Research Article Vol. 31, No. 3 / 30 Jan 2023 / Optics Express 4067
Optics EXPRESS

Speech signal enhancement based on deep learning in distributed acoustic sensing

YING SHANG,[1,4,*] JIAN YANG,[1,4] WANG CHEN,[1] JICHAO YI,[1] MAOCHENG SUN,[1] YUANKAI DU,[1] SHENG HUANG,[2] WENAN ZHAO,[1] SHUAI QU,[1] WEITAO WANG,[1] LEI LV,[1] SHUAI LIU,[1] YANJIE ZHAO,[3] AND JIASHENG NI[1]

[1]*Laser Institute, and International School for Optoelectronic Engineering, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250104, China*
[2]*Key Lab of In-Fiber Integrated Optics of Ministry of Education, Harbin Engineering University, Harbin 150001, China*
[3]*School of Science, Shandong Jianzhu University, Jinan 250101, China*
[4]*Equal contributors*
[*]*shangying@sdlaser.cn*

Abstract: The fidelity of a speech signal deteriorates severely in a distributed acoustic sensing (DAS) system due to the influence of the random noise. In order to improve the measurement accuracy, we have theoretically and experimentally compared and analyzed the performance of the speech signal with and without a recognition and reconstruction method-based deep learning technique. A complex convolution recurrent network (CCRN) algorithm based on complex spectral mapping is constructed to enhance the information identification of speech signals. Experimental results show that the random noise can be suppressed and the recognition capability of speech information can be strengthened by the proposed method. The random noise intensity of a speech signal collected by the DAS system is attenuated by approximately 20 dB and the average scale-invariant signal-to-distortion ratio (SI-SDR) is improved by 51.97 dB. Compared with other speech signal enhancement methods, the higher SI-SDR can be demonstrated by using the proposed method. It has been effective to accomplish high-fidelity and high-quality speech signal enhancement in the DAS system, which is a significant step toward a high-performance DAS system for practical applications.

1. Introduction

The acoustic monitoring technology in the seismic and marine fields has entered a phase of rapid development. Currently, infrasound monitoring is one of the common acoustic monitoring techniques used for seismic [1,2], while the marine sector is more likely to utilize sonar technology to monitor and locate the target [3]. However, infrasound and sonar technology are difficult to meet the needs of long distances and wide-range monitoring. Now, distributed acoustic sensing (DAS) technology has attracted significant interest throughout the world since it can achieve the quantitative analysis of acoustic signals [4–6]. This is made possible to realize acoustic monitoring over long distances and wide-range by utilizing the huge network of optical fibers already deployed over land and sea areas [7–9]. Some researchers have successfully taken near-surface snapshots of entire urban areas by connecting urban underground optical fiber cables through DAS technology [10], as well as successfully improving the accuracy of earthquake monitoring combined with deep learning techniques [11]. On the marine side, the vessel track can be obtained by monitoring acoustic signals in the 5.8-20 km sea area by using DAS technology [12].

The DAS technology and applications described above are for the monitoring of specific events through acoustic signals, while the identification and analysis of specific information from

#477175 https://doi.org/10.1364/OE.477175
Journal © 2023 Received 3 Oct 2022; revised 14 Dec 2022; accepted 28 Dec 2022; published 20 Jan 2023

Optics EXPRESS

speech signals by DAS technology have always been difficult. The main reason for this is that the energy of the Rayleigh back-scattered signal is extremely weak and is inevitably affected by noise during transmission. The main noises in DAS systems include Gaussian white noise in the demodulated phase caused by laser phase noise (LPN), low frequency fluctuations in the differential phase caused by laser frequency drift (LFD), and uniform white noise generated by the environment [13]. Through the medium of air and other media, the speech signal is continuously attenuated before it reaches the optical fiber. At this point, the speech signal received by the optical fiber will be very weak. Then, it is propagated by backward Rayleigh scattering. As a result, the demodulated speech signal is even weaker. Thus, the information in the speech signal is difficult to identify effectively due to the effects of noise. Generally, the study of speech signal enhancement is to weaken the noisy part, which can restore as pure a speech signal as possible and reduce the interference of noise. Traditional speech enhancement methods usually include spectral subtraction (SS) [14], empirical modal decomposition (EMD) [15], and subspace approach (SA) [16,17]. The noise information will be removed and a clean speech signal can be obtained. For traditional speech enhancement methods, noise features can be well estimated without prior knowledge to enhance speech signals in real time. However, it has poor ability to suppress non-stationary noise is the main reason why it is replaced by mainstream speech enhancement algorithms. On the contrary, deep-learning-based neural networks rely on their powerful fitting capabilities to suppress non-smooth noise well enough to achieve speech enhancement. Thus, deep-learning-based approaches hold promise for complete segmental speech enhancement and information recognition in DAS. Recently, the enhancement of acoustic signals by using convolutional neural network (CNN) is researched and digital pronunciation can be recovered with high fidelity in the DAS system [13]. However, it does not achieve speech enhancement and information recognition for complete segments. This is mainly due to the weak ability of CNN to learn prediction of speech context and information recognition. Moreover, the enhancement and information recognition of complete segments of speech signals need to be combined with the complete semantic environment.

