

时延估计的声源定位算法及 MATLAB 实现

陈向阳¹, 刘 松¹, 刘 培², 徐 莹³, 沈 超¹

1. 武汉工程大学计算机科学与工程学院;

2. 邮电与信息工程学院, 湖北 武汉 430074; 3. 湖北省档案局(馆)科技处, 湖北 武汉 430071

摘 要:概述评价了在机器人、视频会议、大型会议记录中等常用声源定位的时延估计和定位方法, 包括广义互相关函数时延估计法、最小方差自适应时延估计法和角度距离定位算法、球形差值定位算法、线性差值法的声源定位方法. 利用仿真软件 Matlab 对各种算法的仿真结果, 综合考虑仿真中使用的信号和噪音的互相关性, 并结合改进后的相位变换加权函数改进了广义互相关函数时延估计法. 结果表明利用改进的广义互相关算法结合球形差值定位技术能够得到更有效、更精确地定位.

关键词:声源定位; 时延估计; 互相关函数; 自适应估计; 插值法

中图分类号: TN643

文献标识码: A

doi: 10.3969/j.issn.1674-2869.2014.010.013

0 引 言

自 20 世纪 70 年代起传声器阵列开始用于语音信号处理以来, 到目前为止已经设计出各种各样的声源定位技术^[1]. 其中尤其是基于到达时间延迟的声源定位技术因其良好的实时性、计算量也不大而成为目前研究的热点之一. 基于到达时间延迟^[2]的声源定位方法需要先估算出声源与传声器阵列各阵元之间的相对时差, 然后根据得到的时差推导计算声源至各阵元之间的距离之差, 最后再利用搜索算法或几何算法得到声源所处的位置. 故其实现可分为两步进行, 即第一步估算出声音信号至传声器阵元之间的相对时间延迟, 候选实现算法有: 广义互相关函数法^[3], 最小均方 (LMS) 自适应滤波算法^[4]; 第二步再利用所得的延迟值进行数学建模并利用阵元空间位置信息进行定位, 具体实现算法有角度距离定位算法^[5]、球形差值定位算法^[6]. 通过对已知各种算法研究, 结果发现选择改进的广义互相关算法并且与球形差值定位技术结合使用能得到更好的定位效果.

1 时延估计

1.1 广义互相关函数法

假定声源信号表达式如式(1)、(2)

$$x_1(n) = a_1 s(n - \tau_1) + w_1(n) \quad (1)$$

$$x_2(n) = a_2 s(n - \tau_2) + w_2(n) \quad (2)$$

这里使用 $s(n)$ 来表示声源信号, 利用 a_1 、 a_2

来表达声波从声源到传声器的传播衰减系数. 而利用 τ_1 、 τ_2 来表达声源传播到两部传声器所需要的时间, 而 $w_1(n)$ 、 $w_2(n)$ 则代表高斯白噪声^[7], 并且 $s(n)$ 与 $w_1(n)$ 、 $w_2(n)$ 是不相关^[8-9]的.

利用求两信号之间的互相关函数来实现广义互相关函数. 两信号之间的相对时延就是在这种互相关函数的峰值位置处. 从声源至两台麦克风的信号 $x_1(n)$ 和 $x_2(n)$ 的互相关函数 $R_{12}(\tau)$ 可表示为

$$R_{12}(\tau) = E(x_1(n)x_2(n-\tau)) \quad (3)$$

将式(1)和(2)代入(3)式后便得到

$$R_{12}(\tau) = a_1 a_2 E(s(n-\tau_1)s(n-\tau_2-\tau)) + a_1 E(s(n-\tau_1)w_2(n-\tau)) + a_2 E(s(n-\tau_2-\tau)w_1(n)) + E(w_2(n)w_2(n-\tau)) \quad (4)$$

因为 $w_1(n)$ 和 $w_2(n)$ 是不相关的高斯白噪声, $s(n)$ 和 $w(n)$ 也是不相关的随机信号, 故由式(4)得

$$R_{12}(\tau) = E(a_1 a_2 s(n-\tau_1)s(n-\tau_2-\tau)) = a_1 a_2 R_s(\tau - (\tau_1 - \tau_2)) \quad (5)$$

利用相关函数的性质就可以得到, 当 $\tau = \tau_{12} = \tau_1 - \tau_2$ 时 $R_{12}(\tau)$ 便能够取得最大值. 因此求得 $R_{12}(\tau)$ 的最大值处所对应的 τ 值, 其实就是两个麦克风之间的时延 τ_{12} .

