IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 5 August 2022, accepted 26 August 2022, date of publication 29 August 2022, date of current version 21 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3202971

RESEARCH ARTICLE

94 GHz Asymmetric Antenna Radar for Speech Signal Detection and Enhancement via Variational Mode Decomposition and Improved Threshold Strategy

FUMING CHEN[1,2], JIANQI WANG[3], AND CHUANTAO LI[4]
[1]The 940th Hospital of Joint Logistic Support Force of Chinese PLA, Lanzhou 730050, China
[2]College of Information Engineering, Gansu University of Chinese Medicine, Lanzhou 730000, China
[3]Department of Biomedical Engineering, Fourth Military Medical University, Xi'an 710032, China
[4]Department of Aviation Physiological and Psychological, Naval Special Medical Center of PLA, Second Military Medical University, Shanghai 200433, China

Corresponding authors: Jianqi Wang (wangjq@fmmu.edu.cn) and Chuantao Li (lichuantao614@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61901515 and in part by the Technology Incubation Fund of Second Military University under Grant 21X0201.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Committee on Ethics of Medicine, Second Military Medical University, China.

ABSTRACT To further improve the detection distance and sensitivity of bio-radar, a 94 GHz asymmetric antenna radar sensor is employed to detect speech signal. However, the radar speech is often mixed with various noise, which will seriously affect the quality and intelligibility of the speech signal. Therefore, a novel method based on variational mode decomposition (VMD) and improved threshold strategy (ITS) is proposed in this paper for improving the quality and intelligibility of the radar speech. VMD is a novel adaptive decomposition method, which overcomes the problem of mode aliasing and end effect in empirical mode decomposition (EMD). ITS can overcome the limitation of traditional wavelet threshold and achieve the best compromise between speech intelligibility and noise reduction. Firstly, EMD is applied to determine the number of decomposition level, and then radar speech is decomposed into several limited bandwidth intrinsic mode functions by VMD. Secondly, ITS is employed to remove noise from useful modes which are determined by Pearson correlation coefficient (PCC). The performance of the proposed method is evaluated by perceptual evaluation of speech quality (PESQ), short-time objective intelligibility (STOI) and composite measures (CMs). The experimental results show that the radar sensor can detect long distance speech signal and the proposed method can effectively improve the quality and intelligibility of the radar speech signal. Due to the good performance, the proposed method will provide a promising alternative for various applications related to radar speech and traditional microphone speech signal enhancement.

INDEX TERMS 94 GHz asymmetric antenna, radar speech, speech enhancement, variational mode decomposition, improved threshold strategy, composite measure.

I. INTRODUCTION

Speech, as one of the important physiological signal of the human body, is the most important and effective means of human communication. At present, the present technologies for obtaining speech signals can be divided into air conduction and non-air conduction detection. However, the shortcomings limit the development of these techniques for speech detection [1], [2], [3], [4], [5]. In recent years, bio-radar technology has been developed for using in a variety of remote sensing applications [6], [7]. This article mainly focuses on the application of bio-radar technology in human speech

The associate editor coordinating the review of this manuscript and approving it for publication was Hasan S. Mir.

signal detection to overcome the drawbacks of the existing speech detection technologies. Unlike traditional speech detection technology, the principle of the speech detection technology is that when the electromagnetic wave signal transmitted by bio-radar, reaches the throat of the human body, it is modulated and reflected by the vibration information of the throat. The reflected signal is received by the receive antenna, and is then sent to the computer for processing to form the speech signals [8], [9], [10].

Nowadays, the human speech signal detection based on bio-radar has certain drawbacks such as short detection range and low intelligibility. Research shows that higher transmission frequencies result in higher detection sensitivity, which improves the intelligibility; however, this will affect the detection range. Very high transmission frequencies will also cause the technology to be very expensive. A bio-radar with a transmission frequency of 94 GHz can provide the best compromise between detection range and detection sensitivity [11]. If we use a small-diameter wide-beam radiation, although it will be easy to aim at the target, the lower antenna gain will limit its detection range. If the antenna gain is increased, it will lead to an increase in the antenna diameter, narrow radiation beam, increased cost, and difficulty in aiming at the detection target. For this reason, some compensation technology must be considered.

In order to solve the above problems, a research on human speech detection technology based on a 94 GHz asymmetric antenna bio-radar was carried out in this study. The purpose of using asymmetric antennas is to minimize the cost of the hardware and to maximize the acquisition detection distance and detection sensitivity. The structure of the asymmetric antenna bio-radar system is as follows: A Cassegrain antenna with a diameter of 600 mm, which has a very high gain and the ability to obtain speech information, is used in the receiver module to overcome the problem of insufficient intelligibility. A Cassegrain antenna with a diameter of 200 mm, which has a relatively wide beam radiation target, is used in the transmitter module to overcome the problem of the narrow beam being difficult to aim at the target. The Schematic of the 94 GHz asymmetric antenna bio-radar is shown Figure 1. During the detection experiment, it was shown that the radar system can effectively detect the long-distance speech signal, and considerably extend the capabilities of traditional speech detection methods [12]. It was also observed that the radar speech signal is disturbed by non-Gaussian additive noise, which is mainly combined by ambient, electromagnetic and electrical circuit noise. This issue may lead to the performance degradation of the radar system for speech signal detection. Due to the special characteristics of the radar speech signal, it is required to apply appropriate enhancement algorithm to improve the quality of the 94 GHz asymmetric antenna bio-radar speech signal.

During the last decades, many speech enhancement algorithms have been proposed to improve the quality of speech signal by reducing noise, including the power spectral subtraction [13], Kalman filtering [14], Wiener filtering [15], wavelet shrinkage [16], minimum mean-square estimator (MMSE) [17]. Afterwards, various improved enhancement algorithms based on the aforementioned method are proposed [18], [19], [20], [21]. These methods promote the development of traditional microphone speech enhancement technology. For the radar speech, the noise sources of signal are more complex than those found in speech acquired using a traditional microphone. In our previous study, we have researched many speech signal enhancement methods aimed at different radar systems. These methods restrain the residual noise and improve the speech quality of the corresponding radar system [22], [23], [24], [25]. Although the noise of radar speech is greatly suppressed, the improvement of intelligibility is not obvious. Therefore, it is necessary find a new method aimed at improving intelligibility when reducing noise.

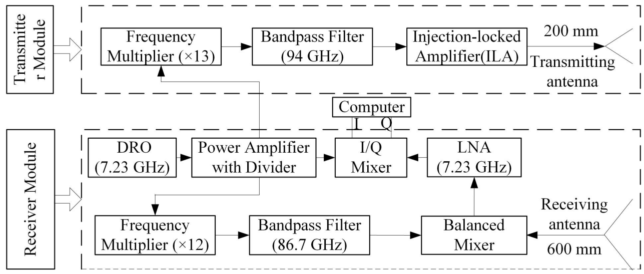

FIGURE 1. Schematic diagram of the 94 GHz asymmetric antenna bio-radar sensor system.

Empirical mode decomposition (EMD) method is an adaptive method for processing nonlinear and nonstationary signals [26]. Khaldi *et al.* used EMD method to enhance the quality of speech for the first time, the results showed that the method is effective for additive white Gaussian noise removal from speech [27]. Chatlani and Soraghan proposed a EMDF method for speech enhancement, it is particularly effective in low frequency noise environments [28]. Zao *et al.* proposed a EMD and Hurst-Based Mode Selection (EMDH) method for speech enhancement, the results show that the EMDH method improves the segmental signal-to-noise ratio and an overall quality composite measure [29]. However, shortcomings of mode aliasing and end effect restrict the development of EMD method in the field of speech enhancement [30], [31]. To overcome the limitations of EMD technique, a new adaptive signal processing method, called variational mode decomposition (VMD) was proposed, this method provides a solution to the decomposition problem that is theoretically well founded and easy to understand, it can decompose complex signal into an ensemble of band-limited intrinsic mode functions (IMFs). The experiments show that VMD method outperforms EMD with regards to tone detection, tone separation, and noise robustness [32]. In recent years, VMD method has been widely used in various fields such as fault diagnosis, underwater acoustic, and signal denoising. Mohanty *et al.* [33] utilized VMD method to decompose bearing vibration signal for fault diagnosis. In [34], VMD method is applied to detect multiple signatures caused by rotor-to-stator rubbing. In [35], a novel method for feature extraction of ship-radiated noise was proposed using VMD and slope entropy. Li *et al.* [36] developed a new denoising algorithm for ship-radiated noise (SN) signals based on VMD and correlation coefficient. An *et al.* [37] presented a denoising method for the hydropower unit vibration signal based on VMD and approximate entropy. These studies demonstrate that VMD approach is superior to traditional decomposition methods in signal processing. However, the VMD method also has its limitations, the capability of it to decompose a signals accurately is totally depends on the settings of its input parameters [38]. Eespecially for decomposition mode number and bandwidth control parameter. In practical applications, the two parameters are always given in advance based on previous experience, which may cause inaccurate analysis results. Therefore, the selection of appropriate parameters is crucial for VMD decomposition. Zhang *et al.* [39] proposed a parameter-adaptive VMD method based on grasshopper optimization algorithm (GOA) to analyze vibration signals from rotating machinery. Ni *et al.* [40] developed a novel bearing fault information-guided VMD (FIVMD) method to extract the weak bearing repetitive transients.