In this paper, a speech signal recognition and reconstruction method based on deep learning technique is proposed to resolve the above shortcomings in the DAS system. An end-to-end complex convolution recurrent network (CCRN) is trained using the simulated speech signal and the real speech signal. The CCRN combines the powerful feature extraction and learning capability of RNN for the contextual relationship of speech information. Experimental results verify that the speech signal can be enhanced greatly. The real speech signal measured by the DAS system is successfully attenuated by 20 dB of noise and the scale-invariant signal-to-distortion ratio (SI-SDR) is improved by 51.97 dB, respectively. Thus, the proposed scheme provides a promising solution to monitor speech information in distributed sensing system.

2. Principle

2.1. Principle of CCRN

The signal under test process based on CCRN is demonstrated in Fig. 1. A complex spectral mapping-based CCRN is built to enhance the speech signal [18,19]. The phase information is extracted by demodulation algorithm according to the located position. Then, during the training period, both the phase signal and original speech signal are supplied into the network in sequence form. After a short-time Fourier transform (STFT), the complex spectrum of the noisy and original speech signal is used as input and training target, respectively. In order to estimate the nonlinear mapping between noisy and original speech signals, a code-and-decode (CED) method is used to map the complex spectrum of the noisy speech signal to that of the original speech signal. The best network model is obtained by training with a large amount of data to enhance the DAS speech signal. In the test phase, the original speech signal in the blue dashed box in

Optics EXPRESS

Fig. 1 will no longer be required. The enhancement of the DAS speech signal can be achieved with the trained network model.

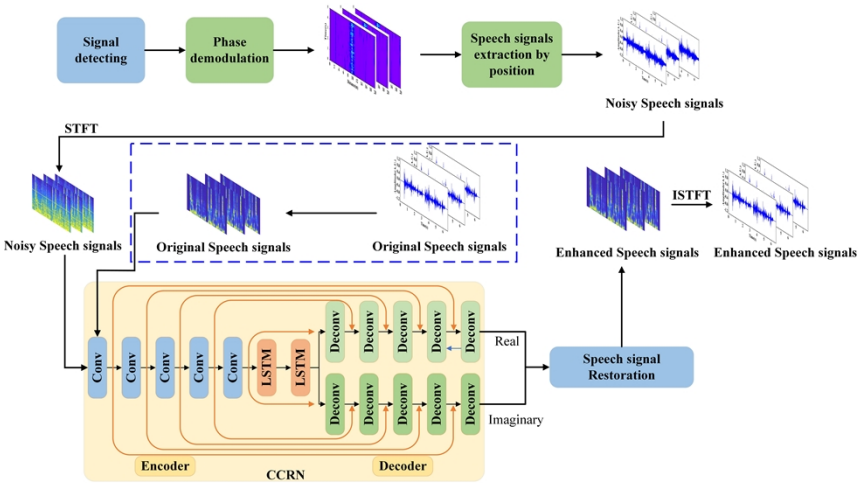

Fig. 1. The data processing flow chart of the CCRN. (The original speech signals in the blue dashed box are used for training and are not required for testing.)

For CCRN, the encoder and decoder consist of five convolutional and five deconvolutional layers, respectively. This symmetrical structure not only ensures that the input and output have the same shape, but also ensures that as many features as possible are extracted and that the loss of information is minimized in the output. Compared with the rectified linear unit (ReLU), the quicker convergence and better generalization ability can be offered by the exponential linear unit (ELU) [19,20]. This research uses ELU for all convolutional and deconvolutional layers since it is crucial for the enhanced processing of speech signals collected by the DAS system in complex situations. In general, the encoder-decoder structure is made up by the convolution-deconvolution layer, batch normalization and activation function. In the meantime, in order to enhance the movement of gradients and information, a skip connection is also added to the convolution and deconvolution layers [21].

Furthermore, compared with amplitude spectra, the quality of DAS data can be preserved better by the complex spectral mapping [18,21]. The purpose of complex spectral mapping is to extract the real and imaginary spectra of a clean speech signal from a noisy speech signal, while improving both the amplitude and phase response of the speech signal. The noisy speech signals could be written as:

$$y = x + n, \tag{1}$$

where y is noisy speech signal, x is clean speech signal, n is noise information.

Then magnitude spectral mapping is doing a STFT on both sides as follows:

$$Y = X + N, \tag{2}$$

where Y, X and N represent the STFT of y, x and n, respectively.

The role of complex spectral mapping is to apply the STFT representation of STFT to the rectangular coordinate system, thus, Eq. (2) can be represented as:

$$(Y_r + Y_i) = (X_r + N_r) + i(X_i + N_i), \tag{3}$$

where r and i represent the real part and the imaginary part, respectively.

Optics EXPRESS

During training, the model is used to estimate the complex ideal ratio mask (CIRM), the process is given as follows [18]:

$$\hat{M} = \hat{M}_r + i\hat{M}_i = \frac{Y_r X_r + Y_i X_i}{(Y_r)^2 + (Y_i)^2} + i\frac{Y_r X_r - Y_i X_i}{(Y_r)^2 + (Y_i)^2}, \tag{4}$$

By estimating $\hat{M}$ from the CIRM for the noisy spectrum, the improved spectrum can be utilized to reconstruct the clean speech information, as follows:

$$X = (\hat{M}_r \times Y_r) + i(\hat{M}_i \times Y_i). \tag{5}$$

Finally, the enhanced speech signal can be obtained after combining the improved real and imaginary spectra and performing an inverse STFT.

