

基于聚焦信号子空间估计导向矢量的干扰声源抑制方法

周 静, 鲍长春, 张 旭

(北京工业大学信息学部语音与音频信号处理实验室, 北京 100124)

摘 要: 针对最小方差无失真响应(Minimum Variance Distortionless Response, MVDR)波束形成器对导向矢量失配较敏感的问题, 本文提出了一种有效的干扰声源抑制方法. 该方法首先将语音信号的频带划分为多个子带, 通过聚焦信号子空间方法估计各子带的声源到达方向(Direction of Arrival, DOA), 并采用统计直方图估计各声源的初始DOA; 其次, 为了减小导向矢量失配, 利用声源的空间稀疏性, 通过Capon功率构建目标声源导向矢量估计的代价函数, 约束目标声源导向矢量远离干扰声源空间; 最后, 根据估计的导向矢量, 估计干扰声源加噪声协方差矩阵, 以获得MVDR波束形成器的权重. 基于TIMIT语料库的实验结果证明, 提出的干扰声源抑制方法的输出信干噪比(SINR)及语音质量感知评价(PESQ)优于参考方法, 具有更佳的抗导向矢量失配性能.

关键词: 语音增强; 麦克风阵列; 波束形成; 聚焦信号子空间; 最小方差无失真响应

基金项目: 国家自然科学基金(No.61831019)

中图分类号: TN912

文献标识码: A

文章编号: 0372-2112(2023)01-0076-10

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20220210

Suppression Method of the Interference Sound Sources by Estimated Steering Vector Based on the Focusing Signal Subspace

ZHOU Jing, BAO Chang-chun, ZHANG Xu

(Speech and Audio Signal Processing Laboratory, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

Abstract: Based on the problem that the minimum variance distortionless response (MVDR) beamformer is very sensitive to the mismatch of the steering vector, an effective method of suppressing the interference sound sources is proposed in this paper. First, the bandwidth of speech signal is divided into multiple sub-bands, and the direction of arrival (DOA) of sound sources at each sub-band is estimated by the focusing signal subspace method. Specially, the initial DOA of each sound source is estimated via statistical histogram. Second, in order to reduce the mismatch of the steering vector, based on the spatial sparsity of sound sources, the cost function used for the steering vector estimation of the target sound source is constructed by Capon power so that the steering vector of the target sound source is constrained away from the space of interference sound sources. Finally, the covariance matrix of interference sound source plus noise is estimated based on the estimated steering vector for obtaining the weights of the MVDR beamformer. The experimental results on the TIMIT corpus show that the proposed method outperforms the reference methods on the tests of the output signal to interference-plus-noise ratio (SINR) and the perceptual evaluation of speech quality (PESQ) and has a better performance for preventing the mismatch of the steering vector.

Key words: speech enhancement; microphone array; beamforming; focusing signal subspace; minimum variance distortionless response

Foundation Item(s): National Natural Science Foundation of China (No.61831019)

1 引言

随着互联网技术与人工智能产业的迅速发展, 远

场语音通信的应用场景越来越多, 如远场人机交互、远场视听会议等, 且质量需求也越来越高^[1,2]. 然而, 由于噪声及干扰声源等因素的影响, 远场语音通信的质量

不如人意,尤其是干扰声源极易导致语音识别率的下降和降低通信中的远端用户听觉质量^[3-5]. 针对这一问题,学者们提出了许多解决方法^[5-8]. 如基于自适应无失真响应波束形成进行空域滤波的方法能够保障目标方向的信号被无失真滤波,因其具有自适应零陷抑制干扰声源的优点^[4,5,8]而得到广泛的研究与应用. 在自适应无失真响应波束形成器中,最小方差无失真响应(Minimum Variance Distortionless Response, MVDR)波束形成器因其约束简洁、优化处理灵活,故常被用于干扰声源及噪声的抑制中. 然而, MVDR 是基于观测信号协方差矩阵求逆的自适应波束形成器,所以,其性能极易受导向矢量失配的影响^[3,9,10]. 造成导向矢量失配的主要因素是声源到达方向(Direction Of Arrival, DOA)失配和阵列结构失配.

DOA 失配是指所估计的声源入射角存在误差,其极易导致主瓣对准上的偏差,甚至出现主瓣自消现象,使得目标声源失真. 针对这一问题,学者们提出了许多有效的 DOA 估计算法,如基于波束形成功率估计的方法、基于极大似然估计的方法、基于稀疏表示的方法和基于子空间的方法等^[11,12]. 其中,基于子空间的方法因其空间高分辨率的优点而备受关注,如经典的多信号分类(Multiple Signal Classification, MUSIC)法^[11]. 建立在 MUSIC 方法的基础上,大量基于子空间的宽带 DOA 估计方法被提出,如 Su 等^[13]提出了非相干信号子空间(Incoherent Signal Subspace, ISS)法,即通过 MUSIC 估计多个窄带的 DOA 并取均值作为宽带 DOA 估计的结果,但其在低信噪比(Signal to Noise Ratio, SNR)子带存在 DOA 估计误差较大及无法有效估计相干声源 DOA 等问题. Wang 等^[14]基于 ISS 提出了相干信号子空间(Coherent Signal Subspace, CSS)法,即通过频率聚焦的方式构建聚焦协方差矩阵,规避了多子带独立 DOA 估计导致误差较大的问题,且在一定程度上提高了满秩协方差矩阵的稳定性,但由于声源信息和噪声的不均匀分布以及需要预估 DOA 信息, CSS 的性能受到一定的限制. 随后,基于频率聚焦的思想, Ma 等^[15]提出了聚焦信号子空间(Focusing Signal Subspace, FSS)法,即通过奇异值分解(Singular Value Decomposition, SVD)构建酉聚焦矩阵,避免了预估 DOA 的问题. Beit-On 等^[16]在 FSS 的基础上,通过分解带宽来减小宽带聚焦误差,并将该方法应用于机器人头部阵列中. 针对噪声的非均匀分布问题,曹等^[17]提出了一种有色噪声下差分聚焦的宽带 DOA 估计方法,减小了“伪峰”影响. 贾等^[18]提出了一种改进的信号子空间聚焦方法,根据波长选取三个子带进行 DOA 估计,进一步提高了 DOA 估计的空间分辨率,但它不适用于非均匀分布的宽带信号. 尽管这些方法做出了不少改进,但在声源定位应用中依然存在较

大误差.

阵列结构失配是指阵元间距存在误差^[19],一般是由工艺生产或人为布置误差造成的,且往往难以人为测量校正. 与 DOA 失配不同,阵列结构失配下的真实导向矢量将不再满足传统的导向矢量形式,即阵元间的相对时延不再固定. 因此,学者们将 DOA 估计误差和阵元间距误差统一归结为导向矢量失配的问题,并期望通过构建不确定集来求解更为准确的导向矢量信息. 如, Stoica 等^[20]通过在不确定集内求解与噪声子空间正交性最强的导向矢量来修正 DOA 与阵列结构失配的误差,但该方法依赖于噪声子空间划分的准确性; Gu 等^[21]利用声源信号的空间稀疏性,通过误差补偿因子来修正导向矢量,并约束补偿结果远离干扰声源空间,改善了依赖于噪声子空间的缺点. Yang 等^[22]在文献[21]的基础上,通过 Capon 功率谱估计干扰声源子空间,并约束估计的导向矢量偏离干扰声源子空间,但其性能也依赖于噪声子空间划分的准确性. 杨等^[10]在文献[21]基础上,通过独立约束估计的导向矢量偏离噪声空间和干扰声源空间,在不确定集内确定了更佳的误差补偿因子,但其增加了算法对约束式所需变量准确性的负荷. 故 Zhang 等^[9]将问题转向于目标声源联合噪声的信号空间,在导向矢量不确定集内求解使得目标声源 Capon 功率最大的导向矢量,但其并不能保障在干扰声源与目标声源的 DOA 较近时所估计到的导向矢量远离干扰声源空间.