In the field of speech enhancement, Gowri *et al.* [41] utilized VMD based approach for enhancing speech signals distorted by white Gaussian noise. In [42], EMD and VMD method are compared for speech enhancement, the results show that the VMD is more efficient than EMD method in speech enhancement. Although these results verified the effectiveness of VMD method in speech enhancement, they are mainly focused on the application of the method in speech enhancement, instead of examining the performance of this method in improving the quality and intelligibility of speech. For radar speech, the decomposition method should also ensure the intelligibility of the speech when reducing noise [11]. If each IMF is filtered, we find that the noise is suppressed, the intelligibility of the radar speech is poor. Moreover, the traditional estimated noise level, is not accurate in estimating IMF noise, which will take effect on the quality of enhanced speech. Furthermore, although the soft thresholding was widely employed to remove the noise of signal, it may cause signal over-processing, resulting in speech signal distortion [27].

To address the above-mentioned issues, a novel method based on VMD and improved threshold strategy (ITS) is proposed for improving the quality and intelligibility of the 94 GHz asymmetric antenna bio-radar speech. VMD is a novel adaptive decomposition method, which overcomes the problem of mode aliasing and end effect in EMD, and is more suitable for dealing with nonlinear and non-stationary signals. ITS can overcome the limitation of traditional wavelet threshold and soft threshold function and achieve the best compromise between the intelligibility of radar speech and noise reduction. First, EMD method is used to determine the mode number by selecting the major information modes, and then VMD algorithm is used to decompose the 94 GHz asymmetric antenna radar speech signal into several limited bandwidth intrinsic mode functions (IMFs). Second, an improved threshold strategy (ITS) is employed to remove noise from useful modes which are determined by the Pearson correlation coefficient (PCC).

(a) (b)

FIGURE 2. Experimental environment of the 94 GHz asymmetric antenna bio-radar speech signal detection: (a) Indoor experiment; (b) Outdoor experiment.

The remainder of this paper is organized as follows. In Section 2, the experimental environment, speech corpus and detection result and noise characteristic analysis of the 94 GHz asymmetric antenna bio-radar speech are presented. In Section 3, the method for improving the quality of 94 GHz asymmetric antenna radar speech is proposed. In Section 4, performance evaluation is provided. In Section 5, experimental results of the proposed algorithm are demonstrated and the performance is evaluated. Finally, the conclusion is provided in Section 6.

II. EXPERIMENT

A. EXPERIMENT ENVIRONMENT

The 94 GHz asymmetric antenna radar system is mainly composed of transmitting antenna, receiving antenna, transmitter module, receiver module. One antenna with a diameter of 600 mm, is used as receiving antenna, the beam width is 0.4 degrees at −3 dB levels, the gain is 50 dB. One antenna with a diameter of 200 mm, is used as transmitting antenna, the beam width is 1 degrees at −3 dB levels, and the gain is 41.7 dB. The more detailed description of the 94 GHz asymmetric antenna radar detection theory and system was shown in our previous work [10], [12].

The target sound source selected in this paper includes human sound source and simulated sound source. Five healthy volunteers including 4 males and 1 female were selected as human sound source participated in the speech detection experiment. All of the volunteers (from 20 to 28 years old) were native speakers of mandarin Chinese, none of them had a history of voice training or voice disorders. All of the experimental procedures were in accordance with the rules of the Declaration of Helsinki, and all volunteers signed the appropriate consent forms.

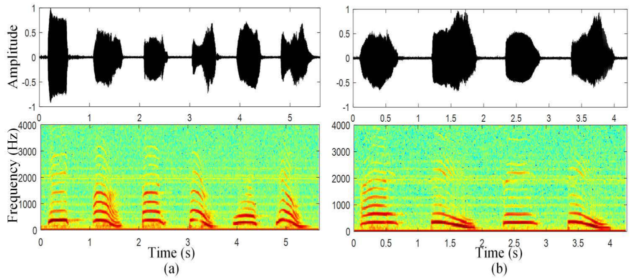

FIGURE 3. **The experimental results of 94 GHz asymmetric antenna radar speech signals for simulated sound source. (a) The time-domain waveforms and the spectrograms of the radar speech "1-2-3-4-5-6"; (b) The time-domain waveforms and the spectrograms "a-b-c-d".**

The experiments were carried out in two scenarios. The first scenario was in an indoor laboratory, which is about 13 meters long and 5.5 meters wide, as shown in Figure 2a. The second scenario is in an outdoor stadium, which is 100 meters long and 100 meters wide, as show in Figure 2b. In these experimental scenarios, a subject was asked to sit in front of the radar system, and the larynx of the volunteer should be at the same height as the center line of the antenna of the 94 GHz asymmetric antenna bio-radar system. A laser pen was used to focus an electromagnetic beam on the larynx area. The radar system were positioned at distances ranging from 1 to 50 m from the sound source.

B. SPEECH CORPUS AND DETECTION RESULT

To ensure the consistency of the pronunciation material and facilitate the experiment operation, a loudspeaker was selected as simulated sound source. One English language letters a-b-c-d, two Mandarin Chinese sentences "1-2-3-4-5-6", "Jun-Yi-Da-Xue" and the standard library from the TIMIT Database [43], such as "She wore warm, fleecy, woolen overalls", "Jane may earn more money by working hard", "The best way to learn is to solve extra problems", etc.were selected as the speech materia for sound sources. In order to facilitate the analysis and processing, the amplitude of the detected speech signal is normalized in this paper. The experimental results of 94 GHz asymmetric antenna radar speech signals for simulated and human sound sources at a distance of 5 m are illustrated in Figure 3 and Figure 4 respectively.

C. NOISE CHARACTERISTIC ANALYSIS OF RADAR SPEECH

The basic principle of bio-radar detecting speech signal is that radar emits electromagnetic waves, when the electromagnetic waves reach the human larynx area, the wave is modulated and reflected by vibration information caused by the human voice, which is received by the radar antenna and sent to the computer for processing to form speech signals. Air is the common medium of traditional microphone speech signal, the air particles vibration is caused by the sound source, and then microphone sensor convert vibration information into electrical signals. Due to the difference of detection principle, the radar speech has some special advantages not offered by

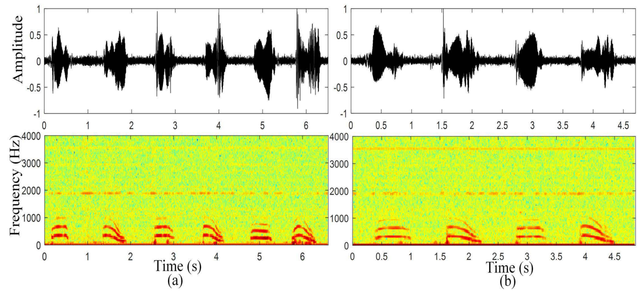

FIGURE 4. **The experimental results of 94 GHz asymmetric antenna radar speech signals for human sound source. (a) The time-domain waveforms and the spectrograms of the radar speech "1-2-3-4-5-6"; (b) The time-domain waveforms and the spectrograms "a-b-c-d."**

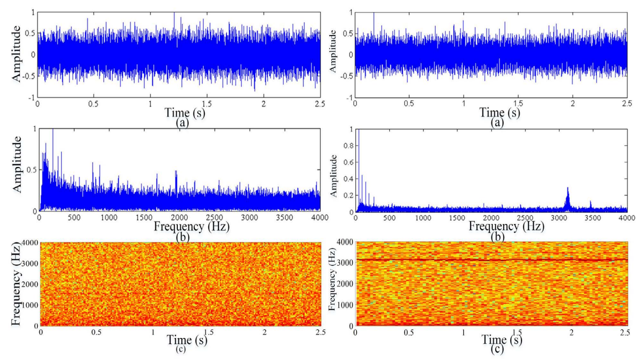

FIGURE 5. **The spectral analysis of noise signal. (a) Waveforms of noise signal; (b) Spectrums of noise signal; (c) Spectrograms of noise signal.**

traditional microphones, such ashigh directional sensitivity, strong acoustical disturbance. Therefore, the noise source of radar speech is different from that of traditional microphone. Radar speech is mainly interfered with electromagnetic noise such as inter-modulation interference and cross modulation interference, circuit noise such as thermal nois and scattered nois, environmental noise such as side lobe clutter and human body swerve noise In order to effectively remove the noise from radar speech, we used 94 GHz asymmetric antenna radar system to detect noise signal at two different time periods in thecase of no sound source. Then the spectral analysis of noise signal is carried out, as shown in Figure 5.

From the Figure 5, we can observe that the noise is concentrated in the whole frequency range, especially true for the low-frequency components of the speech. Besides, there will be single frequency noise in the middle and high frequencies. On the whole, the noise distribution of 94 GHz asymmetric antenna radar speech is non-Gaussian. Therefore, it is of great significance to apply an appropriate enhancement algorithm to improve the quality of the 94 GHz asymmetric antenna bio-radar speech signal

III. METHODOLOGY

A. VARIATIONAL MODE DECOMPOSITION

Variable Mode Decomposition (VMD) is a novel adaptive decomposition method, which decompose a signal $x(t)$ into an ensemble of band-limited intrinsic mode functions (IMFs)

u_k, each mode k to be most compact around a center pulsation ω_k [32]. The VMD decomposition process can be viewed as the solution process to the variational problem. The algorithm mainly includes two parts which are the construction of the variational problem and the solution of the variational problem.