2.2. Performance of simulation

Simulations are performed to ensure that the speech signal obtained by DAS system can be successfully enhanced by the proposed method. Firstly, several different methods including CRN [22], SS [13] and SA [15,16] are used as the comparison. The original speech signal is synthesized with the simulated noise data from DAS to create the simulated speech signal [17]. Table 1 shows the number of samples used for the simulation, including 5000 speech signals for the training set, 1300 speech signals for the validation set, and 10 different types of speech signals used in the test set.

Table 1. SI-SDR improvement of simulated speech signal

Datasets	The number of examples
The training sets	5000
The validation sets	1300
The test sets	10

The spectrogram by STFT of noisy, original, and enhanced speech signals are shown in Fig. 2(a), 2(b), and 2(c). The red, white, and black boxes shown in Fig. 2(a) represent the various types of noise of the speech signal, respectively. The enhanced speech spectrogram processed by the proposed method is given in Fig. 2(c). It can be found that the noise can be effectively removed and the speech features that are drowned in the noise could be well reconstructed compared with the original speech signal shown in Fig. 2(b).

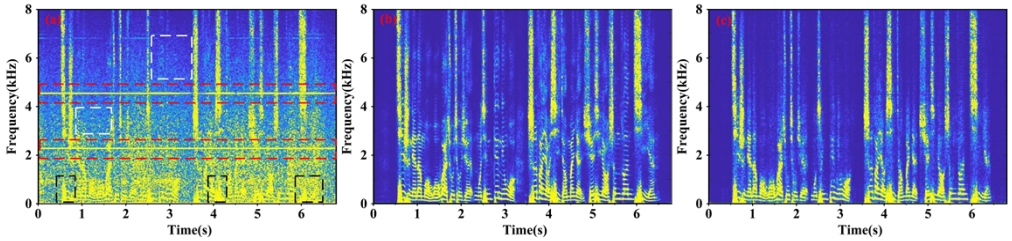

Fig. 2. Spectrogram results of simulation data based on the proposed method. (a) Noisy speech signal. (b) original speech signal. (c) Enhanced speech signal.

Next, to further test the enhancement capability of CCRN on simulated speech signals, Fig. 3(a) and 3(b) respectively show the time and frequency domain comparisons of the simulated signal. The noisy, original, and enhanced speech signals are shown in grey, red, and blue, respectively. It can be found from Fig. 3(a) that the original speech signal is almost identical to the enhanced

Optics EXPRESS

speech signal by CCRN. Moreover, compared with the noisy speech signal in Fig. 3(b), the noise intensity of the enhanced speech signal is decreased by approximately 15 dB. These results show that the proposed method can successfully enhance simulated speech signals and improve recognition of speech messages.

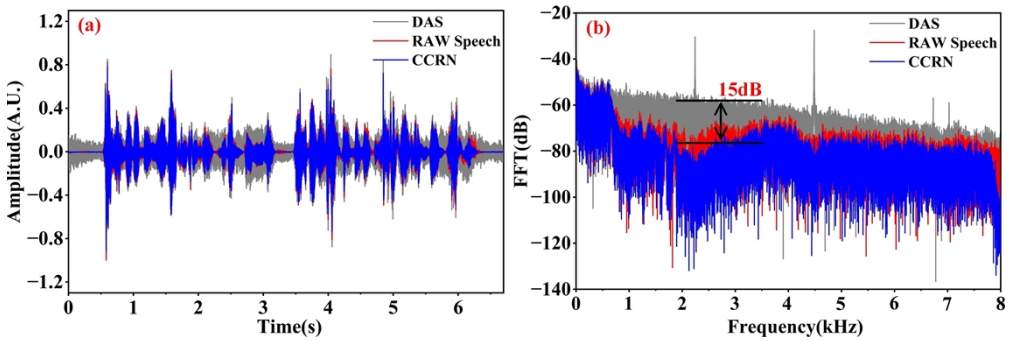

Fig. 3. Time-frequency domain comparison between noisy, original, and enhanced speech signals. (a) Time domain comparison between noisy, original, and enhanced speech signals. (b) Frequency domain comparison between noisy, original, and enhanced speech signals.

Moreover, to further verify the ability of different algorithms to reconstruct the waveform of a speech signal, we have compared part of the waveform before and after enhancement. The simulated speech signals before and after enhancement by these methods are shown in Fig. 4(a), 4(b), 4(c), and 4(d). The results show that all four enhanced methods can be successfully reconstructed to obtain an enhanced speech signal which is similar to the original speech signal in the time domain. But the reconstructed speech signal by CCRN have the best matching lever with the original speech signal.

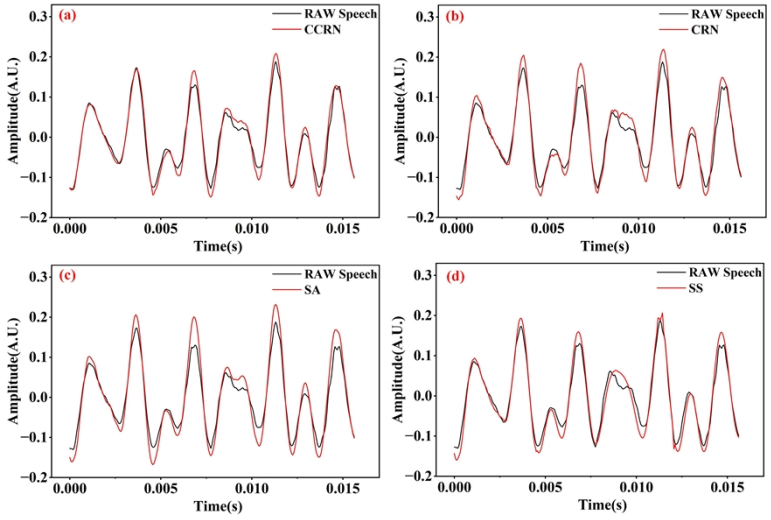

Fig. 4. Time-domain waveform of simulated speech signal collected by the DAS system enhanced by different methods. (a) CCRN. (b) CRN. (c) Spectral subtraction. (d) Subspace approach.

