



ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ ВЫЧИСЛЕНИЙ

D.M. Umurzakova, S.K. Axmedov

PECULIARITIES OF SPEECH INTERACTION WITH A COMPUTER

(Ferghana branch of the Tashkent university
of Information technologies named after Muhammad al-Khwarizmi)

Advances in speech recognition technologies in recent years have led to an increase in the interest of control system developers in speech interfaces. The voice interface is an alternative channel for providing dialogue between the operator and the controlled system (for example, the driver and the car, the pilot and the airplane).

Currently, there are numerous technical means that can perceive (pronounced) spoken speech messages: computers, medical electronic equipment, cars, mobile phones, etc.

Voice interfaces are well suited for information management and access, in the case of a complex and open information space, users of which do not have serious training in the field of interaction with computers. Often, interaction with the information environment is possible only by phone, in this case, one can not do without voice control. The use of voice control does not require special user training and allows you to free your hands and eyes while interacting with your computer.

To implement voice interfaces, automatic speech recognition systems are used, such systems translate speech delivered via input channels into text. Depending on the purpose of the interaction, the recognized text may be the desired result, but may serve as input to the systems for processing and understanding texts in the natural language.

There are two main directions for using voice interfaces:

- Tasks that mainly require speech recognition:
 - ✓ Simple commands and management;
 - ✓ Simple data entry (by phone);
 - ✓ Dictation;
- Tasks that, in addition to speech recognition, require understanding of the text (interactive conversation):
 - ✓ Information kiosks;
 - ✓ Dialogue processing of requests;
 - ✓ Intelligent agents.

The main difficulties encountered in automatic speech recognition:

- The effect of co-articulation; (in natural speech, sounds do not have clear boundaries, in this connection it is very difficult to determine where one



phoneme ends and another begins, the phonemes of speech go smoothly into one another and the sound environment greatly distorts the form of the phoneme)

- The need to customize the system of automatic speech recognition for each speaker separately, thus there may be problems with the diversity of dialects, as well as the difficulty of recognizing the speech of persons speaking a foreign language;
- Free speech: the presence in the speech of words-parasites and words not included in the dictionary;
- The need to create a model of natural language
- Resistant to noise.

The capabilities of automatic speech recognition systems are characterized by the following parameters:

- Speech mode: can vary from separate pronunciation of words to continuous speech;
- Speech style: varies from reading text to spontaneous speech;
- Adjustment: Dependency on the speaker - the user must provide samples of his speech prior to working with the system, on the other hand, independence from the speaker does not presuppose any settings before using the system;
- Vocabulary: a set of words can range from a small volume (<20 words) to a huge (> 50,000 words), the larger the dictionary or the more similar words in it, the more difficult the speech recognition process is;
- Language model: used when the speech is presented as a sequence of words, the simplest language model can be defined as a network with a finite number of states, more complex, but more similar to the natural language model, is described in terms of context-dependent grammars;
- perplexity (degree of uncertainty of the probability model): a popular measure of the complexity of the problem, combining the size of the dictionary and the language model;
- Finally, there are a number of external parameters that can affect the performance of the automatic speech recognition system, including environmental noise characteristics, as well as the placement and characteristics of the microphone.

The main obstacles to the further development of automated speech recognition systems are:

- 1) the need for large volumes of dictionaries;
- 2) noisiness of the speech signal;
- 3) various accents and pronunciation.

The volumes of dictionaries determine the degree of complexity, the requirements for computing power and the reliability of speech recognition systems. It is necessary to continue thorough research. This will solve the problems associated with morphology, accents, pitch, tempo, loudness, confluent words, articulation, linguistic



information, etc. It is expected that the main direction of development will be the modeling of languages for use in speech recognition systems. The problem of separating the speech signal from the noise background has not been finally solved. Currently, users of voice recognition systems are forced to work in conditions of minimal noise background. One of the priority developments in the field of speech recognition is human-machine interactive systems, which are being worked on in many research laboratories of the world. One of such developments is the technical system of AT & T (USA), which is used to recognize speech in the telephone network: the client can request one of five categories of services using any words; he says until one of the five key words is found in his statement. This system currently serves about a billion calls a year.

Such systems "know how" to work with a continuous speech flow and with unknown speakers, understand the meaning of speech fragments of a limited vocabulary and take response actions. The systems operate in real time and are capable of performing five functions:

1. Recognition of speech - the transformation of speech into a text consisting of individual words;
2. Understanding - grammatical analysis of sentences and recognition of meaning;
3. Restoration of information - obtaining data from operational sources on the basis of the obtained semantic meaning;
4. Generation of linguistic information - construction of proposals representing the received data in a user-selected language; the synthesis of speech - the transformation of sentences into computer-synthesized speech.

Dialogue interface in such systems allows a person to talk with the machine, create and receive information, solve their problems. Systems with a dialog interface differ in the level of initiative of a person or computer. Research focused on "mixed initiative" systems in which both humans and computers play an equally active role in achieving the goal through dialogue.

Limitations of the use of speech recognition systems in the framework of the most traditional applications allow us to conclude that it is necessary to search for potentially new solutions in the field of speech recognition. In the next decade, the task of recognizing and understanding natural speech, regardless of language and speaker, will occupy a central place in speech technologies.

References

1. Christopher J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", Kluwer Academic Publishers, Boston, 1998.
2. S. Das, R. Bakis, A. Nadas, D. Hahamoo and M. Picheny, "Influence of background noise and microphone on the performance of the IBM tangora speech recognition system", Proc. of ICASSP, Vol. II, pp. 71-74 (1993).