Design and Implementation of Dual Channel Speech Signal Transceiver System Based on FPGA

Tongbai Yang[1(✉)] and Shuo Shi[1,2]

[1] School of Electronic and Information Engineering, Harbin Institute of Technology, Harbin 150001, Heilongjiang, China
yangtongbai2000@163.com, crcss@hit.edu.cn

[2] Peng Cheng Laboratory, Network Communication Research Centre, Shenzhen 518052, Guangdong, China

Abstract. Real-time speech information has always been the main way of information transmission, but there are still problems in transmission. Firstly, while speech signals are easy to transmit information, they also bring information leakage and malicious tampering. Secondly, speech signals are more or less affected by noise, interference and transmission effects in the transmission process, which will make the recovery of speech signals more complicated. Based on the above, the wireless transmission of speech signals is taken as the background to realize the merging and separation of dual-channel speech signals. The problems of information leakage and distortion during the transmission of speech signals are studied. The encryption and decryption of speech signals are taken as the key points. Realize dual speech signal receiving and receiving based on Field Programmable Gate Array (FPGA). The frequency division multi-plexing method is used to transmit dual speech signals simultaneously, realize speech signal convolution on FPGA and extract two speech signals from the convolution signal. AES algorithm is realized the encryption and decryption of on FPGA. 2FSK modulation and demodulation in digital modulation is selected to complete the design and implementation of the whole system, and the research is completed through the final overall modulation and analysis.

Keywords: speech signal processing · FPGA · AES algorithm

1 Introduction

The rapid development of science and technology drives the progress of society, so does the field of communication. The diversified and multi-scene application of contemporary communication technology enables people to obtain relevant valuable and important information in a timely manner in multiple ways. Strong real-time speech information has always been the main way of human information transmission, but there are still problems in information transmission. First of all, while speech signals are

A. Li et al. (Eds.): 6GN 2022, LNICST 504, pp. 146–157, 2023.
https://doi.org/10.1007/978-3-031-36011-4_13

easy to transmit information, they also bring information leakage and malicious tampering. Secondly, speech signals are more or less affected by noise, interference and transmission effects in the transmission process, which makes it more difficult to separate and extract the source speech signals, that is, speech signals will reach the same receiver through multiple different channel conditions. The received signal will be a convolution mixture of multi-channel speech signals with attenuation, delay and phase effects. This kind of receiving signal is more complex, which makes the subsequent receiving and processing of speech signal more difficult. Against the above background, the background of this topic in wireless transmission of speech signal, two-way road and separation of speech signals as the key point, first of all, the speech signal distortion of information disclosure and transmission process, and then focus on modulation of two-way speech signal encryption and decryption algorithm, selecting the appropriate modulation demodulation method, based on the above research, FPGA based dual speech signal transceiver.

At present, the mainstream speech processing is designed on the basis of software platform, and it is still a minority to realize the speech processing with hardware devices. However, field programmable gate array (FPGA) is very suitable for processing speech signal with its advantages of high parallelism and high flexibility. There are some prevalidated IP cores for signal processing algorithms in FPGA. The design of these IP cores can achieve higher performance of some commonly used signal processing functions, which is very convenient to use in the development process and can greatly shorten the development time. Moreover, the effective integration of multiple signal processing algorithms can effectively reduce the actual cost and reduce the risk. With the development of FPGA technology, the realization of signal proc-essing algorithm in FPGA is becoming a new alternative because of its relatively low implementation cost and high parallel processing speed. By implementing the speech signal processing algorithm in FPGA, the traditional limitation separation between hardware and software design level is gradually overcome [1].

Speech encryption is to transform the information processing, so that in addition to the target receiver of the communication system, can not get the real speech information in the transmission signal. The method of speech signal transformation is speech encryption and decryption algorithm. The most common method is to transform speech signal by key control. Speech decryption is the process of restoring the encrypted signal to the original speech signal using the opposite processing operation. The research of speech encryption has been deepening at home and abroad [2, 3].

In addition to ensuring the security of speech communication, it is also necessary to ensure that the information in speech communication can be accurately and completely transmitted. Otherwise, speech encryption is meaningless when the information cannot be completely transmitted. For the problems that speech communication must face, the requirements of speech encryption generally include the following: with the increasing awareness of information security in modern people, people hope to be able to protect their personal information security through speech encryption; On the other hand, for the country and the enterprise, if the confidential information leakage will bring irreparable huge economic losses to the country and the enterprise, which more reflects the importance of speech encryption [4].