In order to quantitatively assess their errors, MSE in Eq. (6) is used to calculate the difference between original and enhanced speech signals [23]. It can be found from Fig. 5 that the MSE of

Optics EXPRESS

the enhanced speech signals by CCRN, CRN, SS, and SA are shown in black, red, green, and blue boxes, respectively. The results show that the proposed method leads to the lowest difference and the highest for the subspace approach. It means that CCRN can effectively reduce the error between the original and the enhanced speech signal. This allows the enhanced speech signal to be reconstructed with high fidelity. MSE is defined as

$$MSE = \frac{\sum_{i-1}^{r} (n_i - 1)s_i^2}{N - r}. \tag{6}$$

where N represents the total number of data, r represents the number of N data into groups, s_i^2 represents the variance of the sample, that is, the numerator is the sum of squares of errors, and the denominator is the degrees of freedom.

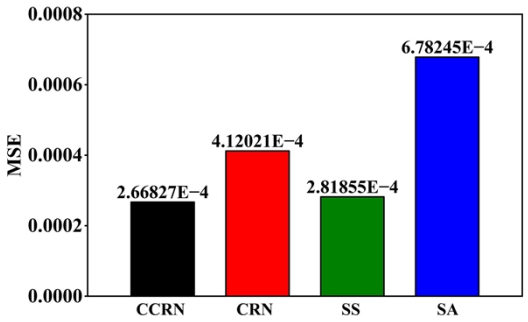

Fig. 5. Results of MSE.

Finally, after the above comparative analysis, in order to quantitatively evaluate the speech signal fidelity and distortion, the SI-SDR of ten different simulated speech signals with respect to the original speech signals is used to be calculated in Fig. 6 [24]. The SI-SDR of the noisy speech signal and enhanced speech signal by CCRN, CRN, SS, and SA are shown in black, red, green, blue, and orange lines respectively. It can be found that the enhanced result by CCRN is the best. The average SI-SDR of the reconstructed speech signal by these methods is improved by 19.1 dB, 17.7 dB, 5.8 dB, and 9.3 dB respectively are shown in Table 2. This indicates that high-fidelity simulated speech signals can be successfully reconstructed by the proposed method. It also shows that the proposed enhancement method has great robustness and stability. SI-SDR is defined as

$$SI - SDR = 10\log_{10}\left(\frac{\left\|e_{t\arg et}\right\|^2}{\left\|e_{res}\right\|^2}\right) = 10\log_{10}\left(\frac{\left\|\frac{\hat{x}^T x}{\|x\|^2}x\right\|^2}{\left\|\frac{\hat{x}^T x}{\|x\|^2}x - \hat{x}\right\|^2}\right). \tag{7}$$

where $\hat{x}$ refers to enhanced speech, x refers to pure speech without noise.

Table 2. SI-SDR improvement of simulated speech signal

Methods	CCRN	CRN	SS	SA
Average improvement(dB)	**19.1**	17.7	5.8	9.3

Optics EXPRESS

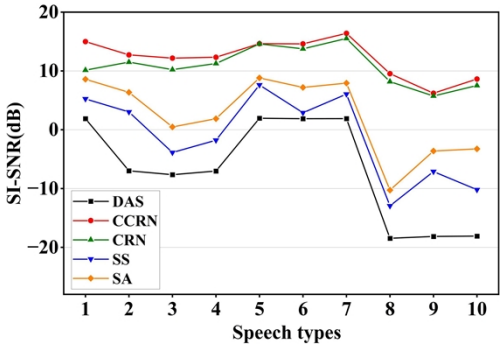

Fig. 6. SI-SDR results of 10 different simulation speech signals using different methods.

3. Experimental setup and results

3.1. Datasets creation

An accurate and high-quality dataset is essential for the enhancement of speech signals in the deep learning approach, which determines the performance of the model. However, it is extremely challenging to manually align the original speech signal with the speech signal collected by the DAS system or to obtain a speech signal without background noise. Hence, a four-channel acquisition solution is used to solve the above problems. The setup of the experiments is shown in Fig. 7. The light from the narrow linewidth laser (NLL) is modulated by the AOM, which modulates the continuous light into pulsed light. The pulsed light is amplified by the front erbium-doped fiber amplifier (EDFA) and then passes through the circulator (Cir) into the fiber under test. The backward Rayleigh scattering signal from the fiber under test then enters the other EDFA via the circulator to be amplified. The amplified backscattered Rayleigh signal then enters the Mach-Zehnder interferometer and split into three paths after passing through the optical coupler (OC) and into three photodetectors (PD). Finally, the optical signal is converted into an electrical signal, which is captured by an acquisition card and sent to the computer for processing. Three of the channels are used to demodulate the voice data in the 3×3 algorithm, and one channel is connected directly to the loudspeaker to obtain the original speech signals. Although there are still many differences with the speech signal acquired by the DAS system, this data should be useful for training. Furthermore, the original speech signals are used as a reference in subsequent quantitative analysis, which can avoid the error of manual alignment operation. The experimental settings of DAS are summarized in Table 3. At 2.4 km of optical fiber, a loudspeaker is used to play the audio messages from the speech dataset THCHS-30 [25].

