Design of Speech Signal Analysis and Processing System

Based on Matlab Gateway

Weidong Li,Zhongwei Qin,Tongyu Xiao

Electronic Information Institute, University of Science and Technology, Shaanxi, China

Abstract: Speech signal processing is an emerging discipline to study the processing of speech signals with digital signal processing technology and phonological knowledge. It is one of the core technologies in the field of information science research. Passing information through voice is the most important, most effective, most commonly used and convenient exchange form of information. The Matlab language is a very powerful computer application software for data analysis and processing. It can transform sound files into discrete data files and then use their powerful matrix computing power to process data such as digital filtering, Fourier transform, domain and frequency domain analysis, sound playback and a variety of graphics. Its signal processing and analysis toolbox for the voice signal analysis provides a very rich function, the use of these functions can quickly and easily complete the voice signal processing and analysis, as well as signal visualization, making human-computer interaction more convenient. Signal processing is one of the important applications of Matlab. This design is aimed at most of the voice processing software content, operation inconvenience and other issues. The use of MATLAB 7.0 integrated GUI interface design, a variety of function calls to achieve the voice signal frequency, amplitude, fourier transform and filtering, concise interface, easy to operate. All these have certain practical significance. Finally, this paper puts forward his own views on the further development of speech signal processing.

Key words: Matlab; speech signal; Fourier transform; signal processing;

1 Introduction

Voice is the language of the acoustic performance, is the most natural human exchange of information, most effective and most convenient means. With the development of social culture and the development of science and technology, mankind has entered the era of information technology. People can generate, transmit, store, and access to voice information more effectively by using modern means to study voice processing technology. This has a great significance on accelerating the social development. Therefore, voice signal processing is getting more people's attention and extensive research.

1.1 Background and significance of the subject

Speech signal processing is a more practical professional course of electronic engineering. Voice is an important source of human access to information and has an important means of using information. Passing messages to each other through language is one of the most important basic functions of mankind. Language is a human-specific function, it is to create and record thousands of years of human civilization. It is the fundamental means of history, there is no today's human civilization without language. Speech is the acoustical representation of language and most important means of transmitting information to each other. It is the most important, most effective, most commonly used and most convenient exchange of information form.

Speech signal processing is a discipline that studies the processing of speech signals with digital signal processing technology. It is a new discipline, comprehensive multidisciplinary field and a wide range of interdisciplinary disciplines.

1.2 Research status at home and abroad

The theory and algorithm of digital signal processing such as digital filter and fast Fourier transform (FFT) are the theoretical and technical basis of digital processing of speech signal in the mid-1960s. With the rapid development of information science and technology, the speech signal processing has made great progress. After entering the 1970s, proposed a linear prediction technology (LPC) for information compression and feature extraction of speech signals. It has become the most powerful tools that widely used in voice signal analysis, synthesis and application areas, as well as for the input voice and reference samples of the time between the dynamic programming methods. In the early 80s, a new clustering analysis based on efficient data compression technology - vector quantization (VQ) is applied to speech signal processing. The generation of speech signal process with HMM is a significant development of speech signal processing technology in the 1980s. At present, HMM has formed modern speech recognition and becoming an important cornerstone of the study. In recent years, artificial neural network (ANN) research has made rapid development; multiple subjects of voice signal processing are important driving force to promote its development. At the same time, many of its results are also reflected in the voice signal processing technology.

1.3 Research contents and methods of this subject

1.3.1 Research contents

This paper mainly introduces the simple processing of speech signals. In this paper, the basic theory of digital signal is used to realize the processing of speech signal. The signal processing is carried out by using signal extraction, amplitude-frequency conversion, Fourier transform and filtering in Matlab7.0 environment. What I have done is prepare a process to deal with voice signals, voice signals can be collected, and its various processing in the matlab7.0 software as to achieve the purpose of simple voice signal processing.

1.3.2 Operating environment

The operating environment mainly introduces the hardware environment and software environment.

Hardware environment:

① Processor: Inter Pentium 166 MX or higher

② memory: 512MB or higher

③ hard disk space: 40GB or higher

④ graphics card: SVGA display adapter

Software Environment:

Operating system: Window 98 / ME / 2000 / XP

1.3.3 Development environment

The development environment mainly introduces the operating system and development language adopted by the system.