基于上述问题,本文提出了一种基于聚焦信号子空间估计导向矢量的方法,并将其用于远场语音通信中干扰声源的抑制,提升了传统 MVDR 波束形成器在 DOA 与阵列结构失配下干扰声源抑制的性能.

2 DOA 估计

2.1 基于 FSS 的 DOA 估计

设自由场内存在 Q 个声源,则 t 时刻 M 元均匀线性阵列(Uniform Linear Array, ULA)的观测信号向量 $\tilde{\mathbf{x}}(t)$ 可表示为:

$$\tilde{\mathbf{x}}(t) = \sum_{q=1}^Q \tilde{\mathbf{s}}_q(t) + \tilde{\mathbf{n}}(t) \quad (1)$$

其中, $\tilde{\mathbf{x}}(t) \in \mathbb{R}^{M \times 1}$, $\tilde{\mathbf{s}}_q(t) \in \mathbb{R}^{M \times 1}$ 表示第 q 个声源入射到 ULA 的信号向量, $\tilde{\mathbf{n}}(t) \in \mathbb{R}^{M \times 1}$ 为噪声向量,且假设声源及噪声信号间均相互独立. 符号“ \mathbb{R} ”表示实数矩阵维度.

将观测信号 $\tilde{\mathbf{x}}(t)$ 经短时傅里叶变换(Short Time Fourier Transform, STFT)变换到频域,可得:

$$\mathbf{x}(k, l) = \sum_{q=1}^Q \mathbf{a}(k, \theta_q) s_q(k, l) + \mathbf{n}(k, l) \quad (2)$$

其中, k 和 l 分别表示频点和语音帧的索引, $\mathbf{a}(k, \theta_q) \in \mathbb{C}^{M \times 1}$ 为第 q 个声源的导向矢量, $s_q(k, l)$ 为第 q 个声源的 STFT,

θ_q 为第 q 个声源相对于 ULA 的入射角, $\mathbf{x}(k, l)$ 和 $\mathbf{n}(k, l)$ 为 $\hat{\mathbf{x}}(t)$ 和 $\hat{\mathbf{n}}(t)$ 的 STFT 变换. 符号“ \mathbb{C} ”表示复数矩阵维度.

观测信号协方差矩阵 $\mathbf{R}(k, l)$ 可由频域观测信号 $\mathbf{x}(k, l)$ 估计如下^[1,4,15,17,21]:

$$\mathbf{R}(k, l) = \frac{1}{J} \sum_{j=0}^{J-1} \mathbf{x}(k, l-j) \mathbf{x}^H(k, l-j) \quad (3)$$

其中, $\mathbf{R}(k, l) \in \mathbb{C}^{M \times M}$, J 表示通过统计平均求解 $\mathbf{R}(k, l)$ 时所用到的帧数, 常被称为快拍数, \mathbf{H} 表示矩阵或向量的共轭转置. 为了简便表达, 在后文中将省略语音帧的索引 l .

聚焦的目的是以乘上聚焦矩阵的形式消除各子带间频率不同带来的影响^[15,16]. 设聚焦矩阵为 $\mathbf{C}(k) \in \mathbb{C}^{M \times M}$, 则式(3)中 $\mathbf{R}(k)$ 聚焦过程可表述如下:

$$\mathbf{R}_k(f_0) = \mathbf{C}(k) \mathbf{R}(k) \mathbf{C}^H(k) \quad (4)$$

其中, f_0 表示聚焦参考频率, $\mathbf{R}_k(f_0)$ 为观测信号的第 k 个子带的聚焦协方差矩阵, $\mathbf{R}_k(f_0) \in \mathbb{C}^{M \times M}$.

为了避免求解聚焦矩阵过程中需预估初始 DOA 的问题, 本文采用文献[15]中聚焦误差最小化的方法来求解聚焦矩阵 $\mathbf{C}(k)$, 即使得式(5)为最小的 $\mathbf{C}(k)$ 为所求的聚焦矩阵.

$$\min_{\mathbf{C}(k)} \left\| \mathbf{V}_s(f_0) - \mathbf{C}(k) \mathbf{V}_s(k) \right\|_2^2 \quad (5)$$

式(5)的一个有效解为

$$\mathbf{C}(k) = \tilde{\mathbf{U}}(k) \tilde{\mathbf{V}}^H(k) \quad (6)$$

其中, $\tilde{\mathbf{U}}(k) \in \mathbb{C}^{M \times M}$, $\tilde{\mathbf{V}}(k) \in \mathbb{C}^{M \times M}$, $\tilde{\mathbf{U}}(k)$ 和 $\tilde{\mathbf{V}}(k)$ 由第 k 个子带的聚焦信号子空间 $\mathbf{V}_s(f_0)$ 和信号子空间 $\mathbf{V}_s(k)$ 的共轭转置的乘积进行 SVD 获得, 即

$$\mathbf{V}_s(f_0) \mathbf{V}_s^H(k) = \tilde{\mathbf{U}}(k) \boldsymbol{\Sigma}_s(k) \tilde{\mathbf{V}}^H(k) \quad (7)$$

这里 $\mathbf{V}_s(k) \in \mathbb{C}^{M \times P}$ 由观测信号协方差矩阵 $\mathbf{R}(k)$ 的前 P 个较大奇异值所对应的右奇异向量构建.

P 往往对应着真实声源的数目^[15,16,18,23], 其可分两步求解. 步骤1, 通过求解 $\mathbf{R}(k)$ 特征值的熵, 判别子带是否存在声源成分^[24], 如式(8)所示:

$$\varpi = \begin{cases} 1, & \left(-\sum_{m=1}^M \hat{e}_m \log \hat{e}_m \right) \leq \xi_1 \\ 0, & \text{其他} \end{cases} \quad (8)$$

步骤2, 若 $\varpi=0$, 则判别子带不存在声源成分, 即 $P=0$; 若 $\varpi=1$, 则判别子带存在声源成分, 可通过如下方法解得 P 值^[25]:

$$P = \arg \min_{\hat{P}} \left\{ \frac{\sum_{m=\hat{P}+1}^M e_m}{\text{Trace}[\mathbf{R}(k)]} \leq \xi_2 \right\} \quad (9)$$

式(8)和式(9)中, \hat{e}_m 和 e_m 分别为 $\mathbf{R}(k)$ 的第 m 个归一化特征值和非归一化特征值(降序排列), “Trace”表

示取迹运算, ξ_1 和 ξ_2 为较小的常数阈值^[24,25], 本文选取 ξ_1 为 0.5, ξ_2 为 0.01. 式(9)中, $\tilde{P} \in [1, 2, \dots, M-1]$, 满足式(9)中不等式关系的最小 \tilde{P} 即为所求的 P 值.

