## 一、本项目学术成果

### 1）本项目发表的期刊论文

[1] Zou Y X, Li B, Ritz C H. Multi-Source DOA Estimation Using an Acoustic Vector Sensor Array Under a Spatial Sparse Representation Framework[J]. Circuits, Systems, and Signal Processing, 2016, 35(3): 993-1020. 【SCI 收录 000370819100014】【EI 收录 20160801981561】；

[2] 邹月娴, 郭轶凡, 郑炜乔. 基于 AVS 和稀疏表示的鲁棒语者声源 DOA 估计方法[J]. 数据采集与处理, 2015, 30(2): 299-306；

[3] Zou Y X, Wang P, Wang Y Q, et al. Speech enhancement with an acoustic vector sensor: an effective adaptive beamforming and post-filtering approach[J]. EURASIP Journal on Audio, Speech, and Music Processing, 2014, 2014(1): 1-12. 【SCI 收录 000347390400001】【EI 收录 20142417806603】【影响因子 38】；

[4] 邹月娴, 王鹏, 王文敏. 一种基于单 AVS 的空间目标语音增强方法[J]. 清华大学学报: 自然科学版. 2013 (6): 883-887. 【EI 收录 20134416914714】；

[5] 胡旭琰, 邹月娴, 王文敏. 基于 MDT 特征补偿的噪声鲁棒语音识别算法[J]. 清华大学学报: 自然科学版, 2013 (6): 753-756. 【EI 收录 20134416914686】；

### 2）本项目发表的会议论文

[1] Yanhan Jin, Yuexian Zou, C. H. Ritz，“Robust Speaker 3-D DOA Estimation Based On The Inter-Sensor Data Ratio Model And Mask Estimation In The Bispectrum Domain”，IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2017)【接收发表】；

[2] Jin Y H, Zou Y X. Robust speaker DOA estimation with single AVS in bispectrum domain[C]//2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2016). IEEE, 2016: 3196-3200. 【EI 收录 20162402488768】；

[3] Zheng W Q, Zou Y X, Ritz C. Spectral mask estimation using deep neural networks for inter-sensor data ratio model based robust DOA estimation[C]//2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2015). IEEE, 2015: 325-329. 【EI 收录 20154501510470】；

[4] Zou Y X, Shi W, Li B, et al. Multisource DOA estimation based on time-frequency sparsity and joint inter-sensor data ratio with single acoustic vector sensor[C]//2013 IEEE International Conference on Acoustics, Speech and Signal Processing （ICASSP2013）, 2013: 4011-4015. 【EI 收录 20135217120852】；

[5] Zou Y, Guo Y, Zheng W, et al. An effective DOA estimation by exploring the spatial sparse representation of the inter-sensor data ratio model[C]// 2014 IEEE China Summit & International Conference on Signal and Information Processing (ChinaSIP2014), 2014: 42-46. 【EI 收录 20152100870017】；

[6] Guo Y, Zou Y X, Wang Y. A robust high resolution speaker DOA estimation under reverberant environment[C]// IEEE 9th International Symposium on Chinese Spoken Language Processing (ISCSLP2014), 2014: 400-400. 【EI 收录】；

[7] Shi W, Zou Y, Liu Y. Long-term auto-correlation statistics based voice activity detection for strong noisy speech[C]//Signal and Information Processing (ChinaSIP), 2014 IEEE China Summit & International Conference on. IEEE, 2014: 100-104. 【EI 收录 20152100870666】；

[8] Zou Y X, Zheng W Q, Shi W, et al. Improved voice activity detection based on support vector machine with high separable speech feature vectors[C]//2014 19th International Conference on Digital Signal Processing. IEEE, 2014: 763-767. 【EI 收录 20153601243014】；

[9] Zou Y X, Wang Y Q, Wang P, et al. An effective target speech enhancement with single acoustic vector sensor based on the speech time-frequency sparsity[C]//2014 19th International Conference on Digital Signal Processing. IEEE, 2014: 547-551. 【EI 收录 20153601242972】；

[10] Wang C, Zou Y, Liu S, et al. An Efficient Learning Based Smartphone Playback Attack Detection Using GMM Supervector[C]//Multimedia Big

Data (BigMM), 2016 IEEE Second International Conference on. IEEE, 2016: 385-389. 【EI 收录】；

[11] Shihan Liu, Yuexian Zou, “Multi-Constraint Nonnegative Matrix Factorization Approach to Speech Enhancement with Nonstationary Noise,” International Conference on Intelligence Science and Big Data Engineering (IScIDE). pp. 181-191, Guangzhou, China, May, 2016；

[12] Chun Wang, Yuexian Zou, Weiqiao Zheng, Wei Shi, “An Efficient Playback Attack Detection Approach Based on Supervised Learning.” IEEE International Conference on Intelligence Science and Big Data Engineering (IScIDE), Guangzhou, China, May 13-15, 2016；

[13] Liu S H, Zou Y X, Ning H K. Nonnegative matrix factorization based noise robust speaker verification[C]//Signal and Information Processing (ChinaSIP), 2015 IEEE China Summit and International Conference on. IEEE, 2015: 35-39. 【EI 收录 20160701912123】；

[14] Wang C, Shi W, Zou Y X. Multi-pronunciation dictionary construction for Mandarin-English bilingual phrase speech recognition system[C]//Signal and Information Processing (ChinaSIP), 2015 IEEE China Summit and International Conference on. IEEE, 2015: 15-19.【EI 收录 20160701912119】

[15] Zheng W Q, Yu J S, Zou Y X. An experimental study of speech emotion recognition based on deep convolutional neural networks[C]//Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on. IEEE, 2015: 827-831. 【EI 收录 20161502238751】

[16] Liu J H, Zheng W Q, Zou Y X. A Robust Acoustic Feature Extraction Approach Based On Stacked Denoising Autoencoder[C]//Multimedia Big Data (BigMM), 2015 IEEE International Conference on. IEEE, 2015: 124-127. 【EI 收录 20153701270922】；

[17] Hu X Y, Zou Y X, Shi W. An effective missing feature compensation method for speech recognition at noisy environment[C]//Signal and Information Processing (ChinaSIP), 2014 IEEE China Summit & International Conference on. IEEE, 2014: 133-137. 【EI 收录 20152100870641】；

[18] Ning H, Zou Y X, Hu X. A new score normalization for text-independent speaker verification[C]//2014 19th International Conference on Digital Signal Processing. IEEE, 2014: 636-639. 【EI 收录 20153601242990】

**3) 本项目发表的专利**

[1] 邹月娴，郭轶凡，石伟，“一种基于 AVS 和稀疏表示的单语者声源 DOA 估计方法”，中国发明专利号：1，申请时间 2013-12-25；授权时间 2016-05-18；

[2] 邹月娴，郑炜乔，王永庆，石伟，王春，郭轶凡，宁洪珂，刘诗涵，“一种基于声学矢量传感器的语音控制智能垃圾桶”，申请时间：2014-09-03，发明专利申请号：8；授权时间 2016-09-10；

[3] 邹月娴，王鹏，“一种基于时频掩膜的单声学矢量传感器目标语音增强方法”，2013，发明专利申请号：0，实审

[4] 邹月娴，郑炜乔，余嘉胜，王毅，柳俊宏，陈锦，黄晓林，金彦含，“一种语音控制拍照软件”，2015，发明专利申请号：1，实审

[5] 邹月娴，金彦含，“一种基于声学矢量传感器和双谱变换的鲁棒单语者声源 DOA 估计方法”，2016，发明专利申请号：5，实审

## 二、本项目主要研究工作

课题组的主要研究内容体现在五个方面：1）语音声源 DOA 估计稀疏表示模型构建方法研究；2）基于语音声源 DOA 估计稀疏模型的稀疏矢量求解算法研究；3）基于 AVSA 阵型和稀疏表示的语音声源 DOA 估计方法研究；4）基于所提出算法的机器人语音声源 DOA 估计实验系统和性能研究；5）服务机器人听觉系统关键技术算法研究。

### （一）基于信源空间稀疏性、稀疏表示模型和 AVS 阵列的 DOA 估计算法研究

课题组开展了基于信源空间稀疏性、稀疏表示模型和 AVS 阵列的 DOA 估计算法的研究。首先将主流的基于信源空间稀疏性和稀疏表示模型的 DOA 估计方法扩展到 AVS 阵列，提出了基于空间稀疏性和稀疏表示的 AVS 阵

列 DOA 估计模型，即 AVS-SS-DOA 方法。与传统的 DOA 估计方法相比，AVS-SS-DOA 方法取得了更高的 DOA 估计精度和对噪声的鲁棒性。不足在于，DOA 估计精度与用以构造过完备字典的空间离散角度网格精度成正相关，即要达到较高的 DOA 估计精度则必须要采用较密的采样网格对空间进行离散，这会导致很高的计算复杂度；课题组开展了保持 DOA 估计高精度不变的前提下，降低 DOA 估计算法复杂度的问题研究，提出了一种新的算法，即 AVS-SS-LF 算法，具体算法思想是首先采用 AVS 阵列中的全向传感器(o)子阵接收信号和 AVS-SS-DOA 方法以较粗略的网格实现 o 子阵流形矩阵的初始估计，再利用其它三个子阵(u，v，w)数据与 o 子阵数据的线性关系进行直线拟合（LF），再结合 AVS 的流形结构进行 DOA 估计，AVS-SS-LF 算法如表 1 所示。研究发现，当加性噪声较弱时，AVS-SS-LF DOA 估计算法能够以很低的复杂度有效的消除网格效应，提高 DOA 估计精度。但噪声强度增强，信噪比小于 10dB 后，AVS-SS-LF 算法性能下降；针对该问题，课题组采用子空间信号处理策略，提出了一种新的 AVS-SS-ST DOA 估计算法。AVS-SS-ST DOA 估计算法的思想是：通过对阵列接收数据 x(t)的自相关矩阵的估计和分解，获得信号子空间，根据 AVS 阵列中各子阵接收数据的信号子空间的不变关系求解信源到达方向角的方向余弦并估计 DOA。AVS-SS-ST DOA 估计算法如表 2 所示。研究发现，AVS-SS-ST 算法能够改善粗略的初始 DOA 估计的精度抗噪鲁性。

表 1 AVS-SS-LF 算法流程

| **AVS-SS-LF 算法流程** |
|---|
| Input: $\mathbf{X}_o$， $\mathbf{X}_u$， $\mathbf{X}_v$， $\mathbf{X}_w$， $\Theta=\{(\tilde{\theta}_1\ \tilde{\phi}_1),\ldots,(\tilde{\theta}_{N_1}\ \tilde{\phi}_{N_1})\}$ |
| 1 Begin |
| 2 构造 $\boldsymbol{\Psi}_o \equiv [\mathbf{q}(\tilde{\theta}_1\ \tilde{\phi}_1),\ldots,\mathbf{q}(\tilde{\theta}_{N_1}\ \tilde{\phi}_{N_1})]$； |
| 3 采用 $\ell_1$-SVD 方法求解式(3-20)得 $\hat{\mathbf{Z}}$； |
| 4 从式(1)估计 $\hat{\mathbf{A}}_o$； |
| 5 计算 $\mathbf{S}_u$, $\mathbf{S}_v$, $\mathbf{S}_w$ 及 $\mathbf{S}_o$ 的最小二乘解(2)； |
| 6 从(3)估计 $\hat{u}_k$ 及相应的估计 $\hat{v}_k$ 和 $\hat{w}_k$； |
| 7 根据式(4)求得最终 DOA 估计 $\hat{\theta}_k$ 和 $\hat{\phi}_k$； |
| 8 End |
| Output: $\hat{\theta}_k$ 和 $\hat{\phi}_k$， $k=1,\ldots,K$ |

$$\hat{\mathbf{A}}_o = \boldsymbol{\Psi}_o(:,\mathbf{I}),\quad \mathbf{I}=[I_1,\ldots,I_K] \tag{1}$$

$$[\hat{\mathbf{S}}_u,\hat{\mathbf{S}}_v,\hat{\mathbf{S}}_w,\hat{\mathbf{S}}_o]=\hat{\mathbf{A}}_o^{\perp}[\mathbf{X}_u,\mathbf{X}_v,\mathbf{X}_w,\mathbf{X}_o] \tag{2}$$

$$\hat{u}_k=\frac{2KT\vec{\mathbf{S}}_u{}^T\vec{\mathbf{S}}_o-\mathbf{1}^T\vec{\mathbf{S}}_u\mathbf{1}^T\vec{\mathbf{S}}_o}{2KT\vec{\mathbf{S}}_o{}^T\vec{\mathbf{S}}_o-\mathbf{1}^T\vec{\mathbf{S}}_o\mathbf{1}^T\vec{\mathbf{S}}_o} \tag{3}$$

$$\hat{\theta}_k=\cos^{-1}\hat{w}_k,\quad \hat{\phi}_k=\tan^{-1}\hat{v}_k/\hat{u}_k \tag{4}$$

