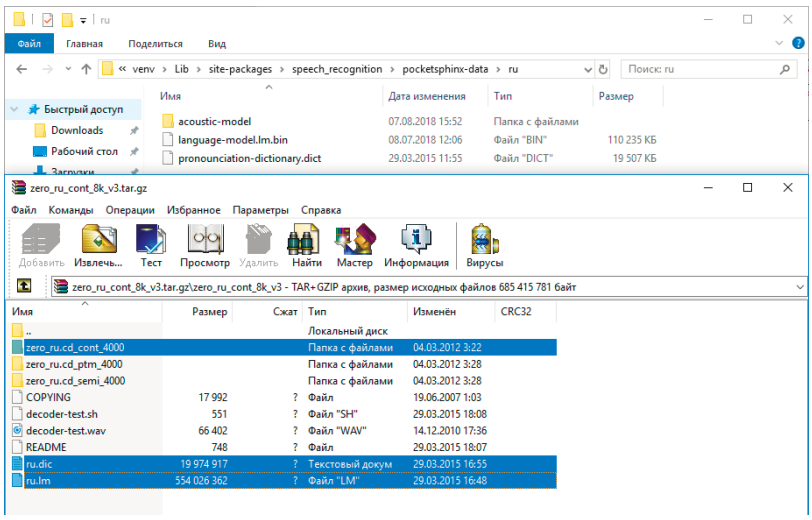

Figure 2. Replacing files

After that, the launch of the following Python script will display the result of speech recognition in both systems (the data from the pre-recorded audio file is recognized - Figure 3).

The GitHub repository of the SpeechRecognition package and installation instructions:

https://github.com/Uberi/speech_recognition

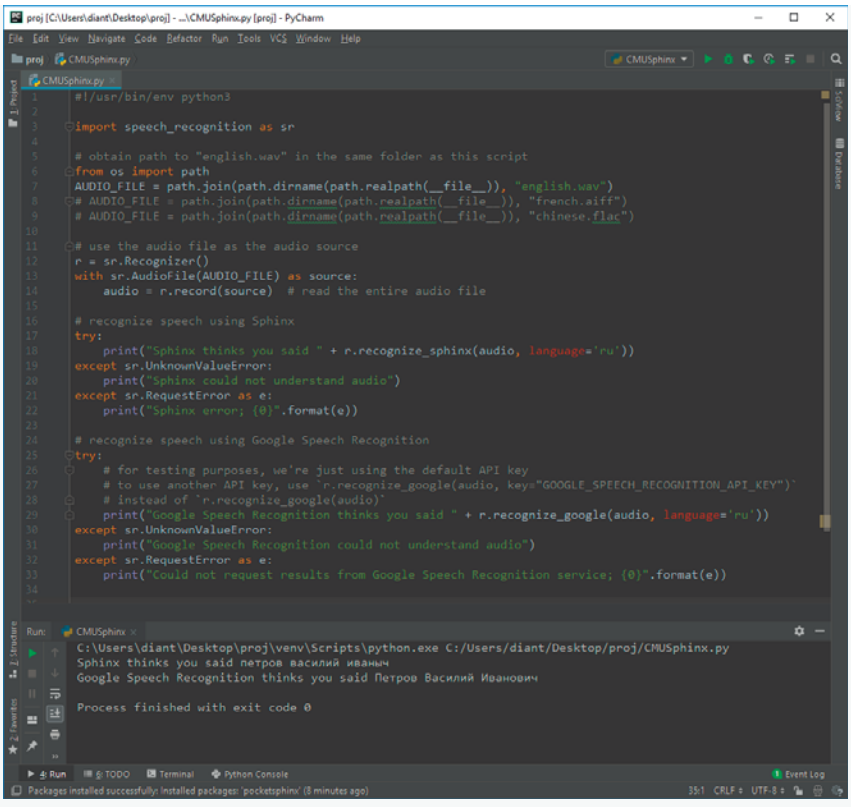

Figure 3. Launch of the script and speech recognition results output

Deploying Yandex.SpeechKit

In order to work with the system, a key must be obtained on the site: https://developer.tech.yandex.ru/. The test key (Figure.4) only works for a month and has a limit on the number of requests.