1) THE CONSTRUCTION OF VARIATIONAL PROBLEM

Firstly, compute the analytic signal of each mode u_k by means of the Hilbert transform to obtain a unilateral frequency spectrum. Secondly, the frequency spectrum of each mode is modulated to the corresponding "baseband" by mixing with an exponential tuned to the center frequency. Finally, the estimated bandwidth corresponding to each mode is obtained by calculating the H^1 Gaussian smoothness of the demodulated signal. Thus, a constrained variational problem is constructed, which is expressed as follows:

$$\min\left\{\sum_{k=1}^{K}\left\|\partial_t[(\delta(t)+\frac{j}{\pi t})*u_k(t)]e^{-jw_k t}\right\|_2^2\right\}$$
$$\text{s.t. } \sum_{k=1}^{K}u_k=x(t) \tag{1}$$

where K is the number of mode number. u_k $(k=1,2,\ldots,K)$ and w_k $(k=1,2,\ldots,K)$ are the ensemble of all modes and their center frequencies, respectively, * denotes the convolution operator, δ(t) is unit pulse function.

2) THE SOLUTION OF VARIATIONAL PROBLEM

By using a quadratic penalty term and Lagrangian multipliers, the constrained variational problem is transformed into an unconstrained variational problem. The quadratic penalty term is used to ensure the accuracy of the reconstructed signal, especially in the presence of additive Gaussian noise, and the Lagrangian multipliers are a common way of enforcing constraints strictly.

The alternate direction method of multipliers (ADMM) [44] is employed to update the u_k^{n+1}, w_k^{n+1} and λ_k^{n+1}. So that the original minimization problem in formula (1) is transformed into the saddle point of the augmented Lagrangian L in a sequence of iterative sub-optimizations.

The Lagrangian L expression is:

$$\begin{aligned}L(\{u_k\},\{w_k\},\lambda)&=\alpha\sum_{k=1}^{K}\left\|\partial_t[(\delta(t)+\frac{j}{\pi t})*u_k(t)]e^{-jw_k t}\right\|_2^2\\&\quad+\left\|x(t)-\sum_{k=1}^{K}u_k(t)\right\|_2^2+\left\langle\lambda(t),x(t)-\sum_{k=1}^{K}u_k(t)\right\rangle\end{aligned} \tag{2}$$

In the formula, α is a bandwidth parameter of the quadratic penalty term, * denotes the convolution operator, $\delta(t)$ is unit pulse function, and $\lambda(t)$ is a Lagrangian multipliers.

The expression updating uk^{n+1} can be expressed as:

$$\begin{aligned}u_k^{n+1}=\underset{u_k\in X}{\arg\min}&\left\{\left\|\partial_t[(\delta(t)+\frac{j}{\pi t})*u_k(t)]e^{-jw_k^n t}\right\|_2^2\right.\\&\left.+\left\|x(t)-\sum_{i=1,i\neq k}^{K}u_i^{n+1}(t)+\frac{\lambda(t)}{2}\right\|_2^2\right\}\end{aligned} \tag{3}$$

By means of Parseval/Plancheral Fourier isometry transform, the equation (3) can be converted from time domain to frequency domain, and the expression of each mode in frequency domain can be obtained as follows:

$$\hat{u}_k^{n+1}(w)=\frac{\hat{x}(w)-\sum_{i=1,i\neq k}^{K}\hat{u}_i(w)+\frac{\hat{\lambda}(w)}{2}}{1+2\alpha(w-w_k)^2} \tag{4}$$

where, $\hat{\lambda}_k^{n+1}$ is updated by formula as follow:

$$\hat{\lambda}^{n+1}(w)\leftarrow\hat{\lambda}^n(w)+\tau|\hat{x}(w)-\sum\hat{u}_k^{n+1}(w)| \tag{5}$$

where, τ is noise tolerance parameter, then, for the updated center frequencies wk^{n+1} of each mode, and the expression of updated wk^{n+1} in frequency domain can be obtained as follows:

$$w_k^{n+1}=\frac{\int_0^\infty w\left|\hat{u}_k^{n+1}(w)\right|^2dw}{\int_0^\infty\left|\hat{u}_k^{n+1}(w)\right|^2dw} \tag{6}$$

where, wk^{n+1} is the center of gravity of the corresponding mode's power spectrum for k mode, $\hat{u}k_{(w)}$ is equivalent to the wiener filter of the current residue $\hat{x}(w)-\sum_{i\neq k}\hat{u}_i(w)$.

On the whole, the VMD algorithm updates each mode in the frequency domain, and then converts it to the time domain by Fourier transform. The mode updating steps are as follows:

(1) Initialize $\{\hat{u}k^1\}$, $\{wk^1\}$, $\hat{\lambda}^1$, and $n=0$;

(2) According to Formula (4) and Formula (6), update $\hat{u}_k$ and w_k.

(3) According to Formula (5), update $\hat{\lambda}$

(4) For a given discriminant precision $e>0$, until the iteration constraint is satisfied:

$$\sum_{k=1}^{K}\left\|\hat{u}_k^{n+1}-\hat{u}_k^n\right\|_2^2\Big/\left\|\hat{u}_k^n\right\|_2^2<e \tag{7}$$

Stop Iteration; otherwise, return to step (2).

B. THE SELECTION OF VMD PARAMETERS

VMD algorithm contains many parameters, there are Lagrangian multipliers λ, decomposition mode number k, and bandwidth constrained equilibrium parameter α. These parameters have a great influence on decomposition result.

Lagrangian multipliers λ: The role of the Lagrangian multiplier is to ensure that the decomposed IMF components can reconstruct the original signal. While exact reconstruction of the original signal is usually not the purpose of VMD method, in particular in the radar speech enhancement, the signal usually has strong noise, the noise contained in radar speech signal is expected to be removed and not included in the decomposition result. Thus, the Lagrangian multiplier should be shut off by keeping its value at zero in this paper. It can be done by choosing its update parameter $\tau = 0$.

Decomposition mode number k: If k is too large, there will be over decomposition, this results in spurious components. If k is too small, there will be under decomposition, this results in mode aliasing. Therefore, Choosing appropriate mode number k is crucial to VMD decomposition. Though mode aliasing and end effect may occur in EMD decomposition, it can adaptively decompose the signal into a series of oscillatory components without any prior knowledge. Thus, we use the advantage of the EMD method to determine the mode number. First, the radar speech signal is decomposed into intrinsic mode functions (IMFs) by EMD, the first few IMFs with high frequency can be regarded as major information modes of the original radar speech signal. The low frequency IMFs mainly are the detailed information modes which is not need to further decompose. Therefore, the number of major modes is determined as the mode number of the VMD method. The amplitude of the IMFs is used to distinguish major information and detailed information. If the maximum amplitude of the decomposed modes by EMD is less than 0.1, the number of mode layers corresponding to the first IMF in which this condition occurs can be regarded as the decomposition mode number of VMD method. The decomposition mode number is given as:

$$\begin{cases} \mathit{If} \ \max[A(IMF_i)] \leq 0.1, 1 \leq i \leq n \\ k = \mathrm{first}(i) \end{cases} \tag{8}$$

where, IMF_i is the i-th intrinsic mode function decomposed by EMD, $A(IMF_i)$ is the amplitude of each IMF.

Bandwidth parameter α: The smaller the value of α, the larger the bandwidth of each IMF component derived from VMD decomposition. If the bandwidth is too large, the frequency spectrum of each mode will overlap, and the number of spectral lines caused by noise in the envelope spectrum will increase. If the bandwidth is too small, not only the noise is suppressed, but also the useful characteristic lines in the envelope spectrum may be filtered out. Therefore, the value of α should not be too large or too small. In addition, we found that this parameter has little effect on the results in the relative range. For the radar speech, we select α value is 1000 in this paper.

C. PEARSON CORRELATION COEFFICIENT

By analyzing the IMFs signal after VMD decomposition, some modes are mainly noise and interference signal without any speech information. These modes are directly removed during reconstruction. Pearson correlation coefficient (PCC) is a statistical metric that measures the strength and direction of a linear relationship between two random variables, the greater the absolute value of the correlation coefficient, the stronger the correlation [45]. In this paper, PCC is used to distinguish between the useful modes and noise modes after VMD decomposition. The PCC values are calculated between the original radar speech signal and each IMF as follows:

$$PCC\,(x, y) = \frac{\sum_{t=1}^{N} [(x_t - \bar{x}]\,[(y_t - \bar{y})]}{\sqrt{\sum_{t=1}^{N} (x_t - \bar{x})^2}\sqrt{\sum_{t=1}^{N} (y_t - \bar{y})^2}} \tag{9}$$

where x_t and y_t are two random variables. $\bar{x}$ and $\bar{y}$ are the average of the two random variables, respectively.

It is concluded that the higher the PCC value, the stronger the correlation between IMF and the original radar speech signal. If the PCC value is high, we can believe the IMF is a useful mode, otherwise is a noise mode. Thus, in order to find the useful modes, a fixed threshold (FT) was defined as 10^{-1}. If the PCC value between the original radar speech signal and IMF is greater than the FT, the IMF can be regarded as a useful mode.

