Frequency-domain distributed multichannel wiener filtering speech enhancement algorithm

Jingxian Tu, *Guijiang* Qin*, and *Haifeng* Lv

School of Big Data and Software Engineering, Wuzhou University,Wuzhou ,China

Abstract. A frequency-domain distributed microphone multi-channel Wiener filter speech enhancement algorithm is proposed in this paper. In this paper, the distributed microphone speech model is considered. First, the speech signal in the time domain is converted into the speech signal in the frequency domain by the discrete Fourier transform method. Then, the unconstrained minimization problem of the noise reduction and speech distortion of the complex linear filter in the frequency domain is established. Simulation results show that the proposed algorithm is superior to some existing multi-channel speech enhancement algorithms.

Keywords: Speech enhancement, Distributed microphone, Complex value filter, Noise reduction, Speech distortion.

1 Introduction

The collected speech signals are often polluted by environmental noise.In order to eliminate or reduce noise, speech enhancement algorithms can be used to process noisy speech[1-2].Multi-channel speech enhancement algorithms can simultaneously use the temporal and spatial information of multiple channel speech signals, thus achieving better denoising performance than single-channel speech enhancement algorithms[3-4].Distributed microphone multi-channel speech enhancement algorithms are suitable for processing distributed microphone speech models.In this model, the room reverberation can often be ignored due to the large size of the room where the microphones are located and the large distance between each microphone [5].Therefore, the distributed microphone speech model is a multi-channel speech model without reverberation.In recent years, scholars have proposed some distributed microphone multi-channel speech enhancement algorithms, which are distributed microphone multi-channel speech enhancement algorithms based on statistical models and distributed microphone multi-channel speech enhancement algorithms based on Kalman filtering[6-7].In this paper, a distributed microphone multi-channel Wiener filtering speech enhancement algorithm in the frequency domain is proposed. In this algorithm, the time domain signal is first converted to frequency domain signal, and the complex value linear filter is applied to the observed speech signal in frequency domain. Then, the measurements of noise reduction and speech distortion in frequency domain are given. Then, the unconstrained minimization optimization problem

* Corresponding author: 568205127@qq.com

of the complex value filter in frequency domain is established. Finally, the expression of the optimal filter is obtained by solving it.Simulation results show the effectiveness of the proposed algorithm.

2 The algorithm proposed in this paper

2.1 Signal model

Consider a distributed microphone system that can accurately compensate the time delay of an observed signal. The distributed multichannel speech model in time domain can be expressed as:

$$y_i(n) = c_i s(n) + v_i(n), \quad i = 1, \cdots, M, \tag{1}$$

where $y_i(n)$ and $v_i(n)$ are respectively the observed speech signal and noise signal of the i-th channel in time n, $s(n)$ is the source speech signal in time n and $c_i \in [0,1]$ is the speech attenuation factor in the i-th channel. The short-time Fourier transform is applied to model (1) to obtain a distributed multichannel speech model in frequency domain

$$Y_i(k,l) = c_i S(k,l) + V_i(k,l), \tag{2}$$

where k and l represent the frequency label and frame label, respectively, and $Y_i(k,l)$, $S(k,l)$ and $V_i(k,l)$ represent the discrete fourier transform coefficients of $y_i(n), s(n)$ and $v_i(n)$, respectively. For simplicity, we remove k and l from the following formula unless otherwise stated. Formula (2) can be written in vector form:

$$\mathbf{Y} = S\mathbf{c} + \mathbf{V}, \tag{3}$$

where $\mathbf{Y} = \left[Y_1, \cdots, Y_M\right]^T$, $\mathbf{V} = \left[V_1, \cdots, V_M\right]^T$ and $\mathbf{c} = \left[c_1, \cdots, c_M\right]^T$. The purpose of speech enhancement is to estimate S from $\left\{Y_i\right\}_{i=1}^M$.

