SURVEY OF SPEECH RECOGNITION APPROACHES

Dmytro Dehtiarov
Supervisor - Ph.D., Associate Professor O. Turuta
Kharkiv National University of Radio Electronics
61166, Kharkiv, Nauky ave, 14, Software Engineering Department,
e-mail: dmytro.dehtiarov@nure.ua

The aim of the work is to analyse modern interfaces and frameworks for application in developing the interface of language conversion into text. The main contribution of this work can be considered that at the moment in computer technologies the tasks of recognizing and understanding the context of speech are very relevant, since this can facilitate the way of communication between a person and a computer, these technologies are currently used in medical and military applications, security systems, automated systems for recognition and identification, etc.

What allows people to recognize the language so well? Interestingly, the human brain works under a completely different computational paradigm than a regular computer. The development of artificial neural networks is closely linked to biology. An artificial neuron is a simplified model of a biological neuron.

Recently, there has been a tendency to increase interest in the use of neural networks for solving various problems and applying them in various spheres of human life. With the use of neural networks, the possibilities of computing in the areas that were only concerned with human intelligence were opened. There were opportunities for creating systems that are capable of learning, memorizing and analyzing information that resembles a person's intellectual abilities.

This all suggests that neural networks can indeed become the basis for a general-purpose speech recognition system and that neural networks offer some obvious advantages over traditional methods.

The purpose of this work is to analyze and research modern interfaces and frameworks for application in the development of the interface of language conversion into the text and the next analysis.

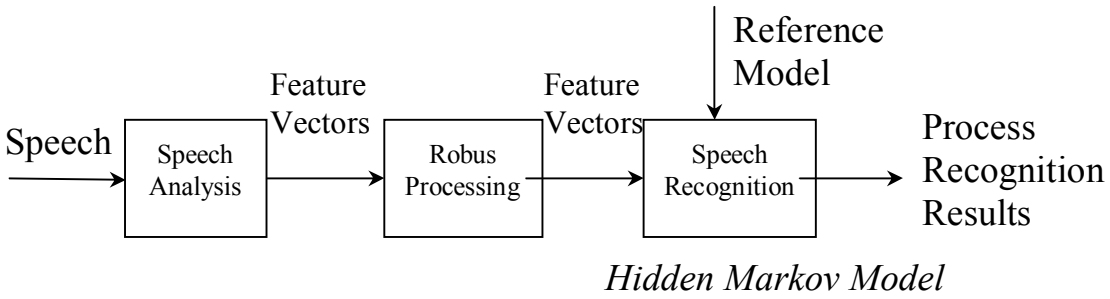
After analyzing existing publications, one can notice that there are many research articles highlighting the topic of recognition and analysis of the text. However, today it still remains an urgent problem, considering that the level of language conversion to text and methods of analysis sometimes do not show an unexpected result, as well as many languages that have not been investigated. That is why the paper reflects theoretical information and provides a practical implementation with the results of experiments.

In recent years, the main trend in research in the field of speech recognition is shifting towards the rejection of the use of latent Markov models(picture 1). According to Markov properties, the next state - in this case, the sound unit of the phoneme type - in the chain depends only on the previous

state and does not depend on all other states in the past. Of course, such a model is very simplified, therefore, for the construction of acoustic models, recurrent neural networks have now been used to maintain long-term dependencies.

The development of modern language technologies is moving toward the implementation of a full cycle of training for spontaneous speech recognition systems without the allocation of separate acoustic and linguistic models. Instead of the prior selection of acoustic signs, all areas of the speech signal are represented by their spectrographs, which are fed to the input of one large neural network. Further, we will dwell in more detail on future systems of recognition and analysis of language in order to outline the relevance of the issue that is being addressed in the work.

There is a huge amount of third-party systems available for speech recognition: Speechmatics, Vocapia Speech to Text API also offline solutions: Speech Engine_IFLYTEK CO, UWP and also open source solutions: CMU Sphinx - Speech Recognition Toolkit and Kaldi. There are also services from such leading companies in the world as Facebook, Google, Microsoft, Yandex and others. They have different functionality and use different solutions to recognition but the main flow is common (picture 1).



Picture 1 - Flow chart of voice recognition system

One of the toughest questions is to realize speech recognition in real time. After many attempts, we were able to teach streaming unidirectional models to process longer sound intervals than those used in "classical" speech recognition models. While calculations themselves do not occur so often. At the same time, the cost of computing resources actually decreased, and the speed of the recognition system has multiplied. Also, we are going to reduce the need for resources so the system can be used on a mobile phone.

We are going to prepare test data to implement and conduct experiments among the leading services of Google Speech Recognition API, Yandex SpeechKit, Microsoft Speech API and find their weakness. In this work, we will define the metrics and methods of evaluation of text analysis systems. And find and consider the causes of errors in text recognition and means of improving the results.