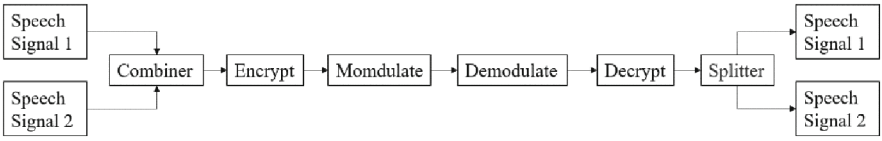

Fig. 1. System flow chart

Based on the overall process, the speech signal is input into the system in the form of digital signal in this study. The two-channel speech signal successively passes through the combination, encryption, modulation, demodulation, and decryption. After separating the complete system, the two-channel signal is extracted without distortion. Figure 1 shows the specific flow chart.

Therefore, it is mainly divided into three parts: speech signal routing and separation, speech signal encryption and decryption, and signal modulation and demodulation. In this study, frequency division multiplexing (FDM) is selected to combine and separate dual speech signals. Because of ensuring the integrity of the signal to the greatest extent, it is easy to operate, easy to implement, and does not occupy too many resources. The AES speech encryption algorithm can protect the two speech messages from being easily deciphered and meet the security requirements. 2FSK modulation and demodulation method is selected to complete and accurately demodulate the original modulated signal.

2 Speech Signal Combination and Separation

In order to reduce the communication cost, the current communication system pursues the high efficiency and low cost scheme of transmitting multiple signals in the same channel. Under the condition that the transmitted signals do not interfere with each other, multiple signals can be merged, so that they can be completely separated after transmission at the same time.

In the process of speech signal convolution transmission, because the two speech signals are simultaneously transmitted, it is difficult to recover the two speech signals from the convolution signal completely and without distortion in the time domain in the processing of the convolution signal transmitted through the same channel. Therefore, considering the time of signal processing and the degree of difficulty of implementation, the frequency division multiplexing processing method in frequency domain is adopted.

2.1 The Realization of Speech Signal Combination

Firstly, one speech signal 1 is processed, its spectrum is transformed to a higher frequency, and then it is transmitted with another speech signal 2 as a closed signal. The two do not affect each other, and the transformed speech signal 1 and speech signal 2 have no overlap in spectrum to complete the closed path.

The realization of speech signal jointing mainly lies in the generation of fixed carrier for spectrum shifting and the multiplication of speech signal and carrier for frequency division multiplexing. The generation of the fixed carrier can be generated by the IP core of the DDS (Direct Digital Synthesis) built into Vivado.

