

# 一种鲁棒的双耳声源方位角定位方法

陈国良, 赵祥瑞

(武汉理工大学 机电工程学院, 武汉 430070)

**摘要:** 在噪声和混响的声学环境中, 基于双耳时间差的声源方位角定位性能会严重降低; 针对这个问题, 提出了一种基于子带选择和 DBSCAN 的双耳声源定位算法, 首先, 采用 Gammatone 滤波器将双耳声源信号分解为若干个子带信号; 其次, 根据子带能量大小进行子带通道数压缩; 然后, 根据子带信噪比大小获取最优子带, 降低无关子带干扰; 接着将子带信号进行分帧, 根据互相关算法获取峰值处的数据点; 最后, 引入 DBSCAN 算法消除噪声点的影响, 获取最优数据点, 从而根据 ITD 定位模型判断目标声源方位角, 实验结果表明, 该算法在复杂的声学环境中, 相较于传统的互相关算法, 可显著提高双耳声源方位角定位性能。

**关键词:** 双耳声源定位; 数据压缩; 子带选择; 互相关算法; DBSCAN

## Method of A Robust Binaural Sound Source Azimuth Angle Location

CHEN Guoliang, ZHAO Xiangrui

(School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan 430070, China)

**Abstract:** In noisy and reverberant acoustic environments, the performance of sound source azimuth angle localization based on binaural time difference is severely degraded. To solve this problem, a binaural sound source angle localization algorithm based on the subband selection and density-based spatial clustering of applications with noise (DBSCAN) is proposed. Firstly, the binaural sound source signal is decomposed into several subband signals by using the Gammatone filter; Secondly, the number of the subband channels is compressed according to the subband energy; Then, the optimal subband is obtained according to signal to noise ratio of the subband, the interference of irrelevant subbands is reduced, the subband signal is divided into frames, and the data points at the peak are obtained according to the cross-correlation algorithm; Finally, the DBSCAN algorithm is introduced to eliminate the influence of noise points and obtain the optimal data points, so the target sound source azimuth angle localization is determined according to the interaural time difference (ITD) positioning model. The experimental results show that, compared with the traditional cross-correlation algorithm in complex acoustic environments, the algorithm can significantly improve the azimuth angle localization performance of the binaural sound sources.

**Keywords:** binaural sound source localization; data compression; subband selection; cross-correlation algorithm; DBSCAN

## 0 引言

机器人的听觉系统是机器人与外部环境互动的一种重要方式。与视觉相比, 机器人听觉系统受障碍物影响较小, 且具有 360 度的识别范围。机器人听觉系统包括许多方面, 例如声源定位、语音识别、讲话者识别、情感识别和语音降噪等<sup>[1]</sup>。其中, 机器人声源定位 (SSL, sound source localization) 作为听觉系统的前端处理模块之一, 在机器人导航、人机交互、视频会议等领域发挥着重要作用<sup>[2-5]</sup>。

目前的声源定位研究主要分为两类: 基于双耳 SSL 算法研究和基于传声器阵列的 SSL 算法研究。与基于传声器阵列的 SSL 相比, 双耳 SSL 具有阵列结构简单, 计算复杂度小, 定位线索较少的特点。双耳 SSL 的线索主要有: 双耳时间差 (ITD, interaural time difference)、双耳强度差 (IID, interaural intensity difference) 和双耳相位差 (IPD, binaural phase difference)<sup>[6-7]</sup>。ITD 是指声波从声源到左右

耳之间的时间间隔, IID 是指两耳接收到的声波之间的强度差, IPD 是 ITD 在频域内的表现。在中低频 (小于 1.5 kHz) 的情况下, ITD 其主要作用, 利用该时延差可以很好地地进行方位的评估, 但在噪声和混响的干扰下, ITD 的性能会严重下降。因此本文主要讨论在噪声和混响环境下基于 ITD 的双耳声源定位。

在正常情况下, 假设声音从声源到麦克风是一个简单的直线传输。然而, 在混响的室内环境中, 麦克风接收到的信号是声源信号与墙壁、地面、家具等物体反射产生的反射信号的叠加。这使得每个麦克风之间的信号不服从理想的时间延迟关系, 最终使得麦克风之间的 ITD 难以判断。此外, 环境噪声是 SSL 无法避免的一个重要问题。当环境噪声较强时, 目标信号与双耳麦克风之间的时间差常常被抑制<sup>[8]</sup>。

针对混响和噪声的定位环境, 基于 ITD 的双耳声源定位性能会严重下降问题, 文献 [9] 对基于互相关函数的

收稿日期: 2022-04-03; 修回日期: 2022-05-07。

作者简介: 陈国良 (1972-), 男, 湖南湘潭人, 博士, 教授, 主要从事智能控制与机器人技术方向的研究。

引用格式: 陈国良, 赵祥瑞. 一种鲁棒的双耳声源方位角定位方法[J]. 计算机测量与控制, 2022, 30(11): 204-212.

ITD 算法进行改进, 提出一种基于 PHAH 加权的广义互相关算法和可控波束算法融合的声源定位算法, 实验证明该方法可提高在噪声环境中的定位准确性。文献 [10] 提出一种基于过零率 (ZC, zero crossing) 的 ITD 估计方法, 该算法首先根据过零点获取各帧的 ITD, 然后通过信噪比估计进行各帧 ITD 的加权, 从而获取最后的 ITD。文献 [11] 提出一种基于混响加权的声源定位算法, 该算法通过信道的混响权重分别降低早期混响和晚期混响的影响, 从而获取更加准确的双耳定位线索, 提高了在混响环境中的声源定位性能。因此, 降低噪声和混响的干扰是双耳声源方位角定位技术的关键。

近年来, 模式识别和机器学习技术在双耳声源定位中运用, 文献 [12] 提出一种模板匹配的方法, 对环境中的不同方位角建立不同的模板, 基于特征和频率加权进行模板匹配。针对低信噪比的双耳 SSL 问题, 文献 [13] 基于深度学习建立双耳定位线索和方位角的映射关系, 进行定位。文献 [14] 将基于模型的声源频谱特征信息与深度神经网络相结合, 解决各种噪声和混响条件下双耳 SSL 问题。这些方法在噪声和混响的环境中可取得较好的定位效果, 但需要训练不同声学环境下的 ITD 和 IID 模型, 通用性差, 且计算量较大。

基于以上算法的不足之处, 本文提出一种信号子频带选择结合具有噪声的基于密度的聚类方法 (DBSCAN, density-based spatial clustering of applications with noise) 的声源定位方法。首先对采集到的双耳语音信号进行 Gammatone 滤波, 分频为不同的子频带; 其次进行数据压缩, 减少无关子带数量, 降低计算复杂度; 然后基于谱减算法进行各个子频带信噪比 (SNR, signal-to-noise ratio) 估计, 选出最优的子频带; 最后对子带信号进行分帧, 基于互相关 (CC, cross correlation) 时延估计算法, 获取每一帧互相关函数峰值对应的数据点, 再根据 DBSCAN 聚类算法消除异常帧的影响, 获取最优点, 从而根据时延和 ITD 定位模型得到方位角。本文提出的算法不仅适用不同的 SSL 环境, 而且提高 SSL 的鲁棒性。所提算法的框图如图 1 所示。

