UDC 004.522

L.N. Bozhukha, M.V. Babenko

THE RESEARCH OF THE VOICE WEB INTERFACE USAGE

Abstract. The article presents the results of the analysis that based on current methods of voice recognition messages. The study represents the requirements for developing the software for voice recognition and developing the algorithm for going the next web page using the voice commands.

The implementing of the standard algorithms can cause the accretion of the development of the software and users' possibilities.

Keywords: voice interface, discrete Fourier transform, neural networks, frame, entropy, perceptron, the algorithm for reverse error propagation.

Formulation of the problem

The increasing of the society needs (the professional fields and the daily life) accelerates the development of the solutions that are more ergonomic and user-friendly. The tools for the interaction of a user with the computer are interfaces in the information technologies. The maintaining of the intuitive and user-friendly interface is the paramount goal of the technology systems.

The language is the natural form of communication. The voice command interface implements the convenient access to the interaction between the user and the computer. The qualitative voice interface probably will avoid the rejection of society and allows people who have problems with hearing and seeing the ability to use new technologies. The critical point of the development of this field is the voice commands that can help most of the people who have problems with the skeletomuscular system.

The voice interface influences the interaction of a user with the computer. The great examples are Google Assistant and Siri from Apple Company. These interfaces show the importance of implementing the voice commands on every platform.

The voice interface is useful while typing the text. There is a question of assisting with searching through the web because of numerous web pages on the Internet.

Formulation of the task

An up-to-date issue is choosing the algorithm of voice interface on the web page.

The paper aims at the analysis of the current solutions of the voice recognition and developing the algorithm of the software for voice commands on the web pages.

The following parts divide the process of voice recognition: splitting a record into frames, selecting frequencies, comparing with a standard database.

The authors use the mathematical transformations of sound waves on the first stage.

They analyze the current libraries of voice recognition for developing the algorithm of voice commands.

Research

There are following forms of interaction of a user with the computer: the command interface, WIMP interface, SILK interface.

The methods depend on the vicinity to the ideal. There are the algorithms that are used for creating and developing the voice interface on the web pages.

In the implementation of the time, the dynamic algorithm (Dynamic Time Warping) from the flow of things isolated separate lexical elements - phonemes and allophones that will be combined into compositions and morphemes. Definition of a word can be made by comparing binary forms of signals or by comparing spectrograms of signals. In the first case, a direct comparison of numeric waveforms is used when for each numerical sequence a new sequence is created, the size of which is much smaller. In the second case, the comparison of two spectrograms consists of dividing the digital signal into a certain number of intervals. The comparison process in both cases should compensate for the different lengths of the sequences and the nonlinear nature of the sound. The DTW algorithm manages to disassemble these problems by finding a deformation that corresponds to the optimal distance between two lines of different lengths.

It is necessary to construct a codebook that contains a set of reference kits for the characteristic features of the language to perform recognition from hidden models of Markov. At the stage of configuring Markov models, it is necessary to apply the Baum-Welch algorithm to

 ISSN 1562-9945

the existing vocabulary and to match each of its words to the matrix of the probabilities of transitions from one minimum language segment to another minimum section of language and the likelihood of dropping in each state of a particular codebook number.

The use of hidden Markov models for speech recognition is based on two assumptions. First, the voice stream can be divided into fragments corresponding to states in the latent Markov model (speech parameters within each piece are considered constant). The second - the probability of each fragment depends only on the current state of the system and does not depend on the previous states.

When creating a model of the voice interface of a web service, it is expedient to accelerate the development of web application software with the ability to use automatic language recognition systems or libraries (Fig. 1).

After analyzing the software requirements for the ability to run a web-based application with a voice interface, you can stay on the classic client-server architecture.

For the development of the software product is selected interpreted object-oriented programming language high-level Python, the benefits of which is the presence of a module for the development of the graphical interface and the ability to work in dialog mode.

Specifications	Горинич ПРОФ 3.0	Voice Navigator	Speereo Speech Recognition	Sakrament ASR Engine	Google Voice Search	Dragon NaturallySpeaking	ViaVoice	CMU Sphinx
Speaker independence	-	+	+	+	+	+	+	+
Noise resistance	-	+	+	+	+	+	+	+
Language independence	Rus, Engl	Rus	+	+	+	+	+	+
Vocabulary	-	100 words	150 thousands words	phrases, sentence	phrases, voice commands	phrases, voice commands	specific terms	continuous speech
Recognition accuracy dependence	+	-	-	-	internet connection	-	-	-
Providing for further use	partially	-	+	+	+	+	-	-
Location of recognition	server	server	device	device	device, browser	device, browser	device	device
The presence of a neural network	-	-	-	+	+	+	-	+
Create documents	-	-	-	-	-	+	-	-
Sending an e-mail	-	-	-	-	-	+	-	-
Manage browsers, applications	-	-	-	-	-	+	-	-

Figure 1 – The properties of the voice recognition systems

When designing the web application software, you can select the necessary classes for implementing the voice interface: the SpeechRecognition interface, which provides recognition of the voice context from the audio input; audio data class for audio processing; The SpeechRNN class is created to interact with the neural network PyTorch (library for machine learning).