3.2. Experimental process

Figure 8 demonstrates the specific data processing scheme. Firstly, 20,000 speech signals are played by loudspeaker and captured by the DAS system to create training, validation, and test sets. Then, a non-linear mapping training is performed with 16,000 speech signals to update the parameters of the network. An additional 4000 speech signals are used for validation to optimize the network parameters. Next, 100 speech signals are used to create test set. The test data are enhanced and reconstructed by the trained CCRN. Finally, test results are used for analysis and discussion.

Table 4 demonstrates the detailed parameters initialization of the CCNN training process. The weight, bias and batch size are initialized to 1, 0 and 16 respectively. The learning rate of CCRN model is initialized to 0.001, and it decays by 0.9 for every 2 epochs. In an 8:2 ratio, 20000 speech signals are added to the training and validation sets, and 100 speech signals are used to be

Optics EXPRESS

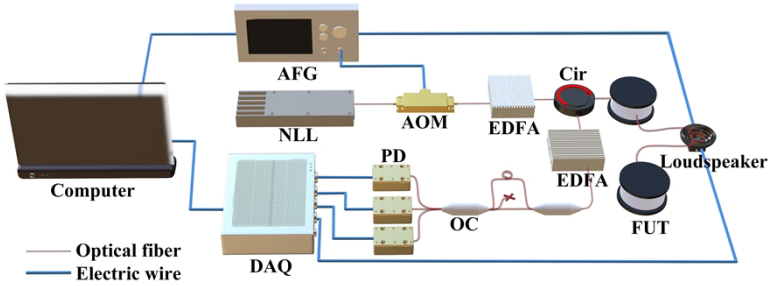

Fig. 7. Experimental setup. NLL, narrow line-width laser (NKT C15); AFG, Arbitrary Function Generator; AOM, acousto-optic modulator (Gooch & Housego); EDFA, erbium-doped fiber amplifier (Beogold Technology); Cir, circulator; PD, photodetector (Thorlabs FPD510); DAQ, data acquisition card; OC, optical coupler; FUT, fiber under test.

Table 3. The experimental settings of DAS

Parameters	Setting Values
Sampling frequency of DAQ card	100MHz
Scanning frequency	16kHz
AOM frequency shift	200MHz
Pulse width	50ns
Spatial resolution	10m
The bandwidth of the photodetector	200MHz
The bandwidth of DAQ card	250MHz
The input Peak power into the fiber	125mw
The linewidth of NLL	<15kHz
The length of optical fiber	5km

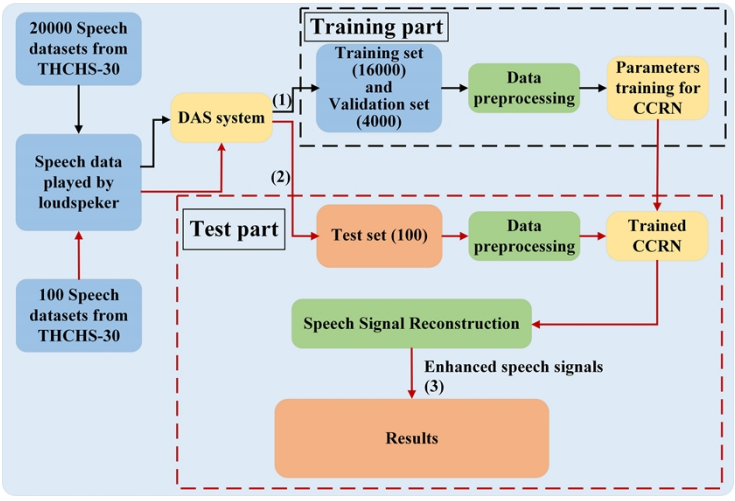

Fig. 8. The processing scheme of speech signal.

Optics EXPRESS

test sets. The training sets is used to train the model and the validation sets is used to find the optimal model. In order to obtain the optimal model for the real speech signal, the period of cross validation is initialized to 300 and training for 150 epochs. The entire training process took 163 hours with NVIDIA GeForce RTX 3090 GPU calculations.

Table 4. Parameters initialization of CCRN

Parameters	Initialized Values
Weight Initialization	1
Bias Initialization	0
Batch Size	16
Epochs	150
Learning Rate	0.001(decay 0.9/2 epochs)
The period of cross validation	300
Number of Training Sets	16000
Number of Validation Sets	4000
Number of Test Sets	100
The time period of each sample	10s

3.3. Experimental results and analyzation

In this section, a speech signal of 10s is used to assess the performance of the proposed enhanced scheme based on CCRN. Firstly, the spectrogram by STFT of noisy, original, and enhanced speech signals are shown in Fig. 9(a), 9(b), and 9(c). The red, white, and black boxes shown in Fig. 9(a) represent the various types of random noise respectively. The quality of the noisy speech signal decreases a lot due to the channel frequency response which can be seen in Fig. 9(a). The voiceprint features of the speech signal are drowned in noise. Compared with Fig. 9(a) and 9(b), Fig. 9(c) shows that the noise can be successfully suppressed and the speech features can be well reconstructed by the proposed method which means the proposed method enables the enhancement of speech signals acquired by DAS system to improve speech recognition.