D. IMPROVED THRESHOLD FOR NIOSE REDUCTION

In order to effectively suppress the noise of the original radar speech, we should select an appropriate threshold T to remove the noise of useful modes before reconstruction. Wavelet threshold has been widely used in noise reduction [46], the threshold is estimated as follow:

$$T = \sigma\sqrt{2\ln(N)} \tag{10}$$

where N is the signal length, σ is the estimated noise level.

In the experiment, It was found that the estimated noise level plays an important role in removing noise from radar speech signal. However, the traditional estimated noise level, is not accurate in estimating IMF noise, which will take effect on the quality of enhanced speech. Furthermore, although the soft thresholding was widely employed to remove the noise of signal, it may cause signal overprocessing, resulting in speech signal distortion.

Therefore, in this paper, we proposed an improved threshold strategy (ITS) for removing noise from the radar speech signal. It includes two aspects: a new noise estimated level and improved soft thresholding function.

A new noise estimated level of each IMF is given by:

$$\sigma_i = \frac{1}{L}\sum |IMF_i(t)|, \quad t = 1, 2, \cdots, L \tag{11}$$

where L is the length of the initial silent segment of the radar speech signal. $IMF_i(t)$ is the i-th band-limited intrinsic mode function decomposed by VMD.

In order to obtain a compromise between noise reduction and intelligibility of radar speech, an improved soft thresholding function is employed to remove the noise of useful modes before reconstruction.

$$IMF_i'(t)=\begin{cases} sign[IMF_i(t)]\cdot[|IMF_i(t)|-T_i], & |IMF_i(t)|\geq T_i \\ IMF_i(t)\times m, & |IMF_i(t)|<T_i \end{cases} \tag{12}$$

where m is compensation factor.

E. THE PROPOSED ALGORITHM FOR RADAR SPEECH ENHANCEMENT

In view of the above section, the flow chart of the proposed algorithm for 94 GHz asymmetric antenna radar speech enhancement is shown in Figure 6. The detailed steps of the proposed algorithm are shown as follows:

Setp1: Decompose the given original radar speech signal $x(t)$ into IMFs using EMD method, then determine the decomposition k using Formula (8).

Step2: According to the value of the alpha and k, decompose the original radar speech signal $x(t)$ into modes using VMD algorithm.

Step3: Compute the PCC value of the original radar speech signal and each IMF using Equations (9) and determine the useful mode.

Step4: Determine the threshold T using Formulas (10) and (11).

Step5: Reconstruct and denoise the useful mode using Formula (12).

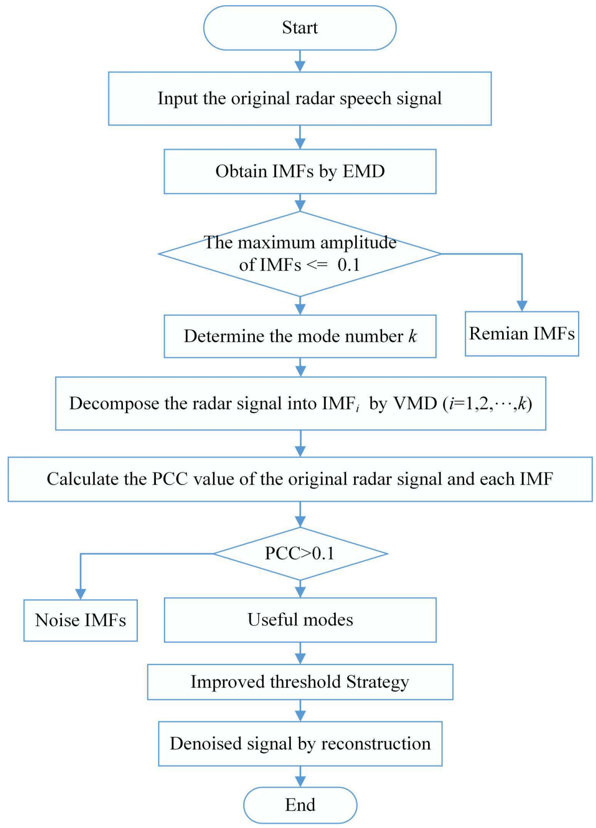

FIGURE 6. The flow chart of the proposed algorithm for radar speech enhancement.

IV. PERFORMANCE EVALUATION

Short time objective intelligibility (STOI), and Perceptual evaluation of speech quality (PESQ, which are widely used for evaluating the performance of speech enhancement algorithms. STOI evaluation is a effective method to test the speech intelligibility. PESQ evaluation is a standard objective measure technique to test the speech quality. However, in [47], the PESQ measure can did not yield as high correlation with speech quality, three different composite measures (CMs) are proposed to test the speech quality. The three CMs can greatly improve the correlations of objective measures and provide more adequate in predicting the subjective quality of noisy speech enhanced by noise suppression algorithms over the conventional objective measures. Therefore, in this paper, STOI, CMs are selected to evaluate the quality and intelligibility of the speech enhancement algorithms, for comparison, PESQ is still being selected as an auxiliarymeasure.

A. PESQ MEASURE

PESQ evaluation was proposed by Rix *et al.* for evaluating the quality of speech enhancement algorithms, which has been defined by the ITU-T recommendation P.862 [48]. The evaluation gives accurate predictions of subjective quality in a very wide range (300-3400 Hz) of conditions, it is suitable for assessing the speech quality of telephone networks and speech signals. A combination of two disturbance parameters: symmetric disturbance (d_{sym}) and asymmetric disturbance (d_{asym}) were used as a predictor of subjective MOS. The final PESQ score is obtained as follow:

$$PESQ = 4.5 - 0.1 \times d_{sym} - 0.0309 \times d_{asym} \tag{13}$$

A higher PESQ score indicates better quality of speech, the highest score is 4.5, it indicates no distortion.

B. STOI MEASURE

The STOI was proposed by Cees *et al.* for predicting the intelligibility of noisy speech [49]. The evaluation results show that STOI has high correlation with the speech intelligibility in listening test. In addition, STOI does work well for additive noise of noisy speech for different noise types and SNRs. The STOI score range between 0 and 1, a higher STOI score indicates better intelligibility of speech.

C. COMPOSITE MEASURE

To further achieve higher correlation with the subjective scores of evaluation method, composite measure (CM) method was proposed for predicting the quality of noisy speech enhanced by noise suppression algorithms [47]. The authors used the ITU-T P.835 methodology to evaluate the speech quality along three dimensions: composite measure for signal distortion (CSIG), composite measure for noise distortion (CBAK), and composite measure for overall speech quality (COVL). CSIG is a measure for signal distortion (SIG) formed by linearly combining the log likelihood ratio

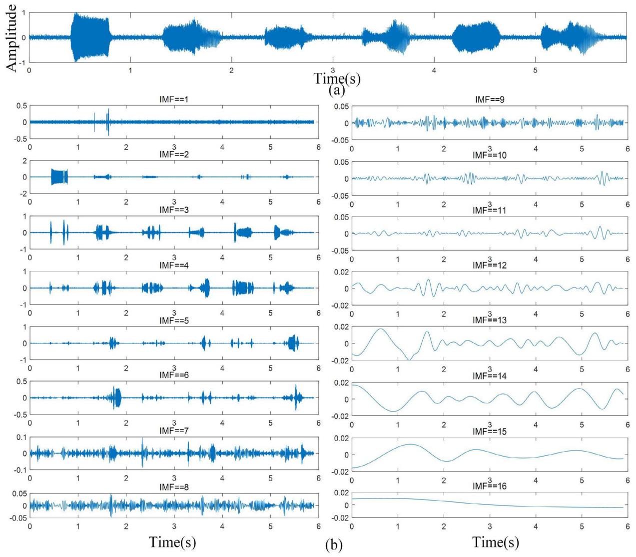

FIGURE 7. (a) The original radar speech signal; (b) the IMFs of the original radar speech signal using EMD.

(LLR), PESQ, and weighted-slope spectral distance (WSS) measures; CBAK is a measure for noise distortion (BAK) formed by linearly combining segmental SNR (segSNR), PESQ, and WSS measures, and COVL is a measure for overall quality (OVRL) formed by linearly combining the PESQ, LLR, and WSS measures. The three new composite measures were defined as:

$$\begin{aligned} CSIG &= 3.093 - 1.029 \cdot LLR + 0.603 \cdot PESQ \\ &\quad - 0.009 \cdot WSS \\ CBAK &= 1.634 + 0.478 \cdot PESQ - 0.007 \cdot WSS \\ &\quad + 0.063 \cdot segSNR \\ COVL &= 1.594 + 0.805 \cdot PESQ - 0.512 \cdot LLR \\ &\quad - 0.007 \cdot WSS \end{aligned} \tag{14}$$

The higher score of the CSIG, CBAK and COVL represents better quality of speech.

V. RESULTS AND DISCUSSION

This section mainly presents the performance of the proposed algorithm in enhancing the quality and intelligibility of the 94 GHz asymmetric antenna radar speech signal. To guarantee high quality and reliable speech signal, in this section, one sentence of Mandarin Chinese ‘‘1-2-3-4-5-6’’ at a distance of 10 meters and one English language letters ‘‘a-b-c-d’’ at a distance of 7 meters for simulated sound source are selected as a representative speech material to evaluate.

