Weighted Spatial Bispectrum Correlation Matrix for DOA Estimation in the Presence of Interferences

Wei Xue, Shan Liang, Wenju Liu

National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

{wxue,sliang,lwj}@nlpr.ia.ac.cn

Abstract

Besides the undirected environmental noise, the surrounding interference also brings great challenges to the robust DOA estimation of the speech source. As conventional DOA estimation methods always assume an undirected noise model, they usually cannot perform reliably when the strong inference exists. In this paper, we propose a novel interference robust DOA estimation method, which is based on the "weighted spatial bispectrum correlation matrix (WSBCM)". The WSBCM contains the spatial correlation information of bispectrum phase difference (BPD), and a new DOA estimator is further derived based on the eigenvalue analysis of the WSBCM. By formulating with WSBCM, the proposed method benefits from the redundant DOA-related information provided by the BPD. In addition, the WSBCM enables bispectrum weighting to highlight the pure speech units in the bispectrum, which further helps to improve the performance. In order to compute the bispectrum weights, a decision-directed approach is derived. The effectiveness of the proposed method is demonstrated by experiments conducted under various kinds of interference-existing scenarios.

Index Terms: direction of arrival estimation, microphone array signal processing, bispectrum, interference

1. Introduction

The knowledge of the direction of arrival (DOA) of the speech source is of great interest in various applications including hands-free communication systems, teleconferencing, automatic camera steering, etc. In these applications, a microphone array is commonly deployed, and the DOA can be estimated by exploiting the spatial diversity of the distributed microphones.

One major concern for the DOA estimation problem is the environmental noise. The noise may be undirected, or emitted by the surrounding interference. However, almost all conventional methods such as high-resolution spectral estimation [1,2,3], steered beamformer response power [4,5], and time difference of arrival (TDOA) estimation [6,7] just assume that the noise signals received by different microphones are undirected. As a result, when the strong directional interference exists, due to the inaccuracy of the noise model, the robustness of these algorithms cannot be guaranteed.

In our previous work [8], an interference robust DOA estimation method had been presented. We proposed a frequency-domain DOA estimator which enabled frequency weighting while avoiding estimating DOAs separately in each band, and the robustness against the interference was improved by selecting only the speech frequency bands using frequency weights. In order to compute the weights, the historical information of DOA estimates and the temporal correlation of the consecutive speech spectra were exploited. Clearly, among the whole set of frequency bands, only a few of them are the speech ones, which contain the cues related to the DOA of the speech source. Then it can be further deduced that, if the interference has a flat frequency distribution, the DOA cues in the speech frequency bands will be more likely polluted by the interference, making the algorithm less robust. This inference has been demonstrated in the experiments of the previous work by comparing between the performances under different interferences.

Finding the DOA cues of the speech source is crucial for the interference robust DOA estimation. For the frequency domain representation, once the cue in one speech band is polluted by the interference, it cannot be found elsewhere. In this paper, we propose a novel method which is formulated in the bispectrum domain. Compared with the frequency domain formulation, the proposed method has a theoretical advantage that the DOA cues are expressed redundantly in the bispectrum, meaning that even the cue of the speech source is polluted in one bispectrum unit, it may be re-found in other unpolluted units. This advantage makes it possible to improve the robustness of the algorithm. We select the speech units with a set of bispectrum weights, and integrate the DOA cues in these units into the "WSBCM". The WSBCM is a function of a hypothesized DOA, and shows interesting property only when the hypothesized DOA equals to the true one. A new DOA estimator is then derived based on the eigenvalue analysis of the WSBCM. In order to compute the bispectrum weights, we propose a decision-directed approach. By conducting experiments under various kinds of interference-existing conditions, we demonstrate the effectiveness of the proposed method.

2. Signal Model

We consider the problem in an environment with an M-element uniform linear microphone array (ULA), a speech source and several interferences. The speech source and interferences are all in the far field [9] and uncorrelated with each other.

Suppose that the signals are equally attenuated from sources to each microphone. If we choose the first microphone as the reference point, the signal received by the mth microphone ($m = 1, \ldots, M$) at time k can be expressed as:

$$\begin{aligned} y_m(k) &= s_m(k) + v_m(k) + n_m(k) \\ &= s_1(k - \tau_{m1}) + v_m(k) + n_m(k), \end{aligned} \tag{1}$$

where $s_m(k)$ is the speech signal received by the mth microphone, which is a delayed version of the speech signal received by the first microphone, and τ_{m1} indicates the time delay. As only the speech signal which is related to the DOA of the speech source is of our interest, we ignore the details of the interference

signal received by the microphone, and simply represent it as $v_m(k)$. The $n_m(k)$ denotes the additive white Gaussian noise.