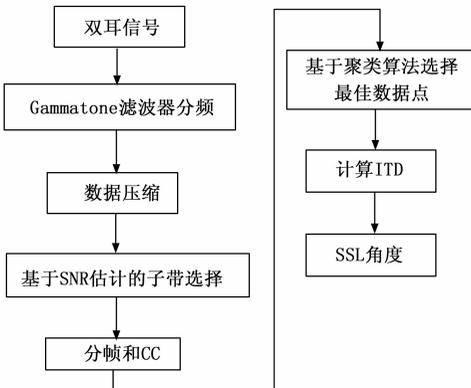


图 1 双耳声源定位算法框图

## 1 基于 Gammatone 滤波器的数据压缩和子带选择

### 1.1 基于 Gammatone 滤波器组的分频

Gammatone 滤波器组可以模拟耳蜗基底膜对于语音信号的处理机制, 耳蜗基底膜可以将信号分解为不同频率的频带信号进行处理, Gammatone 滤波器组将其组中的各个滤波器的带宽参数从低频到高频由窄到宽设置, 可以将含噪语音按频率高低范围进行分频处理。

假设  $s(n)$  为声源语音信号, 左右耳接收到的语音信号为  $x_L(n)$  和  $x_R(n)$ , 则信号模型为:

$$\left. \begin{aligned} x_L(n) &= h_L(n) * s(n) + n_L(n) \\ x_R(n) &= h_R(n) * s(n) + n_R(n) \end{aligned} \right\} \quad (1)$$

式中,  $h_L(n)$  和  $h_R(n)$  分别为声源信号到达左右耳的传递响应函数,  $n_L(n)$  和  $n_R(n)$  分别为左右耳的噪声信号, 包含加性噪声和混响。

Gammatone 滤波器组的响应函数  $g_i(n, f_i)$  为:

$$g_i(n, f_i) = \cos(2\pi f_i n T_s + \varphi) \cdot \exp(-2\pi B n T_s) \cdot B^J \cdot n^{J-1} \cdot U(n) \quad (2)$$

式中,  $f_i$  为第  $i$  个子频带的中心频率,  $T_s$  为采样周期,  $\varphi$  为滤波器的初始相位,  $J$  为滤波器的阶数, 设置  $J = 4$  以模拟人类听力,  $U(n)$  为单位阶跃函数,  $B$  为带宽,  $B$  的计算公式为:

$$B = b \cdot \text{ERB}(f_i) \quad (3)$$

式中,  $b = 1.019$  为衰减系数,  $\text{ERB}(f_i)$  为滤波器等效矩形带宽, 根据大量实验<sup>[15]</sup>, 可得  $\text{ERB}(f_i)$  的计算公式为:

$$\text{ERB}(f_i) = 24.7 + 0.108 f_i \quad (4)$$

将左右耳信号  $x_L(n)$  和  $x_R(n)$  进行分频, 第  $i$  个左右耳子频带信号为  $G_{i,L}(n, f_i)$  和  $G_{i,R}(n, f_i)$ , 即:

$$\left. \begin{aligned} G_{i,L}(n, f_i) &= g_i(n, f_i) * x_L(n) \\ G_{i,R}(n, f_i) &= g_i(n, f_i) * x_R(n) \end{aligned} \right\} \quad (5)$$

### 1.2 数据压缩

语音信号经过分频之后, 不同的频带具有的能量不同, 所包含的信息量也不同, 将分频后的数据进行数据压缩, 压缩后的左右语音信号数据  $y_L(n)$  和  $y_R(n)$  为:

$$\left. \begin{aligned} y_L(n) &= W G_L^T \\ y_R(n) &= W G_R^T \end{aligned} \right\} \quad (6)$$

式中,  $y_L(n) = [y_{1,L}(n), y_{2,L}(n), \dots, y_{i,L}(n)]^T$ ,  $y_R(n) = [y_{1,R}(n), y_{2,R}(n), \dots, y_{i,R}(n)]^T$ , 其中  $y_{i,L}(n)$  和  $y_{i,R}(n)$  分别是加权之后第  $i$  个左右子带信号。  $W$  是基于子带能量的加权矩阵,  $W = \text{diag}(\omega_1, \omega_2, \dots, \omega_i)$ 。  $G_L$  和  $G_R$  分别是分频后的左右子带信号矩阵, 即加权之后第  $i$  个左右子带信号:

$$\left. \begin{aligned} G_L &= [G_{1,L}(n, f_1), G_{2,L}(n, f_2), \dots, G_{i,L}(n, f_i)] \\ G_R &= [G_{1,R}(n, f_1), G_{2,R}(n, f_2), \dots, G_{i,R}(n, f_i)] \end{aligned} \right\} \quad (7)$$

式中,  $G_{i,L}(n, f_i)$  和  $G_{i,R}(n, f_i)$  是经过 Gammatone 滤波器组处理的左右耳第  $i$  个子带的信号。

由于语音信号的能量和信息主要集中在前 2/3 部分的子频带中<sup>[16]</sup>, 所以将  $\omega_1, \omega_2, \dots, \omega_{\lfloor 2i/3 \rfloor}$  的权重设置为 1, 其余的权重设置为 0。通过该数据压缩, 可以提取信号中的重

要子带信息，消除次要信息的干扰，降低计算复杂度。

### 1.3 基于 SNR 估计的子带选择

语音信号是一种非平稳的随机信号，考虑到人类发声器官在发声过程中的变化速度具有一定限度而且远小于语音信号的变化速度，因此可以假定语音信号是短时平稳的。噪声分为加性噪声和非加性噪声，加性噪声通常分为冲击噪声，周期噪声，宽带噪声，语音干扰噪声等；非加性噪声主要是残响及传送网络的电路噪声等<sup>[17]</sup>。

采用谱减法对子带信号进行 SNR 估计，谱减法是语音增强的有效方法之一，其基本思想是假定加性噪声与短时平稳的语音信号相互独立的条件下，从带噪语音的功率谱中减去噪声功率谱，将语音信号和噪声信号分离，从而进行 SNR 估计<sup>[18]</sup>。假定第  $i$  个左右子带信号  $y_{i,L}(n)$  和  $y_{i,R}(n)$  中， $s_{i,L}(n)$  和  $s_{i,R}(n)$  分别为左右子带信号中的纯净语音信号， $n_{i,L}(n)$  和  $n_{i,R}(n)$  分别为左右子带信号中的噪声信号，则有：

$$\begin{cases} y_{i,L}(n) = s_{i,L}(n) + n_{i,L}(n) \\ y_{i,R}(n) = s_{i,R}(n) + n_{i,R}(n) \end{cases} \quad (8)$$

用  $Y_{i,L}(\omega), S_{i,L}(\omega), N_{i,L}(\omega)$  分别表示左通道的  $y_{i,L}(n), s_{i,L}(n), n_{i,L}(n)$  的傅里叶变换， $Y_{i,R}(\omega), S_{i,R}(\omega), N_{i,R}(\omega)$  分别表示右通道的  $y_{i,R}(n), s_{i,R}(n), n_{i,R}(n)$  的傅里叶变换，则可得下式：

$$\begin{cases} Y_{i,L}(\omega) = S_{i,L}(\omega) + N_{i,L}(\omega) \\ Y_{i,R}(\omega) = S_{i,R}(\omega) + N_{i,R}(\omega) \end{cases} \quad (9)$$