而 $\mathbf{V}_s(f_0) \in \mathbb{C}^{M \times P}$ 可通过特征值分解获得, 并期望 $\mathbf{C}(k)$ 满足式(10)关系:

$$\mathbf{V}_s(f_0) = \mathbf{C}(k) \mathbf{V}_s(k) \quad (10)$$

不难发现, 聚焦矩阵 $\mathbf{C}(k)$ 满足酉矩阵的性质, 故聚焦参考频率 f_0 可通过如下的式(11)求解^[15,26]:

$$\begin{cases} \min_{f_0} \sum_{k=1}^K \left\| \mathbf{V}_s(f_0) - \mathbf{C}(k) \mathbf{V}_s(k) \right\|_2^2 \\ \text{s.t. } \mathbf{C}(k) \mathbf{C}^H(k) = \mathbf{I} \end{cases} \quad (11)$$

式(11)中的 K 为频点的个数, 使得式(11)中聚焦误差最小的 f_0 即为所求的聚焦参考频率. 将式 $\mathbf{V}_s(f_0)$ 带入式(11)便可求得聚焦参考频率 f_0 , 再带入式(5)和式(4)中便可求解得到各子带的聚焦协方差矩阵 $\mathbf{R}_k(f_0)$. 通过对各子带聚焦协方差矩阵求均值可得如下的平滑聚焦协方差矩阵:

$$\bar{\mathbf{R}}(f_0) = \frac{1}{K} \sum_{k=1}^K \mathbf{R}_k(f_0) \quad (12)$$

在获得平滑聚焦协方差矩阵 $\bar{\mathbf{R}}(f_0)$ 后, 即可通过其特征值分解构建噪声子空间. $\bar{\mathbf{R}}(f_0)$ 的特征值分解为

$$\bar{\mathbf{R}}(f_0) = \mathbf{V}(f_0) \boldsymbol{\Sigma}(f_0) \mathbf{V}^H(f_0) \quad (13)$$

其中, $\mathbf{V}(f_0) = [\mathbf{v}_1(f_0), \mathbf{v}_2(f_0), \dots, \mathbf{v}_M(f_0)]$ 为特征向量构成的特征矩阵, $\boldsymbol{\Sigma}(f_0)$ 为按特征值大小降序排列的特征值对角阵.

通过取后面的 $(M-P)$ 个特征值所对应的特征向量可获得如下的噪声子空间^[15,22,26]:

$$\mathbf{V}_N(f_0) = [\mathbf{v}_{P+1}(f_0), \mathbf{v}_{P+2}(f_0), \dots, \mathbf{v}_M(f_0)] \in \mathbb{C}^{M \times (M-P)} \quad (14)$$

综上, 可获得如下的 MUSIC 谱估计:

$$P_{\text{MUSIC}}(\tilde{\theta}_i) = \frac{1}{\mathbf{a}^H(f_0, \tilde{\theta}_i) \mathbf{V}_N(f_0) \mathbf{V}_N^H(f_0) \mathbf{a}(f_0, \tilde{\theta}_i)} \quad (15)$$

其中, $\tilde{\theta}_i \in [0^\circ, 180^\circ]$, i 为空间角度取样数的索引. 通过搜索使得式(15)获得前 P 个较大极大值的 $\tilde{\theta}_i$, 可求得所估计的 P 个可能声源的入射角, 即 DOA 结果.

2.2 基于噪声子空间重构的 DOA 估计

当噪声较大时, 信号子空间极易向噪声子空间扩散, 使得噪声子空间与声源导向矢量正交性下降, 从而导致式(15)所求的 MUSIC 谱峰出现分辨率下降问题. 为了解决这一问题, 本文在构建噪声子空间时, 仅选取后 $(M-P-G)$ 个特征值所对应的特征向量来构建噪声子空间, 即

$$\tilde{\mathbf{V}}_N(f_0) = [\mathbf{v}_{P+G+1}(f_0), \mathbf{v}_{P+G+2}(f_0), \dots, \mathbf{v}_M(f_0)] \quad (16)$$

其中, $\tilde{\mathbf{V}}_N(f_0) \in \mathbb{C}^{M \times (M-P-G)}$, G 为减少的噪声特征向量的个数.

尽管式(16)构建的噪声子空间降低了 MUSIC 谱的

平滑性,但其有效提高了 MUSIC 谱峰的分辨率;同时,通过合理选择观测信号协方差矩阵中较大特征值的个数 P ,可以较好地确定真实声源的数目. 所以,该方法有效改善了基于聚焦信号子空间估计 DOA 的分辨率.

然而,按式(5)和式(11)所示在全频带求解最佳聚焦矩阵时,由于频率跨度较大,会导致较大的聚焦误差,文献[16]对这一问题进行了研究,并得出了 500 Hz 左右的带宽能够获得更小聚焦误差的结论. 因此,本文在聚焦处理中,将 0~4 000 Hz 的窄带语音信号划分为 8 个聚焦子带,在每个聚焦子带,通过式(15)独立求解 MUSIC 谱;然后,以一定间隔将 $0^\circ \sim 180^\circ$ 划分为若干区间,通过统计直方图方法获取各聚焦子带 MUSIC 谱的前 P 个极大值所处的角度区间频次,并分别求取频次最高的前 P_{\max} 个角度区间内的 DOA 均值来作为最终的 DOA 估计值. 其中, P_{\max} 为 8 个聚焦子带中 P 值的最大值. 图 1 给出了以 4° 为间隔将 $0^\circ \sim 180^\circ$ 划分为 45 个区间的例子,当选取 $P_{\max}=3$ 时,根据统计频次最高的 3 个角度区间内的 DOA 结果,可有效分辨 3 个声源及其入射角度.

尽管式(16)在一定程度上避免了信号子空间向噪声子空间扩散的问题,但并不能减少噪声向信号子空间扩散的问题,这使得基于 MUSIC 这类子空间方法估计到的 DOA 依然存在误差. 对于 MVDR 这类自适应波束形成器而言,严格依赖于 DOA 信息来构建导向矢量,在导向矢量失配时会导致波束响应出现主瓣自消、偏移及畸变等问题. 这些问题将会导致增强的目标声源信号出现严重的失真,这一问题的解决方法将在下一节给出.

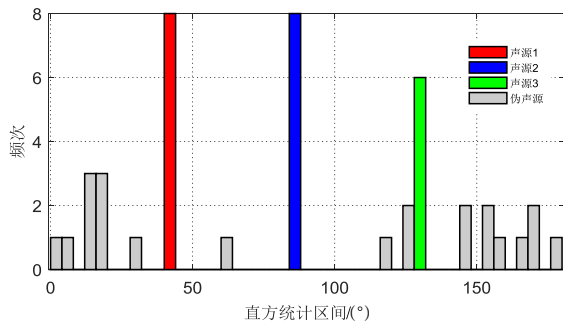


图 1 DOA 统计直方图示例

3 估计导向矢量的干扰声源抑制

针对 DOA 误差和阵列结构误差导致导向矢量失配的问题,学者们提出了基于不确定集的导向矢量估计方法. 本节将对现有方法的优缺点进行分析,并提出新的导向矢量估计方法来构建稳健的 MVDR 波束形成器,提升其在干扰声源抑制中的性能.

3.1 导向矢量的估计

基于第 2 节估计的 DOA 结果,建立目标声源的初始导向矢量 $\hat{\mathbf{a}}(k, \theta_0)$,并通过约束求解不确定集内最佳的目标声源导向矢量,来减少导向矢量失配带来的问题. 如文献[20]以噪声子空间 $\mathbf{U}_N(k)$ 与正确导向矢量 $\mathbf{a}(k, \theta_0)$ 的正交性来构建代价函数,在不确定集内搜索与噪声子空间正交性更强的导向矢量来获得更为稳健的性能,具体优化过程如式(17)所示:

$$\begin{cases} \min_{\mathbf{a}(k, \theta_0)} \mathbf{a}^H(k, \theta_0) \mathbf{U}_N(k) \mathbf{U}_N^H(k) \mathbf{a}(k, \theta_0) \\ \text{s.t. } \|\mathbf{a}(k, \theta_0) - \hat{\mathbf{a}}(k, \theta_0)\|_2 \leq \delta_1 \end{cases} \quad (17)$$

其中, $\mathbf{a}(k, \theta_0) \in \mathbb{C}^{M \times 1}$ 为估计的目标声源导向矢量, δ_1 为不确定集的半径,符号“ $\|\cdot\|_2$ ”表示二范数运算. 式(17)的方法虽然看似简洁合理,但同式(15)一样,依赖于噪声子空间 $\mathbf{U}_N(k)$ 的准确划分,且不确定集半径 δ_1 的值难以确定.