The time delay between the first and mth microphone τ_{m1} can be determined according to the array geometry, which is derived as follows:

$$\tau_{m1} = (m-1)\frac{\sin(\hat{\theta}) f_s d}{c}, m = 1, 2, ..., M, \quad (2)$$

where c is the speed of sound in the air, f_s is the sampling rate, d is the spacing between two adjacent microphones, and $\hat{\theta}$ is the true DOA of the speech source to be estimated.

3. Proposed Method

3.1. Bispectrum Phase Difference

The bispectrum is a kind of high order statistics (HOS) of the signal. A promising property of HOS is that the HOS of the Gaussian signal is always zero. In fact, several HOS based DOA estimation methods have been proposed to improve the robustness against Gaussian noise by utilizing this property [10,11,12]. However, these methods are generally applied for narrowband signals. Moreover, they are not capable to deal with the non-Gaussian directional interferences whose bispectra are non-zero.

In this subsection, we analyze the bispectra of the signals captured by the ULA in a different way, and show that the redundancy provided by the "BPD" helps to improve the performance in the interference-existing scenarios.

The bispectrum is a function of two bi-frequency variables Ω_1 and Ω_2, and it analyzes the frequency interactions between the frequency components at Ω_1, Ω_2 and $\Omega_1 + \Omega_2$. Due to space limitations, the theoretical basis of the bispectrum is not introduced here, and the readers can refer to [13][14]. As most speech signals have asymmetric possibility density functions, their skewness are non-zero [15]. Thus, it is reasonable to analyze bispectrum of the speech signal.

Recall the signals received by first and mth microphone. Following Eq.(1), they can be rewritten as follows:

$$\begin{aligned} y_1(k) &= s_1(k) + v_1(k) + n_1(k) \\ y_m(k) &= s_1(k - \tau_{m1}) + v_m(k) + n_m(k). \end{aligned} \quad (3)$$

Since $n_1(k)$ and $n_m(k)$ are zero-mean Gaussian, their bispectra are identical to zero. According to the derivation in [14], and under the assumption that the speech source is uncorrelated with the interferences, the bispectrum of $y_1(k)$ and the cross-bispectrum between $y_1(k)$, $y_m(k)$ can be derived as:

$$\begin{aligned} B_{y^{111}}(\Omega_1, \Omega_2) &= B_{s^{111}}(\Omega_1, \Omega_2) + B_{v^{111}}(\Omega_1, \Omega_2) \\ B_{y^{1m1}}(\Omega_1, \Omega_2) &= B_{s^{111}}(\Omega_1, \Omega_2) e^{j\Omega_1 \tau_{m1}} + B_{v^{1m1}}(\Omega_1, \Omega_2), \end{aligned} \quad (4)$$

where $B_{x^{abc}}(\Omega_1, \Omega_2)$ stands for the bispectrum of signal $x_a(k)$,$x_b(k)$ and $x_c(k)$. We define the BPD as the ratio between $B_{y^{1m1}}$ and $B_{y^{111}}$, and substitute τ_{m1} using Eq.(2), then:

$$\begin{aligned} I_{m1}(\Omega_1, \Omega_2) &\stackrel{\text{def}}{=} \frac{B_{y^{1m1}}(\Omega_1, \Omega_2)}{B_{y^{111}}(\Omega_1, \Omega_2)} \\ &= e^{j\Omega_1 \frac{(m-1)\sin(\hat{\theta}) f_s d}{c}} (1 - \epsilon_m(\Omega_1, \Omega_2)), \end{aligned} \quad (5)$$

where

$$\epsilon_m(\Omega_1, \Omega_2) = \frac{B_{v^{111}} - B_{v^{1m1}} e^{-j\Omega_1 \tau_{m1}}}{B_{s^{111}} + B_{v^{111}}}. \quad (6)$$

In Eq.(5), the BPD is expressed as a product of two terms. For the fixed ULA geometry and sampling rate, the first term is only related to the DOA of the speech source and the bi-frequency Ω_1, and we call it the "speech DOA cue". The second term reflects how much the speech DOA cue is polluted by interferences. In speech-dominated bispectrum units, according to Eq.(6), the $\epsilon_m(\Omega_1, \Omega_2)$ is close to 0, which makes the BPD approximates the real speech DOA cue.