2.2 The optimal filter

Considering that the target signal is recovered by applying a complex linear filter to the observed signal vector, the output signal can be expressed as:

$$\hat{S} = \mathbf{w}^H \mathbf{Y} = \mathbf{w}^H \mathbf{c} S + \mathbf{w}^H \mathbf{V}, \tag{4}$$

where $\mathbf{w} \in C^{M \times 1}$ is a complex linear filter and ^{H} is a conjugate transpose.The residual signal is defined as:

$$r(\mathbf{w}) = \hat{S} - S = \mathbf{w}^H \mathbf{c} S + \mathbf{w}^H \mathbf{V} - S = \left(\mathbf{w}^H \mathbf{c} - 1\right) S + \mathbf{w}^H \mathbf{V}. \tag{5}$$

Let $r_S(\mathbf{w}) = \left(\mathbf{w}^H \mathbf{c} - 1\right) S$ represents speech distortion and $r_V(\mathbf{w}) = \mathbf{w}^H \mathbf{V}$ represents residual noise, then the energy of speech distortion and residual noise can be written as:

$$J_S(\mathbf{w}) = E\left[\left(\left(\mathbf{w}^H \mathbf{c} - 1\right) S\right)\left(\left(\mathbf{w}^H \mathbf{c} - 1\right) S\right)^H\right] = P_S\left(\mathbf{w}^H \mathbf{c} - 1\right)\left(\mathbf{w}^H \mathbf{c} - 1\right)^H = P_S\left(\mathbf{w}^H \mathbf{c}\mathbf{c}^H \mathbf{w} - \mathbf{w}^H \mathbf{c} - \mathbf{c}^H \mathbf{w} + 1\right), \tag{6}$$

$$J_V(\mathbf{w}) = E\left[\left(\mathbf{w}^H \mathbf{V}\right)\left(\mathbf{w}^H \mathbf{V}\right)^H\right] = \mathbf{w}^H \Phi_V \mathbf{w}, \tag{7}$$

where $P_S = E\left[SS^H\right]$ is the speech signal power spectrum and $\Phi_V = E\left[\mathbf{V}\mathbf{V}^H\right]$ is the noise signal autocorrelation matrix.

By introducing the weight parameter that compromises noise reduction and speech distortion, ,The optimization problem of frequency domain distributed microphone multichannel Wiener filter (FD-DM-MWF) speech enhancement algorithm in frequency domain can be written as:

$$\min_{\mathbf{w}} f(\mathbf{w}) = J_S(\mathbf{w}) + \mu J_V(\mathbf{w}) = P_S\left(\mathbf{w}^H\mathbf{c}\mathbf{c}^H\mathbf{w} - \mathbf{w}^H\mathbf{c} - \mathbf{c}^H\mathbf{w} + 1\right) + \mu\mathbf{w}^H\Phi_V\mathbf{w}, \tag{8}$$

where μ is the weight parameter. This is a convex optimization unconstrained optimization problem. Let the partial derivative of the objective function on **w** be equal to 0, the optimal filter can be obtained as:

$$\mathbf{w}_{FD-DM-MWF} = P_S\left(P_S\mathbf{c}\mathbf{c}^H + \mu\Phi_V\right)^{-1}\mathbf{c}. \tag{9}$$

2.3 Algorithm description

When the algorithm is implemented, the parameters need to be estimated. In the speech frame, Φ_V is not updated. In the noise frame, the first order smoothing method is used to estimate Φ_V as:

$$\Phi_V = \lambda\Phi_V + (1-\lambda)VV^H, \tag{10}$$

where λ is the smoothing factor. c_i is calculated by the following formula:

$$c_i = \frac{\sqrt{\sigma_{y_i}^2 - \sigma_{v_i}^2}}{\sqrt{\sigma_{y_1}^2 - \sigma_{v_1}^2}},\ i = 1,2,\cdots,M, \tag{11}$$

where $\sigma_{y_i}^2$ and $\sigma_{v_i}^2$ are the variance of the observed speech signal and the noise signal in the *i*-thchannel, respectively. By adding the *M* channel signals in formula (2), it can be obtained:

$$\overline{Y} = \overline{c}S + \overline{V}, \tag{12}$$

where $\overline{Y} = \sum_{i=1}^{M} Y_i$, $\overline{V} = \sum_{i=1}^{M} V_i$ and $\overline{c} = \sum_{i=1}^{M} c_i$. P_S is estimated by:

$$P_S = \frac{P_{\overline{Y}} - P_{\overline{V}}}{\overline{c}}, \tag{13}$$

where $P_{\overline{Y}} = E\left[\overline{Y}\overline{Y}^H\right]$ and $P_{\overline{V}} = E\left[\overline{V}\overline{V}^H\right]$. In the speech frame, $P_{\overline{Y}}$ is estimated by:

$$P_{\overline{Y}} = \lambda P_{\overline{Y}} + (1-\lambda)\left|\overline{Y}\right|^2, \tag{14}$$