基于这一问题,有学者将声源信号的空间稀疏特性引入其中,将观测信号的空间划分为目标声源空间 Φ_S 、干扰声源空间 Φ_I 和噪声空间 Φ_N 三部分,这样便可将观测信号协方差矩阵进行拆分,进行更为灵活的约束. 如文献[21]通过 Capon 功率,在 Φ_I 和 Φ_N 的联合空间 Φ_{IN} 上求取干扰声源加噪声的协方差矩阵 $\hat{\mathbf{R}}_{IN}(k)$,然后通过误差补偿因子矢量 \mathbf{e}_\perp 对初始导向矢量 $\hat{\mathbf{a}}(k, \theta_0)$ 进行修正,约束 $\hat{\mathbf{a}}(k, \theta_0)$ 远离 Φ_{IN} 空间,如式(18)所示:

$$\begin{cases} \min_{\mathbf{e}_\perp} [\hat{\mathbf{a}}(k, \theta_0) + \mathbf{e}_\perp(k)]^H \mathbf{R}^{-1}(k) [\hat{\mathbf{a}}(k, \theta_0) + \mathbf{e}_\perp(k)] \\ [\hat{\mathbf{a}}(k, \theta_0) + \mathbf{e}_\perp(k)]^H \hat{\mathbf{R}}_{IN}(k) [\hat{\mathbf{a}}(k, \theta_0) + \mathbf{e}_\perp(k)] \\ \text{s.t. } \leq \hat{\mathbf{a}}^H(k, \theta_0) \hat{\mathbf{R}}_{IN}(k, l) \hat{\mathbf{a}}(k, \theta_0) \\ \hat{\mathbf{a}}^H(k, \theta_0) \mathbf{e}_\perp(k) = 0 \end{cases} \quad (18)$$

其中, $\mathbf{e}_\perp \in \mathbb{C}^{M \times 1}$, $\hat{\mathbf{a}}(k, \theta_0) \in \mathbb{C}^{M \times 1}$, $\hat{\mathbf{R}}_{IN}(k) \in \mathbb{C}^{M \times M}$. $\hat{\mathbf{R}}_{IN}(k)$ 可由 Capon 功率积分求得,即

$$\hat{\mathbf{R}}_{IN}(k) = \int_{\Phi_{IN}} \frac{\mathbf{a}(k, \theta_j) \mathbf{a}^H(k, \theta_j)}{\mathbf{a}^H(k, \theta_j) \mathbf{R}^{-1}(k) \mathbf{a}(k, \theta_j)} d\theta_j \quad (19)$$

由式(18)的不等式约束可以看出, \mathbf{e}_\perp 的作用主要是使得 $\hat{\mathbf{a}}(k, \theta_0)$ 进一步远离干扰声源导向矢量及噪声所处的空间,但其不能确保减小阵列结构失配带来的误差,且 \mathbf{e}_\perp 的解并不能够保证目标声源的能量不被削弱.

考虑上述约束的缺点和复杂性,文献[22]在式(18)的基础上,通过独立约束估计的导向矢量远离噪声子空间和干扰声源子空间,来驱使估计的目标声源导向矢量趋向于真实的目标声源导向矢量,其优化模型如式(20)所示:

$$\begin{cases} \min_{\mathbf{e}_\perp} \|\mathbf{U}_N^H(k)[\hat{\mathbf{a}}(k, \theta_0) + \mathbf{e}_\perp(k)]\|_2^2 \\ \text{s.t.} \quad \|\mathbf{U}_I^H(k)[\hat{\mathbf{a}}(k, \theta_0) + \mathbf{e}_\perp(k)]\|_2^2 \leq \|\mathbf{U}_I^H(k)\hat{\mathbf{a}}(k, \theta_0)\|_2^2 \\ \hat{\mathbf{a}}^H(k, \theta_0)\mathbf{e}_\perp(k) = 0 \end{cases} \quad (20)$$

其中, $\mathbf{U}_I(k) \in \mathbb{C}^{M \times M}$ 是对干扰声源加噪声的协方差矩阵 $\hat{\mathbf{R}}_{\text{IN}}(k)$ 进行特征值分解后的干扰声源子空间^[22]. 然而, 目标声源导向矢量不包含于 $\mathbf{U}_N(k)$ 中, 故其不能够保证求解的导向矢量更加趋近于真实的导向矢量.

鉴于此, 文献[10]重新建立了约束模型, 且约束 $\mathbf{e}_\perp(k)$ 不过度影响 $\hat{\mathbf{a}}(k, \theta_0)$, 即约束 $\mathbf{e}_\perp(k)$ 对目标声源信号的影响, 优化模型如式(21)所示:

$$\begin{cases} \min_{\mathbf{e}_\perp} [\hat{\mathbf{a}}(k, \theta_0) + \mathbf{e}_\perp(k)]^H \mathbf{R}^{-1}(k) [\hat{\mathbf{a}}(k, \theta_0) + \mathbf{e}_\perp(k)] \\ \|\mathbf{U}_N^H(k)[\hat{\mathbf{a}}(k, \theta_0) + \mathbf{e}_\perp(k)]\|_2^2 \leq \|\mathbf{U}_N^H(k)\hat{\mathbf{a}}(k, \theta_0)\|_2^2 \\ \text{s.t.} \quad \|\mathbf{U}_I^H(k)[\hat{\mathbf{a}}(k, \theta_0) + \mathbf{e}_\perp(k)]\|_2^2 \leq \|\mathbf{U}_I^H(k)\hat{\mathbf{a}}(k, \theta_0)\|_2^2 \\ \|\hat{\mathbf{a}}(k, \theta_0) + \mathbf{e}_\perp(k)\|_2 \leq \|\hat{\mathbf{a}}(k, \theta_0)\|_2 + \delta_2 \\ \hat{\mathbf{a}}^H(k, \theta_0)\mathbf{e}_\perp(k) = 0 \end{cases} \quad (21)$$

其中, δ_2 为松弛因子. 式(21)通过第3个不等式约束保障了 $\mathbf{e}_\perp(k)$ 存在非零的其他解. 然而, 式(21)增大了约束复杂度, 对 $\mathbf{U}_I(k)$ 和 $\mathbf{U}_N(k)$ 准确度的要求较高, 且当 $\hat{\mathbf{a}}(k, \theta_0)$ 偏离真实导向矢量较大时, 易因 δ_2 选择不当而无法解得有效的导向矢量.

而文献[9]指出, 对于目标声源导向矢量的估计, 最直接、最有效的方法是通过目标声源加噪声的协方差矩阵, 在导向矢量的不确定集内求解 Capon 功率倒数的最小值来构建代价函数, 如式(22)所示:

$$\begin{cases} \min_{\mathbf{a}(k, \theta_0)} \mathbf{a}^H(k, \theta_0) \hat{\mathbf{R}}_{\text{SN}}^{-1}(k) \mathbf{a}(k, \theta_0) \\ \text{s.t.} \quad \|\mathbf{a}(k, \theta_0) - \hat{\mathbf{a}}(k, \theta_0)\|_2 \leq \delta_3 \end{cases} \quad (22)$$

其中, $\hat{\mathbf{R}}_{\text{SN}}(k) \in \mathbb{C}^{M \times M}$ 为估计的目标声源加噪声的协方差矩阵, 与式(19)类似, 可通过 Capon 功率积分求得; δ_3 为不确定集半径.