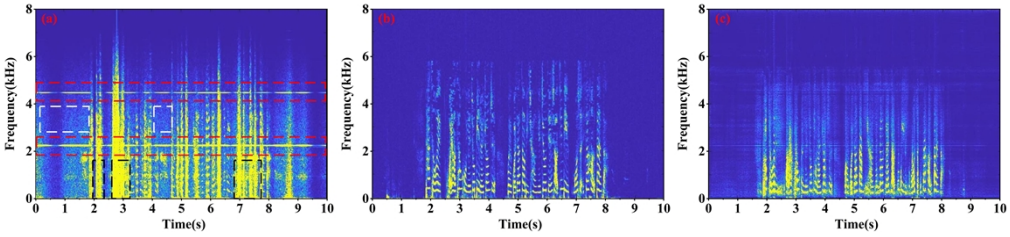

Fig. 9. Spectrogram results based on the proposed method. (a) Noisy speech signal. (b) original speech signal. (c) Enhanced speech signal.

Next, to further verify the performance of the proposed method, we compared the speech signal in the time domain with the frequency domain which are shown in Fig. 10. As can be seen in Fig. 10(a) and 10(b), the grey, red, and blue waveform represent noisy, original, and enhanced speech signals. It can be found from Fig. 10(a) that the enhanced speech signal can maintain high consistency with the original speech signal in the time domain. The noise of the enhanced speech signal is decreased by approximately 20 dB is shown in Fig. 10(b). It illustrates that the proposed method still works for the real speech signal, and the noise from the DAS system is well suppressed.

Optics EXPRESS

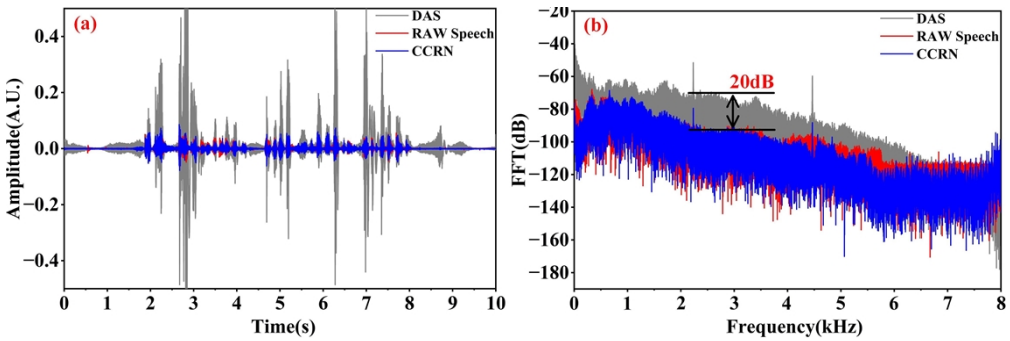

Fig. 10. Time-frequency domain comparison between noisy, original, and enhanced speech signals. (a) Time domain comparison between noisy, original, and enhanced speech signals. (b) Frequency domain comparison between noisy, original, and enhanced speech signals.

Moreover, in order to verify that the proposed method has the same effect on speechless signals. The signal at the 2 km optical fiber without speech is enhanced. Figure 11 demonstrates the spectrograms result for silence positions. There are lots of noise in the silence position before enhancement which is shown in Fig. 11(a). After enhancement, the noise is effectively reduced by the proposed method is shown in Fig. 11(b). This result indicates that the reliability of the proposed enhancement method in terms of noise suppression.

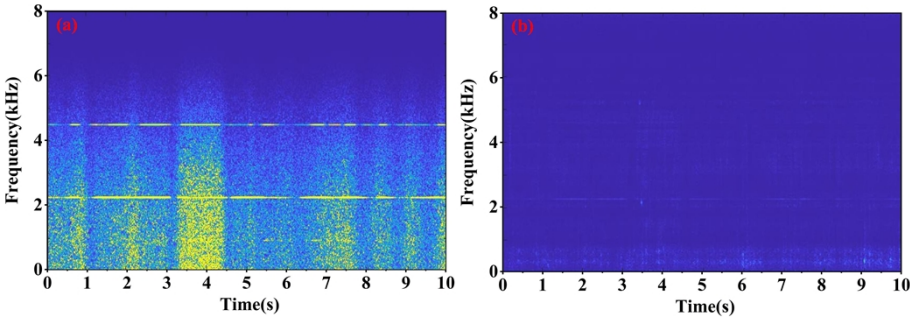

Fig. 11. Spectrograms for silent position. (a) Spectrogram before enhancement. (b) Spectrogram after enhancement.

3.4. Compared with other methods

In order to evaluate the performance of the proposed method in waveform reconstruction, a comparison with other methods was performed. As can be seen in Fig. 12, the black, red, green, and blue boxes represent the average MSE of enhanced speech signals by these methods. It can be found that the lowest MSE are achieved by the proposed method. Therefore, the result proves that CCRN has good robustness and stability, its success in enhancing the speech signal of a new speaker also proves that it has good generalization properties.

