SOFTWARE TOOL FOR SPEAKER RECOGNITION

U.O. Kunitsky, N.S. Lun, V.S. Zaikovsky

The rapid development and widespread dissemination of information systems and technical means necessitates computer processing of speech information because the voice interaction interface seems to be the most convenient. Thus it is advisable to use voice technologies for biometrics namely the verification of the speaker by voice. Verification by voice is a procedure for confirming identity using individual speech characteristics. There are text-dependent and text-independent verification systems. Text-dependent require the pronunciation of a certain phrase and compare it with the standard for each user. Verification in text-independent systems is carried out on the basis of any speech fragment of a given length. The primary task of speaker verification is speech signal recognition. To solve it a software module is proposed that analyzes the acoustic environment and calculates the following signal parameters: root mean square value, average number of signal zero crossings, pitch period. Then the calculated parameter values are compared to user-defined threshold values. If the values of all parameters simultaneously meet the established requirements, then the analyzed fragment of the signal is considered to be a speech. This module is implemented in the C++ programming language in the Visual Studio environment. It is possible to manually set the threshold values of the parameters in order to adapt the algorithm settings and study the features of speech recognition. The user verification is implemented in a separate module which receives signal fragments classified

as speech as input. The verification module is also implemented in the C++ programming language in the Visual Studio environment. The operation of the module is designed in such a way that a text login is required first in order to identify the user and then the pronunciation of password is required to verify the user. This avoids the need for an empirical search for a balance between the possibility of errors of the first and second types while lowering the coefficient of cognitive resistance of the end user [1]. A vector of Mel-frequency cepstral coefficients [2] is calculated from the frequency characteristic of the signal using a discrete cosine transform and is compared to the database of reference user records. As a reference record for each user the average value of vectors of Mel-frequency cepstral coefficients calculated for three pronunciations of the passphrase is used. The comparison is implemented using a self-organizing Kohonen neural network.

Literature

1. Kunitsky U.O., Zelmansky O.B. Verification of the Speaker by voice Based on the Method of Dynamic Time Distortion // Abstracts of the XIX Belarusian and Russian scientific and Technical Conference “Technical means of information protection”, Minsk, June 8, 2021. P. 59.

2. Zapryagaev S.A., Konovalov A.U. Recognition of Speech Signals // Bulletin of VSU. 2009. No. 2. P. 39–48.