显然, 式(22)的约束方法相当简洁, 且较式(17)的不确定集更易确定. 但在噪声功率远大于目标声源功率时, 可能导致无法求解出有效的导向矢量; 并且, 当干扰声源离目标声源较近时, 可能使得估计的导向矢量偏向于干扰声源所处空间. 此外, 在目标声源和噪声的联合空间内求解增大了计算量. 因此, 本文建立了如式(23)所示的优化模型, 仅在 Φ_s 内求解, 同时约束估计的导向矢量远离干扰声源所处空间:

$$\begin{cases} \min_{\mathbf{a}(k, \theta_0)} \mathbf{a}^H(k, \theta_0) \hat{\mathbf{R}}_S^{-1}(k) \mathbf{a}(k, \theta_0) \\ \|\mathbf{a}(k, \theta_0) - \hat{\mathbf{a}}(k, \theta_0)\|_2 \leq \delta_3 \\ \text{s.t.} \quad \|\mathbf{U}_I^H(k) \mathbf{a}(k, \theta_0)\|_2 \leq \|\mathbf{U}_I^H(k) \hat{\mathbf{a}}(k, \theta_0)\|_2 \end{cases} \quad (23)$$

其中, 目标声源协方差矩阵 $\hat{\mathbf{R}}_S(k)$ 可由式(24)估计:

$$\hat{\mathbf{R}}_S(k) = \int_{\Phi_s} \frac{\mathbf{a}(k, \theta_j) \mathbf{a}^H(k, \theta_j)}{\mathbf{a}^H(k, \theta_j) \mathbf{R}^{-1}(k) \mathbf{a}(k, \theta_j)} d\theta_j \in \mathbb{C}^{M \times M} \quad (24)$$

$\hat{\mathbf{R}}_S(k)$ 一般为正定阵^[9], 故式(23)为二阶锥方程. 利用式(23)迭代优化, 可获得迭代收敛后的高精度导向矢量 $\mathbf{a}(k, \theta_0)$.

3.2 干扰声源加噪声的协方差矩阵估计

设噪声为高斯白噪声, 则在解得新的导向矢量 $\mathbf{a}(k, \theta_0)$ 后, 目标声源的协方差矩阵可重新估计估计为:

$$\hat{\mathbf{R}}_S(k) = \frac{\mathbf{a}(k, \theta_0) \mathbf{a}^H(k, \theta_0)}{\mathbf{a}^H(k, \theta_0) \mathbf{R}^{-1}(k) \mathbf{a}(k, \theta_0)} \quad (25)$$

由于 $\mathbf{a}(k, \theta_0)$ 所属空间还包含了一些噪声成分, 使得式(25)估计的 $\hat{\mathbf{R}}_S(k)$ 存在偏差, 即式(25)中的目标声源的 Capon 功率项存在噪声偏差. 若噪声偏差的功率为 σ_N^2 , 则目标声源的 Capon 功率估计 $\hat{p}(k, \theta_0)$ 如下^[27, 28]:

$$\begin{aligned} \hat{p}(k, \theta_0) &= \frac{1}{\mathbf{a}^H(k, \theta_0) \mathbf{R}^{-1}(k) \mathbf{a}(k, \theta_0)} \\ &= \frac{1}{\mathbf{a}^H(k, \theta_0) [\sigma_S^2 \mathbf{a}(k, \theta_0) \mathbf{a}^H(k, \theta_0) + \sigma_N^2 \mathbf{I}]^{-1} \mathbf{a}(k, \theta_0)} \\ &= \sigma_S^2 + \frac{\sigma_N^2}{M} \end{aligned} \quad (26)$$

其中, σ_S^2 为目标声源功率, 可见目标声源的 Capon 功率估计存在偏差项 σ_N^2/M . 要消除此偏差项, 则需进一步估计噪声偏差的功率. 文献[27]中指出, 可以在 Φ_N 内以均匀采样的方式来估计噪声偏差项的功率, 如式(27)所示:

$$\frac{\sigma_N^2}{M} \approx \frac{1}{A_N} \sum_{j=1}^{A_N} \frac{1}{\mathbf{a}^H(k, \hat{\theta}_j) \mathbf{R}^{-1}(k) \mathbf{a}(k, \hat{\theta}_j)}, \quad \hat{\theta}_j \in \Phi_N \quad (27)$$

其中, A_N 为采样总数.

修正后的目标声源的协方差矩阵可表示为

$$\mathbf{R}_S(k) \approx \left[\hat{p}(k, \theta_0) - \frac{\sigma_N^2}{M} \right] \mathbf{a}(k, \theta_0) \mathbf{a}^H(k, \theta_0) \quad (28)$$

则, 干扰声源加噪声的协方差矩阵可重新估计为

$$\mathbf{R}_{\text{IN}}(k) \approx \mathbf{R}(k) - \mathbf{R}_S(k) \quad (29)$$

其中, $\mathbf{R}_S(k) \in \mathbb{C}^{M \times M}$, $\mathbf{R}_{\text{IN}}(k) \in \mathbb{C}^{M \times M}$.

3.3 MVDR 波束形成

在求得干扰声源加噪声的协方差矩阵 $\mathbf{R}_{\text{IN}}(k)$ 后, 可根据式(30)计算 MVDR 空域滤波系数向量 $\mathbf{w}(k)$:

$$\mathbf{w}(k) = \frac{\mathbf{R}_{\text{IN}}^{-1}(k)\mathbf{a}(k, \theta_0)}{\mathbf{a}^{\text{H}}(k, \theta_0)\mathbf{R}_{\text{IN}}^{-1}(k)\mathbf{a}(k, \theta_0)} \in \mathbb{C}^{M \times 1} \quad (30)$$

对各子带进行空域滤波,即可得到频域增强的目标声源信号 $z(k)$:

$$z(k) = \mathbf{w}^{\text{H}}(k)\mathbf{x}(k) \quad (31)$$

$z(k)$ 经逆 STFT (Inverse STFT, ISTFT), 即可得到增强后的时域目标声源信号 $z(t)$.

4 仿真实验

4.1 实验原理

本文提出的干扰声源抑制方法的原理框图如图 2 所示. 首先将观测信号变换到频域,求得各频带协方差矩阵 $\mathbf{R}(k)$,并对 $\mathbf{R}(k)$ 做 SVD 获得声源信号子空间 $\mathbf{V}_s(k)$;其次,以 500 Hz 频率间距划分子带,在各子带内分别求取聚焦频率 f_0 及聚焦矩阵 $\mathbf{C}(k)$,并对平滑聚焦协方差矩阵进行子空间分解,构建出噪声子空间 $\hat{\mathbf{U}}_N(f_0)$,从而通过 MUSIC 谱的统计结果获得各声源的初始 DOA 信息;然后,根据各声源初始 DOA 信息划分角度空间,求得目标声源的协方差矩阵 $\hat{\mathbf{R}}_s(k)$ 及干扰声源子空间 $\mathbf{U}_i(k)$,即按式(22)估计导向矢量,并根据 Capon 功率重新求得目标声源的协方差矩阵,计算出 MVDR 空域滤波权重 $\mathbf{w}(k)$;最后,对频域观测信号进行 MVDR 空域滤波,通过 ISTFT 获得时域增强信号.

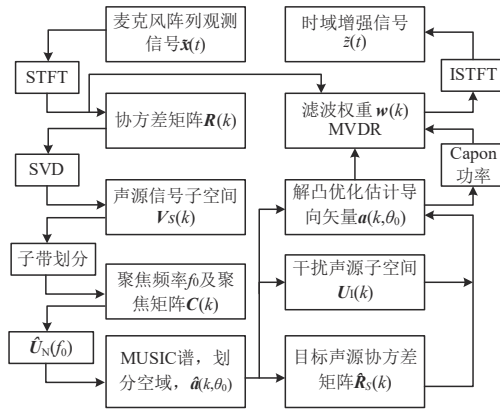


图 2 本文提出的抑制干扰声源方法的原理框图

4.2 实验设置与评测指标

本文的实验使用了 TIMIT 语料库中随机抽取的 600 句语音. 其中, 200 句用作目标声源, 200 句用作干扰声源 1, 200 句用作干扰声源 2. 阵元数目 M 设置为 8, 阵元间距设置为 0.02 m, 并随机生成 0~2 mm 的阵列位置误差, 通过麦克风阵列信号生成器^[29]来生成实验所需的阵列信号. 声速设为 340 m/s, 输入信干比 (Signal to Interference Ratio, SIR)^[30] 分别设置为 0 dB, 5 dB 和 10 dB, 输入信噪比 (Signal to Noise Ratio, SNR) 分别设置为 -5 dB, 0 dB, 5 dB, 10 dB, 15 dB 和 20 dB, 噪声为高斯白噪声, 窗函数为汉明窗, 采样率为 8 kHz, 窗长为

30 ms, 帧间叠接 15 ms. 以 40°、85°、130° 入射角为中心, 在 ±20° 入射角范围内随机生成 3 个入射角, 即 3 个入射角之间的最小间距为 5°; 将 3 个入射角随机排序, 并分别作为目标声源、干扰声源 1 和干扰声源 2 的真实 DOA.