One interesting property which can be seen from Eq.(5) is that, although the BPD is defined in each bispectrum unit (Ω_1, Ω_2), actually, the speech DOA cue in BPD is not a function of the bi-frequency Ω_2. In other words, as long as two bispectrum units have the same Ω_1, whatever the values of their Ω_2 are, they have redundant speech DOA cues in the BPD.

As the bispectra of the speech and interferences distribute differently, in some speech units, the speech DOA cues may be severely polluted, nevertheless, they can be re-found in other speech-dominant units with the same Ω_1. In the speech-dominant units, the speech DOA cue can approximated by the BPD. In contrast, for the frequency domain representation, once the cue in one speech band is polluted by the interference, it cannot be found elsewhere. This property of BPD helps to improve the robustness of the algorithm against the interferences.

3.2. Bispectrum Weights

Since the real DOA cue is close to the BPD only in speech-dominant bispectrum units, it is better to pick out only these units for DOA estimation. Here, a set of non-negative weights are utilized to select these speech units, and the weights are expected to have large values in pure speech units, and zero values in the interference ones.

Actually, the bispectrum weight, which can be viewed as an indicator of how much the speech signal is polluted in one unit, is analogous to the signal-to-noise ratio (SNR) in the frequency domain. Therefore, the ideas for SNR estimation can be exploited to compute the bispectrum weights. In this subsection, the bispectrum weights are computed by a "decision-directed" method, which is similar to the decision-directed *a priori* SNR estimator proposed by Ephraim and Malah [16].

We use the bispectrum of the first microphone signal $B_{y^{111}}(\Omega_1, \Omega_2)$ for the weight calculation. Define the local *a priori* bispectrum signal-to-interference ratio (BSIR) $\xi(\Omega_1, \Omega_2)$ and *a posteriori* BSIR $\gamma(\Omega_1, \Omega_2)$ of the bispectrum unit (Ω_1, Ω_2) as:

$$\xi(\Omega_1, \Omega_2) \stackrel{\text{def}}{=} \frac{|B_{s^{111}}(\Omega_1, \Omega_2)|^2}{\lambda_v(\Omega_1, \Omega_2)}, \quad (7)$$

$$\gamma(\Omega_1, \Omega_2) \stackrel{\text{def}}{=} \frac{|B_{y^{111}}(\Omega_1, \Omega_2)|^2}{\lambda_v(\Omega_1, \Omega_2)}, \quad (8)$$

where $\lambda_v(\Omega_1, \Omega_2) \stackrel{\text{def}}{=} E\{|B_{v^{111}}(\Omega_1, \Omega_2)|^2\}$ is the estimation of interference bispectrum power, which can be obtained and updated during periods of silence. Assuming the speech and interferences are uncorrelated, it is clear that:

$$\xi(\Omega_1, \Omega_2) = \gamma(\Omega_1, \Omega_2) - 1. \quad (9)$$

Following Eq.(7) and Eq.(9), the *a priori* BSIR $\xi(\Omega_1, \Omega_2)$ is estimated as:

$$\begin{aligned} \hat{\xi}^t(\Omega_1, \Omega_2) =& \alpha \frac{|B^{t-1}_{s^{111}}(\Omega_1, \Omega_2)|^2}{\lambda_v(\Omega_1, \Omega_2)} \\ &+ (1 - \alpha) P[\gamma^t(\Omega_1, \Omega_2) - 1]. \end{aligned} \quad (10)$$

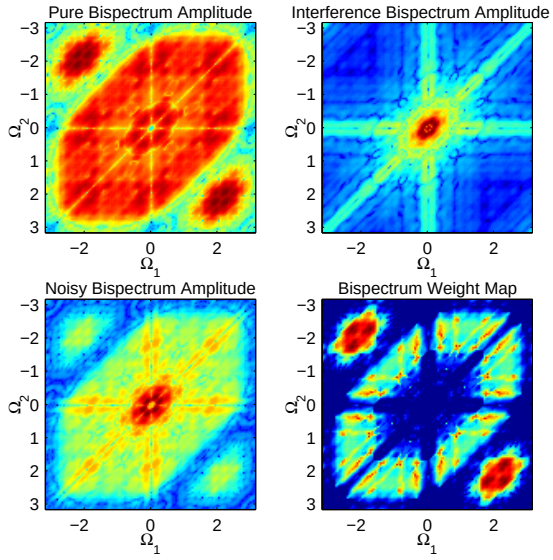

Figure 1: *Example of the calculated bispectrum weights. Noisy environment: car interior noise, SIR=5dB and T60=250ms*