利用互功率谱和互相关函数之间的关系就可以得到式(6):

$$R_{12}(\tau) = \int_0^\pi G_{12}(\omega) e^{-j\omega\tau} d\omega \quad (6)$$

其中 $G_{12}(\omega)$ 则代表麦克信号 $x_1(n)$ 和 $x_2(n)$

收稿日期: 2013-12-04

基金项目: 湖北省教育科学“十二五”规划 2013 年度立项课题(2013B060)

作者简介: 陈向阳(1969-), 男, 河南开封人, 副教授, 硕士. 研究方向: 网络通信工程、复杂网络、生物信息等.

之间的互功率谱.

但是由于语音与噪声的短时间处理可能导致出现 $R_{12}(\tau)$ 的峰值不够突出,由此导致时延 τ 估计精确度的下降.为了能够锐化得到 $R_{12}(\tau)$ 的峰值,利用信号和噪声的先验知识,可以在频域内对互功率谱 $G_{12}(\omega)$ 进行加权处理来抑制噪声与混响部分的不利影响.然后再进行反变换到时域中,如此就能够得到广义互相关函数(GCC),即

$$R_{12}^g(\tau) = \int_0^\pi \varphi_{12}(\omega) G_{12}(\omega) e^{-j\omega\tau} d\omega \quad (7)$$

在应用中,可以为不同的噪声和混响情况选择不同的加权函数 $\varphi(\omega)$,以便能尽量让 $R_{12}^g(\tau)$ 具有相对尖锐的峰值.由于低信噪比以及有限的窗长也会使这种分析出现不稳定的情况,因此合理的加权函数 $\varphi(\omega)$ 会难以选择^[9].

1.2 最小均方(LMS)自适应滤波方法

假设已知在自适应滤波器^[10]中所包含的输入信号分别是 $X(n)$ 和 $d(n)$ 以及输出信号分别为 $y(n)$ 和 $e(n)$,其中 $X(n)$ 能够利用参数可调整数字滤波器中 $w_n(n)$ 的加权分量,以便逐渐逼近所期望的信号 $d(n)$,从而得到输出响应;并且进一步将它与参考信号 $d(n)$ 进行对比,最终产生误差信号 $e(n)$, $e(n)$ 能够利用某种自适应算法对滤波器参数^[11]作出改变,以便使 $e(n)$ 的均方值能够达到最小化.

若把发送信号 $X(n)$ 当作输入信号,而将回波信号 $d(n)$ 当作目标信号的话,那么经过 p 个采样周期延时后再输入滤波器的期望输入端,滤波器便会输出 $y(n)$,再求二者之差便得到误差信号 $e(n)$,假设第 m 阶滤波器的加权量用 $h_m(n)$ 表示,假设收敛步长用 μ 表示,限制 μ 的取值范围为 $0 < \mu < \frac{1}{\lambda_{\max}}$,其中 λ_{\max} 是 $X(n)$ 信号的自相关函数的最大特征值.则有:

$$e(n) = d(n-p) - y(n) \quad (8)$$

$$y(n) = \sum_{m=-p}^p h_m(n) x(n-m) \quad (9)$$

$$h_m(n+1) = h_m(n) + \mu e(n) x(n-m) \quad (10)$$

其中 $m = -p, -p+1, \dots, p-1, p$.

根据 LMS 准则,当 $h(n) = R_x^{-1}P$ 时,便会使得均方误差达到最小,而在该处滤波器也会收敛,然后利用 filter-taps 中加权值 $h_m(n)$ 曲线的最大值处相应的 m 值,减掉 p 个采样周期延时,便得到所需要的输入信号与期望信号之间的时延 t .