仿真实验主要评测两个性能, 即基于 FSS 的 DOA 估计性能和基于导向矢量估计的干扰声源抑制性能. 在 DOA 估计的评测中, 本文选取了 ISS 方法^[13]、CSS 方法^[14] 和聚焦矩阵重构 (Focusing Matrix Reconstruction, FMR) 方法^[17] 作为比较对象, 评测指标为快拍数为 21 时的 DOA 估计结果的均方误差 (Mean Square Error, MSE)^[17]. 在干扰声源抑制性能的评测中, 本文选取文献 [9]、文献 [10]、文献 [20]、文献 [21] 和文献 [22] 中的方法进行对比, 评测指标为输出信干噪比 (Signal to Interference-plus-Noise Ratio, SINR)^[30] 和语音质量感知评估 (Perceptual Evaluation of Speech Quality, PESQ)^[29]. 输入 SIR (SIR_{in})、输入 SNR (SNR_{in}) 及输出 SINR (SINR_{out}) 的定义^[30,31] 如下:

$$\begin{cases} \text{SNR}_{\text{in}} = \sigma_s^2 / \sigma_n^2 \\ \text{SIR}_{\text{in}} = \sigma_i^2 / \sigma_n^2 \\ \text{SINR}_{\text{out}} = \sigma_{zs}^2 / (\sigma_{zi}^2 + \sigma_{zn}^2) \end{cases} \quad (32)$$

其中, σ_s^2 、 σ_i^2 和 σ_n^2 分别为输入信号中目标声源、干扰声源及噪声的功率, σ_{zs}^2 、 σ_{zi}^2 和 σ_{zn}^2 分别为输出信号中目标声源、干扰声源及噪声的功率.

4.3 实验结果与分析

(1) DOA 估计性能的评测

DOA 估计性能评测中设置了单声源、双声源和三声源的情况, 单声源时仅存在目标声源, 双声源时存在目标声源和干扰声源 1, 三声源时存在目标声源、干扰声源 1 和干扰声源 2. 在进行 DOA 估计前, 需要先确定参数 G 的取值. 本文实验中, 麦克风数 M 为 8, 声源数 P 最大为 3, 故 G 的取值范围为 0~4. 图 3 给出了 100 次蒙特卡洛实验下 G 取不同值时的 DOA 估计误差, 可以看出, 当声源数不超过 3 时, 取 $G=2$ 能够获得更小的 MSE.

在确定 G 的取值后, 则可按第 2 节所述方法估计 DOA. DOA 估计的 MSE 结果如图 4 所示. 本文通过划分 500 Hz 的子带宽度进行聚焦, 通过 MUSIC 空间谱极大值的统计直方图来估计 DOA, 避免了误差较大 DOA 值和聚焦误差较大的影响, 在对比的 4 种方法中, 所提方法获得了最佳的性能. 因 ISS 方法是由子带求平均的方式估计 DOA, 所以, 误差较大的子带产生了负面结果; 而传统的 CSS 方法需要估计初始 DOA, 当初始 DOA 误差较大时将带来较大的聚焦误差, 且在全带宽内聚焦时, 也易导致更大的聚焦误差. 因此, ISS 方法和 CSS 方法较提出方法的 DOA 估计性能差. 因 CSS 方法能够提升协方差矩阵满秩的稳定性, 且能够在一定程度上避免误差较大子带的影响, 故在图 4(a) 所示单声源情况

下的性能与ISS方法相近,而在图4(b)~(e)所示多声源情况下的性能明显优于ISS方法.FMR方法通过取正特征值对应的特征向量作为信号子空间,以此降低噪声的影响,来获得重构的聚焦矩阵,但频域下特征值存在复数情况,难以通过区分特征值正负对观测信号进行有效地重构,故性能较ISS差.其次,由图4(a)~4(c)可以看出,随着声源数目的增加,各方法的DOA估计性能均有下降.此外,由图4(c)~4(e)可以看出,在不同 SIR_{in} 下,提出方法仍具备最佳的DOA估计性能.

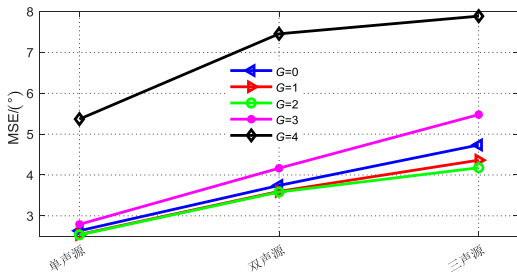


图3 不同声源数目下G取值与DOA估计误差的关系

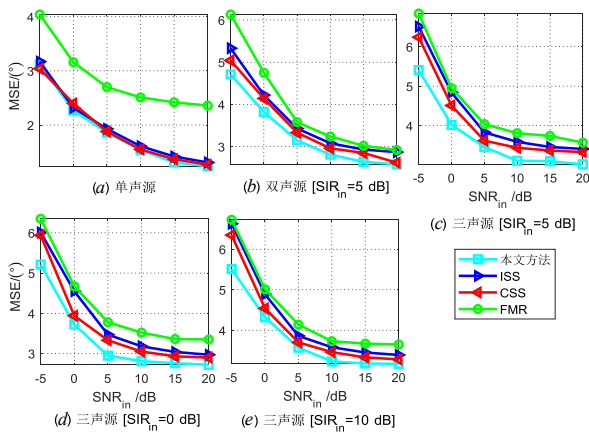


图4 不同 SNR_{in} 下的DOA估计结果比较

(2) 干扰声源抑制性能的评测

存在阵列结构误差的情况下,基于提出的改进FSS方法估计初始DOA,构建了不确定集来估计目标声源导向矢量,并通过MVDR进行干扰声源抑制,不确定集半径 δ_3 取0.1M,式(23)迭代10次.图5给出了 $SINR_{out}$ 和PESQ随 SNR_{in} 变化的比较.其中,各干扰声源是以相同的 SIR_{in} 独立叠加到目标声源上的,导向矢量估计所用的快拍数为21,求取MVDR滤波器权重的快拍数为7.图5中noisy表示参考麦克风的观测语音.需要说明的是,导向矢量估计与MVDR权重所用快拍数并不相同,这是因为导向矢量估计中若快拍数较小易导致各类协方差矩阵非正定或非负半正定,使得式(23)无法通过凸优化过程求解.MVDR滤波器的权重直接影响增强的效果,快拍数的影响将在后面的语谱对比中详细解释.图5结果显示,随 SNR_{in} 的变化,所提方法的 $SINR_{out}$

和PESQ均获得了最好的结果,即便是在-5 dB的 SNR_{in} 环境下也能获得有效的增强.此外,在不同 SIR_{in} 的干扰声源下,各方法均具备稳定的性能,即各导向矢量估计方法受 SIR_{in} 影响较小.