表 2 AVS-SS-ST 算法流程

| **AVS-SS-LF 算法流程** |
|---|
| Input: $\mathbf{X}_o$， $\mathbf{X}_u$， $\mathbf{X}_v$， $\mathbf{X}_w$， $\Theta=\{(\tilde{\theta}_1\ \tilde{\phi}_1),\ldots,(\tilde{\theta}_{N_1}\ \tilde{\phi}_{N_1})\}$ |
| 1 Begin |
| 2 构造 $\boldsymbol{\Psi}_o \equiv [\mathbf{q}(\tilde{\theta}_1\ \tilde{\phi}_1),\ldots,\mathbf{q}(\tilde{\theta}_{N_1}\ \tilde{\phi}_{N_1})]$； |
| 3 采用 $\ell_1$-SVD 方法求解式(3-20)得 $\hat{\mathbf{Z}}$； |
| 4 从式(1)估计 $\hat{\mathbf{A}}_o$； |
| 5 由 $\mathbf{X}$ 近似计算 $\mathbf{R}_{\mathbf{xx}}=\mathbf{X}\mathbf{X}^T/L$； |
| 6 对 $\mathbf{R}_{\mathbf{xx}}$ 特征分解得 $\mathbf{E}_S=[\mathbf{E}_{su}^T,\ \mathbf{E}_{sv}^T,\ \mathbf{E}_{sw}^T,\ \mathbf{E}_{so}^T]^T$； |
| 7 根据式(5)及(6)求得式 $\hat{\boldsymbol{\Lambda}}^u$， $\hat{\boldsymbol{\Lambda}}^v$ 及 $\hat{\boldsymbol{\Lambda}}^u$； |
| 8 根据式(7)求得 $\hat{u}_k$， $\hat{v}_k$ 和 $\hat{w}_k$； |
| 9 根据式(4)求得最终 DOA 估计 $\hat{\theta}_k$ 和 $\hat{\phi}_k$； |
| 10 End |
| Output: $\hat{\theta}_k$ 和 $\hat{\phi}_k$， $k=1,\ldots,K$ |

$$\hat{\boldsymbol{\Lambda}}^u=\hat{\mathbf{A}}_o^{\perp}\mathbf{E}_{su}\mathbf{E}_{so}^{\perp}\hat{\mathbf{A}}_o \tag{5}$$

$$\begin{aligned}\hat{\boldsymbol{\Lambda}}^v&=\hat{\mathbf{A}}_o^{\perp}\mathbf{E}_{sv}\mathbf{E}_{so}^{\perp}\hat{\mathbf{A}}_o\\ \hat{\boldsymbol{\Lambda}}^w&=\hat{\mathbf{A}}_o^{\perp}\mathbf{E}_{sw}\mathbf{E}_{so}^{\perp}\hat{\mathbf{A}}_o\end{aligned} \tag{6}$$

$$\hat{u}_k = [\hat{\mathbf{\Lambda}}^u]_{kk},\ \ \hat{v}_k = [\hat{\mathbf{\Lambda}}^v]_{kk},\ \ \hat{w}_k = [\hat{\mathbf{\Lambda}}^w]_{kk} \qquad (7)$$

课题组采用仿真实验的方式进行了 DOA 估计性能测试，7 个窄带随机空间信源从不同的方向到达由 13 个 AVS 组成的三维阵列，信噪比为 30dB。仿真实验结果如图 1 和图 2 所示。分析图 1 和图 2 的结果，我们得到如下**研究结论**：（1）由图 1 可以看出，相比于 L1-SVD 和 L1-SVD-GR 算法，AVS-SS-LF 和 AVS-SS-ST 在信源间隔小于等于 20° 时具有更高的 DOA 估计精度，且在两信源间隔只有 4°时仍非常出色，表明提出的算法能够极大地提升紧密间隔信源的 DOA 估计精度；（2）由图 2 可以看出，在信噪比等于 0dB 时，AVS-SS-ST 性能要略差于 L1-SVD-GR，但在信噪比大于等于 5dB 时，AVS-SS-ST 算法具有最好的 DOA 估计精度。同时当噪声水平适中，即 SNR 大于 10dB 时，AVS-SS-LF 算法性能与 AVS-SS-ST 算法相当。但随着 SNR 降低到 10dB 以下，AVS-SS-LF 算法的 DOA 估计精度迅速下降。因此，AVS-SS-ST 算法中的子空间技术有效的提升了 DOA 估计对噪声的鲁棒性。

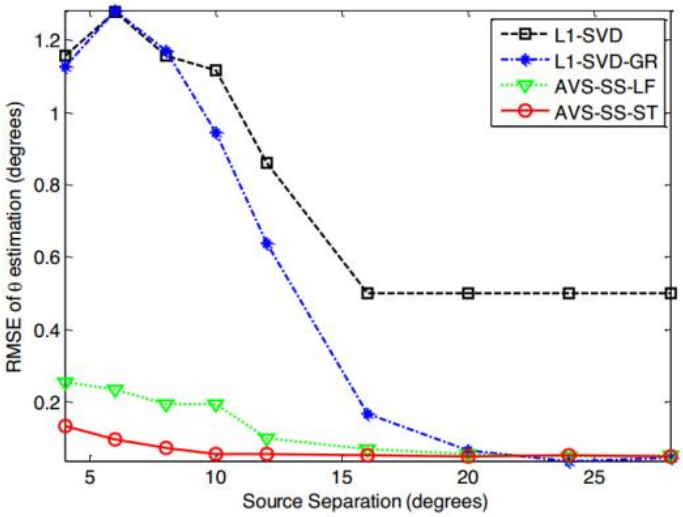


图 1 RMSE versus 信源空间间隔（DOA1=42.5° ，SNR=30dB，离散网格=2° ）

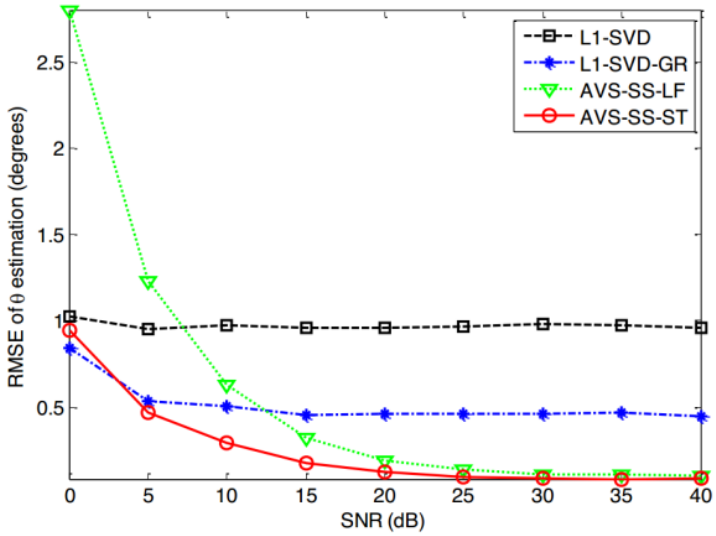


图 2 RMSE versus 信噪比（两信源的 DOA 角度：(63.3°,0°)及(73.3°,0°)；离散网格=2°）

**（二）基于语音时频稀疏性和单个 AVS 的多声源 DOA 估计算法研究**

课题组开展了基于语音时频稀疏性的 AVS 多声源 DOA 估计算法的研究。其研究重点为：首先利用 AVS 各个阵元接收到的信号是同相的，或相位差异很小可以忽略不计这一特点，计算各个阵源输出的信号的幅度比（Inter-Sensor Data Ratio，**ISDR**），利用 AVS 的导向矢量中各个分类的三角函数关系，提取各个声源的方向信息；然后基于多声源的时频稀疏性假设，利用正弦迹法提取满足时频单一声源主导的时频点；最后利用核密度估计的方法对声源主导的时频点进行聚类，通过聚类中心的数值估计多个声源的 DOA。研究过程中，课题组针对四分量 AVS 和三分量 AVS 分别进行了 DOA 估计原理的推导，分别提出了两个 DOA 估计算法，即 AVS4-ISDR 和 AVS3-ISDR，其具体算法流程如表 3 和表 4 所示。

表 3 AVS4-ISDR 算法流程

| **AVS4-ISDR 算法流程** |
| --- |
| **Input:** $o(t)$，$x(t)$，$y(t)$，$z(t)$ |
| **1** **Begin** |
| **2** 对$o(t)$、$x(t)$、$y(t)$、$z(t)$短时傅里叶变换，得$O(\tau,\omega)$、$X(\tau,\omega)$、$Y(\tau,\omega)$、$Z(\tau,\omega)$； |
| **3** 从$O(\tau,\omega)$中提取正弦迹，其上时频点组成的集合记为 S； |
| **4** 对集合 S 中的所有时频数据点，根据式(9)计算 ISDR $I_{uo}(\tau,\omega)$、$I_{vo}(\tau,\omega)$和$I_{wo}(\tau,\omega)$； |
| **5** 采用 KDE 算法分别计算$I_{uo}(\tau,\omega)$、$I_{vo}(\tau,\omega)$和$I_{wo}(\tau,\omega)$的联合概率密度函数； |
| **6** 搜索联合概率密度函数的前 $K$ 个峰，其位置即为 $K$ 个声源 DOA 的方向余弦的估计值； |
| **7** 由式(4)求得最终的 DOA 估计$\hat{\theta}_k$和$\hat{\phi}_k$； |
| **8** **End** |
| **Output:** $\hat{\theta}_k$和$\hat{\phi}_k$，$k=1,\ldots,K$ |

$$I_{io}(\tau,\omega) \triangleq \frac{Y_i(\tau,\omega)}{Y_o(\tau,\omega)} \tag{9}$$

$$I_{ij}(\tau,\omega) \triangleq \frac{Y_i(\tau,\omega)}{Y_j(\tau,\omega)} \tag{10}$$

$$\begin{aligned}\hat{\phi}_k &= \tan^{-1}\left(\hat{v}_k / \hat{u}_k\right)\\ \hat{\theta}_k &= \tan^{-1}(\hat{u}_k / \hat{w}_k / \cos(\hat{\phi}_k))\end{aligned} \tag{11}$$

表 4 AVS3-ISDRWO 算法流程

| **AVS3-ISDRWO 算法流程** |
| --- |
| **Input:** $x(t)$，$y(t)$，$z(t)$ |
| **1** **Begin** |
| **2** 对$x(t)$，$y(t)$，$z(t)$短时傅里叶变换，得$X(\tau,\omega)$、$Y(\tau,\omega)$、$Z(\tau,\omega)$； |
| **3** 从$X(\tau,\omega)$、$Y(\tau,\omega)$、$Z(\tau,\omega)$分别中提取正弦迹，取时频点最多的正弦迹记为 S； |
| **4** 对集合 S 中的所有时频数据点，根据式(10)计算 ISDR $I_{yx}(\tau,\omega)$、$I_{xw}(\tau,\omega)$； |
| **5** 采用 KDE 算法分别计算$I_{yx}(\tau,\omega)$、$I_{xw}(\tau,\omega)$的联合概率密度函数； |
| **6** 搜索联合密度函数的前 $K$ 个峰，其位置即为 $K$ 个声源的 DOA 的方向余弦的估计值； |
| **7** 由式(11)求得最终的 DOA 估计$\hat{\theta}_k$和$\hat{\phi}_k$； |
| **8** **End** |
| **Output:** $\hat{\theta}_k$和$\hat{\phi}_k$，$k=1,\ldots,K$ |

课题组采用仿真实验和实测实验相结合的方式进行了 DOA 估计性能测试，其中，仿真实验的采样率 Fs=8kHz，语音长度 3s，AVS 中的每个分量传感器上的加性高斯白噪声为统计独立，实测实验为实验室自主设计和实现的 DOA 估计实验原型系统，实验室混响时间 $RT_{60}$=30ms，具有电脑和空调等