where $P[\cdot]$ denotes half-wave rectification, and the superscript $(\cdot)^t$ indicates the time frame t. $\alpha \in [0,1]$ is a smoothing factor, which is set to be 0.98. The $|B^{t-1}_{s111}(\Omega_1,\Omega_2)|^2$ in the right side of Eq.(10) is unknown, and it can be approximated as:

$$\begin{aligned}|B^{t-1}_{s111}|^2 &= (|B^{t-1}_{s111}|^2/|B^{t-1}_{y111}|^2) * |B^{t-1}_{y111}|^2 \\ &= [|B^{t-1}_{s111}|^2/(|B^{t-1}_{s111}|^2+\lambda_v)] * |B^{t-1}_{y111}|^2 \\ &\approx [\hat{\xi}^{t-1}/(\hat{\xi}^{t-1}+1)] * |B^{t-1}_{y111}|^2. \end{aligned} \quad (11)$$

Substituting Eq.(11) into Eq.(10), we have:

$$\hat{\xi}^t = \alpha\gamma^{t-1}\frac{\hat{\xi}^{t-1}}{\hat{\xi}^{t-1}+1} + (1-\alpha)P[\gamma^t - 1]. \quad (12)$$

In Eq.(11)(12), "(Ω_1,Ω_2)" is omitted for simplicity.

From the definition of $\xi(\Omega_1,\Omega_2)$ in Eq.(7), it can be seen that when the unit is dominated by the interference, $\xi(\Omega_1,\Omega_2)$ will be less than 1. We define the bispectrum weight as:

$$w(\Omega_1,\Omega_2) \stackrel{\text{def}}{=} |B^t_{y111}(\Omega_1,\Omega_2)|\{[\max(\hat{\xi}^t(\Omega_1,\Omega_2),1)]^p - 1\}, \quad (13)$$

where p is a factor controlling the importance of high B-SIR units, which is set to be 0.8 here. By multiplying with $|B^t_{y111}(\Omega_1,\Omega_2)|$, the units belonging to neither the speech nor interference are excluded. Obviously the bispectrum weights will be equal to zero in the interference-dominated units.

Fig.1 shows an example of the calculated bispectrum weights (plus a small positive number and transform to the log-scale) and the log-scale amplitude of the bispectra corresponding to the clean speech, interference and noisy speech in one frame, respectively. It can be observed the effect of the interference has been almost totally removed.

3.3. Weighted Spatial Bispectrum Correlation Matrix

The speech-dominant bispectrum units can be chosen as units with large bispectrum weights. For each unit (Ω_1,Ω_2), BPDs of multiple microphones are available. In order to express the procedure of "selecting speech units from BPDs of multiple microphones, and integrating the selected speech cues for DOA estimation", into a compact framework, a matrix called "WSBCM" is formulated. The WSBCM reflects the spatial correlations between the phase aligned BPDs for a hypothesized DOA, and shows an interesting property only when the hypothesized DOA equals to the true one.

We define the BPD vector in the unit (Ω_1,Ω_2) as:

$$\mathbf{I}(\Omega_1,\Omega_2) \stackrel{\text{def}}{=} [I_{11}(\Omega_1,\Omega_2),\ldots,I_{M1}(\Omega_1,\Omega_2)]^T. \quad (14)$$

Inspired by the expression of speech DOA cue in Eq.(5), a BPD compensation vector for a hypothesized DOA θ is defined as:

$$\mathbf{C}(\theta,\Omega_1) \stackrel{\text{def}}{=} [1, e^{-j\Omega_1\frac{\sin(\theta)f_s d}{c}},\ldots,e^{-j\Omega_1\frac{(M-1)\sin(\theta)f_s d}{c}}]^T. \quad (15)$$