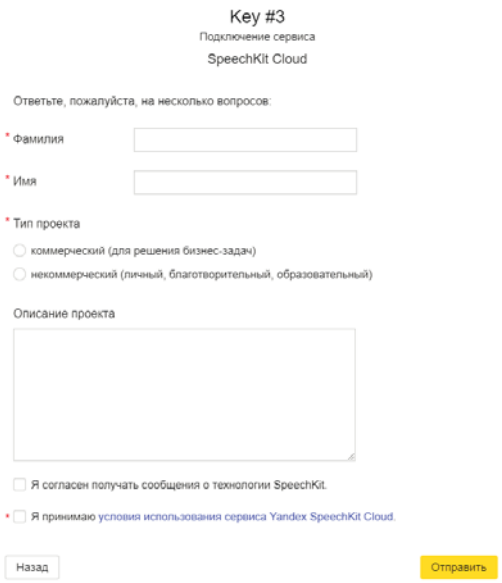

Figure 4. Obtaining a key

Speech Kit Cloud API is a WebAPI, which means that in order to work with the system, a POST request should be sent, and then an XML file with the recognition results will processed (Figures 5-6). Instructions for interacting with Speech Kit Cloud API are presented in [7].

```
import urllib.request as req
from xml.dom import minidom

key = '1e692527-ad23-4fdb-b463-b34e545f9a13'
uuid = 'aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaab'
file = 'english.wav'

defread_file():
with open(file, 'rb') as f:
return f.read()

defget_text_xml(data):
    request = req.Request(url='https://asr.yandex.net/asr_xml?uuid=' + uuid +
'&key=' + key + '&topic=queries',
headers={'Content-Type': 'audio/x-wav', 'Content-Length': len(data)})
    response = req.urlopen(url=request, data=data)
return response.read()

defparse_xml(xml_string):
xmldoc = minidom.parseString(xml_string)
if
xmldoc.getElementsByTagName('recognitionResults')[0].attributes['success'].value
== '1':
return xmldoc.getElementsByTagName('variant')[0].childNodes[0].nodeValue

binary = read_file()
xml = get_text_xml(binary)
best_variant = parse_xml(xml)
print(best_variant)
```

Figure 5. Python script for Yandex.SpeechKit

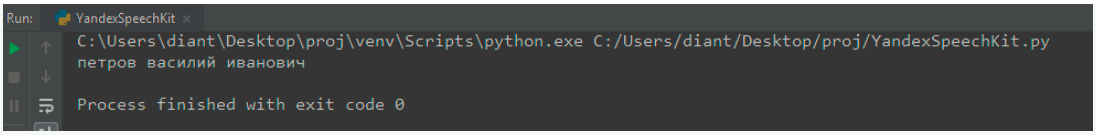

Figure 6. Output result of the script for Yandex.SpeechKit

2. Research and testing of the speech recognition methods and algorithms

This section presents the results of testing speech recognition methods and algorithms using existing systems Sphinx Speech Recognition, Yandex Speech Kit and Google Speech Recognition.

Fragments of the testing results for the Sphinx Speech Recognition system are presented in Tables 1-4.It includes only cases with errors.

Table 1 –Recognition of names

Voiced text	CMU Sphinx result	List of errors (0 – if none is present)	Yandex Speech Kit result	List of errors (0 – if none is present)	Google Speech Recognition result	List of errors (0 – if none is present)
Names						
Евгений	нет не имею	Wrong word	евгений	0	Евгений	0
Борис	боец	Wrong word	борис	0	Борис	0
Коля	о-ля-ля	Wrong word	коля	0	Коля	0
Колян	калям	Wrong word	колян	0	Колян	0
Last names						
Алексеев	алексеев	0	алексей	Wrong ending	Алексеев	0
Иванин	и ваня	Wrong word	иван	Wrong ending	иванин	0
Петренко	ветрянка	Wrong word	петренко	0	Петренко	0
Сидоров	сидорову	Wrong ending	сидоров	0	Сидоров	0
Patronymics						
Андреевич	андреич	Wrong ending	андреевич	0	Андреевич	0
Аркадьевич	акакиевич	Wrong word	аркадьевич	0	Аркадьевич	0
Григорьевич	рекой эридж	Wrong word	григорьевич	0	Григорьевич	0
Иосифович	привозят ходили	Wrong word	иосифович	0	Иосифович	0
Леонидович	да я не дали	Wrong word	леонидович	0	Леонидович	0
Львович	да вловивщо	Wrong word	львович	0	Львович	0
Павлович	палыча	Wrong word	павлович	0	Павлович	0