Table 5 demonstrates that the average SI-SDR of the reconstructed speech signal by the proposed method, CRN, SS, and SA is improved by 51.97 dB, 47.86 dB, 8.24 dB, and 5.39 dB respectively. These results show that CCRN can effectively accomplish high-fidelity and high-quality speech signal enhancement in the DAS system. And the proposed method can well eliminate the non-stationary background noise rather than simply learning and reproducing the results of the training sets.

Optics EXPRESS

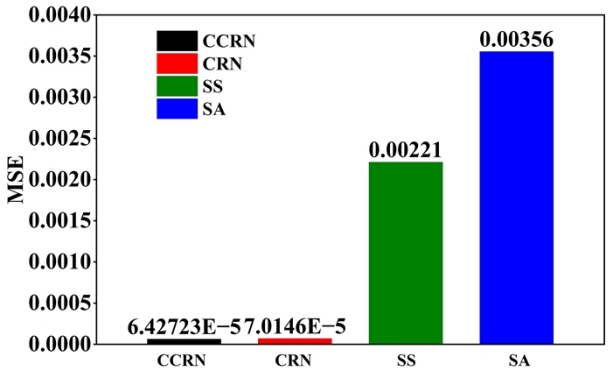

Fig. 12. Average MSE of speech sensing tests using different methods.

Table 5. Average SI-SDR improvement of speech signals collected by the DAS system

Methods	CCRN	CRN	SS	SA
Average improvement(dB)	**51.97**	47.86	8.24	5.39

The average testing time of the four methods for speech enhancement by the DAS system is shown in Table 6. The average test times for testing a sample by the proposed method, CRN, SS, and SA are 20 ms, 18 ms, 42 ms, and 31 ms respectively. The proposed method tests a sample in a faster average time than the traditional method but is slower than CRN due to the addition of the phase part information processing. The results show that the proposed method can achieve real-time enhancement of the speech signal from the DAS system and obtain the best enhancement results.

Table 6. Evaluation index of the methods

Method	Testing time
CCRN	20ms
CRN	18ms
SS	42ms
SA	31ms

3.5. Validation of generalization

In order to further validate the generalization of the proposed method, 100 different speech signals are tested via the DAS system. Both SI-SDR and SNR are used to evaluate the performance of the test speech signals. The test speech signals were treated as high enhancement levels when the SI-SDR increase is larger than 20 dB and the SNR increase is larger than 10 dB. Speech signals with SI-SDR increase greater than 0 dB and less than 20 dB and SNR increase greater than 0 dB and less than 10 dB were treated as low enhancement levels. When there is no improvement in SI-SDR or SNR, the test speech signal is treated as an unenhanced sample. The test results are shown in Table 7 which can be found that 97% of the speech signal can be enhanced, including 95% high enhancement level and 2% low enhancement level. In addition, 3% cannot be enhanced by the proposed method and it is an acceptable error range of the deep learning technique. Therefore, experimental results show that the proposed method has the information recognition capability and great generalization, enabling to improve the performance of speech signals in the DAS system.

Optics EXPRESS

Table 7. Test results of 100 speech signals

Enhancement performance	Number	Results
High level	95	95%
Low level	2	2%
Unenhanced	3	3%

4. Conclusion

In this paper, an end-to-end CCRN is used to enhance the speech signal collected by the DAS system. The performance of speech signals with and without recognition and reconstruction methods based on deep learning techniques has been theoretically and experimentally compared. Experimental results show that the random noise intensity of speech signal collected by the DAS system is decreased by approximately 20 dB and the average SI-SDR is improved by 51.97 dB. The proposed method can achieve high-fidelity and high-quality speech signal enhancement and improve the recognition of speech information in the DAS system.