(1) Operating system: Windows XP

(2) Development environment: Matlab 7.0

2 Voice signal processing of the overall program

2.1 Basic overview of the system

Graphical user interface (GUI) also known as graphical user interface refers to the use of graphical display of computer operating user interface. Compared to the command line interface used by earlier computers, the graphical interface is visually more acceptable to the user. Wide usage of GUI in today's computer development is one of its major achievements. It greatly facilitates the use of non-professional users and people no longer need to rote a lot of orders since it has been replaced by the window, menu, buttons, etc. to facilitate to operate.

2.2 System basic requirements

In this paper, we use Matlab to filter and analyze the noisy speech signal in the time domain and frequency domain simultaneously. We design an easy-to-use graphical user interface (GUI) under MATLAB application software to solve the problem of general application processing of speech signals.

2.3 system framework and implementation

1) Voice signal acquisition

Use the computer's sound card device to collect a voice signal and save it in the computer.

2) Processing of voice signals

The processing of voice signals mainly includes signal extraction, signal adjustment, signal transformation and filtering.

I.Time Domain Analysis of Speech Signals

The voice signal is a non-stationary time-varying signal, which carries a variety of information. In speech processing, speech synthesis, speech recognition, voice enhancement and other voice processing need to extract the voice which contains a variety of information without any exception. The purpose of voice signal analysis is to extract the information carried by the voice signal with the convenience and effective extraction. The speech signal analysis can be divided into time domain and transform domain. The time domain analysis is the simplest method. The time domain waveform of the speech signal is analyzed directly. The characteristic parameters of the speech are mainly short-term energy, average zero-crossing rate, short-time autocorrelation function and so on.

① Extraction: through the graphical user interface on the menu function keys to collect a computer equipment on the audio signal to complete the audio signal frequency, amplitude , other information extraction and get the voice signal waveform.

② Adjustment: the user interface in the design of the input audio signal under a variety of changes such as changes in amplitude and the frequency of operation in order to achieve the adjustment of the voice signal.

II. Frequency domain analysis of speech signals

The Fourier representation of the signal plays an important role in the analysis and processing of the signal. Because of the linear system, it is easy to determine its response to sine or complex exponential sum. Therefore, the Fourier analysis method can solve many of the signal analysis and processing problems. In addition, Fourier indicates that some of the characteristics of the signal become more obvious so it can be more in-depth description of the signal of the red physical phenomenon.

Since the speech signal changes over time, it is generally believed that speech is the output of a linear system that is excited by a quasi-periodic pulse or random noise source. The output spectrum is the product of the channel system frequency response and the excitation source spectrum. The frequency response and the excitation source of the channel system all change with time. Therefore, the general standard Fourier representation is not suitable for speech signal although it is suitable for the representation of periodic and stationary random signals. Since the speech signal can be considered in a short time, the approximation does not change so we can use short-term analysis.

① Transformation: in the user graphical interface under the acquisition of the voice signal Fourier and other transformations then draw the before and after the transformation of the spectrum and the conversion of the cepstrum.

②Filter: filter out the noise part of the voice signal, can be used low-pass filter, high-pass filter, band-pass filter and band resistance filter then compare the effects of various filters.

3) The effect of voice signal display

Play the voice of the processed signal and listen to the processing effect through the user's graphical interface output function.

2.4 System initial flow chart

Figure 2.1 shows the workflow of the entire voice signal processing system:

Figure 2.1 Workflow of voice signal processing systems

The signal adjustment comprises an arbitrary multiple of the amplitude and frequency of the signal as shown in Figure 2.2.

Figure 2.2 Signal adjustment

Signal filtering uses four filtering methods to observe the advantages and disadvantages of various filtering performance:

Figure 2.3 Methods of the voice signal filtering

In the above three figures, we can see that the whole process of voice signal processing system is divided into three steps. First, read into the voice signal to be processed then the voice signal processing including information extraction, amplitude, frequency transformation, voice signal, Fourier transform, filtering and others. Filtering includes low-pass filtering, high-pass filtering, band-pass filtering, band-stop filtering and other means. Finally, the processed voice signal is processed after the effect of display. The above is the work of the system, this paper will start from the voice signal collection to do a detailed introduction.

