

An Ear For Robots: New Approach

A fundamentally new approach to computer identification of words was suggested by Russian scientists. With its help, people will be able to give orders even to the most primitive cellular phones.

A sentient being recognizes without difficulty a familiar word regardless of the voice and intonation it is pronounced with. "Six" or "eight" remain six and eight for a person no matter how they are pronounced – in a loud voice or in a whisper, in an excited or a calm voice, by the voice of an old man or a child, by that of a man or a woman. The brain of a person immediately separates the semantic part from the mass of background sounds.

As for a machine, each variant of voice is unique. That is why the speech recognition program usually has to be taught. As a result of training, an enormous library appears in the memory of the silicon brain, where thousands of possible options of pronunciation of the same words (for example, numerals) are stored. Having heard a word, the computer would look through the library and almost certainly something similar to the heard word will be found in it.

The approach suggested by the scientists from the Institute of Radio Engineering and Electronics, Russian Academy of Sciences, is rather human, than machine one: a computer under the researchers' guidance filters individual peculiarities, i.e. picks out the most basic things and rejects all immaterial ones. As a result, the machine even acquires the ability to discern individual sounds and to put together in its mind familiar words from these sounds.

As a result, memory of only 1 KB would be sufficient for a processor to confidently recognize all numerals and some simple commands, however, pronounced only in Russian yet. Several dozens of human beings – men and women, with irreproachable and far-from-ideal articulation – tried to confuse a quick-witted program, pronouncing numerals either in a whisper or in a voice trembling with excitement. However, the computer successfully rejected emotional frequencies as immaterial.

“The prototype software interface developed and established by our specialists for the system of data and management commands voice input is intended for mass mobile electronic devices, says the project manager, Vyacheslav Anciperov. Perhaps, the most important and fundamentally new about our work is that we have managed to single out essential elements of speech being guided by the notion of hierarchical structure of speech. Like in a musical composition, one can recognize more or less high levels of organization - rhythm, main theme, arrangement, so we have also learned to single out the ranges in the speech flow (i.e. in the wide frequency spectrum), which carry the major semantic loading. It has turned out that this is a very small part of human speech sounds – only up to 1 KHz. All the rest relates to psychophysis. Thus we simplified the task for the computer to the maximum. And one more thing – we have taught the computer to recognize individual sounds, which is sometimes far from easy. As a result, our system wins in processing speed and in processor time and memory consumption as compared to those of all known similar systems. This is the path to efficient speech processors that nobody has passed yet.”

Source: Informnauka (Informscience) Agency

This document is subject to copyright. Apart from any fair dealing for the purpose of private study, research, no part may be reproduced without the written permission. The content is provided for information purposes only.