产生的噪声，DOA 估计实验原型系统的具体实现细节在第(九)部分阐述。仿真实验结果如图 3 和表 5 所示，实测实验的结果如表 6 所示。 分析实验结果得出如下**研究结论**：（1）由图 3 和表 5 可以看出，在欠定条件（声源个数大于阵元个数）下，单颗 AVS 可以同时估计出 7 个空间声源的 DOA 估计结果，说明了 AVS-ISDR 算法对多声源 DOA 估计性能的强大；（2）由表 6 可以看出，实测 DOA 估计结果的绝对值误差小于 10°，说明了 AVS-ISDR 算法在实际应用中的有效性。

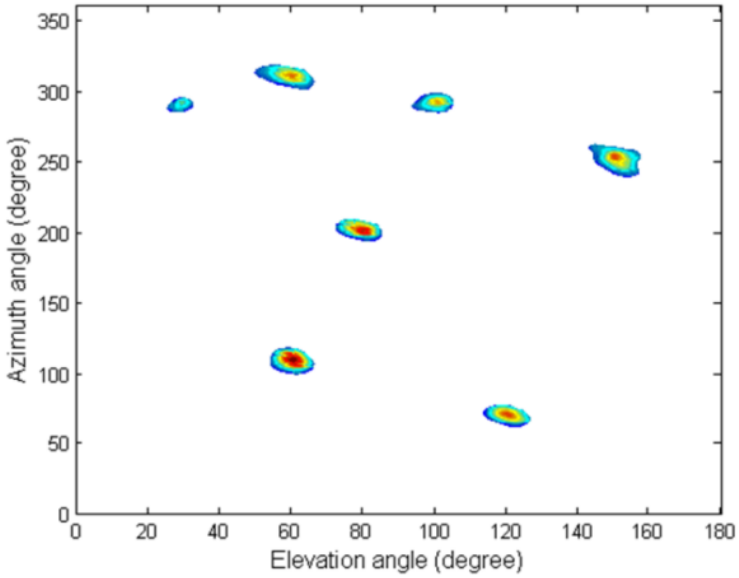


图 3 估计出的声源在角度域的分布结果

表 5 欠定条件下 AVS4-ISDR 方法估计结果

| | | | | |
|---|---|---|---|---|
| 真实DOA (°) | (30, 290) | (60, 110) | (100, 290) | (120, 70) |
| 估计值 (°) | (30.59,289.4) | (59.48, 111.4) | (99.84, 291.2) | (120.3, 70.21) |
| 绝对误差 (°) | (0.59,0.6) | (0.52,1.4) | (0.16,1.2) | (0.3,0.21) |
| 真实DOA (°) | (150, 250) | (60, 310) | (80,200) | |
| 估计值 (°) | (150.7, 251.9) | (60.48, 310) | (78.91,201.3) | |
| 绝对误差 (°) | (0.7,1.9) | (0.48,0) | (1.09,1.3) | |

表 6 AVS4-ISDR 方法实测实验结果

| 真实DOA (°) | 15 | 30 | 45 | 60 | 75 |
|---|---|---|---|---|---|
| 测试结果 (°) | 14.96 | 31.46 | 47.85 | 56.41 | 66.44 |

### (三) 基于单颗 AVS 和稀疏表示理论的鲁棒语者声源 DOA 估计算法研究

课题组开展了基于单颗 AVS 和稀疏表示理论的鲁棒语者声源 DOA 估计算法的研究。其研究重点为：针对混响和加性噪声同时存在的场景，在 ISDR 模型基础上，根据声源的空域稀疏性，将空间等间隔划分并构建过完备字典，进而推导出 DOA 估计的空域稀疏表示模型 AVS-SSR，把 DOA 求解问题转换成了稀疏矢量求解问题；通过重构空间谱获得高精度的 DOA 估计，并利用奇异值分解技术来降低重构问题的运算复杂度。其算法流程及框图分别如表 7 和图 4 所示。

表 7 AVS-SSR 算法流程

| AVS-SSR 算法流程 | |
|---|---|
| | Input: $x_u(t)$、$x_v(t)$、$x_w(t)$、$x_o(t)$ |
| 1 | Begin |
| 2 | 对 $x_u(t)$、$x_v(t)$、$x_w(t)$ 和 $x_o(t)$进行 STFT； |
| 3 | 由式(9)计算 HLSNR 时频点对应的 ISDR 值 $\boldsymbol{I}(\tau,\omega)$； |
| 4 | 构造式(12)中的数据矩阵 $\boldsymbol{A}$； |
| 5 | 构造式(13)中的过完备字典 $\boldsymbol{\Psi}$； |
| 6 | 利用 $l_1$-SVD 方法求解式(14)中的 $\boldsymbol{Z}$; |
| 7 | 通过式(15)计算 $\boldsymbol{P}_Z$ 和 $i_p$； |
| 8 | 通过式(16)从 $i_p$ 中计算语者声源所在网格$(i,j)$，并估计出 DOA。 |
| 9 | End |
| | Output: $(\theta_s,\phi_s)$ |

$$\boldsymbol{A}=[\boldsymbol{I}(\tau_1,\omega_1),\ldots,\boldsymbol{I}(\tau_L,\omega_L)],\boldsymbol{A}\in\boldsymbol{R}^{3\times L} \quad (12)$$

$$\boldsymbol{\Psi}=\left[\boldsymbol{b}(\theta_1,\phi_1),\boldsymbol{b}(\theta_1,\phi_2)\ldots,\boldsymbol{b}(\theta_i,\phi_j),\ldots,\boldsymbol{b}(\theta_{N_1},\phi_{N_2-1}),\boldsymbol{b}(\theta_{N_1},\phi_{N_2})\right],\boldsymbol{\Psi}\in\boldsymbol{R}^{3\times M} \quad (13)$$

$$\boldsymbol{Z} = \arg\min_{\boldsymbol{Z}} \|\boldsymbol{A} - \boldsymbol{\Psi}\boldsymbol{Z}\|_2^2 + \lambda\|\boldsymbol{Z}\|_1 \qquad (14)$$

$$P_{\boldsymbol{Z}}(i) = 10\log\sum_{j=1}^{L} \boldsymbol{Z}_{SV}^2(i,j), i = 1,\ldots,M \qquad (15)$$

$$\theta_s = \theta_i, \phi_s = \phi_j,\ \{\theta_i, \phi_j\} \in \Theta \qquad (16)$$

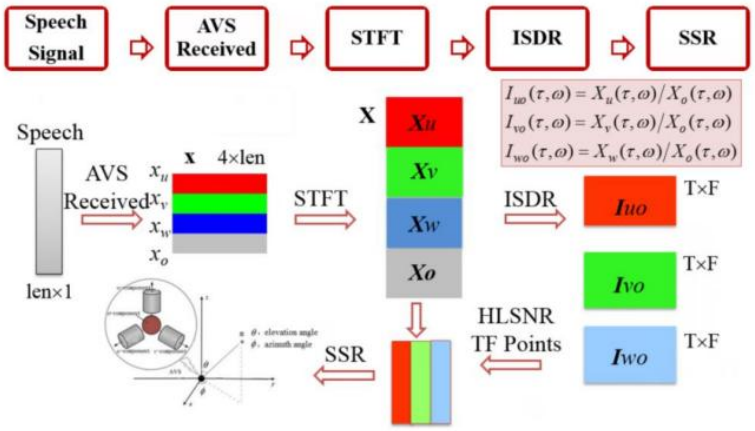


图 4 AVS-SSR 算法流程框图

课题组采用仿真实验和实测实验相结合的方式进行了 DOA 估计性能测试，其中，仿真实验的采样率 Fs=32kHz，语音长度 3s，AVS 中的每个分量传感器上的加性高斯白噪声为统计独立，实测实验为实验室自主搭建的系统，实验室混响时间 RT60=30ms，具有电脑和空调等产生的噪声,具体实现细节见第(九)部分。图 5、图 6、图 7 显示了仿真实验的结果，表 8 显示了实测实验的结果。根据实验结果，我们得出如下**研究结论**：（1）由图 5 可以看出，AVS-SSR 算法在所有的方位角下均具有更低的角度估计误差，且最大误差要小于 1°；（2）由图 6 可以看出，2 个算法的估计精度都随着信噪比的增大而提高，且 AVS-SSR 算法的估计误差要更小；(3)AVS-SSR 算法的 RMSE 基本不随混响时间而变化（均小于 0.1°）且要比 GMDA-Laplace 更小，说明了 AVS-SSR 算法对混响具有很好的鲁棒性。（4）从表 8 可以看出，实

际环境下的 DOA 估计绝对误差均不大于 7°，显示出 AVS-SSR 算法在实际应用中的有效性。

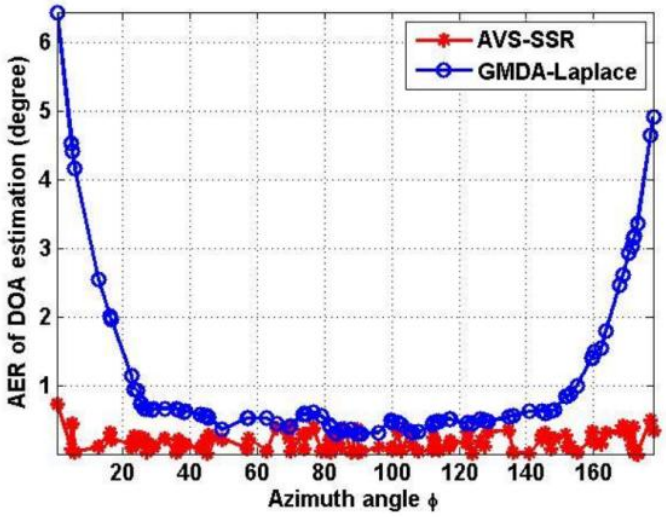


图 5 绝对值误差 AER versus 方位角（信噪比 SNR=10dB，不考虑混响，俯仰角固定为 60°，方位角在 0-180°变化）

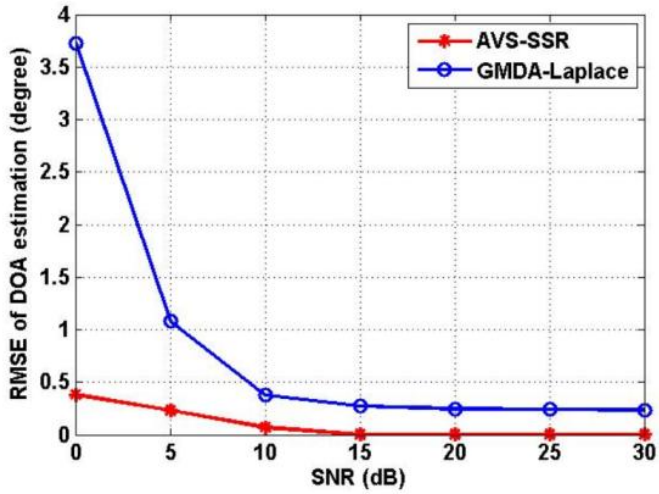


图 6 不同信噪比 SNR 下的 DOA 估计 RMSE（目标语音声源位于（60°，45°），信噪比从 0 到 30dB 变化，间隔为 5dB）

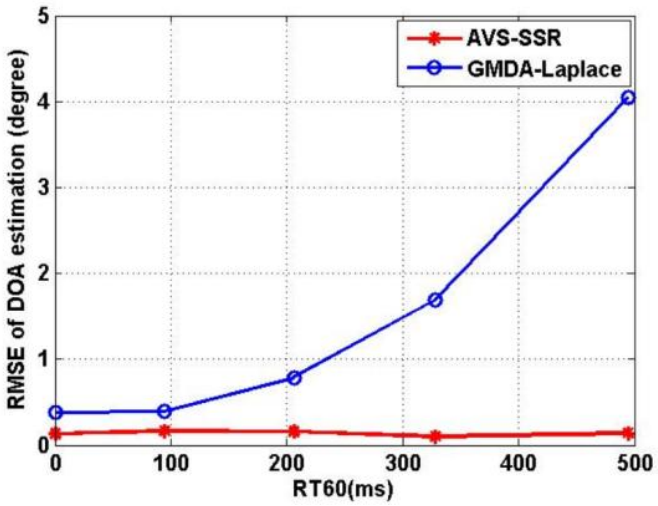


图 7 不同混响条件下的 DOA 估计 RMSE（语者声源在距 AVS 为 2m 的 (60°,45°)方向，SNR 设为 10dB）

表 8 实际环境下的 DOA 估计结果

| $(\theta_s,\phi_s)(°)$ | (90,0) | (90,45) | (90,90) | (90,135) | (90,180) |
|---|---|---|---|---|---|
| DOA 估计 (°) | (87,4) | (90,41) | (92, 89) | (86,135) | (86,180) |
| 绝对值误差 | 7 | 4 | 3 | 4 | 4 |

