UDK 001

Bazarova SHakhzoda Husanboyevna
Master student
Namangan State University

PRE-PROCESSING OF AUDIO SIGNALS

Annotation: This paper is devoted to research in the field of speech technology. The paper presents a description of the software for pre-processing of speech signals using the discrete Fourier transform. This software shell aims to create a tool for studying various algorithms and methods for processing data contained in speech signals. In this paper, based on the recognition system, we investigate the conditions for the applicability of the discrete Fourier transform as a tool for identifying the acoustic characteristics of speech signals.

Keywords: discrete Fourier Transform, the spectrum of the signal, white noise, impulse.

The creation of natural human means of communication with a computer is currently the most important task of modern science, while speech input of information is carried out in the most convenient way for the user. Speech recognition is the task of classifying images of the acoustic characteristics of speech signals. In speech recognition systems based on a neural network, there are two main subsystems:

- a subsystem for pre-processing speech signals, designed to highlight the informative acoustic characteristics of the speech signal and form an acoustic image of the signal as a set of characteristics;

- subsystem of classification of acoustic images using neural networks.

In this paper, we describe the add-ons of the software shell for analyzing speech signals using the fast Fourier transform. This software shell is intended to form a tool for studying various methods and algorithms for analyzing data contained in speech signals.

Subsystem pre-processing of speech signals

The preprocessing of the speech signal includes the following steps:

- the process of entering a speech signal;
- selection of the speech signal boundary;
- digital filtering;
- cutting the speech signal with overlapping frames;
- signal processing in the window;
- spectral transformation;
- normalization of the frequency spectrum.

Let's look at the stages in detail.

The process of entering a speech signal

Audio input is performed in real time via a sound card or via WAV files in PCM encoding. The sampling rate of 8 kHz and the quantization of 16 bits are typical parameters in the systems of transmission, storage and processing of speech information. Working with files was provided to facilitate the repeated repetition of neural network processing, which is especially important in training.

The allocation of the boundaries of the speech signal.

To isolate sections containing only speech from the input signal, the following characteristics of the speech signal are used:

- short-term energy of the speech signal
- the number of intensity zeros (instantaneous frequency);
- the distribution density of the value of the pause reports.

The short-term energy of the audio signal and the number of intensity zeros are simultaneously used to isolate speech from the input signal. In addition, you can remove the pause from the output signal using a method based on the normal (Gaussian) distribution.

Digital filtering.

Together with the useful signal, usually fall into a variety of noises. Noise has a negative impact on the quality of speech recognition systems, so you have to deal with it. To reduce the noise level in the subsystem, two types of digital filter are used:

- pass-through bandpass filter;
- pre-filter.

A pass-through bandpass filter can be thought of as a combination of a low-pass filter and a high-pass filter. Such a filter delays all frequencies below the so-called lower pass frequency, as well as above the upper pass frequency.

Pre-filtering is provided to reduce the impact of local distortions on the characteristic features that will be used for recognition in the future. For spectral alignment of the speech signal, it should be passed through a weighting low-pass filter.

Slicing a speech signal with overlapping frames

In order to get feature vectors of the same length, you need to cut the speech signal into equal parts, and then perform transformations within each frame. Overlap is used to prevent loss of information about the signal at the boundary.

The smaller the overlap, the smaller the dimension of the property vector that is characteristic of the area under consideration. Overlap is sometimes skipped because it saves computing resources, as it significantly slows down the speed of data processing. Usually, the length of the segments corresponding to the time interval of 20-30ms is selected.

Signal processing in the window

Signal processing in the window is presented to reduce the boundary effects resulting from segmentation. To suppress unwanted boundary effects, it is customary to multiply the signal by the window function. There are 4 types of window functions:

- rectangular window;
- window;
- hamming window;
- blackman's window.

The Hamming window is used as a function.

Spectral transformation

Information about the amplitude and shape of the envelope of the speech signal is not enough to distinguish lexical elements from speech. Depending on various circumstances, the shape of the envelope of the speech signal can vary widely. To solve the recognition problem, it is necessary to identify the primary features of speech that will be used in the subsequent stages of the recognition process. The primary features are identified by analyzing the spectral characteristics of the speech signal. Fast Fourier transform (FFT) is used to obtain the frequency spectrum of a speech signal. The FFT is presented to obtain the amplitude spectrum and information about the phase of the signal (in real and imaginary coefficients). The phase information of the signal is discarded and the amplitude spectra are calculated. In this case, the logarithm of this value is more often used.

Normalization of the frequency spectrum

The program shell is implemented in the C# programming language. The input is an audio file in WAV format. The screen displays the signals, the corresponding processing steps, and the conversion parameters. The user can change the parameters to get the results of different algorithms and data processing methods. The output of the program shell gets an array of frames. Each frame corresponds to a set of numbers of equal size that characterize the amplitude spectra of the speech signal.

Conclusion

As a result of this work, a software shell for preprocessing speech signals for a speech recognition system using a discrete Fourier transform is proposed. It is planned to develop an automatic speech recognition system based on a neural network with the output of pre-processing of speech signals.

References:

1. Компьютерное распознавание и порождение речи. [Электронный ресурс]. – Режим доступа: http://speech-text.narod.ru/chap3.html

2. Корицкий, Д.В. Система распознавания речевых команд. [Электронный ресурс]. – Режим доступа: http://www.nsc.ru/ws/show_abstract.dhtml?ru+130+9365

3. Оконное преобразование Фурье [Электронный ресурс]. – Режим доступа: http://ru.wikipedia.org/wiki/Оконное_преобразование_Фурье

4. Фролов, А.В. Синтез и распознавание речи. Современные решения. / А.В. Фролов, Г.В. Фролов. – 186 с.