In the noise frame, $P_{\overline{V}}$ is estimated by:

$$P_{\overline{V}} = \lambda P_{\overline{V}} + (1-\lambda)\left|\overline{V}\right|^2. \tag{15}$$

The input and output of the proposed algorithm are *M* channels of observed speech signals and estimated source speech signals respectively. The key steps of the algorithm proposed in this paper are as follows:

Step 1: The observed speech signals in the time domain are converted into the signals in the frequency domain by the discrete Fourier transform;

Step 2: Using the method based on short-time energy and zero-crossing rate to detect the speech activity of the observed signal in the time domain;

Step 3: If the current frame is a speech frame, $P_{\bar{Y}}$ is estimated by formula (14); Otherwise, $P_{\bar{V}}$ is estimated by formula (15), and Φ_V is estimated by formula (10);

Step 4: Estimate P_S by formula (13);

Step 5: Estimate c_i by formula (11);

Step 6: Estimate $\mathbf{w}$ by formula (9);

Step 7: Estimate S by $\hat{S} = \mathbf{w}^H \mathbf{Y}$;

Step 8: The estimated speech signals are converted into the time domain speech signals through the inverse discrete Fourier transform.

3 Experiment

The source speech signal is spliced by 20 different sentences selected from the NOISEUS speech library[8], while the noise signal is selected from the NOISEX noise library [9], and the sampling frequency of both speech and noise signal is 8000 Hz. Four objective evaluation indexes are selected to evaluate the quality of the speech signal output by the algorithm. These four indicators are segmental signal to noise ratio improvement (SSNRI), log-likelihood ratio (LLR), perceptual evaluation of speech quality (PESQ) and short-term objective intelligibility (STOI). The larger the SNRI, PESQ, and STOI values, the better the output speech quality, and the smaller the LLR values, the better the output speech quality[10-11]. The attenuation factor of the *i*-th channel is set to $c_i = (2M - i + 1)/(2M)$, where. In this paper, the number of distributed microphone channels is increased from 1 to 10 one by one, the noise is babble and factory, and the input SNR is -5 dB and 5 dB. The programming platform is MATLAB. In this paper, the one-rank multi-channel Wiener filtering algorithm (denoted as R1-MWF) proposed in reference[12] and the speech distortion weighted multi-channel Wiener filtering algorithm (denoted as SDW-MWF) proposed in reference[13] are used as a comparison algorithm.

FIG. 1-4 shows the experimental results of three algorithms in SSNRI, LLR, PESQ and STOI under babble noise and factory noise. From these four graphs, at least two conclusions can be drawn. First, when the number of channels is greater than 1, the algorithm proposed in this paper is superior to other two traditional multi-channel speech enhancement algorithms under four objective evaluation indexes. Second, generally speaking, when the number of channels is increased, the four indexes of the algorithm proposed in this paper will become better and better. With the increase of the number of channels, the proposed algorithm can utilize more spatial information about the speech signal, so as to achieve better denoising performance.

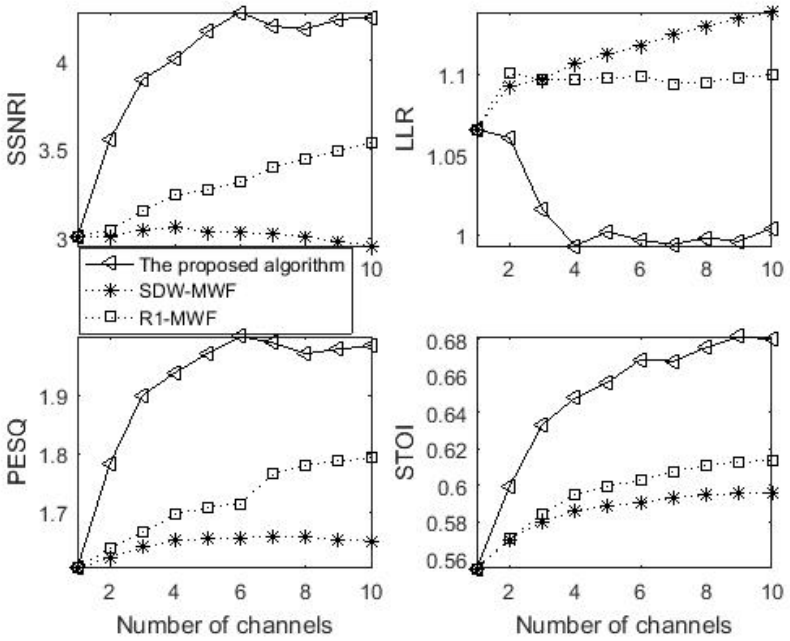

Fig. 1. Comparison of experimental results of the three algorithms with babble noise and an input SNR of -5 dB.