**（四）基于双颗 AVS 和稀疏表示的高精度语音声源 DOA 估计算法研究**

为了将 AVS 更好地应用于智能家居中的服务机器人，获得高精度的语者声源 DOA 估计，课题组开展了基于两颗 AVS 和稀疏表示的语音声源 DOA 估计算法，即 DAVS-SSR DOA 估计算法。其研究重点为：首先分别为每颗 AVS 计算的传感器间数值比 ISDR，根据声源空域稀疏性构建了 ISDR 的稀疏表示模型，将 DOA 估计问题转换为稀疏矩阵的重构问题；接着利用两颗 AVS 中全向麦克风接收信号的关系，定义了相关系数的计算方法，并以此选取了满足时频稀疏性的高局部信噪比时频点；最终获得了高精度的声源 DOA 估计。DAVS-SSR DOA 估计算法如表 9 所示。

表 9 DAVS-SSR 算法流程

| DAVS-SSR 算法流程 |
| --- |
| **Input**: $x_l(t)$、$x_r(t)$ |
| **1** **Begin** |
| **2** 对 $x_l(t)$和 $x_r(t)$进行 STFT 变换得到 $X_l(\tau,\omega)$和 $X_r(\tau,\omega)$； |
| **3** 采用相关系数法法提取 HLSNR TF 点集合{$(\tau, \omega)$\| $\|\gamma(\tau,\omega)\|>c=h_4\cdot\max(\|\gamma\|)$ }； |
| **4** 计算 HLSNR TF 点对应的 ISDR 值$I_l(\tau,\omega)$和$I_r(\tau,\omega)$； |
| **5** 构造式(17)中的数据矩阵 $\boldsymbol{A}$ 和过完备字典 $\boldsymbol{\Psi}$； |
| **6** 利用 OMP 方法稀疏重构式(18)中的 $\boldsymbol{Z}$; |
| **7** 计算 $\boldsymbol{P}_Z$ 和前 $K$ 个峰值，并估计出 $K$ 个声源的 DOA。 |
| **8** **End** |
| **Output**: $(\theta_1,\phi_1)$, ….,$(\theta_K,\phi_K)$ |

$$\boldsymbol{A}=\boldsymbol{\Psi Z}+\boldsymbol{E} \qquad (17)$$

$$\boldsymbol{Z}=\arg\min_{\boldsymbol{Z}}\|\boldsymbol{A}-\boldsymbol{\Psi Z}\|_2^2+\lambda\|\boldsymbol{Z}\|_0^0 \qquad (18)$$

课题组采用仿真实验的方式进行了 DOA 估计性能测试，采样率 Fs=32kHz，语音长度 3s，AVS 中的每个分量传感器上的加性高斯白噪声为统计独立。图 8，图 9 显示了仿真实验的结果。**我们的研究结论如下：**（1）从图 8 可以看出，在所有间隔角度下，基于相关系数和稀疏表示的 DAVS-SSR 算法的 DOA 估计精度均优于基于正弦迹和聚类的 DAVS-Cluster 算法；（2）在所有测试的信噪比下，DAVS-SSR 算法的性能都要优于 DAVS-Cluster 算法，尤其是当 SNR 大于 15dB 时，DAVS-SSR 算法的 DOA 估计 RMSE 结果几乎为 0，当 SNR 为 0dB 时，DOA 估计 RMSE 误差也都小于 0.5°。另外，从图中可以看出，两个算法在信噪比从 0dB 到 30dB 的变换中，RMSE 误差均小于 1°，说明了双颗 AVS 用于声源 DOA 估计很具优势。

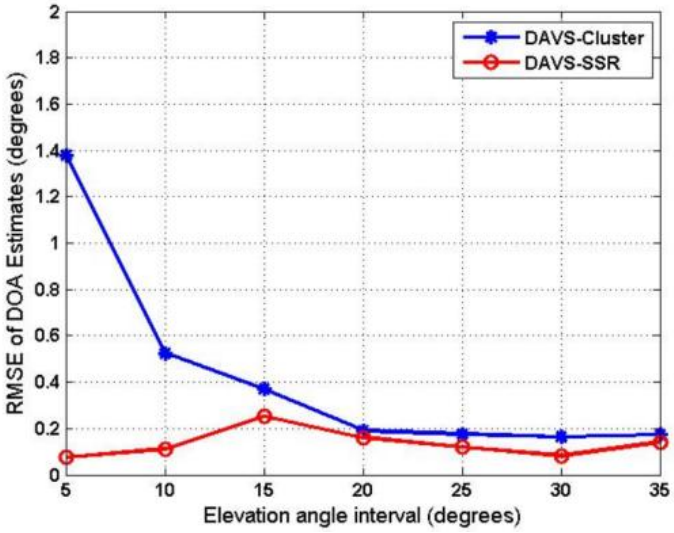


图 8 DOA 估计的 RMSE 随声源间隔的变化曲线（DOA1 为(45°， 60°)，DOA2 则从(50°,60°)改变到(80°， 60°)，信噪比 SNR=10dB，混响时间为 328.1ms。）

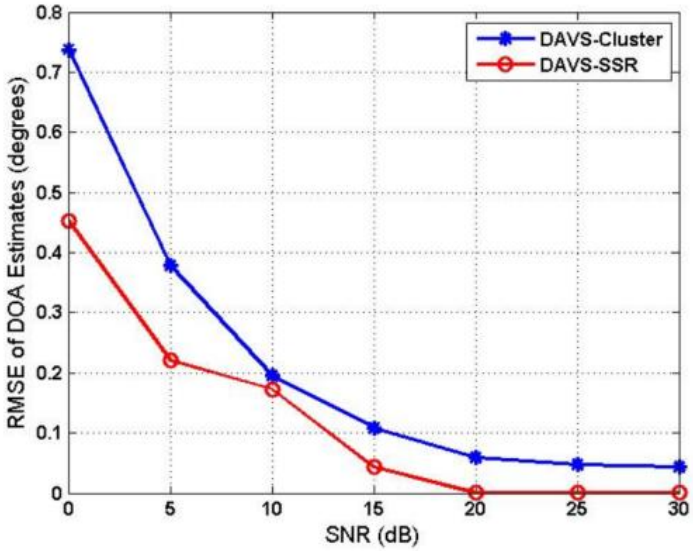


图 9 RMSE versus 信噪比（DOA1=（20°， 60°），DOA2=（70°， 60°），信噪比 SNR 从 0dB 变化到 30dB，间隔 5dB，混响时间为 328.1ms）

**（五）基于深度神经网络的高局部信噪比时频点的提取算法研究**

如上所述，高局部信噪比时频点的提取均为核心步骤。借助深度学习的成果，针对本课题提出的 AVS-ISDR 算法在低信噪比和混响条件中性能不够稳定

的问题，重点开展了基于深度神经网络（DNNs）的谱掩膜估计方法研究，提取高局部信噪比时频点。其研究重点为：通过 DNNs 的非线性映射学习语音与噪声的函数表示关系，提取了主导语音能量的时频点（掩膜单元设为1），同时抑制了噪声干扰的时频点（掩膜单元设为 0），在 AVS-ISDR 框架下，提出了一种鲁棒的多语音声源 DOA 估计算法即 AVS-DNN-ISDR 算法，其算法流程如图 10 所示。

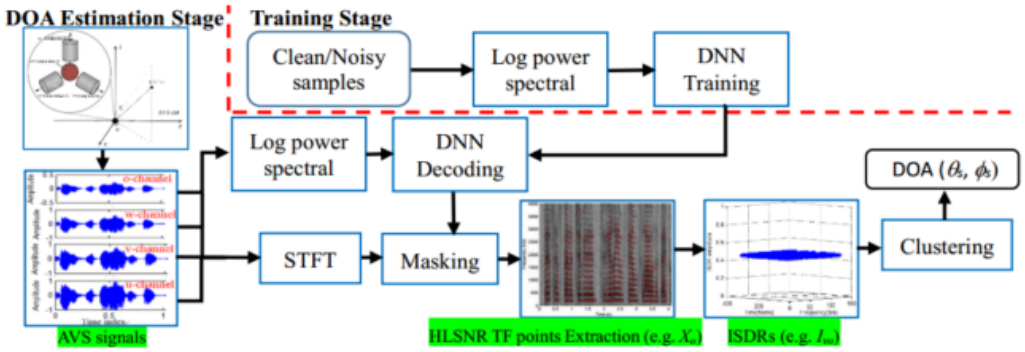


图 10 基于 DNNs 的高局部信噪比时频点提取的 DOA 估计系统框图

课题组采用仿真实验和实测实验相结合的方式进行了 DOA 估计性能测试，所有实验运行在 CPU 主频为 2.53GHz，处理器为 Intel(R) Core(TM)2 Duo CPUP8700，4G 内存的 64 位操作系统上，训练数据来自 TIMIT 数据库，整个网络包括 3 个隐层，每个隐层包括 512 个神经元，其中实测系统见第(九)部分。图 11 显示了基于 DNNs 的高局部信噪比时频点的提取效果，图 12 显示了仿真实验的结果，表 10 显示了实测实验的结果。通过实验结果分析，我们得出如下**研究结论**：实验证明，该方法在强噪声环境下的高局部信噪比时频点的提取性能受噪声影响较小，具有很好的稳定性，且保持了较高的精度，在 SNR=-5dB 时，DOA 估计的 RMSE 仍小于 5°；AVS-DNN-ISDR 算法的整体性表现优于 AVS-ISDR 算法，在低信噪比条件下（小于 10dB）明显优于 AVS-ISDR 算法；此外，AVS-DNN-ISDR 算法的抗混响能力优于

AVS-ISDR 算法，该项研究成果为智能服务机器人的语音声源 DOA 估计技术的实际应用提供了可行性和关键技术。

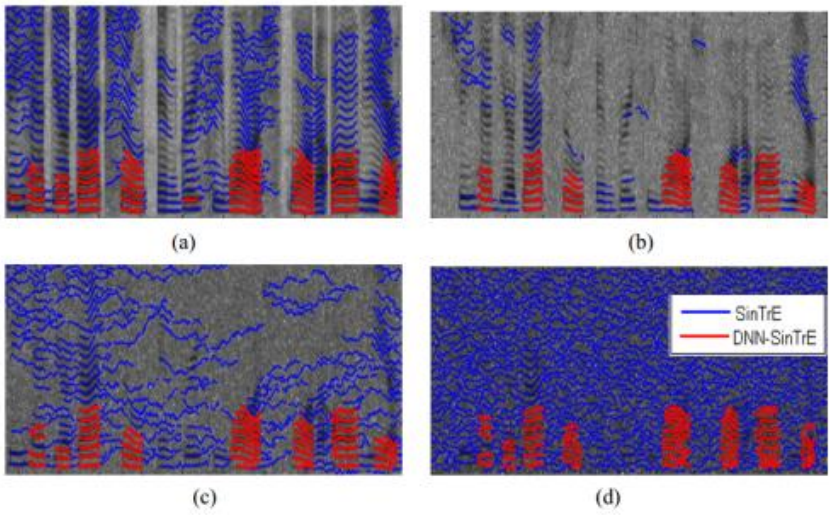


图 11 在不同信噪比条件下的基于 DNNs 的高局部信噪比时频点的提取效果，(a)无噪声环境，(b)SNR=20dB，(c)SNR=10dB，(d)SNR=0dB

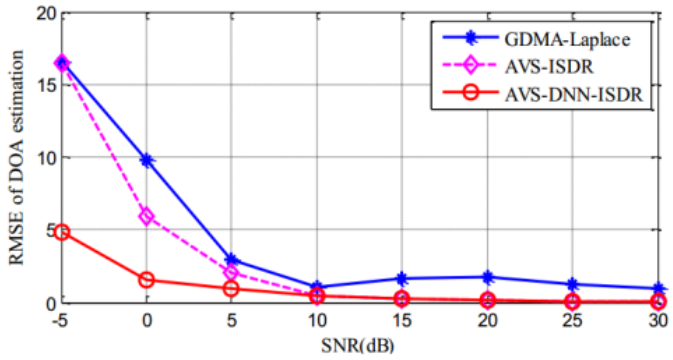


图 12 基于 DNNs 的高局部信噪比时频点提取对 DOA 估计抗噪性的改善（DOA=（60°，45°），信噪比 SNR 以 5dB 的间隔从-5dB 变化到 30dB）