用  $Y_{i,L,angle}(\omega), Y_{i,R,angle}(\omega)$  分别表示相位谱，保留相角，则：

$$\begin{cases} Y_{i,L,angle}(\omega) = \arctan \left[ \frac{\text{Im}(Y_{i,L}(\omega))}{\text{Re}(Y_{i,L}(\omega))} \right] \\ Y_{i,R,angle}(\omega) = \arctan \left[ \frac{\text{Im}(Y_{i,R}(\omega))}{\text{Re}(Y_{i,R}(\omega))} \right] \end{cases} \quad (10)$$

由于假定语音信号与加性噪声是相互独立的，因此有：

$$\begin{cases} |Y_{i,L}(\omega)|^2 = |S_{i,L}(\omega)|^2 + |N_{i,L}(\omega)|^2 \\ |Y_{i,R}(\omega)|^2 = |S_{i,R}(\omega)|^2 + |N_{i,R}(\omega)|^2 \end{cases} \quad (11)$$

用  $P_{i,L,y}(\omega), P_{i,L,s}(\omega), P_{i,L,n}(\omega)$  分别表示  $y_{i,L}(n), s_{i,L}(n), n_{i,L}(n)$  的功率谱，用  $P_{i,R,y}(\omega), P_{i,R,s}(\omega), P_{i,R,n}(\omega)$  分别表示  $y_{i,R}(n), s_{i,R}(n), n_{i,R}(n)$  的功率谱，则有：

$$\begin{cases} P_{i,L,y}(\omega) = P_{i,L,s}(\omega) + P_{i,L,n}(\omega) \\ P_{i,R,y}(\omega) = P_{i,R,s}(\omega) + P_{i,R,n}(\omega) \end{cases} \quad (12)$$

由于平稳噪声的功率谱在发声前和发声期间可以认为基本没有变化，因此可以通过发声前的所谓“寂静段”来估计噪声的功率谱，从而有：

$$\begin{cases} P_{i,L,s}(\omega) = P_{i,L,y}(\omega) - P_{i,L,n}(\omega) \\ P_{i,R,s}(\omega) = P_{i,R,y}(\omega) - P_{i,R,n}(\omega) \end{cases} \quad (13)$$

式中， $P_{i,L,s}(\omega)$  和  $P_{i,R,s}(\omega)$  分别为纯净的语音信号功率谱，为防止出现负功率谱的情况，谱减时当  $P_{i,y}(\omega) < P_{i,n}(\omega)$  时，令  $P_{i,s}(\omega) = 0$ ，然后从这个功率谱通过快速傅里叶逆变换 (IFFT, inverse fast fourier transform)，并借助相位谱，使得较为纯净的语音信号  $\hat{y}_{i,L}(n)$  和  $\hat{y}_{i,R}(n)$  被分离出

来，即：

$$\begin{cases} \hat{y}_{i,L}(n) = \text{IFFT}(P_{i,L,s}(\omega), Y_{i,L,angle}(\omega)) \\ \hat{y}_{i,R}(n) = \text{IFFT}(P_{i,R,s}(\omega), Y_{i,R,angle}(\omega)) \end{cases} \quad (14)$$

因此根据第  $i$  个左右子带信号  $y_{i,L}(n)$  和  $y_{i,R}(n)$  以及式 (14)，可得第  $i$  个左右子带的信噪比  $\text{SNR}_{i,L}, \text{SNR}_{i,R}$ ，即：

$$\begin{cases} \text{SNR}_{i,L}(y_{i,L}(n)) = 10 \log_{10} \frac{\sum_n \hat{y}_{i,L}^2(n)}{\sum_n (y_{i,L}(n) - \hat{y}_{i,L}(n))^2} \\ \text{SNR}_{i,R}(y_{i,R}(n)) = 10 \log_{10} \frac{\sum_n \hat{y}_{i,R}^2(n)}{\sum_n (y_{i,R}(n) - \hat{y}_{i,R}(n))^2} \end{cases} \quad (15)$$

根据子带 SNR 估计，当左右子频带信号 SNR 的均值最大值时，为最优左右子频带  $Y = [y_L(n), y_R(n)]^T$ ，其选择计算公式为：

$$Y = \underset{y_L(n), y_R(n)}{\text{argmax}} \frac{\text{SNR}_{i,L}(y_{i,L}(n)) + \text{SNR}_{i,R}(y_{i,R}(n))}{2} \quad (16)$$

### 1.4 基于子带选择方法的定位对比实验

为了评估本文所提出的子带选择方法的性能，将本文所提出的基于最优子带选择的 SSL 算法 (SS-SSL) 与文献 [19] 的基于通道求和的子带选择定位算法 (GT-PHAT-SSL) 进行比较，实验设置和性能评估标准如 3.1 节所示。将采集到的双耳信号添加信噪比为 -5 dB, 0 dB, 5 dB, 10 dB, 15 dB, 20 dB 的全局白噪声，如图 2~3 展示了不同 SNR 下的两种算法定位的准确率和 RMSE。为了评估算法的运算复杂度，本文使用 Matlab 计算两种子带选择算法的运算时间，如表 1 展示了两种算法的运算时间对比。

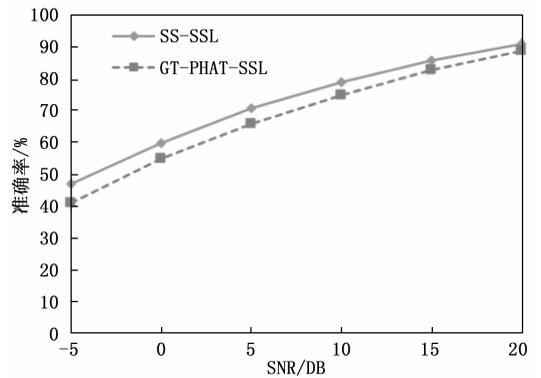


图 2 不同信噪比下的准确率

表 1 运算时间对比

算法	SS-SSL	GT-PHAT-SSL
时间/s	0.21	0.65

图 2 显示了不同 SNR 情况下基于最优子带选择的 SS-SSL 算法的准确率高干基于通道求和的子带选择的 GT-PHAT-SSL 算法，图 3 显示了不同 SNR 情况下基于最优子带选择的 SS-SSL 算法的 RMSE 低于基于通道求和的子带选择的 GT-PHAT-SSL 算法，这说明 SS-SSL 的定位性能优于 GT-PHAT-SSL。表 1 显示了 SS-SSL 的运算时间低于 GT-

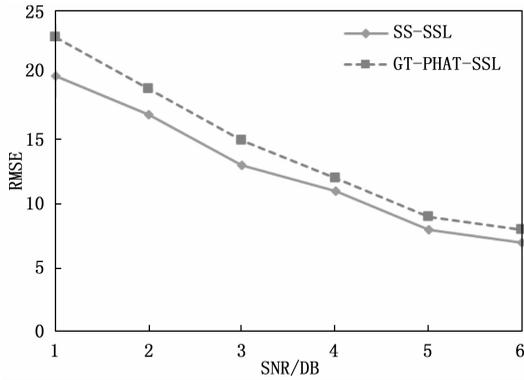


图 3 不同信噪比下的 RMSE

PHAT-SSL, 这说明与 GT-PHAT-SSL 算法相比, 采用最优子带选择的 SS-SSL 可以降低运算复杂度, 提高机器人声源方位角定位的实时性。

## 2 基于互相关的 ITD 估计和基于 DBSCAN 的 SSL

### 2.1 基于互相关的 ITD 估计

ITD 估计算法的方法有很多, 基于互相关的 ITD 算法 (CC-ITD) 具有原理简单, 运算量小的特点。CC-ITD 通过求左右耳两信号  $y_L(n)$  与  $y_R(n)$  之间的互功率谱, 并在频域内给予一定的加权  $\Phi_{y_L y_R}$ , 再反变换到时域得到两信号之间的互相关函数, 互相关函数的峰值处就是两信号之间的相对时延<sup>[20]</sup>。CC-ITD 基本流程如图 4 所示。