Table 2 – Recognition of age, addresses, things

Voiced text	CMU Sphinx result	Errors description (0 – if none is present)	YandexSpeech Kit result	Errors description (0 – if none is present)	GoogleSpeechRecognition result	Errors description (0 – if none is
Age						
3 года 9 месяцев	три года в девять месяцев	Extra word "в"	3 года 9 месяцев	0	3 года 9 месяцев	0
4 года 6 месяцев	четыре года на шесть месяцев	Extra word "на"	4 года 6 месяцев	0	4 года 6 месяцев	0
87 лет 2 месяца	восемь если те моменту баррис	Wrong words	87 лет 2 месяца	0	87 лет 2 месяца	0
Addresses						
Дачная, дом 59	да что там добытый содеять	Wrong words	дачная дом 59	0	дачная дом 59	0
Улица Дорожная, дом 42	улицы дорожная дом сорок два	Wrong ending	улица дорожная дом 42	0	улица дорожная дом 42	0
Улица Гагарина, дом 25	молятся когда дом двадцать пять	Wrong words	улица гагарина дом 25	0	улица гагарина дом 25	0
Улица Ленина, дом 112	улица оленя на дом сто двенадцать	Wrong word	улица ленина дом 112	0	улица ленина дом 112	0
Лесная, дом 6	не знаю дом шесть	Wrong word	лесная дом 6	0	лесная дом 6	0
Улица Мира, дом 32	улица методом идется тако	Wrong words	улица мира дом 32	0	улица мира дом 32	0
Улица Набережная, дом 8	улица набережная довольства	Wrong words	улица набережная дом 8	0	улица набережная дом 8	0
Улица Советская, дом 1	улица совет сказал матя	Wrong words	улица советская дом 1	0	улица советская дом 1	0
Свердловская область город Полевской	тьеро носка область и город по мирской	Wrong words	свердловская область город полевской	0	свердловская область город полевской	0
улица Карла Маркса корпус 2 квартира 87	улица карла маркса корпус впал в парке живых систем	Wrong words	улица карла маркса корпус 2 квартира 87	0	улица карла маркса корпус 2 квартира 87	0
Things (goods)						
Дрель	далее	Wrong word	дрель	0	дрель	0
Мультиварка	ну и чего там	Wrong word	мультиварка	0	мультиварка	0
Наушники	но уже фиат	Wrong word	наушники	0	наушники	0
Ноутбук	но у губ	Wrong word	ноутбук	0	ноутбук	0

Плита	да	Wrong word	плита	0	плита	0
Пылесос	вы отсос	Wrong word	пылесос	0	пылесос	0
Тостер	то вздыхать	Wrong word	тостер	0	тостер	0
Весы напольные	везде на вольная	Wrong word	весы напольные	0	весы напольные	0

3. Conclusion

Based on the information presented above, we can conclude that the Yandex system is good at recognizing short expressive phrases, as well as numerals. On contrary, the Google API is good at recognizing long phrases and terms. The Sphinx system is struggling with recognition of Russian speech. The results of this research are being used in development of the automatic system that calls the customers of TWIN.

4. Acknowledgments

This work is supported by Act 211 Government of the Russian Federation, contract № 02.A03.21.0006.

References

1. Cloud Speech-to-Text https://cloud.google.com/speech-to-text/
2. YandexSpeechKithttps://tech.yandex.ru/speechkit/
3. CMU Sphinx -OPEN SOURCE SPEECH RECOGNITION TOOLKIT https://cmusphinx.github.io
4. Features of TWINhttps://twin24.ai/#features
5. Cloud Speech-to-Text API https://cloud.google.com/speech-to-text/docs/reference/rest/v1/speech/recognize
6. Speech Recognition using Google Speech APIhttps://pythonspot.com/speech-recognition-using-google-speech-api/
7. Speech Kit Cloud https://tech.yandex.ru/speechkit/cloud/
8. Speech Kit Mobile SDK https://tech.yandex.ru/speechkit/mobilesdk/