表 10 实验室环境下 AVS-DNN-ISDR 的 DOA 估计结果

| 真实方位角$\phi_s$ (°) | 0 | 45 | 90 | 135 | 180 |
|---|---|---|---|---|---|
| AVS-DNN-ISDR | 0.78 | 45.18 | 90.35 | 137.10 | 180.26 |
| 绝对误差(°) | 0.78 | 0.18 | 0.35 | 2.10 | 0.26 |
| 运行时间(s) | 0.835 | 0.780 | 0.672 | 0.782 | 0.823 |

### （六）基于单颗 AVS 和双谱的抗干扰鲁棒语者声源 DOA 估计算法研究

除了针对加性噪声的鲁棒性 DOA 估计算法的研究，课题组还深入开展了针对方向性干扰对 DOA 估计的精度会产生负面影响的问题，即基于 AVS 和双谱的抗干扰鲁棒语者声源 DOA 估计算法的研究。其研究重点为：在同时具有方向性噪声、加性噪声和混响的场景中，通过分析 AVS 拾取的多通道语音信号的双频谱特性，利用双频谱域上高斯噪声被抑制等有利特性，推导出一种新的基于通道间的双频谱数据比的鲁棒 DOA 估计算法（**AVS-BISDR** 算法）。此外，为了进一步提高 AVS-BISDR 的鲁棒性，通过分析语音信号和非语音信号的 AVS 双谱特性，提出一种迭代双频谱域上的高语音信噪比时频点（掩模）估计算法，并在此基础上实现了一个鲁棒的 DOA 估计算法（**AVS-MBISDR** 算法），其算法框图如图 13 所示。

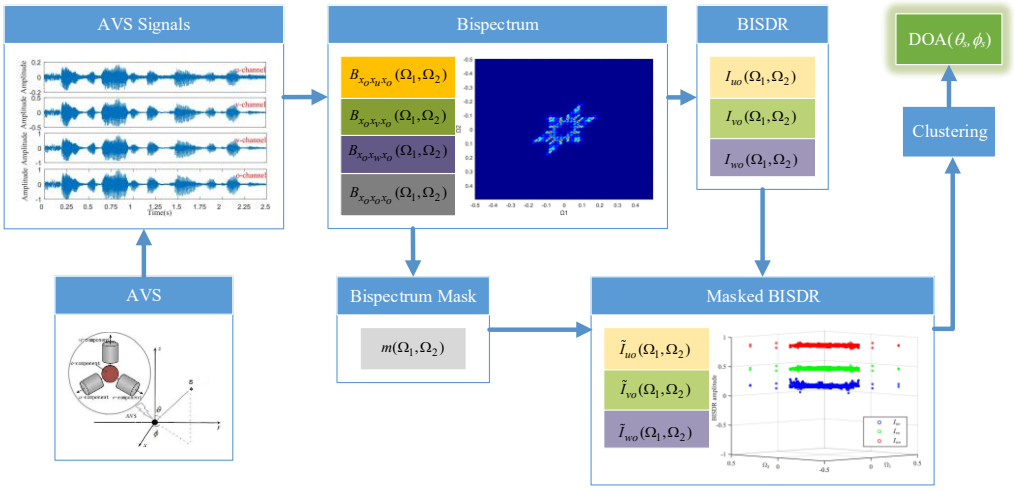


图 13 AVS-MBISDR 算法流程框图

课题组采用仿真实验和实测实验相结合的方式进行了 DOA 估计性能测试，其中，仿真实验的采样率 Fs=8kHz，语音长度 3s，帧长 30ms，20ms 重叠，加窗函数采用汉明窗，窗长为 30ms，干扰噪声的种类包含高斯白噪声，hfchannel 噪声，粉红噪声以及工厂噪声，实测实验为实验室自主搭建的系

统，实验室混响时间 $RT_{60}$=30ms，具有电脑和空调等产生的干扰噪声，具体实现细节见第(九)部分。图 14-16 显示了仿真实验的结果，表 11 显示了实测实验的结果。仿真实验和实测实验结果都表明，AVS-MBISDR 能够有效地抑制加性高斯白噪声以及方向性高斯噪声干扰的影响，获得优于 AVS-BISDR 算法的 DOA 估计性能。通过实验结果分析，我们得出如下**研究结论**：基于单 AVS 和双谱的 DOA 估计算法在信噪比为 5dB 时仍能对于各个方向的声源估计保持较高的精度，RMSE 保持在 5°以下；而且，在强干扰环境下，尤其在 SIR 低于 0dB 的时候，对于不同的干扰噪声，AVS-BISDR 和 AVS-MBISDR 均能保持较强的鲁棒性，得到较高精度的 DOA 估计。且这两个算法的 RMSE 曲线基本不随混响时间而变化，约为 1°。而实测试验也表明，AVS-MBISDR 的俯仰角和水平角的绝对值误差总和为 5°左右，在可接受的误差范围，因此，进一步验证了所研究算法的有效性。

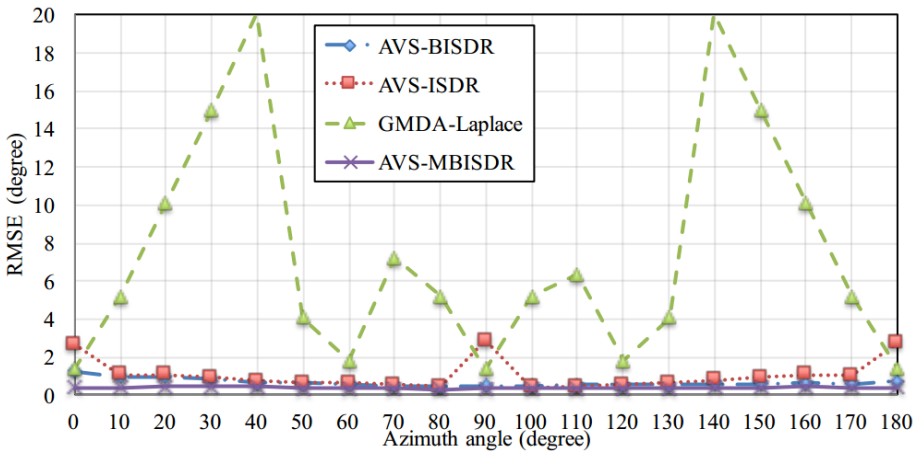


图 14 RMSE 随方位角的变化图（干扰信号：hfchannel 噪声，SIR=5dB，加性噪声，SNR=10dB）

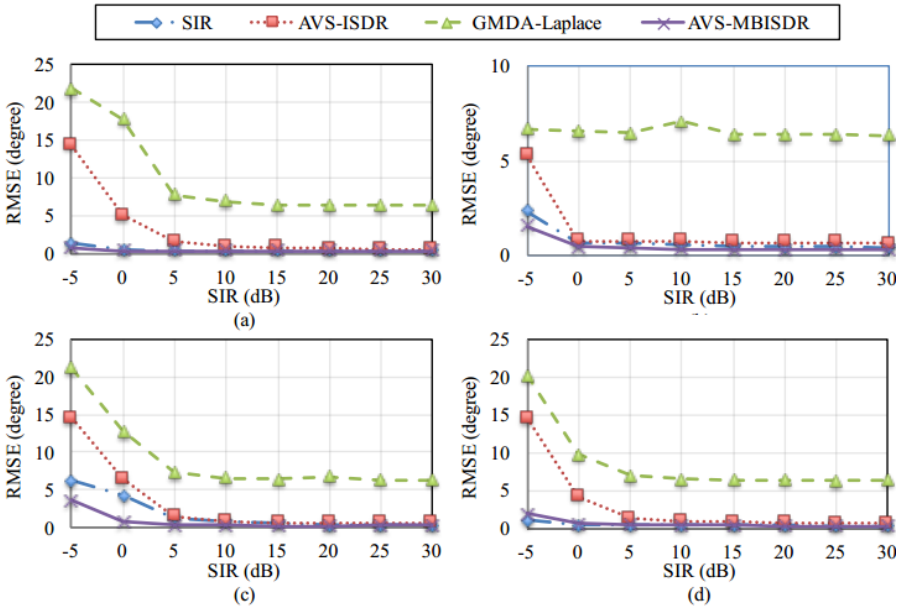


图 15 RMSE 随不同 SIR 和干扰信号的变化图:(a)高斯白噪声；(b)hfchannel 噪声；(c)pink 噪声；(d)工厂噪声。（加性噪声，SNR=10dB）

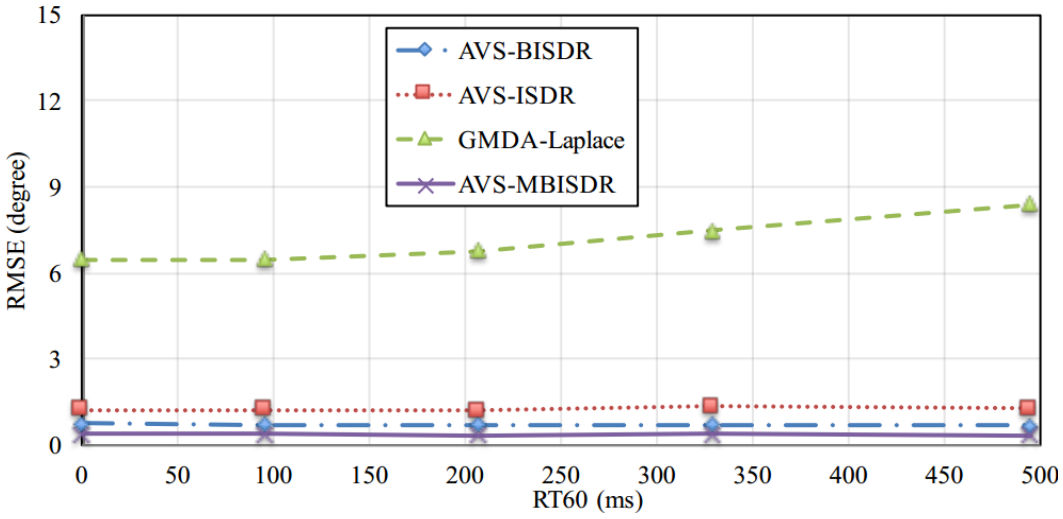


图 16 RMSE 随混响时间的变化图（干扰信号：hfchannel 噪声，SIR=5dB，加性噪声，SNR=10dB）

表 11 基于双谱的 DOA 估计应用于实验室环境的测试结果

| 真实 DOA(°) | (90,0) | (90,45) | (90,90) | (90,135) | (90,180) |
|---|---|---|---|---|---|
| 估计 DOA(°) | (94,1) | (94.44) | (90,90) | (92,133) | (89,180) |

**（七）基于级联语谱块和深度神经网络的鲁棒 DOA 估计算法研究**

同时考虑到低信噪比和强混响对 AVS-ISDR 性能的影响，课题组重点开展了基于深度神经网络的信号主导时频点(TD-TFPs)的提取研究，并利用提取出的 TD-TFPs 进行准确的 DOA 估计(AVS-WISDR-DNN)。其研究重点为：提出了一种新的级联语谱块(TLSB)特征用于有效区分 TD-TFPs 和 ID-TFPs，通过构造大量不同噪声和混响的 TLSB 数据集，利用深度神经网络(DNN)进行学习，准确预测时频点的软性掩膜并有效确定 TD-TFPs。同时软性掩膜可以用于 ISDR 簇的加权平均中心的计算，从而提高 DOA 估计精度。其算法基本流程如图 17 所示。