A. DETERMINE THE NUMBER OF DECOMPOSITION MODE

First, EMD method is used to determine the mode number. Figure 7a shows the waveform of the original radar speech signal ‘‘1-2-3-4-5-6’’. Figure 7b shows the IMFs of the original radar speech signal decomposed by EMD. It can be seen that the original radar speech is decomposed 16 IMFs by EMD, the frequency of IMFs decrease gradually with the increase of IMF value. According to equation (8), the decomposition mode number is equal to seven.

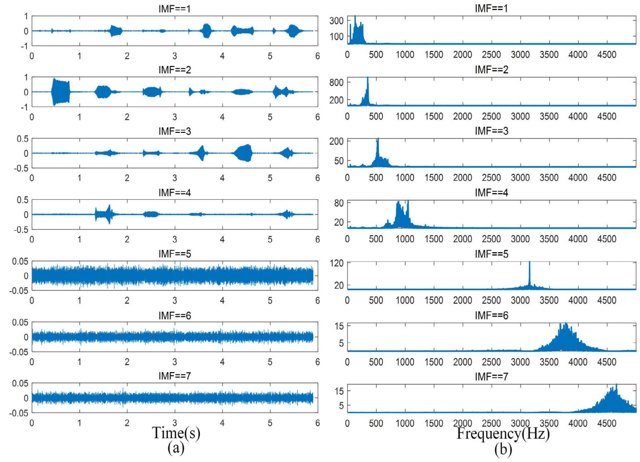

FIGURE 8. (a) the seven IMFs of original radar speech signal decomposed by VMD method; (b) the spectrum of corresponding to each IMF.

According to the decomposition mode number, Figure 8a shows the seven IMFs of the original radar speech signal decomposed by VMD method, Figure 8b shows the spectrum of corresponding to each IMF. The longer the radar detection distance is, the more serious the attenuation of high frequency components of speech is, and the wider the frequency range of speech signal itself is, the higher the speech quality it presents.

The speech signal frequency range of a normal person is 0-4000 Hz, the limit frequency range that determines the intelligibility is 300-1000 Hz. As presented in Figure 8, the frequencies of IMFs increase gradually with the increase of IMF value. It can also be inferred that the waveforms of the former four IMFs are similar to speech signal, and the waveforms of the latter three IMFs are similar to noise signal. We can assume that some modes are mainly noise and interference signal without any speech information which should be rejected before reconstruction.

B. SELECTING THE RECONSTRUCTION MODES

Figure 9 shows the PCC values of original radar (OR) speech signal and each IMF using equation (9). According to the FT, the former four modes are selected as the useful modes for reconstruction.

C. ANALYSIS AND COMPARISION OF SPECTROGRAMS

Figure 10 shows the comparison of the enhanced radar speech signal. Figure 10a presents the waveform and spectrogram of the clean speech signal synchronously acquired. Figure 10b presents the waveform and spectrogram of the original radar speech. Figure 10c-f present the waveform and spectrogram of the original radar speech after processing using the

TABLE 1. Evaluation of Speech Quality by Four Speech Enhancements.

Evaluation measures	Original	Wavelet-hard	Wavelet-soft	EMD shrinkage	Proposed method
PESQ	1.8886	2.0863	2.2289	1.7657	2.2981
STOI	0.6640	0.6857	0.7253	0.6801	0.7024
CSIG	0.9702	0.5771	0.8364	1.9197	2.2756
CBAK	1.0311	0.6702	0.7858	1.0307	1.5143
COVL	1.1760	0.9106	1.1343	1.6028	2.0380

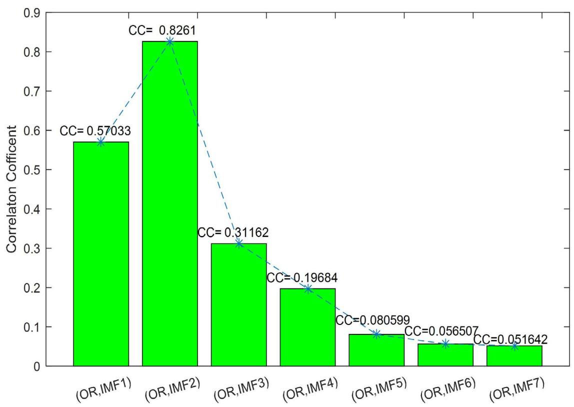

FIGURE 9. Correlation coefficients of original radar speech signal and each IMF.

wavelet-hard, wavelet-soft, EMD-shrinkage, and proposed method, respectively.

As shown in Figure 10, the spectrogram of the original radar speech has excellent consistency in clean speech in low frequency components. Compared with the clean speech signal, the radar speech signal is contaminated by noise components, and the high-frequency information is lost. Its frequency range is 0-2000 Hz. This is due to the difference between radar speech signal and traditional speech signal in transmission media and detection principle. It is also observed that noise reduction is achieved to some degree for speech processed by all of the four algorithms. We can find that the wavelet-hard and the wavelet-soft method can reduce the noise to a certain extent, but these methods introduce some new residual noise to the enhanced speech, and the energy stripe of spectrogram has some distortion, so the quality of the radar speech was not improved. The waveform and the spectrogram of the enhanced radar speech with EMD shrinkage are displayed in Figure 10e, which shows the noise has mostly been removed, the quality of the radar speech signal is greatly improved. However, the clarity of the energy stripe is affected to some extent. For the proposed method, the power of noise is reduced dramatically as shown

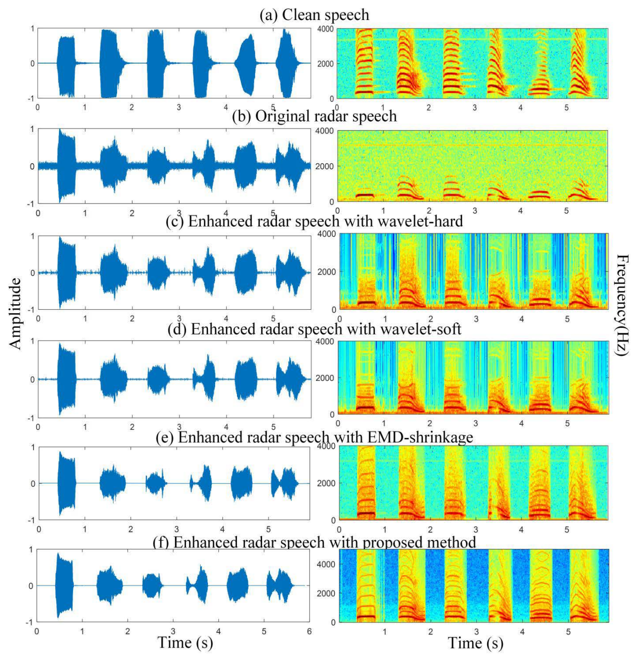

FIGURE 10. The comparison of the enhanced radar speech signal.

in Figure 10f, and the energy stripe of the spectrogram is very clear. The waveforms and spectrograms show that the quality and the intelligibility of the original radar speech is significantly improved by the proposed method.

D. PERFORMANCE EVALUATION

The evaluation measures results of original radar speech and enhanced speech by four speech enhancement methods are presented in Table 1. It can be seen from the table that the proposed method yields the highest scores, which indicates that the enhanced speech signal processed by the proposed method has the highest speech quality and intelligibility.

Despite the PESQ score of enhanced speech by the EMD shrinkage method is lower than the original radar speech, the STOI, CMs scores are all higher than the original radar speech. Combining it with the result of spectrograms, it can

TABLE 2. The Evaluation Measure Results of Radar Speech Corrupted by White Noise Obtained Using Four Enhancement Algorithms.

Evaluation measures	SNR(dB)	Noisy	Wavelet-hard	Wavelet- soft	EMD-shrinkage	Proposed method
PESQ	0	1.1553	0.6654	1.0831	0.8601	1.3684
	5	1.3209	0.9707	1.3946	1.0441	1.5998
	10	1.5552	1.3717	1.7809	1.3635	1.8507
	15	1.7201	1.6866	1.9889	1.5113	2.1978
STOI	0	0.5543	0.4704	0.4818	0.5543	0.6127
	5	0.5905	0.5602	0.5723	0.5649	0.6182
	10	0.6212	0.6229	0.6448	0.5970	0.6772
	15	0.6428	0.6600	0.6892	0.6287	0.7007
CSIG	0	-0.9087	-0.8025	-0.5622	0.2410	0.4841
	5	-0.5200	-0.3272	-0.0647	1.1849	1.0067
	10	-0.0299	0.1188	0.3756	1.3894	1.3664
	15	0.4299	0.3987	0.6122	1.3791	1.8543
CBAK	0	0.5276	-0.3067	-0.1736	0.5293	0.7611
	5	0.6597	-0.0076	0.1386	0.7049	1.0498
	10	0.8152	0.2816	0.4494	0.8405	1.2195
	15	0.9248	0.4851	0.6235	0.8811	1.4353
COVL	0	-0.1843	-0.5956	-0.2895	0.2930	0.5926
	5	0.1145	-0.1506	0.1733	0.8793	0.9975
	10	0.4901	0.3103	0.6333	1.1336	1.3369
	15	0.8128	0.6276	0.8840	1.1924	1.7710