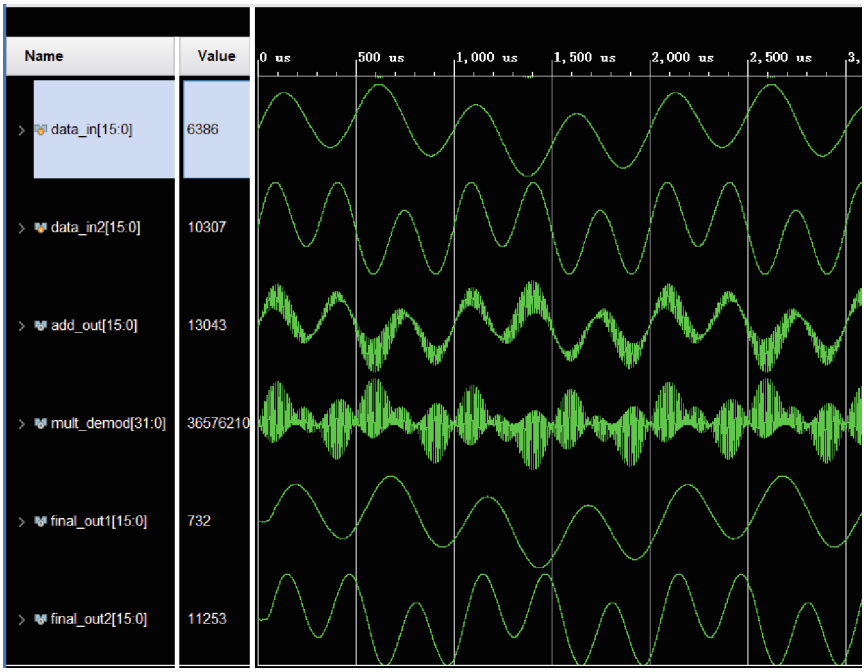

Fig. 2. Results of speech signal combination and separation

Because the carrier with a fixed frequency needs to be generated, it is not necessary to write a frequency control word to change the frequency of the output waveform, but only need to fix the phase increment to generate a single frequency wave with a fixed frequency. A fixed phase increment is selected, which does not change during operation, and a sinusoidal waveform with a fixed frequency is generated. The output frequency of the IP core of DDS is controlled by the frequency control word, clock frequency and Phase Width. The output frequency is:

$$f_{out} = (f_{clk} \cdot f_{word})/2^{PhaseWidth} \tag{1}$$

Subsequent operations only need to connect the input speech signal 1 and DDS output carrier signal to the multiplier, and connect the output of the multiplier and another speech signal 2 to the adder, and the output of the adder is the signal to complete the circuit closing.

2.2 The Realization of Speech Signal Separation

Considering the spectrum characteristics of the two channels of speech signal after processing, the two channels of speech signal can be obtained only by two different processing of the combined channel signal. The first processing is to directly low-pass filter the received combined channel signal to get the original speech signal 1. Of the second processing the received signal to bandpass filter, bandpass filter center frequency

is a way to deal with the carrier frequency, bandwidth of 2 times the signal bandwidth, and after the road processing carrier multiplication, the modulation of the speech signal after 2 spectrum move back to its original position, then the original speech signal 2 is obtained by low pass filter.

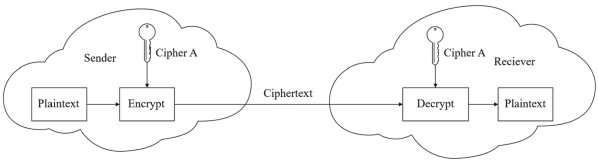

Fig. 3. Symmetric Encryption Algorithm

Using the IP core of the Vivado FIR Compiler. The IP core can be designed according to the Filter Designer in Matlab to generate the required Filter. This greatly simplifies the difficulty of filter design, so the parameters can be directly designed and selected according to the characteristics of speech signals.

In order to facilitate the analysis, data_in and data_in2 are the input two-way speech signals, add_out is the join-way signal, and final_out1 and final_out2 are the separated two-way speech signals. The realization results of speech signal combining and separating FPGA are shown in Fig. 2:

According to the results, the two signals are separated without distortion. It verifies the correctness of the dual speech signal merging and separating module implemented by FPGA.

3 The Implementation of AES Algorithm

3.1 Principle of AES Algorithm

Because the inverse function of the decryption algorithm of the AES algorithm is itself, the encryption and decryption algorithm of the AES algorithm [5] has symmetry. In the algorithm, the same key is used, and both the sender and the receiver process the data with the same key. As shown in Fig. 3:

The AES algorithm mainly has four different stages, which are SubByte layer, Shift Rows layer, Mix Column layer and Add Round Key. The subsequent 128bit key is selected for 10 rounds of processing, and the length of the plaintext and the key are both 128bit.

Most of the processing of AES algorithm is byte processing, so according to the regulations, the data should be arranged by byte, the input order is arranged by column from left to right into 4·4 array. Similarly, the encrypted ciphertext is read in the same order, and the array is changed into the original 128bit form. In the same decryption, the data is also processed in a 4·4 array byte arrangement and read in the same order.

(1) Round key addition: 128bit plaintext and key are processed in round key addition. In the finite field used in round key addition, each element can be represented as:

$$A(x) = a_7x^7 + a_6x^6 + \ldots + a_1x + a_0 \tag{2}$$

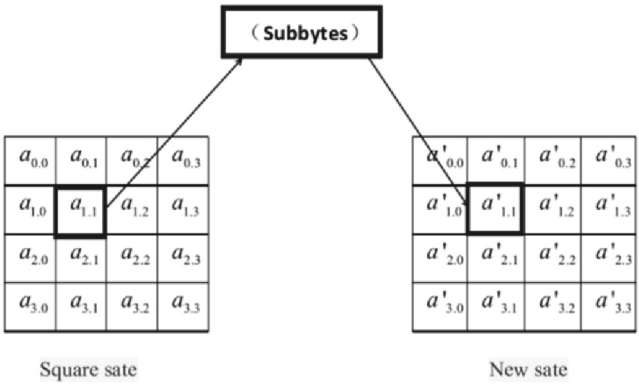

Fig. 4. S-box Processing

The addition and subtraction of two elements give the same result because it is equivalent to an XOR gate in a digital circuit. The object of the operation is the round transformation result of each round except the round key plus itself and the specific generated round key. The final processing result obtained by XOR operation is the output result of this layer.

(2) byte substitution layer: byte substitution layer to the processing of data of the main is to make the input data through the S-box table finish from a byte changes into another completely different, the conversion operation is the input data of each byte by S-box table to find the corresponding new byte as output bytes, S-box table lookup method is: The input of a former four bytes of data value as the table to find S-box, and after the bytes of the value of the four as a table column value to find S box, by the number value and the value of the column to identify the S-box in the table the transformation output of new bytes address, find the corresponding new byte as complete through the address byte substitution transformation output, Fig. 4 abstracts this transformation.

There are 256 8bit long elements in the S-box table, which can be defined as a two-dimensional number table. The method of reading the data in the S-box is to take the first 4 bits of each byte in the input data to represent the row value in the table, and the last 4 bits to represent the column value in the table. Because the corresponding processing can be carried out in FPGA, the S-box is regarded as a two-dimensional number table for subsequent processing and research. The inverse S-box has the same principle as the S-box, which is applied to the data processing of byte replacement during decryption.

(3) Row displacement: Row displacement operation is very easy compared with other processing stages. It is used to process the 16-byte input data of this layer as the state matrix, the input bytes as the units of the state matrix, and then the position of the unit rows in this state matrix is changed. The operation of row shift is to change the position of a single byte to affect other byte positions and thus affect the entire state matrix. Row displacement is the data processing operation only between the rows of

the state matrix, without changing the position of the cell column in the matrix. In the state matrix, for the data with four units in each row, the first row of the state matrix is guaranteed to remain unchanged during processing. The second row is moved to the left by 1 unit, the third row is moved to the left by 2 units, and the fourth row is moved to the left by 3 units. According to the symmetry feature, in decryption and encryption operation opposite, the retrograde displacement processing is still keeping the first row unchanged, the second row to the right one unit, the third row to the right two units, the fourth row to the right three units.

(4) Column confusion: The data processing in the column confusion stage is the most complex in this algorithm. Column obfuscation operation shuffles each column of the input state matrix, so that any byte of the input data will affect the four bytes of the result, which is the most important diffusion element in this algorithm. The operations performed by column obfuscation include matrix multiplication, finite field addition and finite field multiplication. For encryption, column obfuscation right-multiplies a particular matrix by the input state matrix.

(5) AES key generation: Sub-key generation is to process the columns in the key matrix, not the rows. A column contains 4 bytes, and the four columns together constitute the sub-key. Because the first key addition layer also uses sub-keys for XOR addition, the number of generated sub-keys is one more than the round number of AES algorithm, that is, 11. The sub-key data is all in the extension. However, if the column generated by the operation is the first column of the sub-key, the processing transformation of G function should be carried out on the data in the column of the operation.

The G function is to first invert the four bytes of the input column, then replace each byte through the S-box table, and finally XOR calculate the round coefficient with the first byte after replacement. The round coefficient contains 10 data, each of which is 8bit in length. Clearing the symmetry of AES and improving the nonlinearity of the key arrangement process are the reasons for the existence of G function. With these two features, the probability of password cracking attacks is greatly reduced.

AES decryption process and encryption stream operation process relative comparison can find the symmetry of the symmetric encryption and decryption algorithm process, the content of the data processing process is similar, the difference is only that the decryption is the inverse process of encryption, and the use of the same key, so no longer related to the specific introduction.