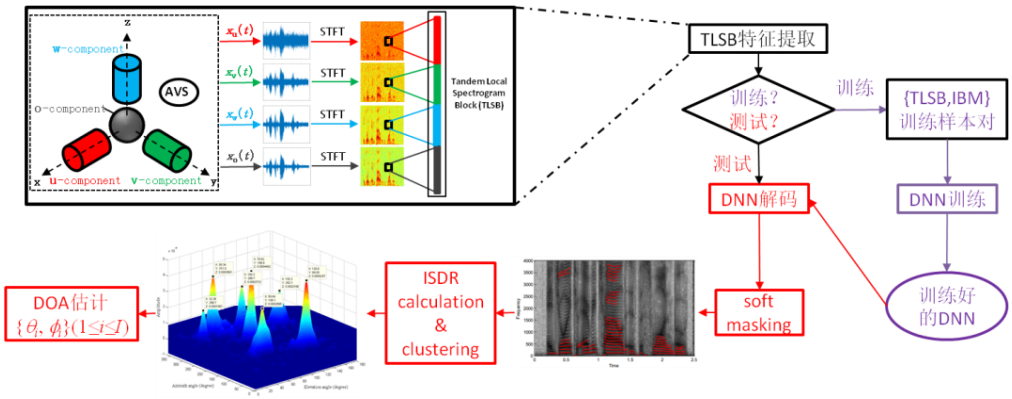


图 17 AVS-WISDR-DNN 算法流程图

课题组采用仿真实验和实测实验相结合的方式进行了 DOA 估计性能测试，训练数据构造来自于 TIMIT 数据库，每句话时长 3s，采用率为 8kHz，窗长为 256，DNN 的结构为输入层维度为 484，有三个隐含层，每层节点数为 512，输出维度为 2，采用 GTX1080 显卡进行训练，其中实测系统实现见第(九)部分。图 18、19 显示了仿真结果，表 12 显示了实测结果。通过实验结果分析，我们得出如下**研究结论**：（1）由图 18 可以看出，相比于 AVS-ISDR 和 AVS-LRSS 算法，AVS-WISDR-DNN 算法在所有的方位角下具有更高的

DOA 估计精度，且在 0° 处具有最低的估计误差；（2）由图 19 可以看出，随着噪声和混响强度的增大，所有算法的性能均在下降，但在所有的噪声和混响条件下，AVS-WISDR-DNN 方法具有更低的估计误差，显示出所提出的算法对噪声和混响的鲁棒性；（3）由表 12 可以看出，AVS-WISDR-DNN 算法在实测实验中的估计误差在 2° 以内，说明了该算法在实际应用中的有效性。

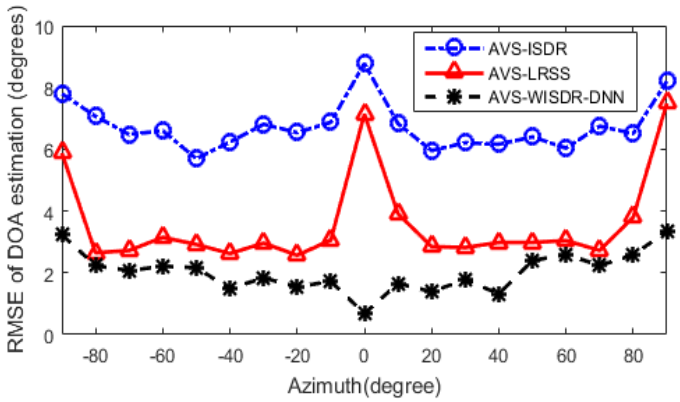


图 18 RMSE 随方位角的变化图（信噪比 SNR=5dB，混响时间为 350ms，俯仰角固定为 60°，方位角从-90° 到 90° 变化，间隔 10°）

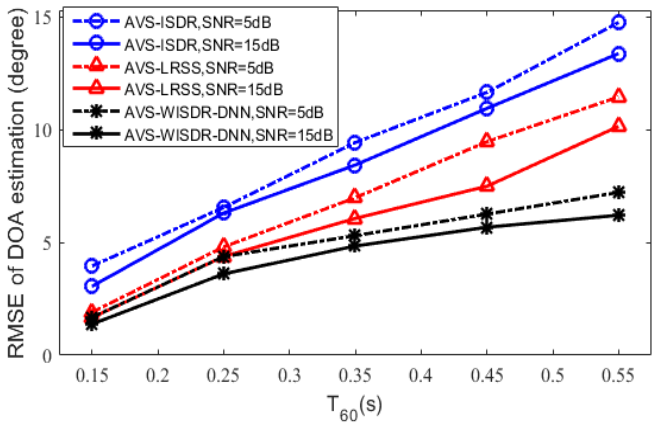


图 19 多源 DOA 估计 RMSE 随信噪比和混响时间的变化图（DOA1=（60°，-45°），DOA2=（80°，120°），信噪比 SNR 为 5dB 或 15dB，混响时间从 0.15s 到 0.55s 变化）

表 12 AVS-WISDR-DNN 算法在实测环境中的 DOA 估计结果

| True DOA (°) | 0 | 45 | 90 | 135 | 180 |
|---|---|---|---|---|---|
| Estimated (°) | -0.26 | 45.47 | 90.36 | 136.05 | 180.65 |

**(八) 语音声源 DOA 估计试验系统设计和实现研究**

DOA 估计试验系统主要需要完成 AVS 传感器的设计和实现、AVS 各通道麦克风的测量校准、AVS 阵列的阵型设计、语音预处理等等关键技术。我们开展了机器人语音声源 DOA 估计试验系统的设计和实现技术研究，具体研究内容可参考石伟毕业论文以及 2014 年度进展报告。现将机器人语音声源 DOA 估计试验系统的设计和实现技术的主要工作进展归纳如下。

1）AVS 传感器的设计和实现（原创性工作）

AVS 从提出至今已分别用于电磁学、水下声学、大气声学等领域。相较于之前在大气中可使用的 AVS 的昂贵价格，本课题组在研究和吸收前人的 AVS 设计和制作成果，设计了面向服务机器人听觉技术的能在大气中使用的价格低廉的声学矢量传感器，并且使其输出尽量接近 AVS 的理想数据模型。我们先后开发了三种 AVS 结构：AVS I，AVS II，AVS III。

AVS I

AVS I 的设计图片如图 20，AVS I 的支架尺寸如图 21(a)所示，其横截面为矩形，边长 3mm，仅比 NR-3158 和 EK-3132 麦克风的宽度（2.1mm）略宽，在保证能够人工正确安装的前提下尽量降低了支架侧面的宽度，以减少支架侧表面对声波的反射进而影响麦克风的输出响应。在 AVS I 中，三颗 NR-3158 压力梯度麦克风和一颗 EK-3132 全向麦克风分别记为 $u$、$v$、$w$ 和 $o$ 传感器，分别粘贴于金属支架的四个侧面。其中，$o$ 传感器位于一

面的顶端；*u* 传感器位于相邻面，比 *o* 传感器低 0.3mm；*v* 传感器位于 *u* 的相邻面，与 *u* 传感器的竖直距离是 0.5mm，约是 15kHz 声波波长的四分之一；*w* 传感器与 *o* 其竖直位置在 *u* 和 *v* 之间。AVS Ⅰ的这种结构设计，既考虑了四个麦克风组成的阵列整体的同位性，也考虑了麦克风之间互相遮挡可能带来的影响。实物所占空间尺寸约为 10mm×10mm×19mm，远远小于传统的包含四颗麦克风的几何阵列。

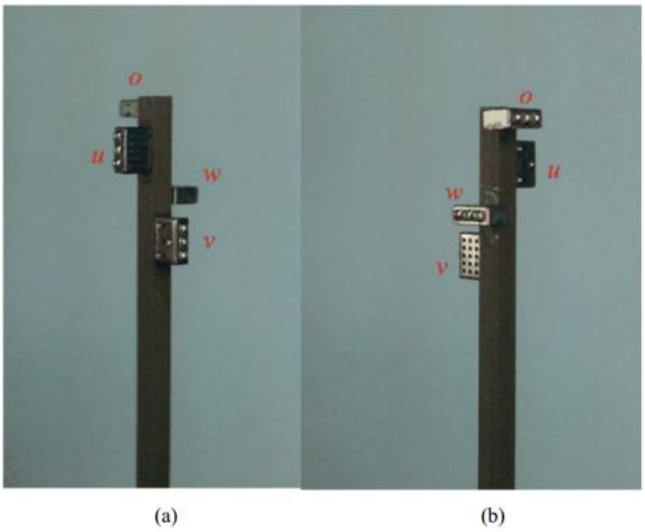


图 20 AVS Ⅰ，(a)正面显示 x 和 y 方向的传感器；(b)背面显示 z 方向的传感器和全向传感器

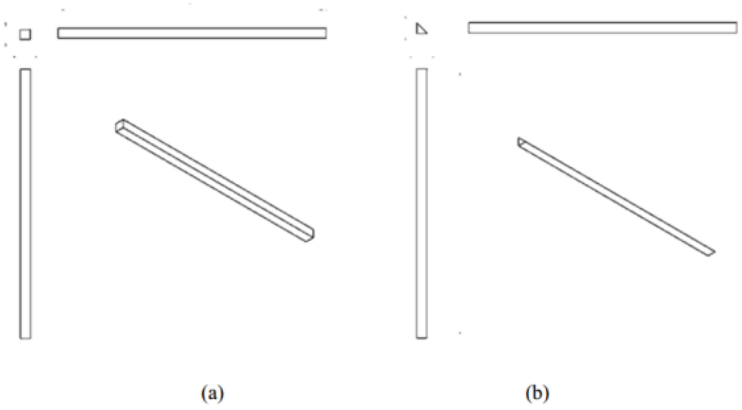


图 21 支架尺寸(单位：毫米)，(a)AVS Ⅰ支架尺寸；(b)AVS Ⅱ支架尺寸

AVS Ⅱ

在设计和制作 AVS Ⅰ的同时，我们注意到，EK-3132 和 NR-3158 测量的物理量不同，虽然二者都能有效地表示声音信号，但在面对相同响度和频率的声源时，其灵敏度差异较大。EK-3132 拥有相对更低的灵敏度，能够更有效地降低环境中响度较低的噪声，适用于声源与麦克风距离较远的场景；而 NR-3158 则适用于声源与麦克风较近的场景，如作为头戴式麦克风、手机麦克风等。可见，两种麦克风的最佳适用场景并不一致，但我们利用这两种麦克风构建 AVS 时，声源距离二者的距离是一样的。另一方面，在 AVS 的理想数据模型中，$u$、$v$、$w$通道的数据与 $o$ 通道的数据呈确定比例关系，即：

$$u^2+v^2+w^2=o^2 \qquad (19)$$

因此，如果令 AVS 仅包含同种类的三颗方向性传感器，可避免由于使用不同种类麦克风带来的增益和频率响应的校正工作，同时也不至于丢失方向性信息。只使用三颗麦克风，对硬件系统的要求更低，AVS 的整体尺寸也必然会更小，记为 AVS Ⅱ。AVS Ⅱ的照片见图 22。AVS Ⅱ仍使用长条形支架，但为使麦克风间距更小、整体结构更紧凑，支架横切面采用等边三角形的形式，如图 21(b)。在 AVS Ⅱ中，由于没有全向麦克风，所以 $u$ 传感器被粘贴在支架侧面的顶端；$v$ 传感器粘贴在于 $u$ 垂直的另一个侧面，竖直距离同样为 0.5mm；$w$ 传感器则在第三个侧表面，水平粘贴，其竖直位置在 $u$ 和 $v$ 之间。需要注意的是，由于 $w$ 传感器所在侧面较宽，约为 4.25mm，为减少支架对 $w$ 传感器的遮挡，可将 $w$ 传感器部分粘贴到支撑面上，如图 22 中标示的 $w$ 传感器的位置。

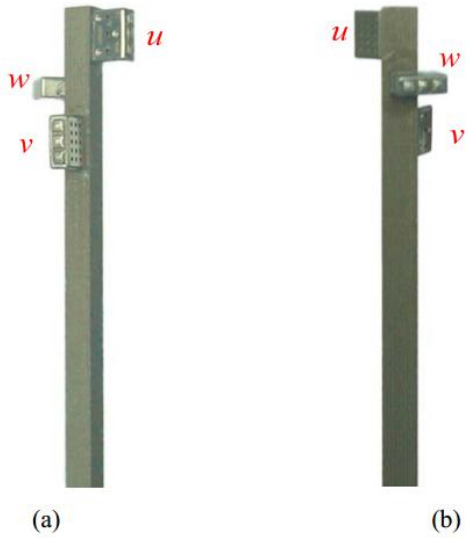


图 22 AVS II，(a)正面显示 x 和 y 方向的传感器；(b)背面显示 z 方向的传感器

## AVS Ⅲ

课题组设计的声学矢量传感器使用的是体积较小的 MEMS 麦克风，两种麦克风都具有相同的贴片封装的焊脚，可贴于 PCB 的表面。但是由于 AVS 的各个分传感器并不在同一个表面上，故之前研究人员都是将麦克风粘贴到设计好的支架上，然后通过导线连接麦克风的各个引脚并将之连入电路。上文提到的 AVS Ⅰ、AVS Ⅱ都是采用这种设置。由于每颗麦克风都有三个焊脚，整个 AVS 需要 10 条左右的导线才能将各个分传感器正确地互联，导致 AVS 整体结构凌乱且不稳固。由于麦克风体积较小、焊点密集，外部连接的导线交杂在一起，势必会对麦克风的输出造成干扰。焊接导线的方式也并不稳固，由于长期导线本身的弹性压迫，外露的金属丝部分容易折断导致断路或接触不良。而且导线质量较差、焊点解除不良等 隐私可能会引入噪声，AVS 的体积也被变相增大。鉴于以上因素，本文考虑使用 PCB