3 Basic knowledge of voice signal processing

3.1 Voice input and opening

In MATLAB, [y, fs, bits] = wavread ('Blip', [N1 N2]); used to read the speech, the sample value is placed in the vector y, fs is the sampling frequency (Hz), bits number. [N1 N2] indicates reading the value from N1 to N2 (if only one N is the point at which the N point is read).

?Sound (x, fs, bits); used for playback of sound. The vector y represents a signal (that is, a complex 'function expression') that means that the sound signal can be processed like a signal expression.

3.2 Sampling bits and sampling frequency

The number of sampling bits is the sampled value. The parameter used to measure the fluctuation of the sound is the number

of binary digits of the digital sound signal used by the sound card when collecting and playing the sound file. Sampling frequency refers to the recording device in a second time on the sound signal sampling times, the higher the sampling frequency of the sound, the more realistic and more natural reduction.

Sampling bits and sampling rate for the audio interface are the two most important indicators. They are also the two important criteria to select the audio interface. Regardless of the sampling frequency, theoretically the number of bits measured determines the maximum velocity range of the audio data. Each additional one bit is equal to an increase of 6dB in the intensity range. The higher the number of sampling bits, the more accurate the captured signal. For the sampling rate you can imagine it is similar to a camera, 44.1kHz means that the audio stream into the computer when the computer will be photographed up to 441000 times per second. Obviously the higher the sampling rate, the more images taken by the computer, the more accurate the reduction of the original audio.

3.3 FFT analysis of time domain signals

FFT is a fast Fourier transform, which is a fast algorithm of discrete Fourier transform. It is based on the singular, even, imaginary and real characteristics of discrete Fourier transform. The algorithm of discrete Fourier transform is improved. Function FFT and IFFT are used in fast Fourier transform and inverse transform in MATLAB's signal processing toolbox. The function FFT is used for the sequence fast Fourier transform, whose call is in the form y = fft (x), where x is the sequence, y is the FFT of the sequence, x can be a vector or matrix, if x is a vector, y is x. FFT and the same length as x, if x is a matrix, then y is the FFT of each column vector of the matrix. If the x-length is a power of 2, the function fft performs a high-speed base-2FFT algorithm, otherwise fft performs a hybrid-based discrete Fourier transform algorithm that computes slower. The other call format of the function FFT is y = fft (x, N), where x, y is the same as before, and N is a positive integer. If the length of the vector x is greater than N, the function truncates x so that the length equal to N; if x is a matrix, the function is truncated to the length of N; if x is a vector and the length is less than N, handle x in the same way.

3.4 Digital filter design principles

The function of the digital filter is to use the characteristics of discrete time system to process the input signal waveform (or spectrum), or use the digital method to change the signal according to the predetermined requirements.

A digital filter can be understood as a computational program or algorithm that converts a digital time series representing an input signal into a digital time series representing the output signal and causes the signal to change in a predetermined form during the conversion process. Digital filters are classified according to the time domain characteristics of the digital filter impulse response. The digital filter can be divided into two types which are infinite impulse response (IIR) filter and finite impulse response (FIR) filter. In terms of performance, the poles of the IIR filter transfer function can be located anywhere within the unit circle, so that a higher order can be used to achieve high selectivity and the memory cells used are less economical and efficient. However, this high efficiency is at the expense of phase nonlinearity. The better the selectivity, the more serious the phase nonlinearity. In contrast, the FIR filter can obtain a strict linear phase but because the pole of the FIR filter transfer function is fixed at the origin, it can only achieve high selectivity with higher order. For the same filter design, FIR filter required 5 to 10 times higher than that of the IIR filter. As a result, the cost is high and the signal delay is large. If the same selectivity and the same linearity

are required, the IIR filter must be added to the whole network for phase correction, in addition to increase the number of filters and complexity of the filter.

Overall, IIR filter to achieve the same effect of small order and delay but there are stability problems and nonlinear phase. FIR filter has no stability problem, linear phase, but the larger order number and delay.

3.5 The concept of cepstrum

Definition: The cepstrum is defined as the logarithmic inverse Fourier transform of the short-time amplitude spectrum of the signal.