2 定位算法

2.1 角度距离定位算法

角度距离定位算法^[12]最重要内容的一部分就是:利用声源与两个麦克风之间的距离差,以及两

麦克风之间的距离估计出声源与这两麦克风所在直线的夹角^[13]由式(11)表示

$$\theta = \cos^{-1} \left(\frac{\tau_i \cdot c}{|m_{i1} - m_{i2}|} \right) \quad (11)$$

然后,由于麦克风摆放成一定的几何形状,根据几何形状中的几何公式,以声源到麦克风之间的距离差以及麦克风之间的距离作为输入变量,就可以求出声源到坐标原点之间的距离.最后再根据所得出的距离与角度即可以求出声源在坐标轴中的坐标.模拟实验中至少需要有 4 个麦克风,坐标原点放置 1 个,至少有 1 个坐标轴上除了原点的麦克风之外还应该放置有 2 个相对原点对称的麦克风,这样做是为了能够计算出声源到原点之间的距离,而且计算量很小.除了麦克风的摆放问题,模拟试验中声源到麦克风之间的距离差是有正负之分的,所以在仿真实验中一定要使之一一对应,不然就会得到错误的结果.

2.2 球形插值法

球形插值法^[14]中的主要计算公式如式(12):

$$(R_s + d_{ij})^2 = R_j^2 - 2\mathbf{r}_i^T \mathbf{r}_s + R_s^2 \quad (12)$$

其中 $R_i = |\mathbf{r}_i|$, $R_s = |\mathbf{r}_s|$,利用 \mathbf{r}_i 表示 \mathbf{m}_j 至 \mathbf{m}_i 之间的矢量,利用 \mathbf{r}_s 表示 \mathbf{m}_j 至声音源 S 之间的矢量,利用 d_{ij} 表示声音源 S 至传声器 $\mathbf{m}_i, \mathbf{m}_j$ 之间的距离差值.在该算法中使用了 9 个麦克风,在坐标原点处放置的麦克风作为参照麦克风,在前边得出了另外 8 个麦克风与参照麦克风之间的声源距离差之后,就会有 8 个参数不同的公式(12)组成的方程组.在解方程组时需要使用最小二乘法,通过它可以直接计算出声源的三维坐标.该算法得到的精确度非常高,在 1 万米的数量级上的相对误差是百分之零点零几,在 10 万米时产生的误差也只有 1.4% 左右,这个距离和误差程度是可以接受的.此外,在仿真中针对不同的麦克风数量进行了大量实验,结果发现至少要使用 5 个麦克风才会有相对精确的测量值.总之仿真结果说明了该方法对麦克风的数量是具有一定限制的.

3 广义互相关函数时延估计法改进后的相位变换加权函数

针对前述难以选择加权函数问题,下面对广义互相关函数(GCC)时延估计法提出改进方法.首先考虑在 $G_{w_1 w_2}(\omega)$ 不为零时,应用中就能够估计出无音阶段的噪声互相关函数 $R_{(w_1 w_2)}(\tau)$,也就是 $R_{12}(\tau) - R_{w_1 w_2}(\tau)$,然后再将噪声的互相关函数从信号的互相关函数中去除掉,由此削弱因噪声带来的不利影响.其次针对在房间中出现的混响

现象,可以将 $G_{12}(\omega)$ 分解为两个成分: 直接成分和反射成分^[9]. 参见式(13)

$$G_{12}(\omega)=a_1a_2G_{ss}(\omega)e^{-j\omega\tau_{12}}+G_{ss}(\omega)\sum_{i=1,\infty}^{\infty,\infty}a_ia_je^{-j\omega d_{ij}}\tag{13}$$

由于 PHAT 加权函数的分母 $|G_{12}(\omega)|$ 不等于 $a_1a_2G_{ss}(\omega)$,因此不能确保对 $e^{-j\omega\tau_{12}}$ 分量的白化作用. 因此,考虑修正 PHAT 加权函数,即乘以加权因子 γ ,以保证 $|\gamma|G_{12}(\omega)=a_1a_2G_{ss}(\omega)$,由此一来便能起到较好的抗混响效应. 加权因子 γ 实质上就是直接分量占功率谱的比例. 可以利用房间的混响时间来估计出直接分量占据整个功率谱的比例,由此便得到加权因子 γ .