来固定和支撑麦克风。仿造 AVS Ⅰ的结构，设计并制作了四块 PCB，分别对应 AVS Ⅰ的四个支撑面。最后制成的 AVS 如所示，被命名为 AVS Ⅲ。与 AVS Ⅰ和Ⅱ相比，AVS Ⅲ可表面贴焊到 PCB 上，而使用导线连接。分立的四块 PCB 被粘接在一起，组成需要的长方体形支架。由此构成的 AVS 结构紧凑，麦克风由焊点固定，电气特性牢固。装上麦克风专用的海绵套后， AVS Ⅲ成为一个整体，可看做一个单一的传感器，如图 23(c)所示。AVS Ⅲ的结构设计的缺点在于其支撑面的宽度更宽，约为 6mm，支架对麦克风的反射和遮挡作用更大。

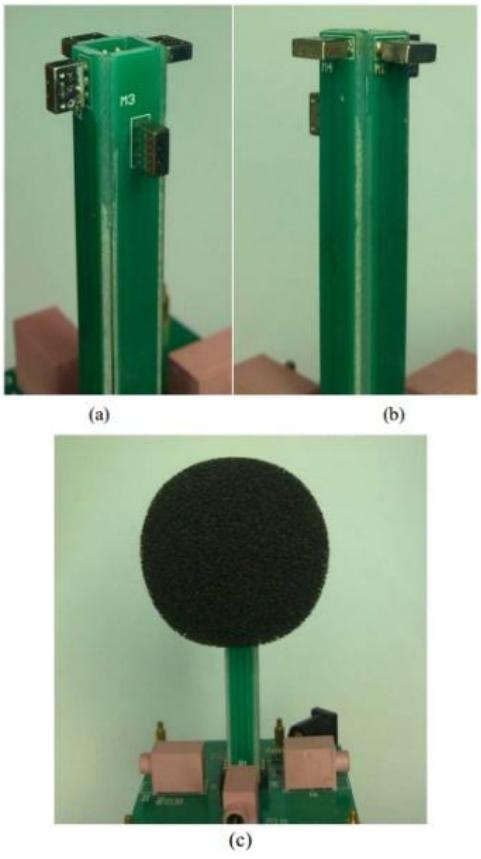


图 23 AVS Ⅲ，(a)正面显示 x 和 y 方向的传感器；(b)背面显示 z 方向的传感器和全向传感器；(c)带海绵套的 AVS Ⅲ

2）声学矢量传感器的校准

通过一系列实验测试了构建 AVS 使用的麦克风的响应特性，重点测试和比较了全向麦克风 EK-3132 和双向性麦克风 NR-3158 的增益、频率响应和方向性响应。根据实验结果，总结利用分立麦克风构建 AVS 的软件校准方法，主要包括两个方面：

频域校准：对 NR-3158 的输出信号进行去加重，利用下式的滤波器对信号进行时域滤波，使得 NR-3158 的频率响应更为平坦，与 EK-3132 接近。

$$y(t) = y(t) + 0.96 \times y(t-1) \qquad (20)$$

增益校准：调节 EK-3132 和 NR-3158 的增益，使得输出尽量满足理想 AVS 数据模型；典型的如信源处于 90° 时，EK-3132 和 NR-3158 的输出应相等；使用 ESS 信号作为信源，测量 EK-3132 和 NR3158 输出信号的能量之比 α ；以α作为 NR-3158 输出信号的软件增益。特别地，由于不同器件之间存在差异，故对每颗构建的 AVS 均需独立进行增益校准。

3）实时语音端点检测算法

考虑实际场景的应用需求，人们对于其他语者说话内容的感知是需要时间的。很多时候，当一句话说完时，其含义才能够被完全确定。因此，在实时 DOA 估计系统中，对连续语音流进行分段、提取其中包含语音的部分是首要工作，故语音端点检测（Voice Activity Detection, VAD）是本系统的另一个重点。其基本思想是借鉴基于长时频谱散度（Long-term Spectral Divergence, LTSD）VAD 中的“长时”思想，根据语音浊音的谐波特性，提出基于长时自相关系数统计量（Long Term Auto-correlation Statistics, LTACS）

VAD。课题组采用仿真实验的方式进行了 VAD 性能测试，说话人为男性，使用中文；录音包含多段短语，伴有撞击、磕绊等突发性噪声；采样率为 8kHz。最终的 ROC (receiver operating characteristic)曲线如图 24 所示。根据实验测试结果分析，我们得出如下**研究结论**：实验证明课题组提出的 VAD 算法能够有效地区分语音和突发地、不规则噪声，且计算简单，容易实时化。

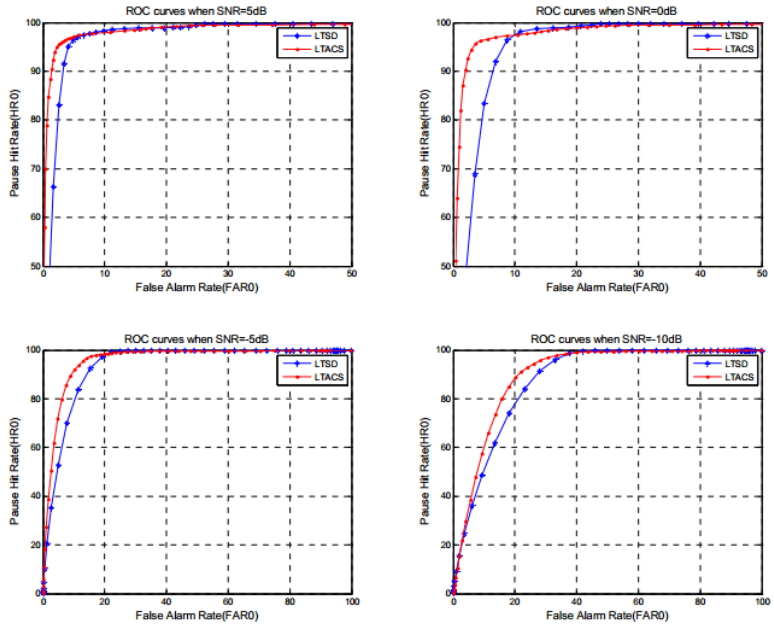


图 24 LTSD-VAD 和 LTACS-VAD 在不同信噪比下的 ROC 曲线

**（九）服务机器人听觉系统关键技术研究**

课题组针对智能家居中的服务机器人听觉系统的关键技术进行了多项研究，具体研究内容可参考 2014 年度、2015 年度以及 2016 年度进展报告。现将服务机器人听觉系统的多项关键技术的主要工作进展归纳如下。

1）语音增强算法研究

- 在基于 AVS-DOA 估计技术的基础上，提出了一个有效的利用 AVS 多通道信息的频域语音增强算法，结合最小方差无失真响应空间滤波

和后置维纳滤波技术，实现对空间干扰声源的抑制和对背景噪声的消除；

- 开展了基于语音时频稀疏性的空间目标语音增强算法研究，提出了基于单个 AVS 的时频掩膜语音增强方法，实验结果表明可以有效的抑制空间干扰噪声和背景噪声，同时很好的保持了目标语音的可懂度和清晰度；
- 在实际应用环境中，由于非稳态噪声的存在使得说话人识别的准确率有所下降，因此课题组相应开展了基于非负矩阵分解的语音增强算法研究，非负矩阵分解算法对于非稳态噪声和低信噪比环境下的信号有很好的增强效果；
- 基于非负矩阵分解的单通道语音增强技术在低信噪比条件下利用现有模型不能对语音和噪声进行很好的表达的背景下，根据语音信号的稀疏性和突发噪声信号的低秩性，提出新的基于多重限定的非负矩阵分解的单通道语音增强模型，提高了语音增强的性能，通过实验分析了本算法在低信噪比非稳态噪声环境下的性能，验证了所提算法的有效性和鲁棒性；
- 基于非负矩阵分解的语音增强技术，在非稳态噪声环境及低信噪比条件下，现有模型不能对语音和噪声进行很好的表达，于是利用语音信号的稀疏性和突发噪声信号的低秩性，提出新的基于多重约束的非负矩阵分解的单通道语音增强模型，提高了语音增强的性能，通过实验分析了本算法在低信噪比非稳态噪声环境下的性能，验证了所提算法的有效性和鲁棒性。

2）VAD 算法研究

- 噪声环境中的有效语音活动检查是 DOA 估计和语音增强与识别系统的关键技术之一，课题组开展了基于希尔伯特变换的 VAD 估计算法研究，提出了基于瞬时频率的 VAD 估计算法，通过实验验证了算法在噪声环境中的有效性；
- 为了进一步提高噪声环境中 VAD 算法的性能，课题组还开展了基于长时自相关统计的带噪语音的语音端点检测 VAD 算法研究，并通过实验验证了所研究算法在低信噪比环境下具有很好的鲁棒性；
- 语音 VAD 可以看做是语音与非语音的二分类问题，课题组基于模式识别和机器学习的方法，开展了基于支持向量机和高度可分的语音特征的 VAD 改进算法研究，通过实验分析了多种语音特征对于语音与非语音的可分性，也验证了所研究算法的有效性和鲁棒性。

3）说话人识别算法研究

- 说话人识别也是机器人语音交互的关键技术之一，说话人确认技术已经有较长的研究历史，在纯净语音条件下，主流说话人确认技术已经可以取得较高的识别率。实验测试显示，工作在实际环境中的服务机器人，需要面对机器人自身噪声、环境噪声、混响等不可避免的干扰；基于传统的人工特征的说话人识别方法在有标签训练数据较少、有噪声等情况下识别效果很差，课题组提出了一种新的基于 stacked denoising Autoencoders(SDAE)的声学特征学习方法，SDAE 结合了无监督学习和有监督学习两种训练方式，以大量的带噪语音和干净语音作为输入，通过监督学习和非监督学习结合的方

式训练 SDAE 网络，获取表征说话人个性信息的区分性特征，公开数据集和自建数据集测试结果验证了所提算法对于未知噪声有较好的泛化性能；

- 基于机器学习的语者确认技术的训练和实际使用是分开的，因此带来训练语音数据来自 A 设备，而确认语音数据来自 B 设备的信道失配问题，对说话人确认算法的性能产生明显的负面影响。针对信道失配问题，课题组以 ivector-PLDA 框架作为基线系统，提出了模型域基于类内协方差规整（WCNN）和扰动属性投影（NAP）的信道补偿算法；借鉴深度学习的成果和思想，提出了基于受限玻尔兹曼机（Restrict Boltzmann Machine，RBM）和深度置信网络（Deep Belief Network，DBN）的说话人身份确认算法，以不同信道下的语音作为输入，说话人的 ID 作为标签，采用监督学习与非监督学习结合的方式训练一个 DBN，获得输入语音的分层特征表示，利用训练所得的特征，开展特征选择研究，通过定义区分度比值来选择可区分性最好的特征作为学习到的声纹特征；在数据量充足的条件下，课题组提出的说话人确认方法可以较好地学到不同信道的区别，提取反应说话人个性信息的区分性特征，提高说话人系统的性能。

4）中英文混合语音识别算法研究

- 为了实现双语语音识别，课题组开展了融合基于混淆矩阵的两遍搜索算法和基于专家知识的混淆矩阵预定义算法来获取中文声韵母和英文音素的映射集，并基于此映射集产生多发音词典，从而实现鲁棒的中英文（双语）语音识别系统，通过实验比较了中文、英文