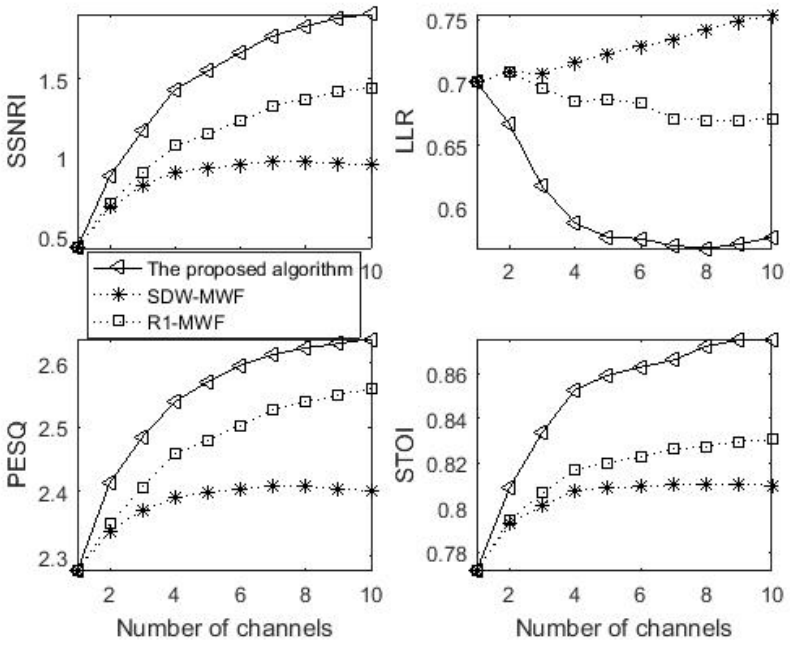

Fig. 2. Comparison of experimental results of the three algorithms with babble noise and an input SNR of 5 dB.

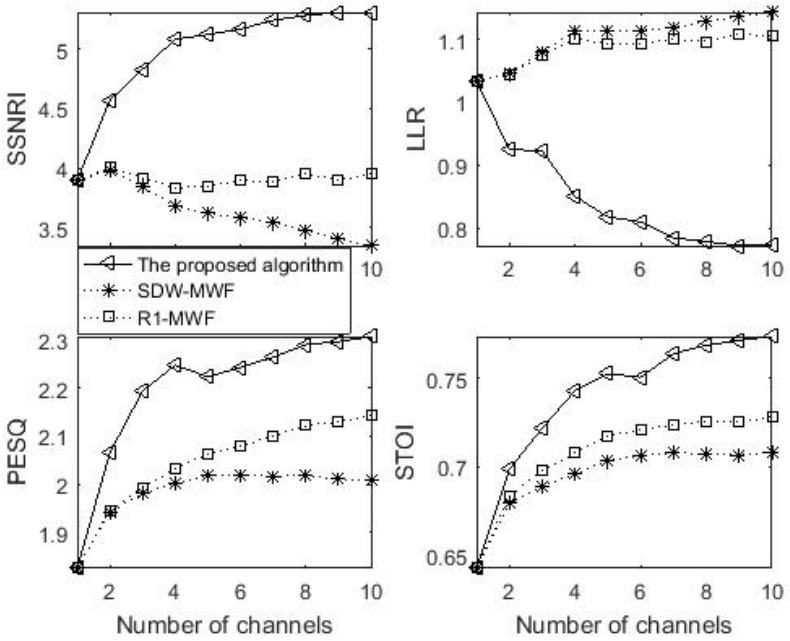

Fig. 3. Comparison of experimental results of the three algorithms with factory noise and an input SNR of -5 dB.