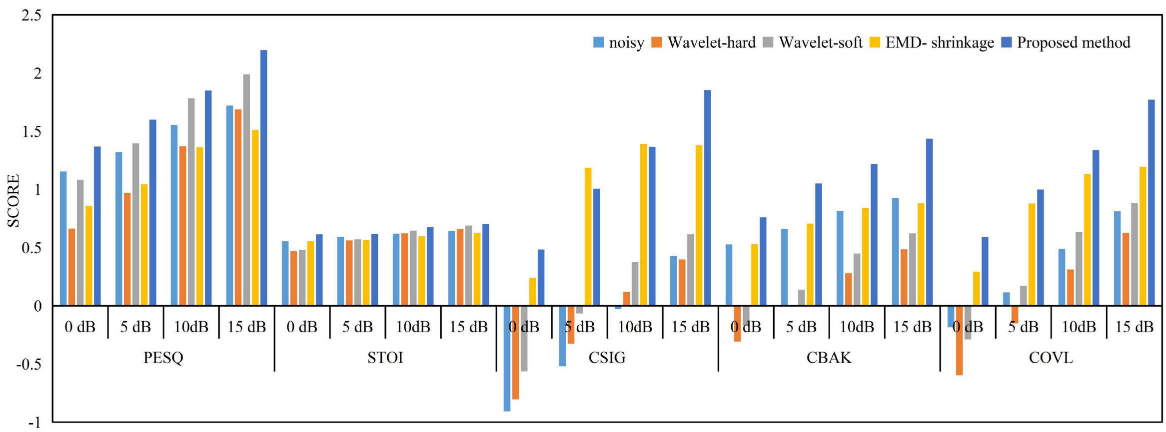

FIGURE 11. The bar chart of the evaluations measure results of radar speech corrupted by white noise obtained using four enhancement algorithms.

be indicated that the PESQ may not yield as high correlation with speech quality. This suggests that the EMD shrinkage method can improve the quality and intelligibility of the radar speech. For the other wavelet-hard and wavelet-soft methods, though the PESQ and STOI scores are higher than the original radar speech, the CMs scores are all lower than the original radar speech. It is illustrated that wavelet-hard and wavelet-soft method is invalid to improve the intelligibility at the sacrifice of the quality of speech.

To further verify the effectiveness and reliability of the results, white noise, pink noise and hfchannel noise are selected from the NOISEX-92 database [50], and added to

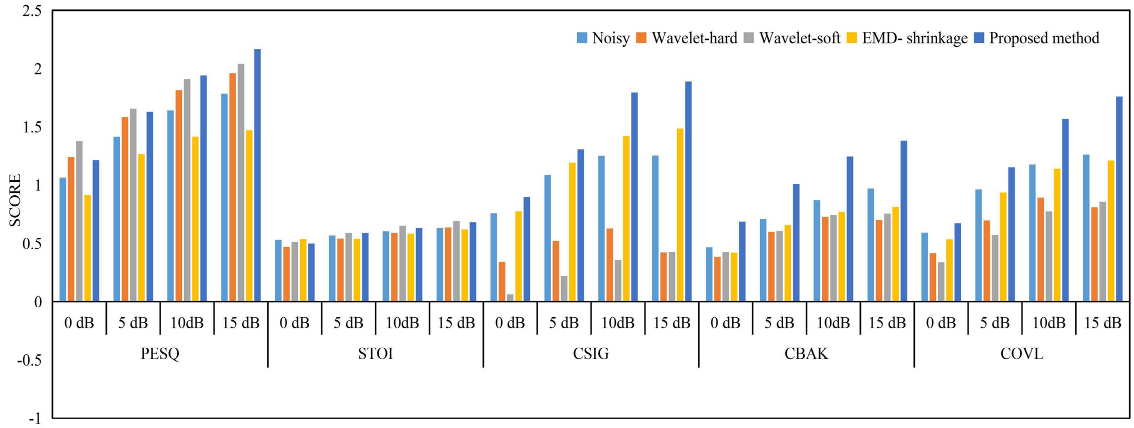

FIGURE 12. The bar chart of the evaluations measure results of radar speech corrupted by pink noise obtained using four enhancement algorithms.

TABLE 3. The Evaluation Measure Results of Radar Speech Corrupted by Pink Noise Obtained Using Four Enhancement Algorithms.

Evaluation measures	SNR(dB)	Noisy	Wavelet-hard	Wavelet- soft	EMD-shrinkage	Proposed method
PESQ	0	1.0623	1.2406	1.3789	0.9176	1.2125
	5	1.4158	1.5865	1.6546	1.2644	1.6302
	10	1.6405	1.8133	1.9098	1.4149	1.9407
	15	1.7847	1.9591	2.0412	1.4733	2.1657
STOI	0	0.5309	0.4704	0.5109	0.5366	0.4981
	5	0.5682	0.5412	0.5892	0.5400	0.5886
	10	0.6029	0.5890	0.6502	0.5845	0.6328
	15	0.6314	0.6360	0.6913	0.6204	0.6818
CSIG	0	0.7574	0.3406	0.0642	0.7756	0.8974
	5	1.0871	0.5212	0.2197	1.1915	1.3056
	10	1.2513	0.6284	0.3576	1.4212	1.7946
	15	1.2531	0.4235	0.4256	1.4861	1.8885
CBAK	0	0.4666	0.3850	0.4278	0.4219	0.6864
	5	0.7116	0.5982	0.6049	0.6558	1.0087
	10	0.8715	0.7287	0.7448	0.7710	1.2451
	15	0.9720	0.7008	0.7551	0.8146	1.3812
COVL	0	0.5918	0.4170	0.3382	0.5347	0.6731
	5	0.9617	0.6961	0.5712	0.9371	1.1513
	10	1.1771	0.8919	0.7751	1.1405	1.5702
	15	1.2609	0.8084	0.8563	1.2093	1.7590

the original radar speech with SNR 0, 5, 10 and 15 dB. Table 2 shows the evaluation measure results of the original radar speech corrupted by white noise obtained using four enhancement algorithms, and the bar chart as shown in Figure 11. From the Table 2 and Figure 11, we can find that the proposed method yields the highest scores than the original radar speech and the other three methods. It is suggested that the effectiveness of the proposed method in removing the

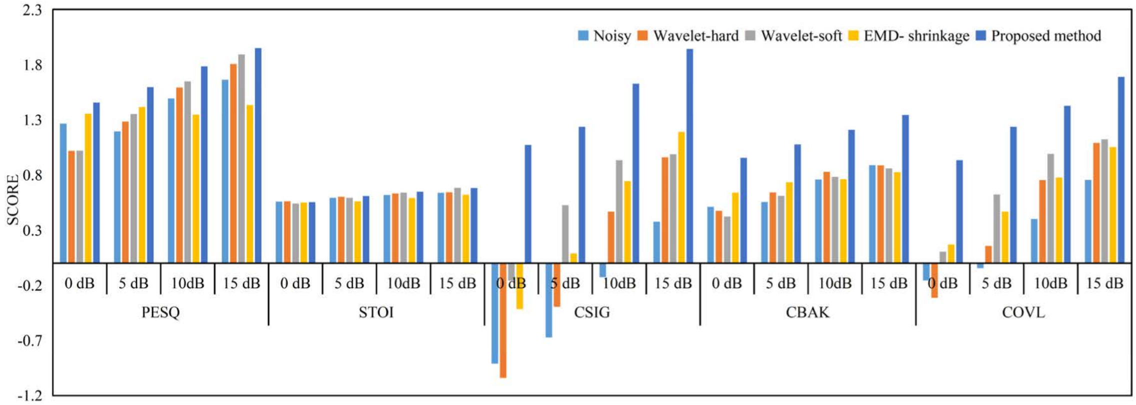

FIGURE 13. The bar chart of the evaluations measure results of radar speech corrupted by hfchannel noise obtained using four enhancement algorithms.

TABLE 4. The Evaluation Measure Results of Radar Speech Corrupted by Hfchannel Noise Obtained Using Four Enhancement Algorithms.

Evaluation measures	SNR(dB)	Noisy	Wavelet-hard	Wavelet- soft	EMD-shrinkage	Proposed method
PESQ	0	1.2645	1.0189	1.0204	1.3553	1.4579
	5	1.1967	1.2846	1.3520	1.4182	1.5968
	10	1.4926	1.5931	1.6493	1.3474	1.7830
	15	1.6635	1.8048	1.8933	1.4359	1.9514
STOI	0	0.5589	0.5624	0.5402	0.5501	0.5545
	5	0.5935	0.6036	0.5936	0.5612	0.6107
	10	0.6209	0.6336	0.6412	0.5907	0.6503
	15	0.6408	0.6456	0.6820	0.6229	0.6813
CSIG	0	-0.9116	-1.0395	-0.1563	-0.4154	1.0716
	5	-0.6747	-0.3956	0.5271	0.0908	1.2358
	10	-0.1275	0.4678	0.9340	0.7443	1.6274
	15	0.3784	0.9588	0.9872	1.1916	1.9438
CBAK	0	0.5115	0.4753	0.4225	0.6411	0.9547
	5	0.5554	0.6432	0.6119	0.7324	1.0774
	10	0.7594	0.8289	0.7838	0.7622	1.2084
	15	0.8910	0.8896	0.8591	0.8264	1.3442
COVL	0	-0.1552	-0.3137	0.1072	0.1686	0.9343
	5	-0.0441	0.1557	0.6258	0.4671	1.2358
	10	0.4001	0.7548	0.9891	0.7766	1.4288
	15	0.7562	1.0914	1.1240	1.0533	1.6902

white noise of the radar speech. The quality and intelligibility of the radar speech signal is greatly improved. It also can be found that the STOI score is not increased for the other three methods, but the CMs score is increased for wavelet-soft and EMD shrinkage methods, especially for the EMD shrinkage method. It suggests that the EMD shrinkage method can obtain a good tradeoff between the intelligibility and noise reduction, but the results are not entirely satisfactory. For the proposed method, the quality and intelligibility of the radar speech signal can be greatly improved.