Then we compensate the BPDs using $\mathbf{C}(\theta,\Omega_1)$ as follows:

$$\mathbf{I}^C(\theta,\Omega_1,\Omega_2) \stackrel{\text{def}}{=} \mathbf{I}(\Omega_1,\Omega_2)\circ\mathbf{C}(\theta,\Omega_1), \quad (16)$$

where $\mathbf{I}^C(\theta,\Omega_1,\Omega_2)$ is the phase-compensated BPD vector, and the symbol "$\circ$" stands for the element-wise product. Once θ is equal to $\hat{\theta}$, according to Eq.(5)(14)(15)(16),

$$\mathbf{I}^C(\theta,\Omega_1,\Omega_2) = \mathbf{\Gamma} + \mathbf{E}(\theta,\Omega_1,\Omega_2), \quad (17)$$

where $\mathbf{\Gamma} = [1,1,\ldots,1]^T$, which indicates that BPDs are phase aligned, and $\mathbf{E}(\theta,\Omega_1,\Omega_2)$ is an error term which equals to $-[\epsilon_1(\Omega_1,\Omega_2)\ldots\epsilon_m(\Omega_1,\Omega_2)]^T\circ\mathbf{C}(\theta,\Omega_1)$.

With the bispectrum weights computed using Eq.(13), we define the WSBCM which contains the spatial correlation information of the phase aligned BPDs for θ as follows:

$$\mathbf{R}(\theta) \stackrel{\text{def}}{=} \sum_{(\Omega_1,\Omega_2)} w(\Omega_1,\Omega_2)[\mathbf{I}^C(\theta,\Omega_1,\Omega_2)](\mathbf{I}^C(\theta,\Omega_1,\Omega_2))^H. \quad (18)$$

We assume that in the speech units, the error term in Eq.(17) can be ignored. Then, once the hypothesized DOA θ is equal to $\hat{\theta}$, according to Eq.(17)(18),

$$\begin{aligned}\mathbf{R}(\theta) &= \sum_{(\Omega_1,\Omega_2)\in\mathbf{\Omega_S}} w(\Omega_1,\Omega_2)\mathbf{\Gamma}\mathbf{\Gamma}^H \\ &= \eta\cdot\begin{bmatrix}1 & \ldots & 1\\ \vdots & \ddots & \vdots\\ 1 & \ldots & 1\end{bmatrix}, \end{aligned} \quad (19)$$

where $\mathbf{\Omega_S}$ denotes the set of speech units, and $\eta = \sum_{(\Omega_1,\Omega_2)\in\mathbf{\Omega_S}} w(\Omega_1,\Omega_2)$ is a constant in a certain frame given $w(\Omega_1,\Omega_2)$. In this case, $\mathbf{R}(\theta)$ is a matrix of rank 1. If $\theta\neq\hat{\theta}$, $\mathbf{R}(\theta)$ will be semi-definite, and its rank will be greater than 1.

3.4. DOA estimator

In this subsection, we define a new DOA estimator based on the eigenvalue analysis of the WSBCM. Let us perform the eigenvalue decomposition of $\mathbf{R}(\theta)$ and let $\lambda_1(\theta)\geq\lambda_2(\theta)\geq\ldots\geq\lambda_N(\theta)$ denote the N eigenvalues of $\mathbf{R}(\theta)$. If the hypothesized DOA θ equals to $\hat{\theta}$, $\mathbf{R}(\theta)$ is of rank 1, $\lambda_2(\theta)=\ldots=\lambda_N(\theta)=0$. Therefore, if we form the following cost function

$$\mathbf{J}(\theta) \stackrel{\text{def}}{=} \frac{1}{\sum_{i=2}^N|\lambda_i(\theta)|}, \quad (20)$$

the cost function reaches the maximum if $\theta=\hat{\theta}$. Thus the estimated DOA $\tilde{\theta}$ is calculated as:

$$\tilde{\theta} \stackrel{\text{def}}{=} arg\max_{\theta}\mathbf{J}(\theta). \quad (21)$$

4. Experiment

We conduct experiments on the synthetic data to evaluate the performance of the proposed algorithm. The SRP-PHAT [5] and broadband MUSIC algorithm[3], and our previously proposed interference robust method [8] are used for comparison.