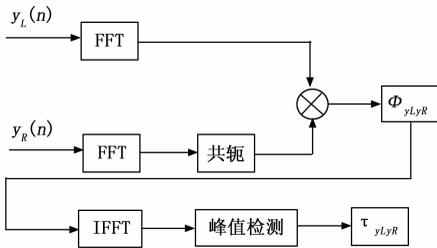


图 4 CC-ITD 基本流程

$y_L(n)$  与  $y_R(n)$  是左右两耳的最优子频带信号, 其傅里叶变换分别为  $Y_L(\omega)$  和  $Y_R(\omega)$ , 两路滤波器的系统函数分别为  $F_L(\omega)$  和  $F_R(\omega)$ , 则两信号之间的互相关函数  $R_{y_L y_R}(\tau)$  可表示为:

$$R_{y_L y_R}(\tau) = \int_0^{2\pi} \varphi_{y_L y_R}(\omega) Y_L(\omega) Y_R^*(\omega) e^{-j\omega\tau} d\omega \quad (17)$$

式中,  $\varphi_{y_L y_R}(\omega)$  为互相关函数的加权函数, 针对不同的噪声和混响环境可以选择不同的加权函数, 其计算公式为:

$$\varphi_{y_L y_R}(\omega) = F_L(\omega) F_R^*(\omega) \quad (18)$$

当  $\varphi_{y_L y_R}(\omega) = 1$  时表示基本广义互相关法的加权函数。 $R_{y_L y_R}(\tau)$  的峰值处是两信号  $y_L(n)$  与  $y_R(n)$  的相对时延  $\tau_{y_L y_R}$ , 即:

$$\tau_{y_L y_R} = \underset{\tau}{\operatorname{argmax}} R_{y_L y_R}(\tau) \quad (19)$$

根据  $\tau_{y_L y_R}$  可以通过 ITD 定位模型求得所需的方位角  $\theta$ , ITD 定位模型如图 5 所示, 图中 A, B 是左右双耳麦克风用于接收声源信号, 以 O 为圆心的圆半径为  $r$ ,  $\theta$  为声源的水

平方方位角。

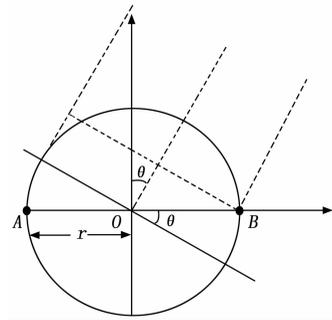


图 5 ITD 定位模型

假定声速传播速度为  $c$ ,  $\tau_{y_L y_R}$  与  $\theta$  的关系式可以表示为:

$$\tau_{y_L y_R} = \frac{r(\sin\theta + \theta)}{c} \quad (20)$$

对于不同的信号频率, ITD 模型有一定的变化规律, 其参数化形式表示为:

$$\tau_{y_L y_R} = \alpha_f \frac{r(\sin\theta + \theta)}{c} \quad (21)$$

式中,  $\alpha_f$  是与  $f_i$  相关的尺度因子。反转模型就可以得到水平方位角  $\theta$  为:

$$\theta = g^{-1}\left(\frac{c\tau_{y_L y_R}}{r\alpha_f}\right) \quad (22)$$

式中,  $g^{-1}$  为  $g(\theta) = \sin\theta + \theta$  的反转函数,  $g^{-1}$  近似表示为:

$$g^{-1}(x) \approx \frac{x}{2} + \frac{x^3}{96} + \frac{x^5}{1280} \quad (23)$$

### 2.2 基于 DBSCAN 的 SSL

采用 CC-ITD 法获取左右子带第  $i$  帧的双耳时间差为  $\tau_{y_L y_R}^i$ , 如图 6 展示基于 CC-ITD 法不同帧的双耳时间差  $\tau_{y_L y_R}^i$ 。

图 6 显示了基于 CC-ITD 法获取  $\tau_{y_L y_R}^i$  存在少量的离群点, 这将会给最终的结果带来干扰。分析其原因可知, 由于最优子带中还残存部分的噪声和混响, 语音信号采集过程中可能存在的短暂“中断期”, 以及前期信号处理带入的噪声, 使得少量  $\tau_{y_L y_R}^i$  为离群值。

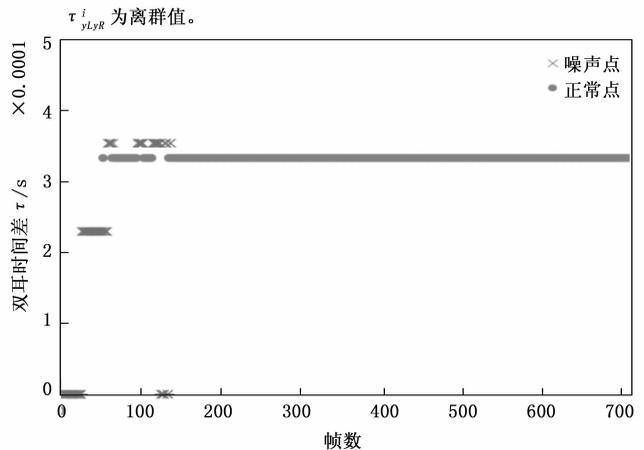


图 6 不同帧的双耳时间差

引入 DBSCAN 算法解决 CC-ITD 的噪声问题, DB-

SCAN 算法使用基于密度的方法来计算数据中任意形状的簇和离群值(噪声), 并且不需要事先知道簇的数量<sup>[21-22]</sup>, 所以引入 DBSCAN 用来解决异常问题, 消除噪声的干扰。

DBSCAN 算法通过两个参数  $Eps$  (半径) 和  $MinPts$  (最小样本点) 识别数据集中的各个簇, 以及判断噪声点。通过 GCC-ITD 法获取数据集  $D = [\tau_{y_1, y_k}^1, \tau_{y_1, y_k}^2, \dots, \tau_{y_1, y_k}^m]$ ,  $m$  是分帧数量, DBSCAN 算法的相关概念如下:

1) 核对象 一个样本  $p$  以  $Eps$  为半径的圆内的有超过一定数目 ( $\geq MinPts$ ) 的样本, 则样本  $p$  称为核对象。

2)  $Eps$  邻域 领域内的点定义为  $N_{Eps}(p) = \{q \in D, dist(p, q) \leq Eps\}$ , 其中  $dist(p, q)$  为  $p, q$  之间的距离。

3) 密度直达对象  $p$  为核对象, 并且  $q$  为  $p$  的  $Eps$  邻域, 则称对象  $q$  从对象  $p$  密度直达。

4) 密度相连 若存在一个对象  $o$ , 使得对象  $p$  和  $q$  都从  $o$  密度可达, 则称对象  $p$  和对象  $q$  密度相连。