当信号能量比较小时, $|G_{12}(\omega)|$ 便会接近于零,这是 PHAT 加权函数 $\varphi_{12}(\omega)=\frac{1}{|G_{12}(\omega)|}\varphi_{12}(\omega)$ 所存在的问题. 此时很大的 PHAT 加权函数 $\varphi_{12}(\omega)$ 值可能引起错误的结果. 因此,可以为信噪比较低时添加一个加权因子来解决该问题,参见式(14):

$$\varphi_{12}(\omega)=\frac{1}{\gamma|G_{12}(\omega)|+(1-\gamma)|\omega_{12}(\omega)|^2}\varphi_{12}(\omega)\tag{14}$$

即在信噪比较低时,就对频域给予一定的信噪比加权. 在利用两路信号求解某个时延时, PHAT 加权函数互功率谱法中需计算 3 次快速傅里叶变换. 因此其运算量相对还是比较低的. 但具有抑制噪声作用,因此在实际定位系统中,得到改进后的 PHAT 加权函数法得到了很好的应用.

由于在仿真模拟实验中针对声源信号的噪音是人为加上的,所以要得出只有噪声的信号是很容易的,那么在得到了噪音的自相关函数^[13]之后,就可以在信号的互相关函数中减掉该部分,经过仿真之后发现结果的确有所改进,但是效果还不够明显. 考虑到信号与噪音可能并不是完全的不相关,也就是说信号和噪音的互相关性可能也会产生一定的影响,于是可用信号互相关函数减去一个分量,即

$$k\cdot r_{\text{noise}}+\frac{(1-k)}{2}\cdot r_{\text{xn}}+\frac{(1-k)}{2}\cdot r_{\text{yn}}\tag{15}$$

其中 r_{noise} 是噪音的相关函数, r_{xn} 与 r_{yn} 是噪音和两个信号的互相关函数,是一个比例系数体现的是这三个分量之间的比例. 经过多次仿真试验发现当 k 为 0.3 时消除噪音的效果最佳.

4 模拟仿真结果

改进前广义互相关函数算法的仿真结果如图 1 所示. 改进前 10db 的结果如图 2 所示.

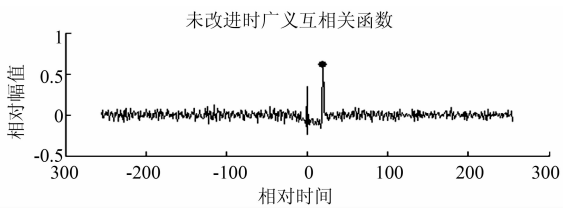


图 1 信噪比为 20db 时的互相关函数图
Fig. 1 The cross-correlation function diagram with signal-to-noise ratio=20dB

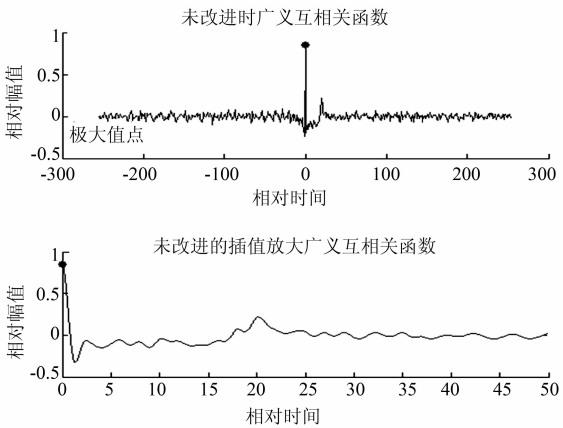


图 2 信噪比为 10db 时的互相关函数图
Fig. 2 The cross-correlation function diagram with signal-to-noise ratio=10dB

在 10db 时算法经改进后的结果如图 3 所示. 由图 2 与图 3 的比较可以得知改进算法后的方法对噪音的抑制效果更好.

