SPEECH AND VOICE RECOGNITION SYSTEM BASED ON MACHINE LEARNING METHODS

[1]*Educational institution «The Belarusian State University of Informatics and Radioelectronics», Minsk, Republic of Belarus*

Person's unique biometric identity can be used to distinguish different people and to augment and upgrade the current regular PIN and password systems for gaining access to computers, phones, or restricted access rooms and buildings [1].

The most popular examples of these systems are iPhone's fingerprint and Facebook's facial recognition technologies. Biometric security systems are already in use across many industries incorporating and affecting directly the lives of virtually all people on the planet e.g. personal ID cards or e-Passports throughout the world. In the United States, all e-passports have a chip that contains a digital photograph of one's face, fingerprints, iris, as well as the sophisticated technology that prevents the chip from being read or the data from being skimmed by other unauthorized data-readers. Voice recognition is strengthening other biometric login solutions. The USAA banking app, for example, uses facial recognition and voice recognition to provide easy and secure multi-factor biometric security, the voice component adding an extra level of liveness detection to the process.

Each of the main AI assistants – Apple's Siri, Microsoft's Cortana, and Android's OK Google – can perform searches and basic tasks based on voice command. Siri was upgraded in 2015 to be able to recognize who is speaking

Speech and voice recognition are contactless, software based technologies, and as such are counted among the most convenient biometrics in regular use. Voice recognition commonly referred to a voiceprint, is the identification and authentication arm of the vocal modalities. Speech recognition, on the other hand, is a user interface technology. In today's increasingly mobile and connected world, having hands free interface options is critical. Speech recognition technology, also called voice command, allows users to interact with and control technologies by speaking to them.

Currently, machine learning methods are widely used in the field of speech technologies. Supervised learning systems generally make use of Artificial Neural Networks (NN), sets of artificial neuron-like points connected together with each other to form networks. The Artificial Neural Network processing points are trained by presenting them with some inputs (for instance, voice samples) each of which is already labeled by human trainers with an output result. This set of inputs and matched outputs is called a training dataset.

After training a NN with the prepared training dataset should be able to extract the features and thus identify which template is presented, or be able to reject the given sample if it wasn't present in the original training data set. Neural Networks have the ability to help derive valuable meaning from complex and imprecise data through varies different method and techniques, and by detecting and extracting trends and patterns which are way too detailed and complex for other computational techniques or directly for humans to extract from. Generally NN can be subcategorized into 6 different broad categories based on the functions which they can perform on the given data sets, amongst them we are now interested in the Recurrent Neural Network (RNN) type which is especially useful with voice samples [2].

The Recurrent Neural Network allows for a bi-directional flow of data which is especially useful for purpose of speech recognition. In our work, we will use TensorFlow – a framework for machine learning.

TensorFlow allows building neural network models to recognize spoken words. Module development includes 4 stages: create (or use a ready-made one) dataset; build NN; train the NN; test NN.

REFERENCES

1. What is Biometrics Security: How It Works [Electronic resource] / link https://usa.kaspersky.com/resource-center/definitions/biometrics / Date of access: 11.09.2020

2. Al Smadi, T. Artificial Intelligence for Speech Recognition Based on Neural Networks // T., Al Smadi, H.A., Al Issa, E., Trad, K.A., Al Smadi // Journal of Signal and Information Processing, 6, P. 66-72, 2015