TABLE 5. Evaluation of Speech Quality by Four Speech Enhancements.

Enhancement Algorithms	Original	Wavelet-hard	Wavelet-soft	EMD- shrinkage	Proposed method
PESQ	2.4025	2.2800	2.3039	2.2841	2.8298
STOI	0.4914	0.4973	0.5214	0.5065	0.5344
CSIG	1.6331	1.1943	0.9344	2.6539	3.0012
CBAK	1.3053	1.1339	1.1231	1.5608	2.4599
COVL	1.7756	1.4548	1.3295	2.2977	2.7640

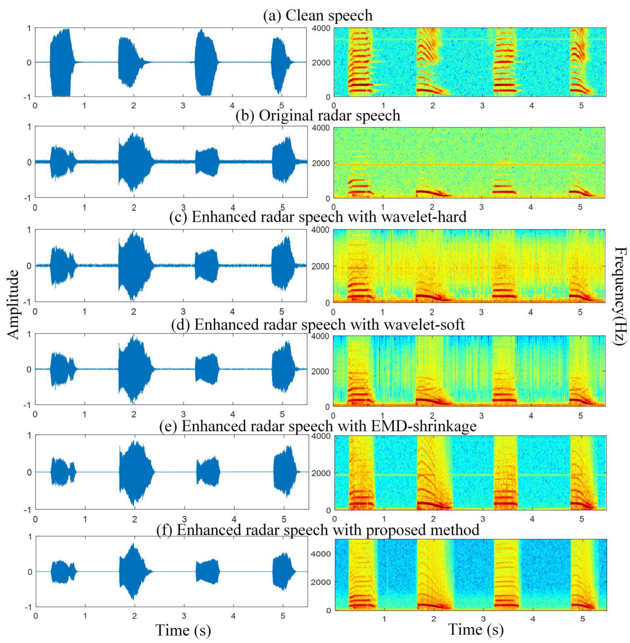

FIGURE 14. The comparison of the enhanced radar speech signal.

For pink noise, the evaluation measure results are shown in Figure 12 and Table 3. For hfchannel noise, the evaluation measure results are shown in Figure 13 and Table 4. We can find that the proposed method is superior not only in white noise but also in pink and hfchannel noise conditions. However, we can observe that all the scores are not increased to some degree for speech processed by the wavelet-hard, wavelet-soft and EMD shrinkage methods in pink noise condition. For hfchannel noise, the results are consistent with that in white noise described above. It also can be indicated that the proposed method is quite effective in hfchannel noise condition.

E. FURTHER EXPERIMENTS

In order to further test the performance of the proposed algorithm in improving the quality and intelligibility of the original radar speech signal. One English language letters "a-b-c-d" detected by 94 GHz asymmetric antenna radar is enhanced by four speech enhancement methods. Figure 14 shows the comparison of the enhanced radar speech signal for the "a-b-c-d". Figure 14a presents waveform and spectrogram of the clean speech signal synchronously acquired. Figure 14b shows the waveform and spectrogram of the original radar speech signal. Figure 14c shows that the noise of original radar speech has not been effectively reduced, and some new noise signal is introduced, this results in severe radar speech distortion. Figure 14d shows the EMD soft method is effective in reducing the combined noise of the radar speech, but there is still too much remnant noise, so the quality of the radar speech was not improved. From Figure 14e, It can be seen that the noise has mostly been removed. However, the clarity of energy stripe is affected to some extent. Figure 14f shows the proposed method can effectively reduce the noise across all of the frequency components, the quality and the intelligibility of the radar speech signal are greatly improved. These results can be further proved in the Table 5.

VI. CONCLUSION

In this paper, a 94 GHz asymmetric antenna bio-radar is employed to detect speech signal detection. The structure of the asymmetric antenna bio-radar system has a high gain and the ability to obtain speech signal from remote distance. However, the original radar speech signal is always disturbed by complex noise, which include ambient, electromagnetic and electrical circuit noise. These types of noise greatly degrade the quality of the radar speech. Due to the special characteristics of the radar speech signal, a novel method based on VMD, EMD and ITS is proposed to improve the quality and the intelligibility of the original radar speech signal. In our experiments, we show that the proposed method clearly outperforms wavelet-hard, wavelet-soft and EMD shrinkage methods. Furthermore, the PESQ, STOI and CMs scores indicate that the proposed method can effectively enhance the quality and the intelligibility of the original radar speech signal.

In conclusion, the proposed method is more suitable than the other above mentioned methods for 94 GHz asymmetric antenna radar speech signal enhancement under different noise conditions. In addition, we also conduct a large number of comparative experiments to further verify the effectiveness of the proposed method in 94 GHz asymmetric antenna radar speech signal enhancement.

The detection experiments demonstrate that the 94 GHz asymmetric antenna bio-radar system can effectively detect the long-distance speech signal, and extend the capabilities of traditional speech detection methods. In the future, due to the different detection principles of the radar system, it will provide some possibilities for a wide range of applications in speech signal processing, such as pitch detection, speech production and speech recognition. Moreover, detection the vital sign of human through the radar system would also be one of the further research directions

It should be noted that the proposed algorithm is not only an effective way to improve intelligibility of the radar speech signal while reducing noise. It also has important guiding significance and application value for other millimeter wave and centimeter wave radar speech signal and traditional microphone speech signal enhancement.

REFERENCES

[1] H.-S. Kwon and K.-C. Lee, "Double-chip condenser microphone for rigid backplate using DRIE and wafer bonding technology," *Sens. Actuators A, Phys.*, vol. 138, no. 1, pp. 81–86, Jul. 2007.

[2] J. F. Holzrichter, G. C. Burnett, L. C. Ng, and W. A. Lea, "Speech articulator measurements using low power EM-wave sensors," *J. Acoust. Soc. Amer.*, vol. 103, no. 1, pp. 622–625, Jan. 1998.

[3] T. Yanagisawa and K. Furihata, "Pickup of speech signal utilization of vibration transducer under high ambient noise," *J. Acoust. Soc. Jpn.*, vol. 31, no. 3, pp. 213–220, 1975.

[4] Y. Ma, H. Hong, H. Li, H. Zhao, Y. Li, L. Sun, C. Gu, and X. Zhu, "Non-contact speech recovery technology using a 24 GHz portable auditory radar and webcam," *Remote Sens.*, vol. 12, no. 4, p. 653, 2020.

[5] Y. Avargel and I. Cohen, "Speech measurements using a laser Doppler vibrometer sensor: Application to speech enhancement," in *Proc. Joint Workshop Hands-Free Speech Commun. Microphone Arrays*, 2011, pp. 109–114.

[6] H. Lee, B.-H. Kim, J.-K. Park, and J.-G. Yook, "A novel vital-sign sensing algorithm for multiple subjects based on 24-GHz FMCW Doppler radar," *Remote Sens.*, vol. 11, no. 10, p. 1237, 2019.

[7] Y. Rong and D. W. Bliss, "Remote sensing for vital information based on spectral-domain harmonic signatures," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 6, pp. 3454–3465, Dec. 2019.

[8] S. Li, Y. Tian, G. Lu, Y. Zhang, H. Lv, X. Yu, H. Xue, H. Zhang, J. Wang, and X. Jing, "A 94-GHz millimeter-wave sensor for speech signal acquisition," *Sensors*, vol. 13, no. 11, pp. 14248–14260, 2013.

[9] S. Li, J. Q. Wang, T. Liu, and X. J. Jing, "Wavelet packet entropy for millimeter wave conducted speech enhancement," *Noise Control Eng. J.*, vol. 57, no. 5, pp. 543–550, 2009.

[10] F. Chen, S. Li, Y. Zhang, and J. Wang, "Detection of the vibration signal from human vocal folds using a 94-GHz millimeter-wave radar," *Sensors*, vol. 17, no. 3, p. 543, Mar. 2017.

[11] F. Chen, S. Li, C. Li, M. Liu, Z. Li, H. Xue, X. Jing, and J. Wang, "A novel method for speech acquisition and enhancement by 94 GHz millimeter-wave sensor," *Sensors*, vol. 16, no. 1, p. 50, Dec. 2015.

[12] C. Fuming, L. Sheng, A. Qiang, Z. Ziqi, and W. Jianqi, "Advancements in bio-radar speech signal detection technology," *Journals Radars*, vol. 5, no. 5, pp. 477–486, 2016.