图6给出了 $SINR_{out}$ 和PESQ随着计算MVDR空域滤波器权重所用的快拍数变化的比较,noisy同样表示参考麦克风的观测语音.其中,各干扰声源以0 dB的 SIR_{in} 叠加,噪声以5 dB的 SNR_{in} 叠加,快拍数为3~51.结果显示,随拍数的增长,各方法增强信号的 $SINR_{out}$ 先呈现上升趋势,后开始逐渐下降,而各算法增强信号的PESQ则随快拍增大均呈现下降趋势.此外,从图6可以看出,所提方法在快拍数为7时可获得最高的 $SINR_{out}$,而在较小的快拍数时能够获得更高的PESQ.

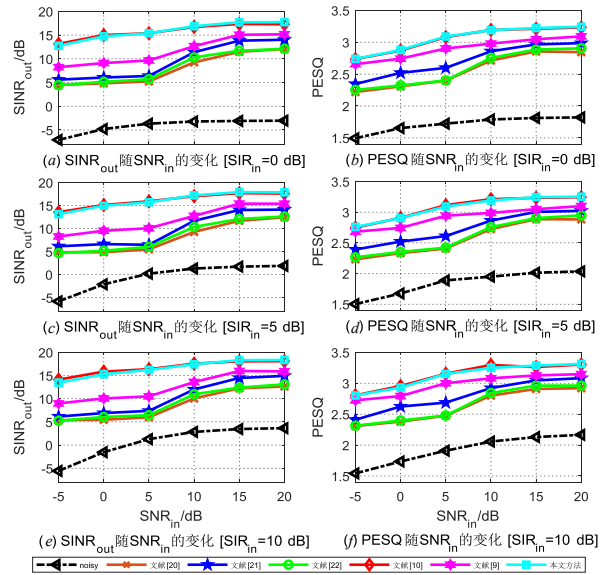


图5 不同 SIR_{in} 下 $SINR_{out}$ 和PESQ随 SNR_{in} 的变化比较

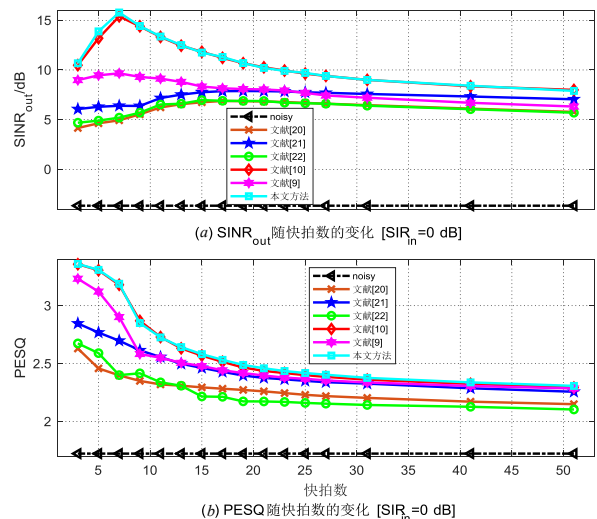


图6 $SINR_{out}$ 和PESQ随快拍数的变化比较

(3) 抑制干扰声源的语谱图比较

为了进一步解释 $SINR_{out}$ 和 PESQ 呈现上述表现的原因,本文对增强语音的语谱图进行了分析.图7给出了一组增强结果的语谱图对比.其中,两干扰声源叠加的 SIR_{in} 为 0 dB,白噪声叠加的 SNR_{in} 为 5 dB,导向矢量估计快拍数为 21, MVDR 权重求取快拍数为 7. 增强结果的语谱图显示,在各声源及噪声均相互独立的假设下,当快拍数较小时,各方法均能够有效抑制噪声和有效零陷干扰声源,但由于估计的导向矢量存在差异,使得增强后的目标声源存在失真上的差异.通过对比图7标注的红色圈内的频谱信息,不难发现文献[20]方法导致的目标声源失真最严重,即其导向矢量估计误差最大;文献[21]方法较文献[20]方法的目标语音失真小,但仍存在一定的目标语音失真;文献[22]方法性能介于文献[20]与文献[21]方法性能之间,原因可能是噪声子空间估计不准确;文献[9]和文献[10]方法较前三种方法更好地保留了目标语音成分,且文献[10]方法较文献[9]方法的目标语音失真更小.所提方法因仅在目标声源空间内求解,且进一步约束了估计的导向矢量远离干扰声源空间,故获得了较文献[9]更好的效果,其增强的语谱图与文献[10]相近.

图8给出了另一组增强结果的语谱图对比,除求解 MVDR 权重时快拍数变为 15 外,其余条件均保持相同.与图7相比,可看出当快拍数增大后,求得的 MVDR 波束形成器的去噪性能降低,但目标语音失真问题获得了改善;此外,部分频带出现干扰声源残留,如图8中黑色圈内的频谱信息.幸运的是,快拍数的不同仅影响了

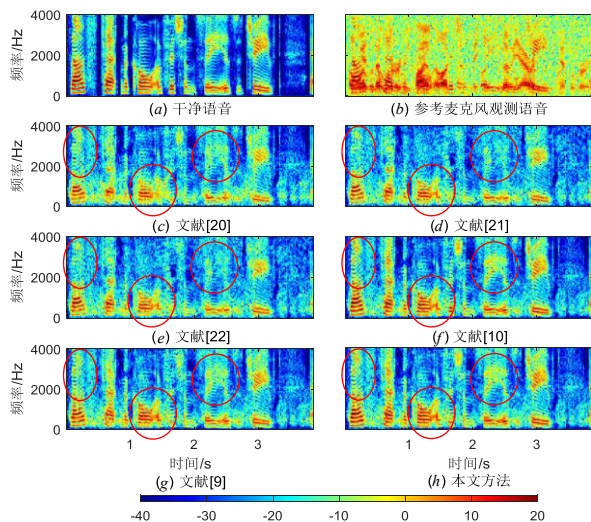


图7 增强语音的语谱图对比(快拍数为7)

式(29)估计的干扰声源加噪声的协方差矩阵 $R_{IN}(k)$.可见, PESQ 在快拍数较小时提高的原因是估计的 $R_{IN}(k)$ 较为精准,从而获得了更好的干扰声源及噪声抑制性能.

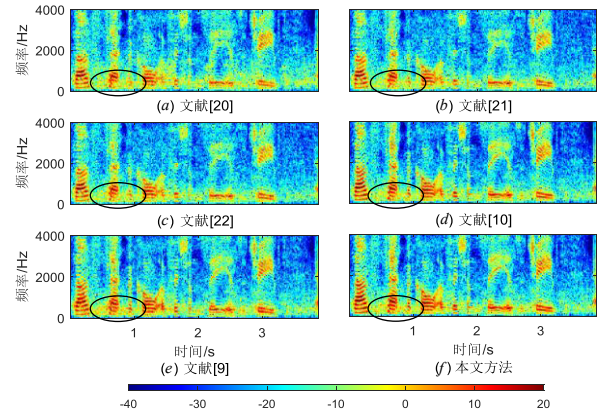


图8 增强语音的语谱图对比(快拍数为15)

随着快拍数的增大, $R_{IN}(k)$ 因取平均的原因使得精确性下降,从而导致干扰声源及噪声抑制性能下降.此外,精确的 $R_{IN}(k)$ 对导向矢量失配也更为敏感,更易出现目标语音失真的问题;而 $R_{IN}(k)$ 精确度随快拍数增大而降低,其在统计意义上更为平滑,使得各方法对导向矢量失配的敏感性下降,从而能够减小目标语音失真.因此, $SINR_{out}$ 随快拍数的增加表现为先增大后降低的趋势.

上述实验结果均表明,本文提出的方法抵抗导向矢量失配的能力更强,较参考方法具备更优的干扰声源抑制性能.

5 结论

针对远场语音通信下 MVDR 波束形成器对导向矢量失配敏感的问题,本文提出了一种基于 FSS 估计导向矢量的干扰声源抑制方法.该方法通过划分带宽及 FSS 构建分辨率更高 MUSIC 谱,更为准确地估计出初始 DOA, 在不确定集内搜索 Capon 功率最大值并约束导向矢量方向远离干扰声源空间,更为准确地估计了导向矢量,提高了 MVDR 波束形成器抵抗 DOA 及阵列结构失配的性能.实验结果表明,在不同的输入 SIR、输入 SNR 及快拍数情况下,所提方法较参考方法有更好的输出 SINR 和 PESQ,具备了更好的干扰声源抑制性能.