Features: Features that can be roughly separated then can extract spectral envelope information and fine structure information

use: ① extraction channel feature information: extract the spectral envelope characteristics as a description of the phonetic characteristics of the parameters used in speech recognition.

② extract the source information: extract the pitch feature, as a description of the phonological characteristics of the auxiliary parameters used in speech recognition.

Seeking:

A: short time signal; B: short time spectrum; C: logarithmic spectrum; D: cepstral coefficient; E: logarithmic spectral envelope; F: Basic period

4 Example of Speech Signal Processing
4.1 Graphical user interface design

In the main window of MATLAB, select the New menu item in the File menu, and then select the GUI command which will then display the graphical user interface design template. For the GUI design, MATLAB prepared a total of four kinds of templates which are Blank GUI (default), GUI with Uicontrols (GUI template with control object), GUI with Axes and Menu (with axis and menu GUI template) and Modal Question Dialog (GUI template with Mode Interrogation dialog box).

(GUI) SoundProcess, which mainly includes File, Process and Output three main parts. File menu consists of Input, Save, Quit and other functions. The Process menu mainly includes the Extract, Transform and Filter. Extract menu consists of Range and Frequency while the Filter menu contains Lowpass Filter, Highpass filtre, BandpassFilter, band resistance filter (BandstopFilter) and other functional menu.

4.2 Signal acquisition

The system is a short voice signal as an analysis of the sample, through the computer system will be a 'master, information received' voice signal saved to the computer and save in the format of '* .wav'.

4.3 Speech signal processing design

4.3.1 Extraction of voice signals

Using the Wavread function in Matlab, the sampling frequency of the signal is 22500 and the sound is monaural. Use the Sound function can clearly hear the 'master, the information received' voice. Collect the data and draw the waveform.

Where the sampling frequency of the sound Fs = 22050Hz, y is the sampling data, NBITS said quantization order.

Some of the procedures are as follows:

```
???Fn = input ('Enter WAV filename:', 's');% get a * .wav file
[X, fs, nb] = wavread (fn);
ms2 = floor (fs * 0.002);
    ms10 = floor (fs * 0.01);
ms20 = floor (fs * 0.02);
ms30 = floor (fs * 0.03);
```

t = (0: length (x) -1) / fs;   % Calculate the sample time
Subplot (2,1,1);   % Determines the display position
Plot (t, x);   % Draw the waveform
Legend ('Waveform');
xlabel ('Time (s)');
ylabel ('Amplitude');

Operation after the pop-up voice signal processing system interface shown in Figure 4.1:

Figure 4.1 Operation interface of the voice signal processing system

And then click the File menu in the sub-menu Input, back to the Matlab software input interface shown in Figure 4.2:

Figure 4.2 Enter the interface

Enter the name of the voice signal to be processed, you can get the voice of the voice waveform as shown in Figure 4.3:

Figure 4.3 Waveform of voice speech

As shown in the waveform of the speech extracted in the figure, the sound intensity fluctuation in the whole audio data is substantially the same as the input sound signal, and it is observed that some of the high frequency noise is included.

4.3.2 Adjustment of voice signal

In the study of speech signals, often required to converse or adjust the voice signal for multiple frequencies and multiple amplitudes. In the daily application, this change is often used to adjust. Therefore, this function was also added in the design and can observe the adjusted signal waveform changes but also through the voice processing interface output function audition processing voice signal.

4.3.2.1 Frequency adjustment of voice signals

In the design, the sampling signal frequency can be increased or decreased as to achieve the voice signal adjustment and get the ideal voice signal. For example, the sampling frequency can be doubled, you can get the voice signal frequency of the original frequency of 2 times the new voice signal. Run Process→Adjust→Frequency, get the signal waveform as shown in Figure 4.4, and listen to the adjusted effect.

Figure 4.4 Waveforms after frequency adjustment

Compared with the original voice signal, after adjusting the signal cycle into the original 1/2 the speed is significantly faster, that is to achieve the signal 2-fold function.

4.3.2.2 Amplitude adjustment of speech signal

In the design, the amplitude of the speech signal can be improved or reduced to achieve the adjustment of the voice signal as to get the sound volume of different voice signals such as the original voice signal amplitude doubled to get the following Figure 4.5 signal waveform. Through the GUI interface, we can listen to the output function after the adjustment of the effect.