4.1. Experimental setup and evaluation

A rectangular room with size $6 \times 4 \times 3$ meters is modeled. We employ a ULA consisting of eight omni-directional microphones, with the inner spacing as 10 cm. The microphones at two ends of ULA are at (2.5,2.0,1.5), (3.2,2.0,1.5) respectively. Although we don't limit the number of interferences, in order to facilitate the test, we assume that only one interference exists in each tested scenario. The speech source and interference are both located on a horizontal plane (x,y,1.5) with a distance of 2m to the center of the microphone array. In different scenarios, the speech DOAs range from -90° to 90° with a step of 20°, and three possible DOAs of the interference, which are 20°, 40° and 60°, are considered. As the ULA is symmetric to the normal line, it is enough to only consider the scenarios that the interference appears in one side.

The room impulse response from the source to each microphone is modeled by image-source method [17]. We set the reverberant time T60 to be 250ms. The speech source is of 10 seconds, with 8KHz sampling rate. We utilize three different types of noises, white Gaussian noise (WGN), car interior noise (CAR) and F16 cockpit noise (F16), taken from Noisex92 [18] as the interference signals. The received speech signal and interference signal are separately generated, and mixed together after being scaled to control the signal-to-interference ratio (SIR). The SIR changes from -10dB to 20dB, with a step of 5dB. For all evaluated algorithms, the frame size is set to be 256 samples with 50% overlap. 50 Monte Carlo simulations are conducted for each scenario (interference type, SIR).

4.2. Experimental results

Two frame level metrics, denoted as Accuracy and Root Mean Square Error (RMSE), are used to evaluate the performance of different algorithms. The estimated DOA is considered to be correct if $|\tilde{\theta} - \hat{\theta}| < 5^\circ$. In the experiment, we only consider the speech frames for evaluation, and the speech frames are labeled manually in advance on the clean speech. It should be pointed out that these labels are never used by any of the four algorithms.

The performances of all evaluated algorithms under different interferences are illustrated in Fig.2. It can be seen that the proposed method can achieve the highest accuracy and the lowest RMSE in almost all conditions considered. Even though in some scenarios, the proposed method can not perform as reliably as the method in [8], the performance degradation is relatively small compared with the improvement in other scenarios.

Among the three kinds of interferences, the WGN has the broadest frequency distribution, while the CAR has the narrowest one. When changing the interference from the CAR to the WGN, the degradation of robustness is still observed for the proposed method. This is understandable for CAR and F16 cases. Because the bispectrum reflects the interaction between different frequencies, broad frequency distribution always results in broad bispectrum distribution, which eventually makes more speech units polluted. But it seems surprising for the WGN case, as the bispectrum of WGN is zero in theory. This may be explained as follows. In fact, its bispectrum is not exactly zero

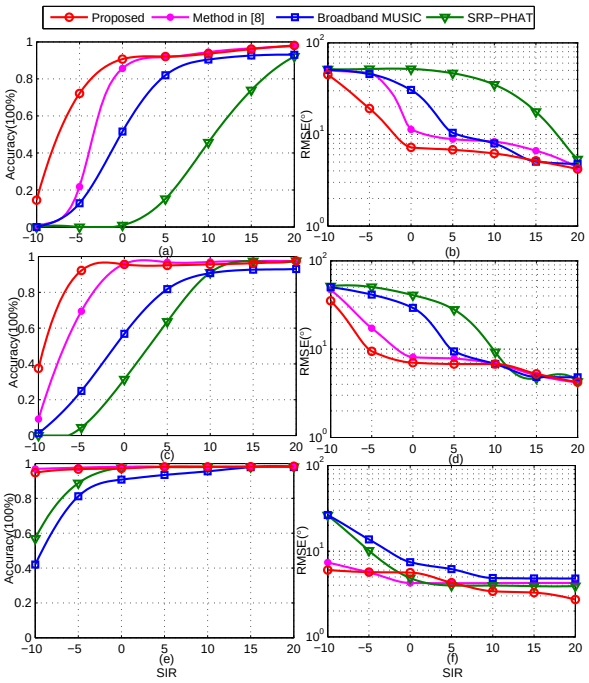

Figure 2: *Estimation performance of different algorithms under different SIRs and interference signals. Accuracy and RMSE with interference as: WGN (a)(b), F16 (c)(d), CAR (e)(f).*

practically, and in low SIR conditions, the residual error may be also large. Moreover, as the all-zeros distribution is the most flat one, the bispectrum distribution of the residual error will be the broadest compared with other interferences. Therefore, despite that the theoretical bispectrum of WGN is zero, in low SIR WGN conditions, the proposed algorithm can not perform so well as in other interference conditions. However, although all algorithms suffer from performance degradation, by exploiting the redundancy of DOA cues expressed in WSBCM, the proposed method degrades least. In the WGN conditions, when the SIR is higher than -5dB, the proposed method can achieve the accuracy higher than 70%.