Funding. Innovation project of Computer Science and Technology of Qilu University of Technology (2021JC02006); Innovation Project of Science and Technology SMES in Shandong Province (2022TSGC2049); Colleges and Universities Youth Talent Promotion Program of Shandong Province (Precision Instrument Science and Technology Innovation Team); Supported by the Taishan Scholars Program; Science, education and industry integration innovation pilot project of Qilu University of Technology (2022PX002, 2022PY008); Joint Natural Science Foundation of Shandong Province (2021KJ049, ZR2021LLZ014); Key R&D Program of Shandong Province (Major Technological Innovation Project) (2021CXGC010704); Colleges and Universities Youth Innovation and Technology Support Program of Shandong Province (2019KJJ004); Natural Science Foundation of Shandong Province (ZR2020LLZ010, ZR2020QF092); National Natural Science Foundation of China (62005137).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. D. Fee, L. Toney, K. Kim, R. W. Sanderson, A. M. Lezzi, R. S. Matoza, S. D. Angelis, A. D. Jolly, J. J. Lyons, and M. M. Haney, "Local explosion detection and infrasound localization by reverse time migration using 3-D finite-difference wave propagation," Front. Earth Sci. **9**(9), 620813 (2021).
2. F. K. D. Dugick, P. S. Blom, B. W. Stump, C. T. Hayward, S. J. Arrowsmith, J. C. Carmichael, and O. E. Marcillo, "Evaluating the location capabilities of a regional infrasonic network in Utah, US, using both ray tracing-derived and empirical-derived celerity-range and backazimuth models," Geophys. J. Int. **229**(3), 2133–2146 (2022).
3. X. Pan, Z. Liu, P. Zhang, Y. Shen, and J. Qiu, "Distributed MIMO sonar for detection of moving targets in shallow sea environments," Appl. Acoust. **185**, 108366 (2022).
4. P. Jousset, G. Currenti, B. Schwarz, A. Chalari, F. Tilmann, T. Reinsch, L. Zuccarello, E. Privitera, and C. M. Krawczyk, "Fibre optic distributed acoustic sensing of volcanic events," Nat. Commun. **13**(1), 1753 (2022).
5. Q. Liu, T. Liu, T. He, H. Li, Z. Yan, L. Zhang, and Q. Sun, "High resolution and large sensing range liquid level measurement using phase-sensitive optic distributed sensor," Opt. Express **29**(8), 11538–11547 (2021).
6. D. Chen, Q. Liu, and Z. He, "Distributed Fiber-optic Acoustic Sensor with Long Sensing Range over 100 km and Sub-nano Strain Resolution," J. Lightwave Technol. **37**(18), 4462–4468 (2019).
7. C. Fan, H. Li, T. He, S. Zhang, B. Yan, Z. Yan, and Q. Sun, "Large dynamic range optical fiber distributed acoustic sensing (DAS) with differential-unwrapping-integral algorithm," J. Lightwave Technol. **39**(22), 7274–7280 (2021).
8. X. Chen, N. Zou, L. Liang, R. He, J. Liu, Y. Zheng, F. Wang, X. Zhang, and Y. Zhang, "Submarine cable monitoring system based on enhanced COTDR with simultaneous loss measurement and vibration monitoring ability," Opt. Express **29**(9), 13115–13128 (2021).
9. J. Xiong, Z. Wang, J. Jiang, B. Han, and Y. Rao, "High sensitivity and large measurable range distributed acoustic sensing with Rayleigh-enhanced fiber," Opt. Lett. **46**(11), 2569–2572 (2021).
10. Z. Song, X. Zeng, B. Wang, J. Yang, X. Li, and H. F. Wang, "Distributed Acoustic Sensing Using a Large-Volume Airgun Source and Internet Fiber in an Urban Area," Seismol. Res. Lett. **92**(3), 1950–1960 (2021).
11. P. D. Hernández, J. A. Ramírez, and M. A. Soto, "Deep-Learning-Based Earthquake Detection for Fiber-Optic Distributed Acoustic Sensing," J. Lightwave Technol. **40**(8), 2639–2650 (2022).
12. D. Rivet, B. Cacqueray, A. Sladen, A. Roques, and G. Calbris, "Preliminary assessment of ship detection and trajectory evaluation using distributed acoustic sensing on an optical fiber telecom cable," J. Acoustical Soc. of America **149**, 2615 (2021).

Optics EXPRESS

13. F. Jiang, Z. Zhang, Z. Lu, H. Li, Y. Tian, Y. Zhang, and X. Zhang, "High-fidelity acoustic signal enhancement for phase-OTDR using supervised learning," Opt. Express **29**(21), 33467–33480 (2021).
14. S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process. **27**(2), 113–120 (1979).
15. B. Xue, H. Hong, S. Zhou, G. Chen, Y. Li, Z. Wang, and X. Zhu, "Morphological Filtering Enhanced Empirical Wavelet Transform for Mode Decomposition," IEEE Access **7**, 14283–14293 (2019).
16. S. Surendran and T. K. Kumar, "Oblique Projection and Cepstral Subtraction in Signal Subspace Speech Enhancement for Colored Noise Reduction," IEEE/ACM Trans. Audio Speech Lang. Process. **26**(12), 2328–2340 (2018).
17. Y. Ephraim and H. L. V. Trees, "A signal subspace approach for speech enhancement," IEEE Trans. Speech Audio Process. **3**(4), 251–266 (1995).
18. D. S. Williamson, Y. Wang, and D. Wang, "Complex Ratio Masking for Monaural Speech Separation," IEEE/ACM Trans. Audio Speech Lang. Process. **24**(3), 483–492 (2016).
19. K. Tan and D. Wang, "Learning Complex Spectral Mapping With Gated Convolutional Recurrent Networks for Monaural Speech Enhancement," IEEE/ACM Trans. Audio Speech Lang. Process. **28**, 380–390 (2020).
20. P. Wang, K. Tan, and D. Wang, "Bridging the Gap Between Monaural Speech Enhancement and Recognition With Distortion-Independent Acoustic Modeling," IEEE/ACM Trans. Audio Speech Lang. Process. **28**, 39–48 (2020).
21. W. Yuan, "A time–frequency smoothing neural network for speech enhancement," Speech Commun. **124**, 75–84 (2020).
22. X. Le, T. Lei, K. Chen, and J. Lu, "Inference Skipping for More Efficient Real-Time Speech Enhancement With Parallel RNNs," IEEE/ACM Trans. Audio Speech Lang. Process. **30**, 2411–2421 (2022).
23. H. Marmolin, "Subjective MSE Measures," IEEE Trans. Syst., Man, Cybern. **16**(3), 486–489 (1986).
24. J. L. Roux, S. Wisdom, H. Erdogan, and J. R. Hershey, "SDR – Half-baked or Well Done?" *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 626–630 (2019).
25. D. Wang, X. Zhang, and Z. Zhang, "THCHS-30: A Free Chinese Speech Corpus," (2015).