Implementation of noise reduction optimization process of language lab recording based on RNN model

Yuzhu Zhang, Wei Zeng, Qiang Liao
Sichuan University, College of foreign languages, Sichuan ,Chengdu, 610065

Abstract: Speech signal noise reduction involves speech signal processing technology, which is one of the core technologies in the field of information application. Speech signal optimization analysis and processing is also a popular direction in the field of artificial intelligence. In this paper, a speech signal optimization processing method based on the recurrent neural network RNN model is proposed for the language laboratory recording signal as the research object. The key technologies in the field of deep learning are introduced for audio noise reduction processing. By using experimental simulation, the signal-to-noise ratio of the recording signal is improved, so as to obtain a more ideal language laboratory recording audio file.

Key words: Speech signal; Deep learning; Recurrent neural network RNN ; Signal-to-noise ratio

1.Introduction

The recording equipment of the language laboratory undertakes the functions of school foreign language classroom teaching, college foreign language public course examination and professional foreign language examination listening recording, which provides a strong guarantee for the operation of school foreign language teaching and examination. As we all know, foreign language recording has high requirements for audio fluency, clarity, fidelity, etc. Foreign language laboratories are equipped with professional recording studios and recording equipment, which can minimize the impact of noise on audio recording. In the daily recording process, it is impossible to record without noise. In the case of good early physical environment noise reduction work, what can be done is to reduce the noise of the recording file through post-processing. This involves the audio noise reduction optimization processing algorithm and speech signal processing optimization technology.

Figure1 The recording environment of foreign language laboratory

2.Research Method

In the past, we mostly used some common traditional noise reduction methods. For example: sound insulation room, installation of muffler, sound insulation barrier, vibration reduction, filtering, traditional late noise reduction algorithm, etc. Traditional algorithms estimate noise by statistics and other methods, and can play a good role in reducing steady-state noise. However, in the case of non-steady-state noise, transient noise and other noise types, traditional noise reduction algorithms often cannot play a good role. In recent years, with the continuous development of artificial Intelligence technology, a lot of noise reduction algorithms based on AI (Artificail Intelligence) model have appeared in audio processing fields such as noise reduction. These AI algorithms have a great improvement in noise reduction ability compared with traditional algorithms. In this paper, the RNN cyclic neural network model is used to optimize the noise reduction of the language laboratory recording. The audio signal of the foreign language training center is taken as the research object to reduce the noise interference.

Compared with traditional noise reduction methods, the process achieved by AI noise reduction algorithm is more accurate, more robust, less redundant and more optimized. Common types of artificial Neural networks include Deep Neural Network (DNN), Convolutional Neural Network (CNN), and Recurrent neural Network (Recurrent Neural Network, RNN), etc.

The traditional neural network is composed of input layer, hidden layer and output layer. One obvious shortcoming of this neural network is that neurons in the same layer do not transmit to each other, which leads to poor performance in audio noise reduction. Because in recorded audio, the next word is related to the previous word, and the sentences are related to each other. RNN can let the neurons of the hidden layer communicate with each other and store the output result of the previous paragraph in the hidden layer in the form of information. When the next paragraph is denoised, the output of the previous paragraph will also have an impact on it. This transmission mechanism of RNN makes the denoised audio statements more coherent and the effect more fidelity.

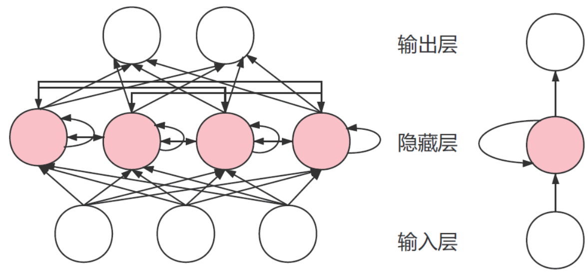

Figure2 A typical RNN network structure

It can be seen from the typical RNN network structure that each neuron of the hidden layer in the RNN network is related to not only the signal of the input layer, but also the state of the hidden layer itself. This kind of autoregressive structure is the characteristic of RNN. Common RNN networks include LSTM, GRU and so on. The hidden features of each time point in the RNN structure can be extracted not only from its own signals, but also from the information of the time points before and after. Therefore, RNN is a commonly used method in sequential modeling.

3.Research content

In this paper, the GRU/LSTM model in RNN is adopted. After understanding the basic structure of RNN, let's see how the AI noise reduction model is trained. A large number of parameters in the AI model, such as the calculated weight of each neuron in the RNN, need to rely on training to get. After several training, the model can fit the mapping relationship between output and input. Pay attention to avoid overfitting and underfitting.

AI model training can be classified into supervised training and unsupervised training according to whether there is labeled data. In the training of noise reduction model in this paper, supervised training is used, and the method of back propagation combined with gradient descent is used to continuously improve the similarity between model prediction and pure speech. This degree of similarity is generally expressed by the Loss function Loss. The smaller the loss is, the closer the result obtained by the model is to the pure speech.