参考文献

- [1] BENESTY J, CHEN J, HUANG Y. Microphone Array Signal Processing[M]. Berlin, Heidelberg: Springer, 2008.
- [2] BRANDSTEIN M, WARD D. Microphone Arrays: Signal Processing Techniques and Applications[M]. Berlin, Heidelberg: Springer, 2001.
- [3] LIN J, PENG Q, HUANG Q. Adaptive beamforming with robustness against both finite-sample effects and steering vector mismatches[J]. IEICE Transaction on Fundamentals of Electronics, Communications and Computer Sciences,

- 2006, 89(A9): 2356-2362.
- [4] HENDRIKS R, GERKMANN T. Noise correlation matrix estimation for multi-microphone speech enhancement[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2012, 20(1): 223-233.
- [5] 何礼, 周翊, 刘宏清. 利用相位时频掩蔽的麦克风阵列噪声消除方法[J]. *信号处理*, 2018, 34(12): 1490-1498.
HE L, ZHOU Y, LIU H. Microphone array noise cancellation method using phase time-frequency masking[J]. *Journal of Signal Processing*, 34(12): 1490-1498. (in Chinese)
- [6] QIAN F, VAN VEEN B. Coherent interference suppression via partially adaptive beamforming[C]//*IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. San Francisco: IEEE, 1992: 441-444.
- [7] HERBORDT W, BUCHNER H, Nakamura S, et al. Multi-channel bin-wise robust frequency domain adaptive filtering and its application to adaptive beamforming[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2007, 15(4): 1340-1351.
- [8] ZHANG K, WEI Y, WU D, et al. Adaptive speech separation based on beamforming and frequency domain independent component analysis[J]. *Applied Sciences*, 2020, 10(7): 2593.
- [9] ZHANG P, YANG Z, LIAO G, et al. An RCB-like steering vector estimation method based on interference matrix reduction[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2021, 57(1): 636-646.
- [10] 杨志伟, 张攀, 陈颖, 等. 导向矢量和协方差矩阵联合迭代估计的稳健波束形成算法[J]. *电子与信息学报*, 2018, 40(12): 2874-2880.
YANG Z, ZHANG P, CHEN Y, et al. Steering vector and covariance matrix joint iterative estimations for robust beamforming[J]. *Journal of Electronics and Information Technology*, 2018, 40(12): 2874-2880. (in Chinese)
- [11] DMOCHOWSKI J, BENESTY J, Affes S. Broadband music: Opportunities and challenges for multiple source localization[C]//*IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. New York: IEEE, 2007: 18-21.
- [12] HERZOG A, HABETS E. Eigenbeam-ESPRIT for DOA-vector estimation[J]. *IEEE Signal Processing Letters*, 2019, 26(4): 572-576.
- [13] SU G, MORF M. The signal subspace approach for multiple wide-band emitter location[J]. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1983, 31(6): 1502-1522.
- [14] WANG H, KAVEH M. Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources[J]. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1985, 33(4): 823-831.
- [15] MA F, ZHANG X. Wideband DOA estimation based on focusing signal subspace[J]. *Signal, Image and Video Processing*, 2019, 13: 675-682.
- [16] BEIT-ON H, RAFAELY B. Focusing and frequency smoothing for arbitrary arrays with application to speaker localization[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2020, 28: 2184-2193.
- [17] 曹司磊, 曾维贵, 王磊. 色噪声下基于差分聚焦的宽带 DOA 估计方法[J]. *哈尔滨工业大学学报*, 2021, 53(2): 140-145.
CAO S, ZENG W, WANG L. DOA estimation of wide-band array with differential focusing under colored noise [J]. *Journal of Harbin Institute of Technology*, 2021, 53(2): 140-145. (in Chinese)
- [18] 贾思宇, 路茗, 丁华泽, 等. 一种改进的信号子空间聚焦宽带 DOA 估计算法[J]. *计算机工程*, 2022, 48(1): 175-181.
JIA S, LU M, DING H, et al. A modified wideband DOA estimation algorithm for focusing signal subspace[J]. *Computer Engineering*, 2022, 48(1): 175-181. (in Chinese)
- [19] LIU J, LI J. Robust detection in MIMO radar with steering vector mismatches[J]. *IEEE Transactions on Signal Processing*, 2019, 67(20): 5270-5280.
- [20] STOICA P, WANG Z, LI J. Robust Capon beamforming [J]. *IEEE Signal Processing Letters*, 2003, 10(6): 172-175.
- [21] GU Y, LESHEM A. Robust adaptive beamforming based on interference covariance matrix reconstruction and steering vector estimation[J]. *IEEE Transactions on Signal Processing*, 2012, 60(7): 3881-3885.
- [22] YANG Z, ZHANG P, LIAO G, et al. Robust beamforming via alternating iteratively estimating the steering vector and interference-plus-noise covariance matrix[J]. *Digital Signal Processing*, 2020, 99: 102620.
- [23] ZHOU Z, LIU Y, CHRISTENSEN M G, et al. A robust approach to the order detection for the damped sinusoids based on the shift-invariance property[C]//*15th IEEE International Conference on Signal Processing (ICSP)*. Beijing: IEEE, 2020: 472-477.
- [24] HIGUCHI T, ITO N, ARAKI S, et al. Online MVDR Beamformer based on complex Gaussian mixture model

with spatial prior for noise robust ASR[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017, 25(4): 780-793.

- [25] ZHOU M, MA X, SHEN P, et al. Weighted subspace-constrained adaptive beamforming for sidelobe control[J]. IEEE Communications Letters, 2019, 23(3): 458-461.
- [26] LI J, LIN Q, KANG C, et al. DOA estimation for underwater wideband weak targets based on coherent signal subspace and compressed sensing[J]. Sensors, 2018, 18(3): 902.
- [27] ZHU X, XU X, YE Z. Robust adaptive beamforming via subspace for interference covariance matrix reconstruction [J]. Signal Processing, 2020, 167: 107289.
- [28] KE Y, ZHENG C, PENG R, et al. Robust adaptive beamforming using noise reduction preprocessing-based fully automatic diagonal loading and steering vector estimation [J]. IEEE Access, 2017, 5: 12974-12987.
- [29] CHENG R, BAO C, CUI Z. MASS: Microphone array speech simulator in room acoustic environment for multi-channel speech coding and enhancement[J]. Applied Sciences, 2020, 10(4): 1484(1-17).
- [30] WAX M, ANU Y. Performance analysis of the minimum variance beamformer in the presence of steering vector errors[J]. IEEE Transactions on Signal Processing, 1996, 44(4): 938-947.
- [32] 鄢社锋. 优化阵列信号处理: 波束优化理论与方法[M]. 北京: 科学出版社, 2018.
- YAN S. Optimal Array Signal Processing: Beamforming Design Theory and Methods[M]. Beijing: Science Press, 2018. (in Chinese)



张 旭 女, 1995 生于内蒙古乌兰察布. 现为北京工业大学博士研究生. 主要研究方向为语音分离.

E-mail: zhangxu@emails.bjut.edu.cn

作者简介



周 静 男, 1993 年生于四川广安. 现为北京工业大学博士研究生. 主要研究方向为语音增强.

E-mail: zhoujing@emails.bjut.edu.cn



鲍长春(通讯作者) 男, 1965 年生于内蒙古赤峰, 中国电子学会会士, IEEE 高级会员. 现为北京工业大学教授、博士生导师. 主要研究方向为语音编码与语音增强.

E-mail: chchbao@bjut.edu.cn