Methods of working with voice are SpeechFreum method, which implements data splitting into frames; a SpeechEntropy method for calculating entropy; method Generate_mfcc calculation of frequency coefficients; SpeechMel method of frequency conversion; The CALLBACK_activityLevel method checks the noise level and save the file with the extension*.wav; The CALLBACK_stopped method sends the file to the server for further processing of the file.

Input data is split into small time intervals - frames [1]. The construction of frames is based on their intersection: the end of one frame is the beginning of another. The frames are a more suitable unit of data analysis than the specific signal values because it is much more convenient to carry out wave analysis at a specified interval than at particular points.

The presence of some pauses in the language (the intervals of silence) leads to the definition of some threshold value. You can consider several options for defining a threshold: a constant (the output signal is always generated under the same conditions, in the same way); clustering the signal value (silence takes up a significant part of the output signal) and entropy analysis.

When designing a web application for implementation of the voice interface, the entropy value was selected as the threshold, which shows the amplitude of the signal fluctuation within a given frame. Calculate the entropy of a particular frame by the algorithm: normalize the value of the signal to the interval [-1; 1]; construction of the histogram of the signal of the frame; calculation of entropy by the formula $E = \sum_{i=0}^{N-1} P[i] * log_2 P[i]$.

To represent the energy of the spectrum of the signal, we use the Mel-frequency cepstral coefficients (Mel-frequency cepstral coefficients). The advantages of choosing these characteristics are

- the use of the frequency of the signal (decomposition from orthogonal sinusoidal functions)

 ISSN 1562-9945

- the design of the range of the signal on a special mel-scale (the selection of the most important for the human perception of frequency)

- the ability to limit the number of calculated coefficients by any value (compression of the frame).

The input data for a discrete Fourier transform is the sequence of amplitudes at the following moments of time with the same reasonably short intervals [2]. The discrete Fourier transform consists in finding such points that.

$$x[i] = \sum_{k=0}^{N/2} Re\, X[i] \cos(2\pi k i/N) + \sum_{k=0}^{N/2} Im\, X[i] \sin(2\pi k i/N),$$

where for calculation of coefficients the formula of direct Fourier transform is used:

$$X[k] = \sum_{m=0}^{N-1} x[m] \cos(2\pi km/N) + i \sum_{m=0}^{N-1} x[m] \sin(2\pi km/N) = \sum_{m=0}^{N-1} x[m]\, e^{-i2\pi km/N}$$

A neural network can build the next stage of implementation of the voice interface [3]. The use of the Backpropagation method results in the calculation of the gradient used to update the weight of a multilayer perceptron. The main steps are:

1. the representation of a set from the training sample to the input of the neural network;

2. the calculation of the output signal of the network;

3. determination of the values of the errors of the neurons of the output layer;

4. determination of the benefits of failures of the neurons of the hidden layers;

5. calculation of the network error.

The primary purpose of the work was to demonstrate the ability to control voice on web-resources.

When developing a web-based application with a voice interface, embedded standardized libraries of the selected software were used. In the future, you can implement several methods of speech recognition and conduct a comparative analysis of the work with the voice interface for arbitrary characteristics. For example, make a comparative study of the work of the web application using methods that depend on the availability of an Internet connection.

The result of the study of using the voice interface is a software product that allows voice control of web pages. Software product testing

was carried out in conditions of different noise levels and use of different browsers.

Conclusions

The problem of using the model of the voice interface of the web application is studied. The strategy of constructing a model of the voice interface of a web application using methods of speech recognition is presented.

An analysis of existing language recognition software, analysis of software development requirements regarding the use of existing approaches for voice stream processing, and analysis of the underlying algorithms for the implementation of the voice control web-pages. The main stages of voice recognition and selected mathematical methods of sound stream processing are highlighted.

Designing a web-based voice-interface application has been conducted using existing standardized language recognition libraries and can be used on the web pages.

REFERENCES

1. Дегтярев, Н.П. Параметрическое и информационное описание речевых сигналов / Н.П. Дегтярев // Минск: Объединенный институт проблем информатики НАН Беларуси, 2003, 216 с.
2. Чучупал В. Я., Чичагов А.С., Маковкин К.А. Цифровая фильтрация зашумлённых речевых сигналов. М.: ВЦ РАН, 1998.
3. Осовский С. Нейронные сети для обработки информации /Пер. с польского И.Д. Рудинского М.: ФиС, 2002. – 343 с.