DBSCAN 算法的处理流程如表 2 所示。

表 2 DBSCAN 算法处理流程

序号	操作
Input:	数据集 $D, Eps, MinPts(\lceil m/2 \rceil)$
Output:	各个不相交的簇
1:	从数据集 $D$ 选择第一个未标记的观测值 $\tau_{y_1, y_k}^i$ 作为当前点, 并将第一个聚类标签 $C$ 初始化为 1;
2:	找到当前点 $Eps$ 的邻域内的点集, 如果邻居数小于 $MinPts$ , 则将当前点标记为噪声点, 转到步骤 4; 否则将当前点标记为属于簇 $C$ 的核对象;
3:	迭代每个邻近点(新的当前点)并重复步骤 2, 直到找不到可以标记为属于当前簇 $C$ 的新邻近点;
4:	选择下一个未标记的点作为当前点, 并将簇数增加 1。

### 2.3 基于 DBSCAN 的 SSL 定位性能分析

为了评估本文提出的基于 DBSCAN 的声源定位算法 (DBS-SSL) 的定位性能, 将 DBS-SSL 与基于均值互相关的声源定位算法 (MEAN-SSL) 进行对比实验, 实验条件和定位性能标准如 3.1 节所示。将采集到的双耳信号添加信噪比为  $-5 \text{ dB}$ ,  $0 \text{ dB}$ ,  $5 \text{ dB}$ ,  $10 \text{ dB}$ ,  $15 \text{ dB}$ ,  $20 \text{ dB}$  的全局白噪声, 如图 7~8 展示了不同 SNR 下的两种算法定位的准确率和 RMSE。

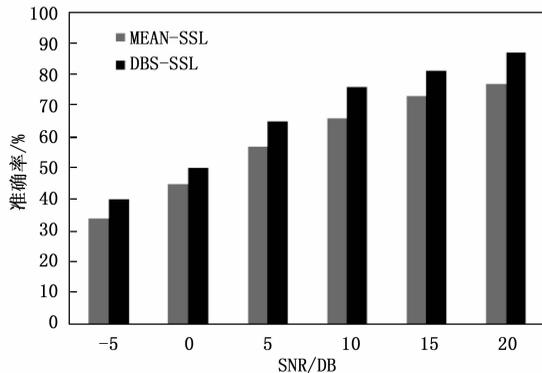


图 7 不同信噪比下的准确率

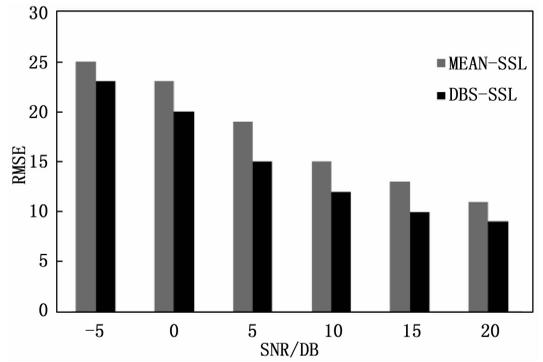


图 8 不同信噪比下的 RMSE

由图 7 可知, 不同信噪比的情况下 DBS-SSL 定位准确率都高于 MEAN-SSL 算法, 由图 8 可知, 不同信噪比的情况下 DBS-SSL 的 RMSE 都要低于 MEAN-SSL 算法, 这说明引入 DNSCAN 算法可以降低基于 GCC-ITD 法获取  $\tau_{y_1, y_k}$  时的噪声影响, 提高定位鲁棒性。

## 3 实验与结果分析

### 3.1 声源定位实验

声源定位系统主要包括麦克风模块, 数据采集模块, 计算机处理模块等。两个麦克风阵列搭载在 3D 打印的平台上, 两个麦克风之间的距离为  $15 \text{ cm}$ , 麦克风传感器连接到信号数据采集卡上, 通过数据线将计算机和信号数据采集卡进行连接, 使用 LabView 对声源信号进行采集, 使用 MatLab 对采集到的左右耳声源信号进行数据分析和处理, 声源定位实验平台如图 9 所示。



图 9 实验平台

本文选用的声音传感器型号为 HJ-386, 该型号麦克风具有全指向性、灵敏度高、抗干扰能力强以及阻抗值低等特点。数据采集卡型号为 NI USB-6009, 其技术参数如表 3 所示。

表 3 数据采集卡技术参数

参数	数值	参数	数值
测量类型	电压	采样率	48 kS/s
通道数	8	电压范围	$-10 \text{ V}, 10 \text{ V}$
分辨率	14 bits	电压精度	$37.5 \sim 138 \text{ mV}$

实验在  $5\text{ m} \times 4\text{ m} \times 3.5\text{ m}$  和  $6\text{ m} \times 8\text{ m} \times 3.5\text{ m}$  的室内环境中进行, 混响时间  $T_{60}$  分别为  $0.3\text{ s}$  和  $0.6\text{ s}$ , 室内噪声主要来自计算机散热和室外车辆行驶噪声。声源采用 CHAINS Speech Corpus 语音库 SOLO 中的单声道女声、男声信号。噪声采用 Noisex-92 和 Nonspeech 噪声库的噪声语音。采样率为  $44.1\text{ kHz}$ , 滤波器的频率范围为  $[0, 8\ 000]\text{ Hz}$ , 最大的中心频率为  $8\ 000\text{ Hz}$ , 根据听觉阈值范围内的临界带<sup>[23]</sup>, 选择滤波器个数为 22 个, 所以子带数量为 22 条, 由于语音信号的能量和信息主要集中在前 15 个子频带中, 所以将  $w_1, w_2, \dots, w_{15}$  的权重设置为 1, 其余的权重设置为 0。帧长为  $20\text{ ms}$ , 帧移为  $10\text{ ms}$ , 规定声源在双耳麦克风右侧时方位角为  $0^\circ$ , 与双耳麦克风垂直时且垂直点为双耳麦克风 midpoint 时方位角为  $90^\circ$ , 在双耳麦克风左侧时方位角为  $180^\circ$ 。

为了评估基于 SNR 估计的子带选择和 DBSCAN 算法对声源定位性能的影响, 本文采用 4 种不同算法做对比声源定位实验, 分别是本文所提出的基于子带选择和 DBSCAN 的 SSL 算法 (SS-DBS-SLL), 基于互相关函数的 SSL 算法 (CC-SSL), 基于最优子带选择的 SSL 算法 (SS-SSL) 和基于 DBSCAN 的 SSL 算法 (DBS-SSL)。算法性能由定位准确率和均方根误差 (RMSE) 评估, 其中定位准确率定义为估计方位角和实际方位角之间的误差在  $\pm 5^\circ$  之内。

### 3.2 不同信噪比下的声源定位性能研究

为了评估算法在不同噪声环境中的声源定位性能, 本实验将采集之后的左右耳语音信号添加信噪比为  $-5\text{ dB}$ ,  $0\text{ dB}$ ,  $5\text{ dB}$ ,  $10\text{ dB}$ ,  $15\text{ dB}$ ,  $20\text{ dB}$  的全局白噪声, 测试算法性能。如图 10~11 展示了不同 SNR 下的算法定位性能。