Figure 4.5 Amplitude adjusted waveform

At this time to listen to the tone after the adjustment of high tone although it is not very obvious. However, you can set the change of amplitude large then you can get the effect of the obvious voice signal.

4.3.3 Fourier transform of speech signal

Cepstrum analysis refers to the inverse of the signal's short-term amplitude spectrum for inverse Fourier transform. It has the characteristics of approximate separation and extraction of spectral envelope information and fine structure information.

To?the spectral analysis of speech signal, fft signal fast Fourier transform can be used in Matlab function to get the speech cepstrum then cepstrum analysis to get signal cepstrum.

Some of the Fourier transform procedures are as follows:
x = y (44101: 55050,1);   % Extracts part of the original speech signal
t = (0: length (x) -1) / fs;   % Calculate the sample time
Subplot (3,1,1);   % determines the display position
Plot (t, x);   % draw the waveform
('Waveform');
xlabel ('Time (s)');
ylabel ('Amplitude');
Y = fft (x, hamming (length (x)));   % is the windowed Fourier transform
fm = 5000 * length (Y) / fs;   % limit frequency range
f = (0: fm) * fs / length (Y);   % Determines the frequency scale
Subplot (3,1,2);
Plot (f, 20 * log10 (abs (Y (1: length (f))) + eps));
Legend ('Spectrum');% Draw Spectrum
ylabel ('amplitude (db)');
xlabel ('frequency (Hz)');
C = fft (log (abs (x) + eps));   % cepstrum calculation
ms1 = fs / 1000;
ms20 = fs / 50
q = (ms1: ms20) / fs;   % Determines the chattering scale
Subplot (3,1,3);
Plot (q, abs (c (ms1: ms20)));   % Draw the cepstrum
Legend ('Cepstrum');
xlabel ('bass (s)');
yrawel ('cepstrum amplitude (Hz)');

Run Process→Transform, the Fourier transform a part of the speech signal then cepstrum analysis to get figure as shown in Figure 4.6

Figure 4.6 Sound sample waveform, spectrum and cepstrum

From the above cepstrum can be seen. When reading 'master, information received', the corresponding frequency is about 200Hz. This is consistent with the human voice signal frequency concentrated between 200 Hz and 4.5 kHz. In the unvoiced period, the relative small high frequency part (200500Hz) should belong to the background noise.

4.3.4 Filtering of voice signals

As shown in Figure 4.4, the speech signal contains background noise, which is generally higher in frequency. So, you can use the MATLAB software filter for filtering to get the ideal voice signal.

4.3.4.1 Low-pass filtering of speech signals

The system has designed a cut-off frequency of 200Hz Chebyshev-I type low-pass filter, its amplitude-frequency characteristics as shown in Figure 4.7:

Figure 4.7 Amplitude-frequency characteristics of low-pass filters

Low-pass filter performance indicators: wp = 0.075pi, ws = 0.125pi, Rp = 0.25; As = 50dB;

After the low-pass filter processing, compare the changes before

and after the waveform diagram, as shown in Figure 4.8:

Figure 4.8 Changes in low-pass filtered waveform and spectrum

After passing the low-pass filter, the sound became slightly boring and low, because the high-frequency components are low-pass filter attenuation but very close to the original voice.

4.3.4.2 High-pass filtering of speech signals

The Chebyshev-II digital high-pass filter is used to filter the speech signal. High-pass filter performance indicators: wp = 0.375pi, ws = 0.425pi, Rp = 0.25; As = 50dB; and then compare it with the original signal as shown in Figure 4.9:

Figure 4.9 Waveform and spectrum changes after high pass filtering

After passing the high-pass filter, this time only a little noise because the low-frequency components are high-pass filter attenuation while the human voice part is just a low frequency part. Therefore, only left the noise or issued high-frequency noise but cannot hear by the human ears

4.3.4.3 Bandpass filtering of speech signals

The use of elliptical digital bandpass filter function, the voice signal after filtering its comparison with the original signal as shown in Figure 4.10:

Figure 4.10 Variation of waveform and spectrum after Bandpass Filtering

4.3.4.4 Band-stop filtering of speech signals

Using the Chebyshev-II type digital band-stop filter, the comparison of the speech signal after filtering with the original signal is shown in Figure 3.11:

Figure 4.11 Variation of waveform and spectrum after Bandage Filter

From the above various digital filters, the sound slightly boring after filtering the speech signal compared to low-pass filter but very close to the original sound. After filtering by high-pass filter it cannot hear the human voice while the sound is a bit like a sound from a robot jingle after passing band-pass filter. With blocking filter, the sound is closer to the original sound. From the spectrum we can see that the energy of the sound is mainly concentrated in the low frequency (0.2pi or 2204.5Hz or less) part.

4.4 Output of voice signal

You can first play the processed voice signal in the Matlab software to experience the effect of voice signal processing. You can also save the processed voice signal on your computer.

Run File→Save to save the processed voice signal. If no voice signal is processed, the system will appear as shown in Figure 4.12:

Figure 4.12 Save the prompt interface

If there is a voice signal is processed, run File→Save, the system will appear as shown in Figure 4.13:

Figure 4.13 Save the interface

The entire operation is completed after saving the files.

5 Conclusions

In this paper, the design of the speech signal processing system is introduced in detail. A series of image analysis and processing techniques are used to realize the basic processing function of the speech signal. After the test run, the design of the speech signal is completed. It is better to complete the spectrum analysis of the speech signal through the fft transform to obtain the spectrum of the voice signal. In the piece of the filter, the main project is to start from the digital filter to design the filter. This basically achieves the filter and completed a variety of filter effect comparison and very consistent with the requirements of the project.

In this paper, the design of the voice signal processing system is introduced in detail. A series of image analysis and processing techniques are used to realize the basic processing function of the speech signal. After the test run, the basic purpose is achieved.

The design has the following advantages:

1) The interface is concise. In the operation of the voice signal processing interface, the menu button clear and each function corresponds to only one button which eliminates the need for large-scale operation of complex software steps.

2) Processing faster. Since there are multiple steps in the entire operation and each step is not very close so the speed of each step is very fast.

3) Occupies less memory space. The entire process only takes up to dozens of KB of physical space which eliminates the need for software installation.

The design also has many places to improve, mainly in the following areas:

1) This procedure can only be carried out under the general application of the voice signal processing, the function is relatively simple, cannot carry out complex voice signal processing.

2) As the system is located in the general conditions of the voice signal processing, the calculation accuracy is relatively low, cannot carry out more accurate voice signal processing.

3) The program is made from simple preparation, the operation interface is small, cannot process the larger physical memory voice signal file.

REFERENCES
[1] Li Yong, Xu Zhen. MATLAB auxiliary modern engineering digital signal processing. Xi'an University of Electronic Science and Technology Press.
[2] Chen HuaiChen. Digital signal processing tutorial - Matlab interpretation and implementation. The Electronics Industry Press.
[3] Wang Yishi. Digital signal processing. Beijing Institute of Technology Press.
[4] Chen Jie. Matlab Collection, the electronics industry press.
[5] Liu Yancheng, Song Tingxin. Speech recognition and control technology. Science Press.
[6] Gao Xiquan, Ding Yumei. Digital signal processing. Xi'an University of Electronic Science and Technology Press.
[7] Cheng Peiqing. Digital signal processing tutorial. Tsinghua University Press.
[8] Zhao Li. Voice signal processing. Machinery Industry Press.
[9] Han Jiqing, Zhang Lei. Speech signal processing. Tsinghua University Press.
[10] Hu Hang. Speech signal processing. Harbin Institute of Technology Press.
[11] Su Jinming, Zhang Lianhua, Liu Bo. MATLAB toolbox application. Electronic Industry Press.
[12] Wang Jinghui. Speech signal processing technology research papers. Shandong University.
[13] Yu Junfeng. Speech signal recognition and conversion research [Degree thesis]. Chengdu University of Technology.
[14] Gersho A, Gr R M. Vector quantization and signal compression. Boston, Kluwer Academic Publishers.
[15] Q. Zhang, A. Benveniste. Wavelet Networks, IEEE Trans, Neural Networks.