Paul Klymenko
L.M. Korotenko research supervisor
I.I. Zuyenok, language adviser
SHEI “National Mining University”, Dnipropetrovsk

Voice Interfaces

In the last few years our world has changed dramatically, the devices we use to access the Internet became more mobile, portable and light, so all of us have modern smart phones and tablets with large screens and powerful processors.

As a result, Interfaces are also changing, if we have been using mainly the keyboard, mouse, now the main interface - is sensitive, that need attach. The touch to interfaces can be not only physical touch, but gesture commands, voice control etc., which are becoming more natural. Touch interface in different form is familiar to us from birth.

Without any doubt, voice is our primary interface to communicate with each other. To understand the topic, we need to be understood as a general arranged voice recognition. One of the main parts of the speech recognition - an acoustic model. Its main task is to take speech to the input and present it in the form of a sequence of phonemes at the output. A phoneme is the sound of a letter in the alphabet. That is, the output we get a phonetic transcription of what was said. And then you just have to understand, what is the word or phrase that represents the phonetic transcription, for this you need a language model.

It is clear that the main task of a language model - is to get the input of phonetic transcription and the output to give you a particular word with a certain probability. There is a very extensive use of the theory of probability. To ensure a high level of quality recognition, it requires a very well trained system of recognition. First of all we need to be trained acoustic model. There is a need to use a great quality and quantity of voice information. But we need a marked speech. In addition it should be noted the special events that took place in the speech. For example, if there was A background noise, who is saying: man, woman, or child, they spoke with an accent or without. The next task is to train the language model. It is much more easy, because for this you just need to have a lot of text, which will be used by the user. After you made huge amount of work, you can expect a good level of quality.

There is a various array of language interfaces. Thus, Yandex Speech kit is a very effective tool for speech recognition, mainly for the Russian language. Google Now is made on a low level. U.S. version is Siri, is a very powerful platform that has moved to a new level, and can not only recognize speech, but also to answer, imitating artificial intelligence.