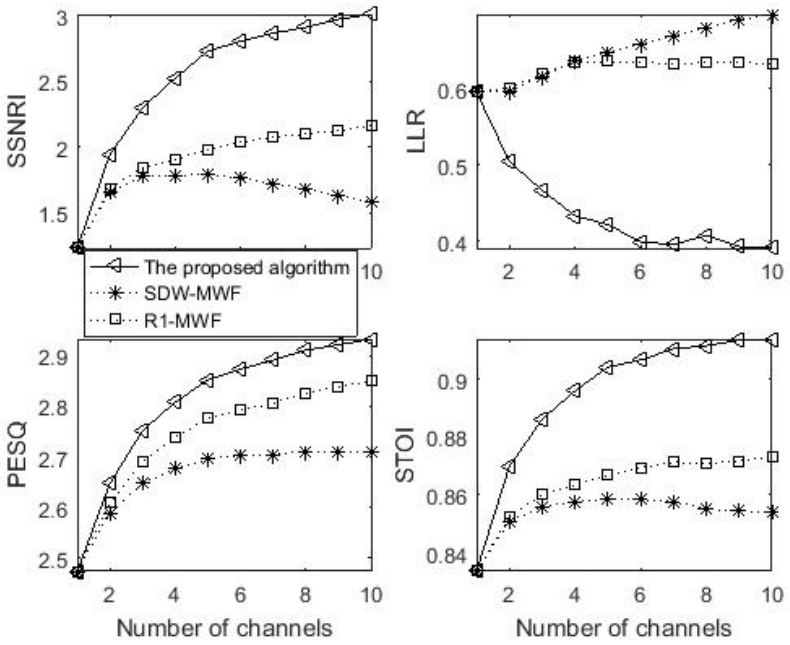

Fig. 4. Comparison of experimental results of the three algorithms with factory noise and an input SNR of -5 dB.

4 Conclusions

A frequency-domain distributed microphone multi-channel Wiener filter speech enhancement algorithm is proposed in this paper. For the distributed microphone speech model, the time domain signal is transformed into frequency domain signal by discrete Fourier transform method, and then the observed signal in frequency domain is filtered by complex value linear. In order to obtain the optimal complex-valued linear filter, the unconstrained minimization problem about the tradeoff between noise reduction and speech distortion of complex-valued linear filters is established. Simulation results show that the proposed algorithm is superior to some existing multi-channel speech enhancement algorithms.

This work is supported by Scientific Research Fund of Guangxi Zhuang Autonomous Region under Grant No. 2018JJB170034.

References

1. Wenhao Yuan 2021 Incorporating group update for speech enhancement based on convolutional gated recurrent network *Speech Communication* vol 132 pp 32–39
2. Junqin Wu and Yingfu Wang 2022 A low-delay speech enhancement algorithm with improved window function *Computer Simulation* vol 39 pp 203-211
3. Zhiheng Cui Jiye Jiao and Zhentian Zhu 2022 Research and implementation of Dual microphone speech enhancement Algorithm *Electronic Design Engineering* vol 30 pp 109-114
4. Yatao Zhu Fei Chen and Yuchen Zhang et al. 2021 Recurrent neural network-based speech enhancement algorithm for binaural hearing aids Chinese Journal of Sensing Technology vol 34 pp 1165-1172
5. Trawicki M B and Johnson M T 2012 Distributed multichannel speech enhancement with minimum mean-square error short-time spectral amplitude,log-spectral amplitude,and spectral phase estimation *Signal Processing* vol 9 pp 345-356
6. Jingxian Tu and Youshen Xia 2015 Fast distributed multichannel speech enhancement using novel frequency domain estimators of magnitude-squared spectrum *Speech Communication* vol 72 pp 96-108.
7. Jingxian Tu and Youshen Xia 2018 Effective kalman filtering algorithm for distributed multichannel speech enhancement *Neurocomputing* vol 275 pp 144-154
8. Loizou P C 2007 Speech Enhancement: Theory and Practice (Boca Raton:CRC)
9. Varga A and Steeneken H J M 1993 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 Communication* vol 12 pp 247-251
10. Hu Y and Loizou P C 2008 Evaluation of objective quality measures for speech enhancement *IEEE Transactions on Audio, Speech and Language Processing* vol 16 pp 229-238
11. Taal C H Hendriks R C and Heusdens R 2011 An algorithm for intelligibility prediction of time-frequency weighted noisy speech *IEEE Trans. Audio Speech Lang. Process.* vol 19 pp 2125–2136
12. Serizel R Moonen B and Van Dijk M et al. 2013 Rank-1 approximation based multichannel wiener filter algorithms for noise reduction in cochlear implants *in Proc. 38th Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2013)* pp 8634-8938

13. Doclo S Spriet A and Wouters J et al. 2007 Frequency-domain criterion for the speech distortion weighted multichannel wiener filter for robust noise reduction *Speech Communication* vol 49 pp