MODERN APPROACHES TO SYNTHESIS AND SPEECH RECOGNITION

Konovalenko O.E.[1]**, Brusentsev V.A.**[2]
[1]*National Technical University «Kharkiv Polytechnic Institute»,*
[2]*Kharkiv State Academy of Culture,*
Kharkiv

Attempts to teach computers to communicate with people using a natural voice interface have been made since the first years of the history of computer technology. In the course of many years of research, it was found that to solve the problem it is necessary to involve not only programmers, but also specialists in linguistics, radio engineers, mathematicians, biologists, and even psychologists.

Successes in the development of natural language technologies promise wide access to on-line information and electronic services. Since almost everyone speaks and understands speech, the development of natural language systems will allow a person without special skills to communicate with a computer at any time and anywhere without additional training, using devices such as a mobile phone, to access information or manage devices.

Unfortunately, despite the tremendous development of computer technology, today the problem of equipping a computer with a full-fledged, natural human voice interface is still far from complete.

Sounds of continuous speech contain a constantly changing spectrum of harmonic frequencies, as well as noise. The volume and pace of speech are also constantly changing. Moreover, the same phrase spoken by different people, or even by one person who is in different mental states, can have a different spectral-temporal color. This makes it very difficult to create universal recognition systems that "understand" the speech of different people.

As a rule, two fundamentally different approaches are used in existing systems: voice tag recognition and lexical element recognition. The first approach involves the recognition of fragments of speech from a pre-recorded pattern. When implementing the second approach, separate lexical elements – phonemes and allophones, which are then combined into syllables and morphemes.

In order to distinguish linguistic constructions from the digitized sound, various mathematical methods are used in combination with special computer equipment, such as, for example, hardware or software neural networks. Throughout the history of speech recognition systems, these methods have been constantly changing. At the same time, some methods died out due to their inefficiency, while others were developed and improved.

Despite a wide range of developments, the problem of speech synthesis is still considered to be solved only satisfactorily. There is still no common opinion which of the existing approaches gives the best results, which models of speech synthesis are the most promising. The study of speech synthesis systems has confirmed that deep neural networks and hidden Markov models using differential-style mathematical modeling give promising results.