[13] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, no. 2, pp. 113–120, Apr. 1979.

[14] K. Paliwal and A. Basu, "A speech enhancement method based on Kalman filtering," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 1987, pp. 177–180.

[15] T. V. Sreenivas and P. Kirnapure, "Codebook constrained Wiener filtering for speech enhancement," *IEEE Trans. Speech Audio Process.*, vol. 4, no. 5, pp. 383–389, Sep. 1996.

[16] D. L. Donoho and J. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[17] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 6, pp. 1109–1121, Dec. 1984.

[18] M. K. Hasan, S. Salahuddin, and M. R. Khan, "A modified a *priori* SNR for speech enhancement using spectral subtraction rules," *IEEE Signal Process. Lett.*, vol. 11, no. 4, pp. 450–453, Apr. 2004.

[19] N. Upadhyay and R. K. Jaiswal, "Single channel speech enhancement: Using Wiener filtering with recursive noise estimation," *Proc. Comput. Sci.*, vol. 84, pp. 22–30, Jan. 2016.

[20] A. Nicolson and K. K. Paliwal, "Deep learning for minimum mean-square error approaches to speech enhancement," *Speech Commun.*, vol. 111, pp. 44–55, Aug. 2019.

[21] Y. Ghanbari and M. R. Karami-Mollaei, "A new approach for speech enhancement based on the adaptive thresholding of the wavelet packets," *Speech Commun.*, vol. 48, no. 8, pp. 927–940, Aug. 2006.

[22] S. Li, J. Q. Wang, M. Niu, T. Liu, and X. J. Jing, "Millimeter wave conduct speech enhancement based on auditory masking properties," *Microw. Opt. Technol. Lett.*, vol. 50, no. 8, pp. 2109–2114, 2008.

[23] Y. Tian, S. Li, H. Lv, J. Wang, and X. Jing, "Smart radar sensor for speech detection and enhancement," *Sens. Actuators A, Phys.*, vol. 191, pp. 99–104, Mar. 2013.

[24] S. Li, Y. Tian, G. Lu, Y. Zhang, H. J. Xue, J.-Q. Wang, and X.-J. Jing, "A new kind of non-acoustic speech acquisition method based on millimeter waveradar," *Prog. Electromagn. Res.*, vol. 130, pp. 17–40, 2012.

[25] M. Jiao, L. Lou, J. Hu, X. Geng, W. Zhang, P. Zhang, and J. Wang, "A new speech enhancement algorithm for millimeter-wave radar speech sensor," *Microw. Opt. Technol. Lett.*, vol. 56, no. 5, pp. 1184–1189, May 2014.

[26] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London. A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, 1998.

[27] K. Khaldi, A.-O. Boudraa, A. Bouchikhi, and M. T.-H. Alouane, "Speech enhancement via EMD," *EURASIP J. Adv. Signal Process.*, vol. 2008, no. 1, pp. 1–8, Dec. 2008.

[28] N. Chatlani and J. J. Soraghan, "EMD-based filtering (EMDF) of low-frequency noise for speech enhancement," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 4, pp. 1158–1166, May 2011.

[29] L. Zão, R. Coelho, and P. Flandrin, "Speech enhancement with EMD and hurst-based mode selection," *IEEE Trans. Audio, Speech, Language Process.*, vol. 22, no. 5, pp. 899–911, May 2014.

[30] E. Deger, M. K. I. Molla, K. Hirose, N. Minematsu, and M. K. Hasan, "Subband DCT and EMD based hybrid soft thresholding for speech enhancement," *Adv. Acoust. Vibrat.*, vol. 2014, pp. 1–11, May 2014.

[31] Y. Gao, G. Ge, Z. Sheng, and E. Sang, "Analysis and solution to the mode mixing phenomenon in EMD," in *Proc. Congr. Image Signal Process.*, 2008, pp. 223–227.

[32] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2013.

[33] S. Mohanty, K. K. Gupta, and K. S. Raju, "Comparative study between VMD and EMD in bearing fault diagnosis," in *Proc. 9th Int. Conf. Ind. Inf. Syst. (ICIIS)*, 2014, pp. 1–6.

[34] Y. Wang, R. Markert, J. Xiang, and W. Zheng, "Research on variational mode decomposition and its application in detecting rub-impact fault of the rotor system," *Mech. Syst. Signal Process.*, vol. 60, pp. 243–251, Aug. 2015.

[35] Y. Li, B. Tang, and Y. Yi, "A novel complexity-based mode feature representation for feature extraction of ship-radiated noise using VMD and slope entropy," *Appl. Acoust.*, vol. 196, Jul. 2022, Art. no. 108899.

[36] Y. Li, Y. Li, X. Chen, and J. Yu, "Research on ship-radiated noise denoising using secondary variational mode decomposition and correlation coefficient," *Sensors*, vol. 18, no. 1, p. 48, Dec. 2017.

[37] X. An and J. Yang, "Denoising of hydropower unit vibration signal based on variational mode decomposition and approximate entropy," *Trans. Inst. Meas. Control*, vol. 38, no. 3, pp. 282–292, Jul. 2016.

[38] M. F. Isham, M. S. Leong, M. H. Lim, and M. K. Zakaria, "A review on variational mode decomposition for rotating machinery diagnosis," in *Proc. MATEC Web Conf.*, 2019, p. 2017.

[39] X. Zhang, Q. Miao, H. Zhang, and L. Wang, "A parameter-adaptive VMD method based on grasshopper optimization algorithm to analyze vibration signals from rotating machinery," *Mech. Syst. Signal Process.*, vol. 108, pp. 58–72, Aug. 2018.

[40] Q. Ni, J. C. Ji, K. Feng, and B. Halkon, "A fault information-guided variational mode decomposition (FIVMD) method for rolling element bearings diagnosis," *Mech. Syst. Signal Process.*, vol. 164, Feb. 2022, Art. no. 108216.

[41] B. G. Gowri, S. S. Kumar, N. Mohan, and K. P. Soman, "A VMD based approach for speech enhancement," in *Advances in Signal Processing and Intelligent Recognition Systems* (Advances in Intelligent Systems and Computing). Switzerland: Springer, 2016, pp. 309–321.

[42] R. Ram and M. N. Mohanty, "Comparative analysis of EMD and VMD algorithm in speech enhancement," *Int. J. Natural Comput. Res.*, vol. 6, no. 1, pp. 17–35, Jan. 2017.

[43] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, "Getting started with the DARPA TIMIT CD-ROM: An acoustic phonetic continuous speech database," in *National Institute of Standards and Technology*, vol. 107. Gaithersburgh, MD, USA: National Institute of Standards and Technology (NIST), 1988, p. 16.

[44] R. T. Rockafellar, "A dual approach to solving nonlinear programming problems by unconstrained optimization," *Math. Program.*, vol. 5, no. 1, pp. 354–373, Dec. 1973.

[45] J. Benesty, J. Chen, and Y. Huang, "On the importance of the Pearson correlation coefficient in noise reduction," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 4, pp. 757–765, May 2008.

[46] A. O. Boudraa, J. C. Cexus, and Z. Saidi, "EMD-based signal noise reduction," *Int. J. Signal Process.*, vol. 1, no. 1, pp. 33–37, Jan. 2004.

[47] Y. Hu and P. C. Loizou, "Evaluation of objective measures for speech enhancement," in *Proc. 9th Int. Conf. Spoken Lang. Process.*, 2006, pp. 1–4.

[48] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ)—A new method for speech quality assessment of telephone networks and codecs," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2001, pp. 749–752.

[49] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," *IEEE Trans. Audio. Speech. Lang. Process.*, vol. 19, no. 7, pp. 2125–2136, Dec. 2011.

[50] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Commun.*, vol. 12, no. 3, pp. 247–251, Jul. 1993.

FUMING CHEN was born in 1989. He received the Ph.D. degree from Fourth Military Medical University (FMMU), Xi'an, in 2017. He is currently a Senior Engineer with The 940th Hospital of Joint Logistic Support Force, China. His research interests include biomedical signal detection and processing.

JIANQI WANG was born in Xi'an, Shaanxi, China, in 1962. He received the B.E. degree from Xi'an Jiaotong University, Xi'an, in 1984, the M.E. degree from the National University of Defense Technology, Changsha, China, in 1990, and the Ph.D. degree from the Key Laboratory of the Ministry of Education of China, Xi'an Jiaotong University, in 2006. Since 1990, he has with the School of Biomedical Engineering, Fourth Military Medical University, Xi'an. He is currently working as a Professor and the Director of the Department of Electronics. He pioneered radar-based human being detection in China, in 1998. He has published more than 100 articles on the technology. His research interests include bioradar technology, including signal processing, human being detection, and imaging.

CHUANTAO LI received the B.E. degree from the School of Electrical Engineering, Zhengzhou University, Zhengzhou, China, in 2010, the M.E. degree from the School of Electrical Engineering, Ordnance Engineering College, Shijiazhuang, China, in 2013, and the Ph.D. degree from Fourth Military Medical University (FMMU), Xi'an, in 2016. He is currently an Assistant Research Fellow with the Aviation Physiological and Psychological Training Team, Second Military Medical University (SMMU). His research interests include the application of deep learning algorithms in the medical field and biomedical processing.