图 10 表明, 不同信噪比的情况下, 本文所提的 SS-DBS-SSL 算法的定位准确率都要优于 CC-SSL 算法, 并且 SS-SSL 和 DBS-SSL 算法的准确率也都要优于 CC-SSL, 这说明子带选择和 DBSCAN 都可以有效提高定位准确率, 其中子带选择对于提高 SSL 的准确率有更积极作用。在不同信噪比的情况下, 4 种算法的定位性能都随 SNR 的提高而提高。图 11 显示了 SS-DBS-SSL 的 RMSE 在各个信噪比条件下也是最好的, 并且随着信噪比的增大 RMSE 越来越小, 定位性能越来越好。

图 10 和图 11 说明子带选择和 DBSCAN 都可以提高定位性能, 其中子带选择的作用更大。分析其原因, 基于子带选择可以消除噪声频带的影响, 保留主要信号能量, 引入 DBSCAN 可以消除在信号处理引入的异常帧, 以及最优子带中残留的部分噪声的影响, 从而从两个方面提高了定位性能。

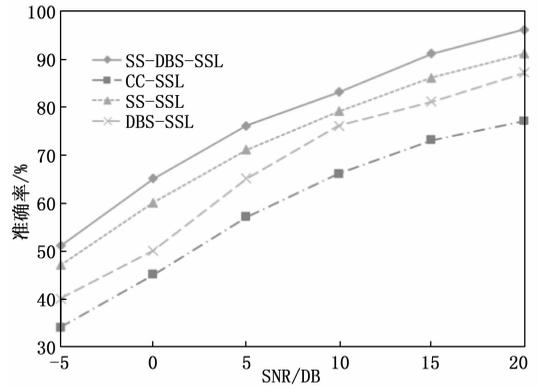


图 10 不同信噪比下的定位准确率

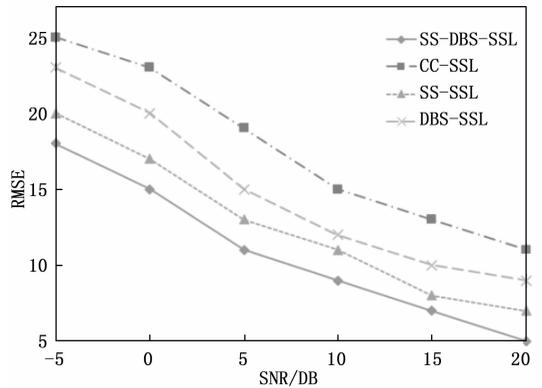


图 11 不同信噪比下的 RMSE

### 3.3 不同噪声环境下的声源定位性能研究

为了评估算法在不同噪声环境下的定位性能, 将采集到的双耳信号添加 SNR 为  $15\text{ dB}$  不同背景噪声, 选择了 12 种不同的背景噪声类型, 其如表 4 所示。

表 5 和表 6 显示了不同噪声环境下的定位准确率和 RMSE。从表 5 可以看出, N2 和 N11 的定位准确率较低, 表 6 也可以看出, N2 和 N11 的 RMSE 较高, 定位性能较低, 这说明算法在 babble 和 Crowd noise 的背景噪声下影响较大, 分析其原因主要是该噪声的主要频率与信号中的主要频率相似, 通过子带选择无法有效消除噪声干扰, 从而影响定位性能。在不同的噪声环境下, SS-DBS-SSL 的定位性能最好, 这也说明了 SS-DBS-SSL 具有较高的鲁棒性。

### 3.4 不同混响条件下声源定位性能研究

为了评估算法在不同混响环境下的定位性能, 分别在混响参数  $T_{60}$  为  $0.3\text{ s}$  和  $0.6\text{ s}$  的房间中进行方位角定位实验。如图 12 展示了算法在不同混响条件下的定位性能。

表 4 背景噪声类型

符号	N1	N2	N3	N4	N5	N6
噪声类型	White	babble	buccaneer1	buccaneer2	destroyerengine	destroyerops
符号	N7	N8	N9	N10	N11	N12
噪声类型	f16	volvo	leopard	hfchannel	Crowd noise	Alarm siren

表 5 不同噪声环境下的定位准确率

符号	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	平均
SS-DBS-SSL	91	86	92	90	93	90	89	93	95	92	85	94	90.83
CC-SSL	73	71	70	67	69	72	71	69	70	72	67	70	70.08
SS-SSL	86	83	86	85	89	85	84	90	90	88	80	84	85.83
DBS-SSL	81	78	83	78	80	81	79	77	82	82	79	82	80.17

表 6 不同噪声环境下的 RMSE

符号	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	平均
SS-DBS-SSL	7	10	7	8	6	8	9	6	6	7	9	5	7.33
CC-SSL	13	14	14	16	13	12	14	13	14	12	15	13	13.58
SS-SSL	8	9	8	8	7	8	8	7	7	8	10	8	8.00
DBS-SSL	10	11	9	12	10	9	11	12	9	9	11	9	10.17

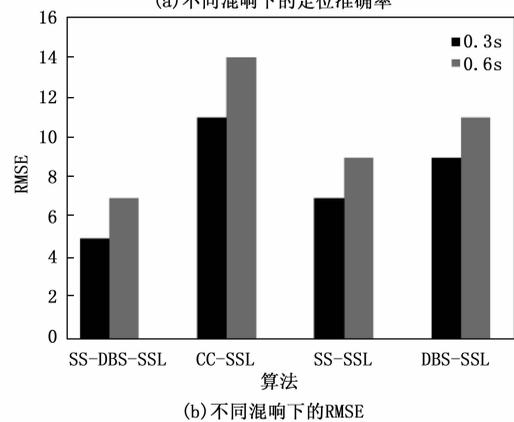
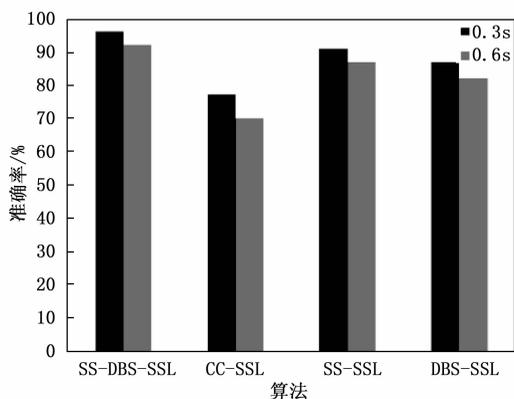


图 12 不同混响条件下的定位性能

图 12 显示了 4 种算法在不同混响条件下的定位准确率和 RMSE。由图 12 可知，4 种算法在低  $T_{60}$  的混响条件下的定位准确率表现更好，并且 RMSE 也更低。这  $T_{60}$  说明越大，定位性能越低。在不同的混响条件下，SS-DBS-SSL 的表现优于其它算法。这也说明了本文所提算法在混响环境中具有一定的鲁棒性。

### 3.5 不同距离和角度下的声源定位性能研究

为了评估算法在不同角度下的定位性能，将声源放置在实验平台的前半面，声源与双耳麦克风的中点距离为

1 m，角度依次为  $0^\circ$ ， $30^\circ$ ， $60^\circ$ ， $90^\circ$ ， $120^\circ$ ， $150^\circ$ ， $180^\circ$ 。图 13 展示了不同角度的情况下定位的准确率和 RMSE。从图 13 可以看到，不同角度的情况下，各个算法的定位性能没有明显的变化趋势，基本保持稳定，其中 SS-DBS-SSL 的定位性能最好。