3.2 AES Algorithm Implements the Results on FPGA

The setup of the pin is described first. Input pin: clock is the clock signal, the clock frequency is 100 MHz, resetn is the reset signal, the low level is effective; Enc_dec is the choice of encryption and decryption algorithm, low level means encryption, high level means decryption; Start is to start the operation of the encryption and decryption algorithm (valid when key_val is 1). The rising edge is triggered, and the next clock starts the operation. Key_in indicates the input value of the 128-bit key value. Text_in indicates the input value of 128-bit plaintext or encrypted data value. Key_val is a module enabled interface. The module can be executed only when the value is 1. Output pin: text_val changes to high level, indicating that the encryption and decryption operation

has been completed; Busy indicates that the module is performing operations. Text_out represents the calculated output of 128-bit encrypted or decrypted data [6].

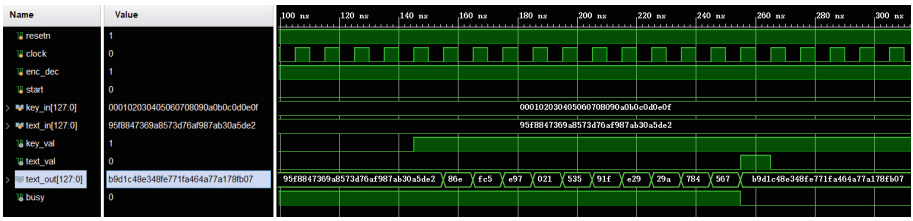

Fig. 5. AES Decryption Result

The AES encryption algorithm is running, where the input value is B9D1C48E348FE771FA464A77A178FB07 (in hexadecimal format, and the subsequent data is also displayed in hexadecimal format). The key input is 000102030405060708090A0B0C0D0E0F, and the final output is 95F8847369A8573D 76AF987AB30A5DE2.

As shown in Fig. 5, AES decryption algorithm is carried out, and the input encrypted data is the output value of the above encryption algorithm, namely 95F8847369A8573D76AF987AB30A5DE2, The key is also 0001020 3040506070809 0A0B0C0D0E0F, and it can be found that the decrypted output is B9D1C48E348FE771FA464A77A178FB07. It can be clearly found that although the corresponding data generated by each round of encryption and decryption is different, the final output result is correct.

4 2FSK Modulation and Demodulation Implementation

2FSK can use the IP core of DDS to complete the generation of carrier on FPGA. Only two different frequency control words are needed to complete the modulation of 2FSK. The keying of frequency control words can be completed by the value of the input digital signal. Then, different frequencies are generated for the IP core control of the DDS that generates the modulation signal to complete the modulation of 2FSK.

Demodulation using a relatively simple non-coherent demodulation, that is, envelope detection. The central frequencies of the passband frequencies of the two bandpass filters are respectively those of the two frequencies generated by keying. The power of the corresponding frequency signal in the received signal can be output by bandpass filtering, rectification and low-pass filtering. The output of two low-pass filters is subtracted to find the difference between the two frequency components. When the difference is large enough, the decision can be considered to receive a signal of “0” or “1” [7].

Bpf1_m_data_out is a signal that passes through a bandpass filter, where BPF1_abs is the output after rectification of one of the channels, and rectification is completed by taking the inverse. Bpf1_m_data_valid indicates that the output of the bandpass filter is enabled. If there is data output, set this parameter to 1. The results of 2FSK demodulation on FPGA are shown in Fig. 6:

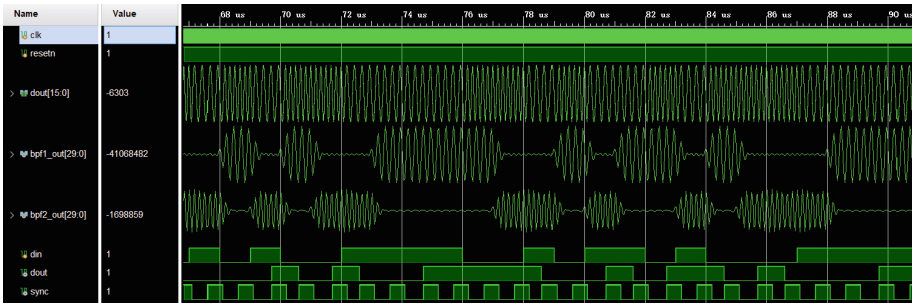

Fig. 6. 2FSK Demodulation Is Implemented on FPGA

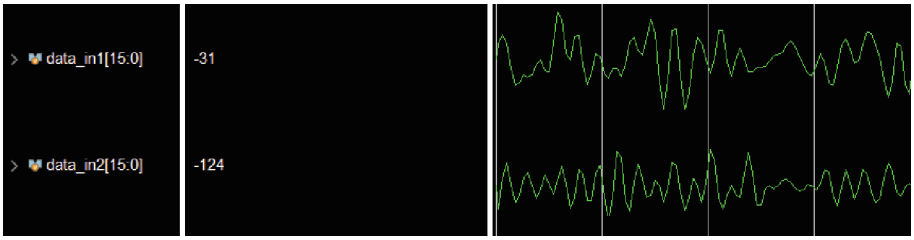

Fig. 7. FPGA Speech Signal Input

In the figure, bpf1_out and bpf2_out are respectively the outputs of two bandpass filters. Approximately, it can be seen that there are outputs only when the sequence signal value is the center frequency corresponding to the filter, dout is the final output sequence, sync is the positioning signal, and the length of a cycle is the length of a symbol, which is used for positioning and output signals. By comparing the input sequence din with the output sequence dout, it is found that only the time is changed, and the information carried by the sequence signal is not changed, which proves the accuracy of 2FSK demodulation [8].

5 System Testing and Analysis

After all the main modules of the system are built, the system is coordinated to correspond to the quantized data of FPGA input, as shown in Fig. 7. In the figure, data_in1 represents the input speech signal 1, and data_in2 represents the input speech signal 2. Ensure that the information of the speech signal is input into the system for subsequent processing [9].

The output of the whole system is shown in Fig. 8: Final_out1 and final_out2 are 16-bit signed speech signals 1 and 2 output by the system. The reason why the waveforms become "smooth" is that the low-pass filter is used to process the signals when separating the speech signals. By comparing the signals in the time domain, it is obvious that the final output is almost the same as the original speech signals. The reason for the error is the quantization error caused by quantization, which is unavoidable. In order to more

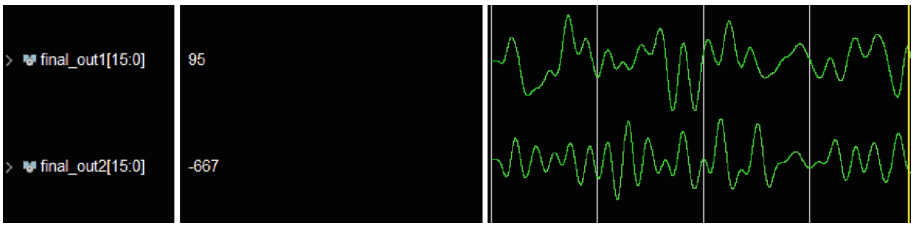

Fig. 8. FPGA System Output

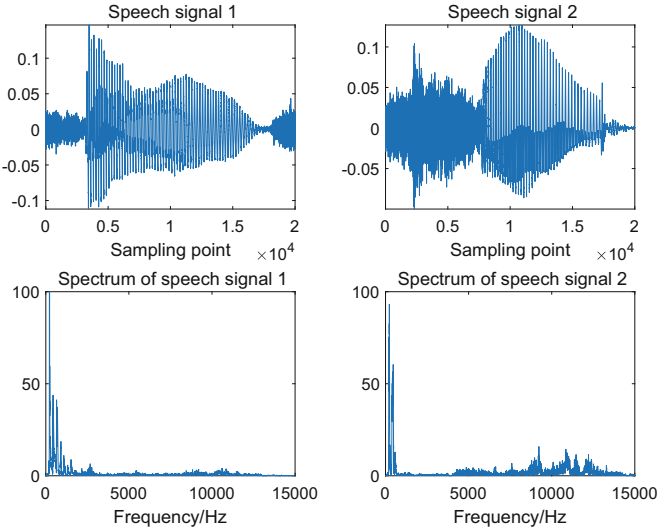

Fig. 9. Spectrum of Two Speech Signals

accurately determine whether to achieve distortion-free demodulation signal, the system output is sampled according to the sampling rate of the original speech signal, and then converted into floating point numbers for the convenience of subsequent operations.

Fourier transform is applied to the original speech signal to obtain the spectrum of two speech signals [10], as shown in Fig. 9. Then Fourier transform is applied to the two signals output by the system to obtain the spectrum of speech signals, as shown in Fig. 10.