语音识别系统的性能，证明了所研究的双语语音识别系统具有一定的优越性。

5）抗录音回放攻击算法研究

- 抗录音回放攻击可以看做是原语音与回放语音的二分类检测问题，课题组采用基于模式识别和机器学习的方法，开展了基于高斯超矢量特征与支持向量机的录音回放攻击检测算法算法研究，实验分析了多种特征对于原语音与录音回放语音的可分性，验证了所提出的算法能有效区分原语音与回放录音，从而实现抗录音回放攻击；
- 为了进一步提升性能，课题组提出了基于有监督学习的抗录音回放攻击算法研究，利用支持向量机作为分类器以获得更高的分类准确率，同时提取 Mel-Frequency Cepstral Coefficients(MFCC)的统计特征来表征语音的特性，借此区分原语音与录音回放语音，根据录音回访的机理，提取了信道模式噪声的统计特征来表征录音信道噪声；课题组为此项研究建立了一个抗录音回放攻击的数据库，通过大量实验，验证了所提出的方法的抗录音回放攻击的有效性；

6）语音情感识别算法研究

- 语音情感识别目前依然是一个没有获得较好解决的问题，一方面是语音情感数据库相对较少和规模小，另外一方面，采用传统的统计和信息处理方法获得的语音特征并不能很好地表达与反映语音情感，采用深度学习的方法是一个新的研究领域。本课题组开展了基于深度卷积神经网络的语音情感识别研究，采用对数语音频谱为训

练输入，情感标签为输出，采用标记数据进行训练，我们的实验结果表明基于 CNN 的语音情感识别性能好于传统方法。

7）基于单 AVS 的实时 DOA 估计系统实现

- 为了验证课题组提出的多个 DOA 估计算法的性能，我们设计了基于单 AVS 的实时 DOA 估计系统，其系统框架如图 25 所示，采用 AVS III 结构，通过外接多通道同步前端放大器和多通道声卡，将 AVS 的输出实时地采集到 PC 中，这里采用百灵达 ADA8000 作为音频采集器(如图 26 所示)。利用接收到的信号，即可用实时 DOA 估计算法对角度进行估计和显示。这里采用的实时 DOA 估计系统软件界面如图 27 所示，软件界面的上半部分为角度指示盘，仅指示方位角的估计结果。俯仰角的估计结果会同步打印到 MATLAB 命令窗口中。下半部分为设置和波形显示区域，可实时显示声卡采集到的时域信号。实时 DOA 估计场景如图 28 所示。

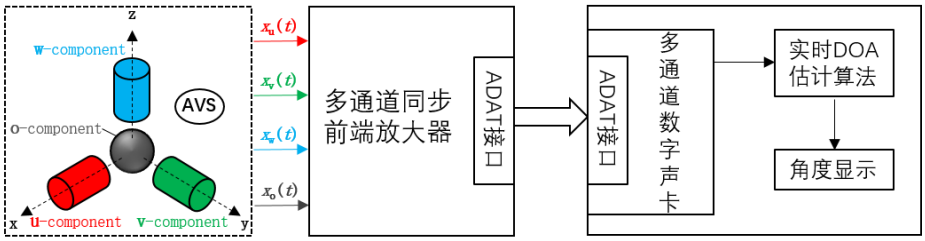


图 25 基于单 AVS 的实时 DOA 估计系统框架


图 26 百灵达 ADA8000

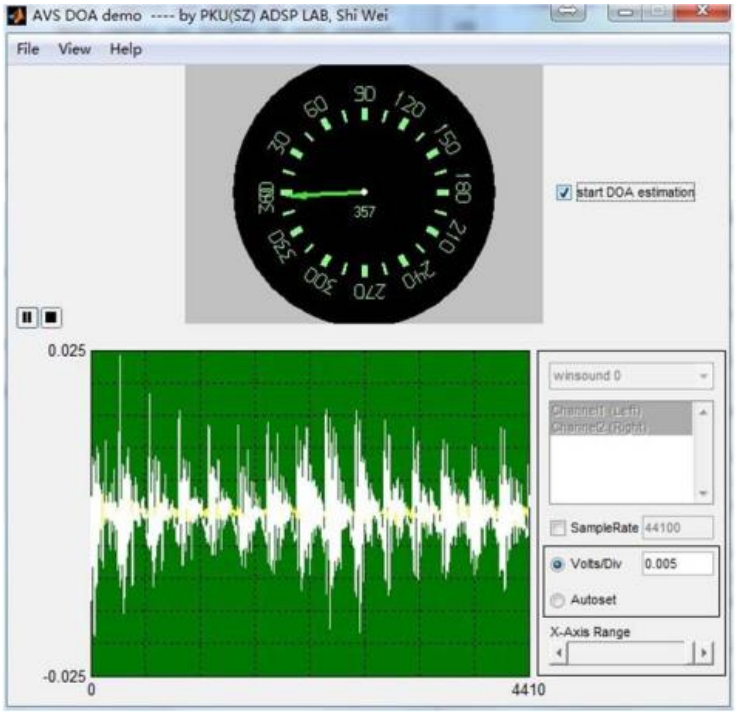


图 27 实时 DOA 估计系统软件界面

图 28 实时 DOA 估计系统实测场景

8）嵌入式移动机器人实验平台开发

- 为进一步测试课题组所提算法在真实环境下的性能，项目组研发了基于移动机器人平台的机器人语音声源 DOA 估计实验系统。该系统具备自主移动能力，可接收 DOA 估计模块给出的方向信息，转向目标说话人所处的特定反向。该系统安装了 Linux 操作系统，提供

了方便的录音 API 给上层应用，方便集成后端的系统，如语音识别、说话人识别、情感识别等，可作为机器人智能听觉系统关键技术的综合实验平台。该平台实物如图 29 所示，其主要特性包括：

- 易于使用 — 由整机及其附件组成，到货后无需再做繁杂的装配；
- 可靠耐用 — 坚固耐用；
- 精度高 — 工业级编码器，运动定位精准；
- 软件开发工具包 — 提供配套的 ROS 开发包，帮助客户加快机器人项目的开发；
- 可定制化 — 轻松地从数十种支持和测试的配件中选择并使用合适的配件；
- 多媒体接口丰富 — 提供简单方便的音频和视频接口，有完善的麦克风阵列信号获取 API。

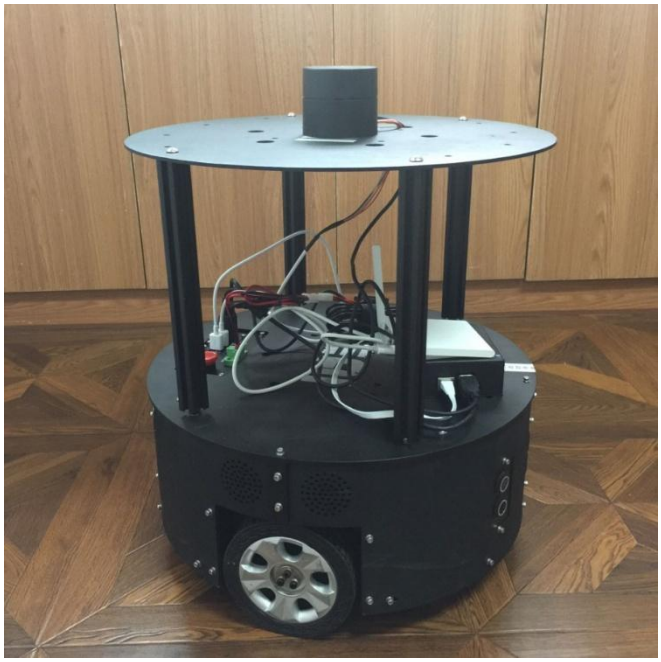

图 29 嵌入式移动机器人 DOA 估计实验平台

## 三、其他相关信息

### 1）国内外学术合作交流等情况

### 2）项目成果转化及应用情况

本课题的研究成果已申请了 5 项中国发明专利，其中 2 项授权，研究成果获得华为、海尔、广州视源股份有限公司、优必选、深圳市海岸技术有限公司、深圳市尔木科技有限公司等高科技公司的关注，已有企业与本课题组进行洽谈购买专利事宜。

本课题培养的北京大学优秀硕士毕业生石伟已经在 2016 年 3 月创建海岸语音技术有限公司，重点借助本课题的知识产权，开展成果转化和产品化工作，期望能够为智能家庭服务机器人提供声源 DOA 估计关键技术。

此外，本课题培养的北京大学硕士研究生王永庆加入了乐视语音小组、郭轶凡加入了腾讯、郑炜乔加入了中国著名语音专业公司思必驰、胡旭琰加入了网易语音小组、任梦琪和李波出国攻读博士等，可以认为 NSFC 项目的支持不仅仅局限于技术，对人才的培养也起到了重要的作用。

### 3）人才培养情况

在自然科学基金项目的支持下，已经培养了 12 名硕士研究生，其中 11 名已经获得北京大学理学硕士学位。

| 研究生姓名 | 专业/研究方向 | 硕士论文题目 | 导师姓名 | 答辩时间 |
|---|---|---|---|---|
| 任梦琪 | 集成电路与系统/嵌入式系统与 DSP 技术 | 小孔径麦克风阵列语者定位技术研究与实现 | 邹月娴 | 2012 年 7 月 |
| 李波 | 集成电路与系统/嵌入式系统与 DSP 技术 | 基于信号稀疏性的声学矢量传感器 DOA 估计方法研究 | 邹月娴 | 2012 年 7 月 |
| 石伟 | 集成电路与系统/多媒体技术 | 基于声学矢量传感器的鲁棒 DOA 估计方法研究与实现 | 邹月娴 | 2013 年 7 月 |
| 郭轶凡 | 计算机应用技术/多媒体信息处理技术 | 基于 AVS 和稀疏表示的鲁棒语者声源 DOA 估计算法研究 | 邹月娴 | 2015 年 7 月 |

| | | | | |
|---|---|---|---|---|
| 郑炜乔 | 计算机应用技术/多媒体信息处理技术 | 面向智能服务机器人的语音声源 DOA 估计技术研究 | 邹月娴 | 2016 年 7 月 |
| 金彦含 | 计算机应用技术/多媒体信息处理技术 | 基于 AVS 和双谱的鲁棒语者声源 DOA 估计算法研究 | 邹月娴 | 2017 年 7 月 |
| 胡旭琰 | 计算机应用技术/多媒体信息处理技术 | 基于带噪语谱补偿的鲁棒语音识别算法研究 | 邹月娴 | 2014 年 7 月 |
| 王鹏 | 计算机应用技术/多媒体技术 | 基于声学矢量传感器的语音增强算法研究 | 邹月娴 | 2013 年 7 月 |
| 王永庆 | 计算机应用技术/多媒体信息处理技术 | 基于时频掩膜的空间目标语音增强算法研究 | 邹月娴 | 2015 年 7 月 |
| 宁洪珂 | 计算机应用技术/多媒体信息处理技术 | 信道鲁棒的说话人确认算法研究 | 邹月娴 | 2015 年 7 月 |
| 刘诗涵 | 计算机应用技术/多媒体信息处理技术 | 基于非负矩阵分解的单通道语音增强算法研究 | 邹月娴 | 2016 年 7 月 |
| 王春 | 计算机应用技术/多媒体信息处理技术 | 基于监督学习的录音回放攻击检测方法及应用 | 邹月娴 | 2016 年 7 月 |

**4) 其他需要说明的成果**

a) 2013.12.30 北京大学深圳研究生院院长基金创意创新竞赛一等奖（项目负责人为郑炜乔），《懒人垃圾桶—基于 AVS 语音定位和识别的智能垃圾桶设计与实现》，该项目采用了课题组开发的二项语音关键技术：DOA 估计-确定语音声源方向、关键词识别-识别命令词进行控制动作；

b) 北京大学信息工程学院 2013 级学生团队（项目负责人为郑炜乔）的参赛作品“Vcamera—语音相机 app”经过决赛答辩，荣获本届“挑战杯”五四青年科学奖竞赛一等奖，该项目采用了课题组开发的一项语音关键技术：关键词识别-识别命令词进行拍照；

c) 北京大学信息工程学院 2013 级学生团队（项目负责人为王春）的参赛作品“基于智能机器人的非特定人中英文混合命令短语识别系统”，荣获本届“挑战杯”五四青年科学奖竞赛二等奖，该项目采用了课题组开发的一项语音关键技术：混合语音识别技术，与 SIRI 对比，取得了在固定命令词集中的更好的中英文识别短语识别结果；

d) 北京大学信息工程学院 2013 级学生团队（项目负责人为郑炜乔）的参赛作品“基于声纹识别的考勤管理平台”荣获中国联通深圳分公司奖学金创新奖竞赛三等奖，该项目采用了课题组开发的一项语音关键技术：说话人确认技术。