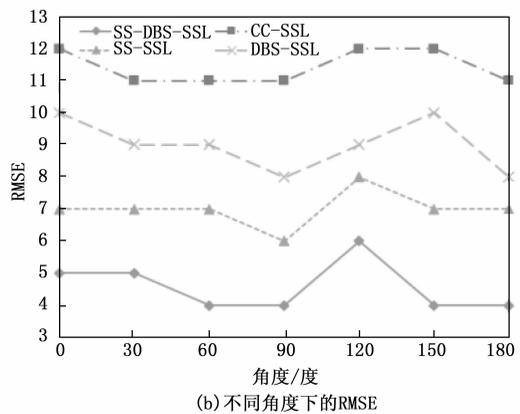
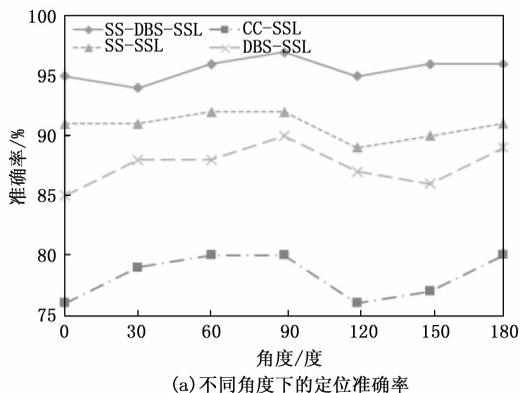
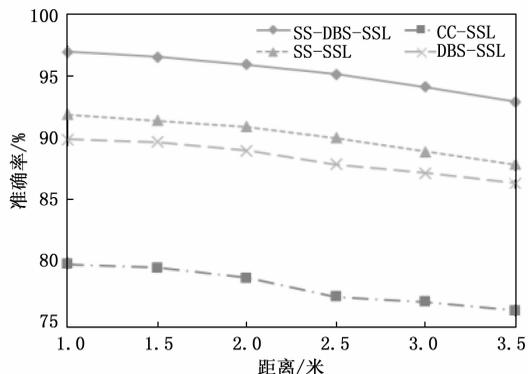


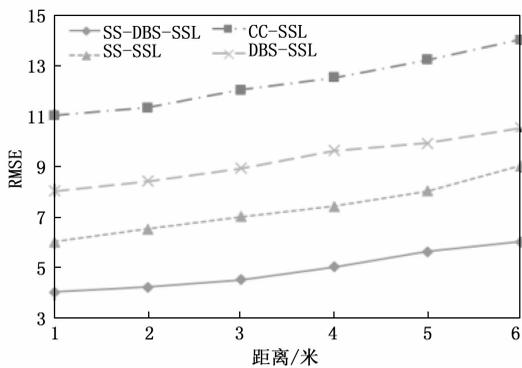
图 13 不同角度下的定位性能

为了评估算法在不同距离下的定位性能，将声源放置在距离实验平台 1 m，1.5 m，2 m，2.5 m，3 m，3.5 m 处，角度为  $90^\circ$ 。图 14 展示了不同距离的情况下定位的准确率和 RMSE。从图 14 可以看到，算法的定位性能随着距离的增大而降低。分析其原因，随着距离的增大，采集到

的信号中的噪声能量越来越高, 声源信号能量越来越低, 从而导致 SNR 越来越低, 所以定位性能下降。



(a) 不同距离下的准确率



(b) 不同距离下的RMSE

图 14 不同距离下的定位性能

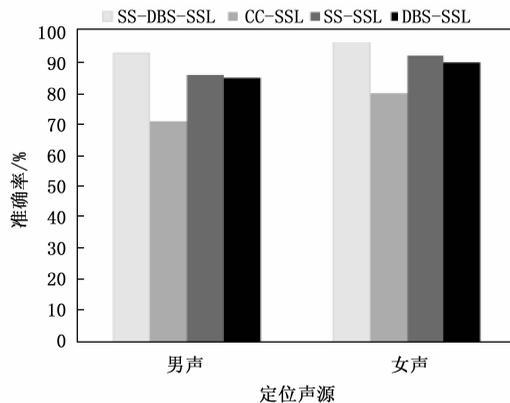
在不同角度和距离的情况下, SS-DBS-SSL 的定位性能都要优于其它算法, 这也体现了本文算法具有较高的鲁棒性和稳定性。

### 3.6 不同声源下的声源定位性能研究

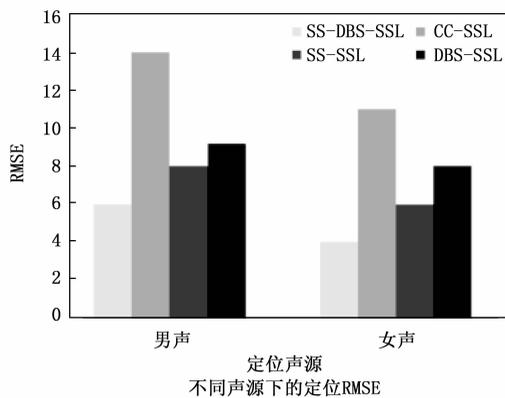
为了评估不同声源对定位性能的影响, 采用男声、女声作为不同的声源, 分析男声和女声情况下的定位性能。图 15 显示了男声和女声的情况下的定位准确率和 RMSE。从图 15 可以看到, 在声源为女声的情况下的定位准确率要优于声源为男声的情况, 在声源为女声的情况下, RMSE 也较小, 并且本文所提算法的表现优于其它算法。这说明声源为女声的定位性能要优于男声, 在不同的声源条件下, SS-DBS-SSL 的定位性能也更好。分析其原因可知, 女声信号的能量大, 在相同的实验环境中, 其定位性能也越好。

## 4 结束语

本文基于 Gammatone 滤波原理对双耳语音信号进行分频, 并通过数据压缩降低计算复杂度, 然后基于谱减法的 SNR 估计选择最优子带, 减少无关子频带影响, 并引入 DBSCAN 算法减少噪声点, 降低异常帧对定位结果的干扰。从实验结果和分析可以看出, 本文提出的算法通用性强, 与基于互相关的传统算法相比, 可以有效改善在混响和噪声的复杂环境中双耳声源的定位精度, 提高声源定位的鲁



(a) 不同声源下的定位准确率



不同声源下的定位RMSE

图 15 不同声源下的定位性能

棒性。进一步, 该实验平台可与移动机器人结合, 研究机器人与声源相对运动时的双耳声源定位问题。

### 参考文献:

- [1] 赵力. 语音信号处理—第 2 版 [M]. 北京: 机械工业出版社, 2009.
- [2] HONG L, JIE Z. A binaural sound source localization model based on time-delay compensation and interaural coherence [C] // IEEE International Conference on Acoustics. IEEE, 2014.
- [3] ARGENTIERI S, DANES P, SOUERES P. A Survey on Sound Source Localization in Robotics: from Binaural to Array Processing Methods [J]. Computer Speech & Language, 2015, 34 (1): 87-112.
- [4] HE W, MOTLICEK P, ODOBEZ J. Joint Localization and Classification of Multiple Sound Sources Using a Multi-task Neural Network [C] // Interspeech 2018, 2018.
- [5] YANG B, LIU H, PANG C, et al. Multiple Sound Source Counting and Localization Based on TF-Wise Spatial Spectrum Clustering [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2019 (99): 1-1.
- [6] WILLERT V, EGGERT J, ADAMY J, et al. A probabilistic model for binaural sound localization. [J]. IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics A Publication of the IEEE Systems Man & Cybernetics Society, 2006, 36

(5): 982-94.