The original speech signal and the system output speech signal spectrum is compared, found that the output of the speech signal of high frequency component has been filter, this is because the shunt processing used by low pass filter, is the low pass filter is not only completed the separation effect of signal, and to filter out the noise of the original speech signal.

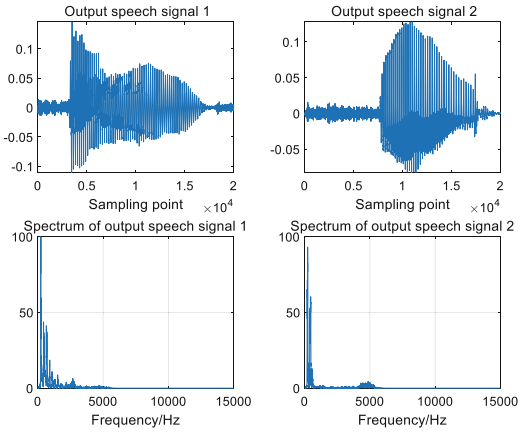

Fig. 10. Spectrum of System Outputs

6 Conclusion

This paper first explains the background and significance, and then introduces and analyzes the current situation.In this paper, the wireless transmission of speech signals as the background, the realization of dual speech signal convolution and separation, first of all, the study of information leakage and distortion in the process of speech signal transmission, AES algorithm, the selection of 2FSK modulation and demodulation method, based on the above research, the realization of dual speech signal transceiver based on FPGA. In the final process of the system, through the analysis of the results, it is found that the whole system can achieve distortion free separation of two speeech signals, which verifies the correctness of the system. On the basis of ensuring the integrity of the signal to the greatest extent, it is easy to operate, easy to implement, does not occupy too many resources, and can be correctly implemented in FPGA.

Acknowledgement. This work is supported by the National Natural Science Foundation of China under Grant 62171158 and Research Fund Program of Guangdong Key Laboratory of Aerospace Communication and Networking Technology under Grant 2018B030322004.

References

1. Kajur Renuka, V., Prasadm K,: Design and analysis of optimized CORDIC based GMSK system on FPGA platform. Int. J. Electric. Comput. Eng. **10**(5) (2020)
2. Renza, D., Ballesteros, D.M., Martinez, E.: Spreading-based voice encryption by means of OVSF codes. Appl. Sci. **10**(1), 112 (2019). https://doi.org/10.3390/app10010112
3. Sathiyamurthi, P., Ramakrishnan, S.: Speech encryption using chaotic shift keying for secured speech communication. EURASIP J. Audio Speech Music Process. **2017**(1), 1–11 (2017). https://doi.org/10.1186/s13636-017-0118-0
4. Bagwe, G.R., Apsingekar, D.S., Gandhare, S., Pawar, S.: Voice encryption and decryption in telecommunication. In: 2016 International Conference on Communication and Signal Processing, pp. 1790–1793 (2016)

5. Noorbasha, F., Divya, Y., Poojitha, M.: Koteswara Rao,K Hari Kishore. FPGA design and implementation of modified AES based encryption and decryption algorithm. Int. J. Innov. Technol. Explor. Eng. **8** (2019)
6. Strachacki, M., Szczepański, S.: Power equalization of AES FPGA implementation. Bull. Polish Acad. Sci. Techn. Sci. **58**(1) (2010). https://doi.org/10.2478/v10175-010-0013-7
7. Guo, F., Xi, L., Xing, G.: Design of low-power 2FSK demodulation circuit. MATEC Web Conf. **232**, 04069 (2018). https://doi.org/10.1051/matecconf/201823204069
8. Xie, L., Tang, M., Li, H.: Joint design of physical-layer network coding and LDPC code modulated by 2FSK. J. Phys. Conf. Ser. **1792**(1) (2021)
9. He, Z., Liu, Y., Ye, X.: A new method of image encryption/decryption via voice features. In: 2009 2nd International Congress on Image and Signal Processing, pp. 1–4. Tianjin, China (2009)
10. Malakooti, M.V., Dobuneh, M.R.N.: A lossless digital encryption system for multimedia using orthogonal transforms. In: 2012 Second International Conference on Digital Information and Communication Technology and It's Applications, pp. 240–244. Bangkok, Thailand. (2012)