5. Conclusion

In this paper, a new interference robust DOA estimation method for speech source is proposed. The method is based on the "WSBCM". The WSBCM contains the spatial correlation information of BPDs, and the redundant speech DOA cues expressed in BPDs helps to improve the robustness against the interference. In addition, the WSBCM enables bispectrum weighting to select bispectrum units from BPDs in which speech DOA cues are less polluted. In order to compute the bispectrum weights, an "decision directed" method is then proposed. We also derive a new DOA estimator based on the eigenvalue analysis of the WSBCM. The effectiveness of the proposed method is demonstrated by experiments under various kinds of interferences in different SIRs.

6. Acknowledgements

This research was supported in part by the China National Nature Science Foundation (No.91120303, No.61273267 and No.90820011).

7. References

[1] R. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Trans. Antennas and Propagation AP-34, vol.3, pp. 276-280, 1986.

[2] M. McCloud and L. Scharf, "A new subspace identification algorithm for high resolution DOA estimation," IEEE Trans. Antennas Propag., vol.50, pp.1382-1390, 2002.

[3] J. P. Dmochowski , J. Benesty and S. Affes, "Broadband MUSIC: opportunities and challenges for multiple source localization," in Proceedings IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp.18-21, 2007.

[4] J. Dmochowski, J. Benesty, and S. Affes, "A generalized steered response power method for computationally viable source localization," IEEE Transactions on Audio, Speech, and Language Processing, vol.15, no.8, pp.2510-2526, 2007.

[5] M. Brandstein, D. Ward, "Microphone Arrays: Signal Processing Techniques and Applications," Springer, 2001.

[6] T. G. Dvorkind and S. Gannot, "Time difference of arrival estimation of speech source in a noisy and reverberant environment," Signal Processing, vol.85, no.1, pp.177-204, 2005.

[7] M. Brandstein, J. E. Adcock, and H. Silverman, "A practical time-delay estimator for localizing speech sources with a microphone array," Computer Speech and Language, vol.9, no.2, pp.153-169, 1995.

[8] Wei Xue, Shan Liang, and Wenju Liu, "Interference Robust DOA Estimation of Human Speech by Exploiting Historical Information and Temporal Correlation," in Proceedings 14th Annual Conference of the International Speech Communication Association, 2013.

[9] Jacob Benesty, M. Mohan Sondhi, and Yiteng Huang, "Springer Handbook of Speech Processing," Springer, 2008.

[10] Forster, Philippe, and Chrysostomos L. Nikias, "Bearing estimation in the bispectrum domain," IEEE Transactions on Signal Processing, vol.39, no.9, pp.1994-2006, 1991.

[11] Shi Zhenghao, and Frederick W. Fairman, "DOA estimation via higher-order cumulants: a generalized approach," in Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing, 1992.

[12] Yuen Norman, and Benjamin Friedlander, "DOA estimation in multipath: an approach using fourth-order cumulants," IEEE Transactions on Signal Processing, vol.45, no.5, pp.1253-1263, 1997.

[13] W. B. Collis, P. R. White, and J. K. Hammond, "Higher-order spectra: the bispectrum and trispectrum," Mechanical Systems and Signal Processing, vol.12, no.3, pp.375-394, 1998.

[14] C. L. Nikias, and R. Pan, "Time delay estimation in unknown Gaussian spatially correlated noise," IEEE Transactions on Acoustics, Speech and Signal Processing, vol.36, no.11, pp.1706-1714, 1988.

[15] J. W. A. Fackrell, S. McLaughlin, "The higher-order statistics of speech signals," IEE Colloquium on Techniques for Speech Processing and their Application, 1994.

[16] Ephraim Yariv, David Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," IEEE Transactions on Acoustics, Speech and Signal Processing, vol.32, no.6, pp.1109-1121, 1984.

[17] E. Lehmann and A. Johansson, "Prediction of energy decay in room impulse responses simulated with an image-source model," Journal of the Acoustical Society of Amesrica, vol. 124, no. 1, pp. 269-277, 2008.

[18] A. Varga, H. J. M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to studythe ePect of additive noise on speech recognition systems," Speech Communication, 12(3): 247-251, 1993.