[7] KNAPP C. The generalized correlation method for estimation of time delay [J]. IEEE Trans. Acoust. Speech and Signal Processing, 1976, 24.

[8] HECKMANN M, RODEMANN T, JOUBLIN F, et al. Auditory Inspired Binaural Robust Sound Source Localization in Echoic and Noisy Environments [C] // Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on. IEEE, 2006.

[9] DIBIASE J. A High-Accuracy, Low-Latency Technique for Talker Localization in. [J]. European Journal of Biochemistry, 2000, 216 (1): 281-91.

[10] KIM Y, KIL R. Estimation of Interaural Time Differences Based on Zero-Crossings in Noisy Multisource Environments [J]. IEEE Transactions on Audio Speech & Language Processing, 2007, 15 (2): 734-743.

[11] PANG C, LIU H, ZHANG J, et al. Binaural Sound Localization Based on Reverberation Weighting and Generalized Parametric Mapping [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017: 1-1.

[12] LIU H, SUN Y, YANG G, et al. Binaural sound source localization based on weighted template matching [J]. CAAI Transactions on Intelligence Technology, 2021 (8).

[13] ZHAO F, LI R, PAN D. Deep Learning for Binaural Sound Source Localization with Low Signal-to-noise Ratio [J]. Journal of Physics: Conference Series, 2021, 1828 (1): 012017 (8pp).

[14] NING M, GONZALEZ J, BROWN G. Robust Binaural Localization of a Target Sound Source by Combining Spectral Source  
收稿日期: 2022-08-10; 修回日期: 2022-08-10; 录用日期: 2022-08-10; 网络出版时间: 2022-08-10; 网络出版地址: https://doi.org/10.16675/j.cnki.1001-908x.2022.05.011  
 (上接第 203 页)

[5] 叶强强, 郑明魁, 邱鑫. 基于 ROS 的室内自主导航移动机器人系统实现 [J]. 传感器与微系统, 2022, 41 (2): 90-93.

[6] 洪晓玮, 陈勇, 杨超淞, 等. 有机蔬菜大棚除草机器人研制 [J]. 制造业自动化, 2021, 43 (5): 33-36.

[7] 李会宾, 韩伟, 史云. 果园作业机器人的自主行间导航系统研究 [J]. 中国农业信息, 2019, 31 (4): 51-64.

[8] 兰天, 李端玲, 张忠海, 等. 智能农业除草机器人研究现状与趋势分析 [J]. 计算机测量与控制, 2021, 29 (5): 1-7.

[9] 杨三永, 曾碧. 基于机器视觉的目标定位与机器人规划系统研究 [J]. 计算机测量与控制, 2019, 27 (12): 161-165.

[10] 孙君亮, 闫银发, 李法德, 等. 智能除草机器人的研究进展与分析 [J]. 中国农机化学报, 2019, 40 (11): 73-80.

[11] 杨萍萍, 黄晓诗, 边晓蓉. 农业机器人的现状与未来发展趋势 [J]. 时代农机, 2015 (7): 8-9.

[12] 曹扬. 除草机器人的结构及控制系统的设计与研究 [D]. 郑州: 华北水利水电大学, 2019.

[13] 农元君, 王俊杰. 基于嵌入式的遥感目标实时检测方法 [J]. 光学学报, 2021, 41 (10): 179-186.

[14] 张前. 基于深度学习的 SAR 图像目标检测算法加速研究 [D]. 北京: 北京交通大学, 2021.

[15] 张玉将. 基于多传感器技术的移动机器人定位与建图研究 [D]. 广西: 广西大学, 2018.

[16] 王建平, 卢杉, 武欢欢. TB6612FNG 在直流电机控制设计 Models and Deep Neural Networks [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2018, 26 (11): 1.

[15] LI R, SHUANG Z, YI X. Scene recognition algorithm based on multi-feature and weighted minimum distance classifier for digital hearing aids [C] // 2016 International Conference on Audio, Language and Image Processing (ICALIP). IEEE, 2017.

[16] 李如玮, 潘冬梅, 张爽, 等. 基于 Gammatone 滤波器分解的 HRTF 和 GMM 的双耳声源定位算法 [J]. 北京工业大学学报, 2018, 44 (11): 1385-1390.

[17] 胡航. 现代语音信号处理 [M]. 北京: 电子工业出版社, 2014.

[18] BOTCHEV V. Speech enhancement: theory and practice (2nd ed.) [J]. Computing reviews, 2013, 54 (10): 604-605.

[19] 胡章芳, 乐聪聪, 张毅, 等. 混响环境下基于分频的广义互相关时延估计算法 [J]. 计算机工程, 2018, 44 (9): 269-273.

[20] 万新旺, 吴镇扬. 基于双耳互相关函数的声源定位算法 [J]. 东南大学学报: 自然科学版, 2011, 41 (5): 903-906.

[21] 杨旭辉, 张帆, 胡涛. 大数据环境下基于改进网格单元的 DBSCAN 算法 [J]. 信息工程大学学报, 2021, 22 (3): 351-358.

[22] 冯少荣, 肖文俊. DBSCAN 聚类算法的研究与改进 [J]. 中国矿业大学学报, 2008, 37 (1): 105-111.

[23] DIEPOLD K, DURKOVIC M, ROTHBUCHER M, et al. HRTF customization using multiway array analysis [C] // 2010 18th European Signal Processing Conference. 0.

收稿日期: 2022-08-10; 修回日期: 2022-08-10; 录用日期: 2022-08-10; 网络出版时间: 2022-08-10; 网络出版地址: https://doi.org/10.16675/j.cnki.1001-908x.2022.05.011  
 中的应用 [J]. 电子设计工程, 2010, 18 (6): 65-67.

[17] 屈梦瑶, 易艺, 陆泽青, 等. 基于电容式传感器的非接触式液位测量系统 [J]. 传感器与微系统, 2021, 40 (9): 81-84.

[18] 尹胜楠, 崔学荣, 李娟, 等. 基于快速 ACE 算法的视觉里程计图像增强方法 [J]. 电子测量与仪器学报, 2021, 35 (6): 27-33.

[19] 曾彪. 基于多传感器信息融合的机器人建图研究与实现 [D]. 成都: 电子科技大学, 2021.

[20] 李苗恒, 李明明, 迟长春. 霍尔传感器的无刷直流电动机转速计算法 [J]. 上海电机学院学报, 2018, 21 (2): 28-33.

[21] 李涛, 关永, 王瑞, 等. 机器人操作系统 ROS 通信层的弱终止性验证 [J]. 小型微型计算机系统, 2016, 37 (9): 2140-2144.

[22] 曹洪鑫, 张向慧, 张忠海, 等. 基于 ROS 机械臂实时控制系统的研究现状和趋势分析 [J]. 计算机测量与控制, 2021, 30 (3): 1-7.

[23] 胡锦. 基于激光与视觉信息融合的 SLAM 方法及路径规划研究 [D]. 长沙: 湖南大学, 2020.

[24] 刘延飞, 彭征, 王艺辉, 等. 基于改进的遗传算法的有刷直流电机 PID 参数整定 [J]. 计算机应用, 2022, 42 (5): 1634-1641.

[25] 凤丽洲, 王友卫, 韩琳琳, 等. 双重驱动的果蝇优化算法及其在 PID 控制器中的应用 [J]. 控制与决策, 2021, 36 (9): 2225-2233.