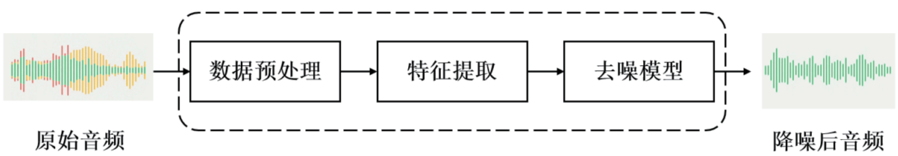

Figure3 Audio noise reduction flow chart

In data preprocessing, pure speech should be used as the target or label, and some noise should be added to the pure speech to generate noisy signals as the input of the model. The noise here mainly refers to environmental noise, which is generally additive noise, so when preparing training data, it is necessary to prepare a pure speech library and a noise library, and noisy data can be directly obtained by adding pure speech and noise signal.

```
RNNoise is a noise suppression library based on a recurrent neural network.

To compile, just type:
% ./autogen.sh
% ./configure
% make

Optionally:
% make install

While it is meant to be used as a library, a simple command-line tool is
provided as an example. It operates on RAW 16-bit (machine endian) mono
PCM files sampled at 48 kHz. It can be used as:

./examples/rnnoise_demo <noisy speech> <output denoised>

The output is also a 16-bit raw PCM file.

The latest version of the source is available from
```

```
void rnnDeNoise(char *in_file, char *out_file) {
    uint32_t in_sampleRate = 0;
    uint64_t in_size = 0;
    int16_t *data_in = wavRead_int16(in_file, &in_sampleRate, &in_size);
    uint32_t out_sampleRate = 48000;
    uint32_t out_size = (uint32_t) (in_size * ((float) out_sampleRate / in_sampleRate));
    int16_t *data_out = (int16_t *) malloc(out_size * sizeof(int16_t));
    if (data_in != NULL && data_out != NULL) {
        resampleData(data_in, in_sampleRate, (uint32_t) in_size, data_out, out_sampleRate);
        denoise_proc(data_out, out_size);
        resampleData(data_out, out_sampleRate, (uint32_t) out_size, data_in, in_sampleRate);
        wavWrite_int16(out_file, data_in, in_sampleRate, (uint32_t) in_size);
        free(data_in);
        free(data_out);
    } else {
        if (data_in) free(data_in);
        if (data_out) free(data_out);
    }
}
```

Figure4 Part of the code situation

The rough steps of RNN model noise reduction: the audio signal with noise is transformed by STFT to obtain the signal in frequency domain; Frequency domain signal as input, using artificial neural network RNN algorithm to obtain frequency domain mask; The frequency domain mask is multiplied by the STFT to get the frequency domain signal, and then the frequency domain signal after noise reduction is obtained. The frequency domain signal after noise reduction is transformed into the inverse STFT to obtain the relatively pure speech signal we need.

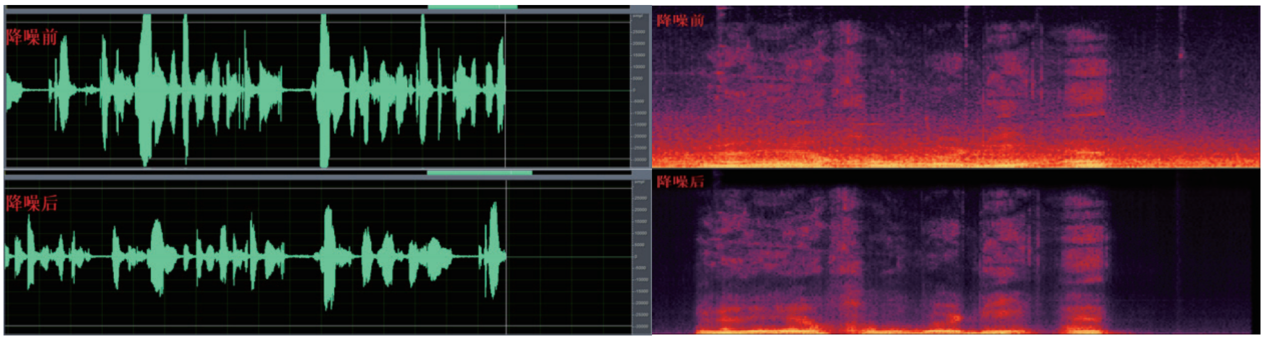

Figure5 Noise reduction experiment results

4.Conclusion

Combined with deep learning algorithm for noise reduction processing of audio in foreign language laboratory, RNN cyclic neural network is used for training modeling processing of audio files, which can effectively reduce the influence of noise on audio, and finally get ideal experimental results. At present, we are in the era of continuous development of artificial intelligence, and the integration of artificial intelligence and all walks of life has become the main development direction in the field of artificial intelligence. The rapid development of AI technology will drive the increasing scale of intelligent voice application, and effectively promote the practical application and development of voice noise reduction algorithm model.

Reference

[1] Ma C, Li D, Jia X. Optimal scale-invariant signal-to-noise ratio and curriculum learning for monaural multi-speaker speech separation in noisy environment[C]//2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). IEEE, 2020: 711-715.

[2] Luo Y, Mesgarani N. Tasnet: time-domain audio separation network for real-time, single-channel speech separation[C]//2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018: 696-700.

[3] Jiang Kaiyu. Research on Correlation Analysis of Audio Noise Reduction [D]. University of Chinese Academy of Sciences,2014.

[4] Han Changjun. Application of Speech Signal Denoising Method Based on MATLAB [J]. Journal of Liaodong University (Natural Science Edition),2017,24(01):72-77.

[5] Jin Zequn. Abnormal Audio Detection Method Based on Deep Learning [D]. Shenyang Jianzhu University,2019.