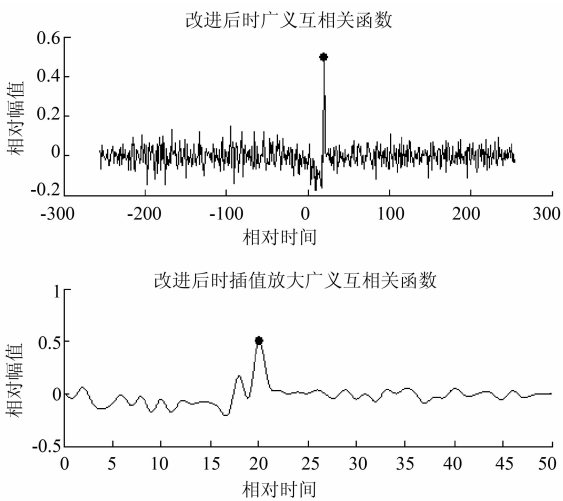


图 3 改进后信噪比为 10db 互相关函数图
Fig. 3 The improved cross-correlation function diagram with signal-to-noise ratio=10dB

在 20db 时使用最小均方(LMS)自适应算法得到的结果如图 3~4 所示. 从图 3~4 中可以看出在 20db 时,时延 5 处出现明显的峰值. 上述仿真结果验证了该方法的正确性.

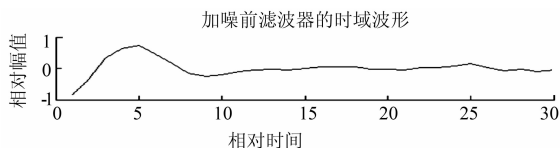


图 4 信噪比为 20 的权值时域图

Fig. 4 Weighted graph in time domain with signal-to-noise ratio=20dB

在 10db 时该方法的结果如图 5 所示。由图 5 可以看出,自适应算法在信噪比为 10db 时已经出现较多的峰值,而且在正确的时延处峰值处略有偏移。所以通过以上实验仿真的结果可以看出利用改进后的广义互相关函数算法要比自适应算法具有更好的抑制噪音效果。

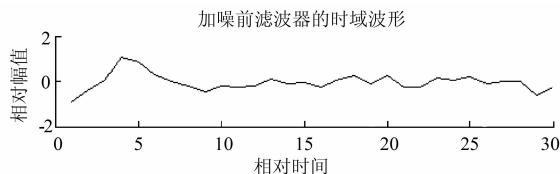


图 5 信噪比为 10 的权值时域图

Fig. 5 Weighted graph in time domain with signal-to-noise ratio=10dB

5 总结与展望

通过对不同声源定位中的到达时间延迟算法以及定位算法的对比,以及在 Matlab 仿真软件上的模拟,说明改进的互相关函数法相对于自适应算法具有更好的抗噪能力,球形插值法也具有更精确的定位能力,所以建议实际应用中使用这两种算法能够实现更精确的声源定位。同时由于诸多限制,仿真模拟试验中还有诸多问题有待解决,如在互相关函数中减去噪音成分的比例具体应该是多少,傅里叶变换对互相关函数的影响到应该是什么样的,傅里叶变化的点数与信号的采样点数的取值是否有一定的要求,才能使得最后的峰值更加明显等问题。

致 谢

本文得到了湖北省教育科学“十二五”规划 2013 年度立项课题:地方性高校新兴交叉学科建设的研——以网络科学为例(项目编号 2013B060)的资助,在此表示衷心的感谢。

参考文献:

[1] 李承智,曲天书,吴玺宏.一种改进的 AEDA 声源定位及跟踪算法[J]. 北京大学学报:自然科学版, 2005,41(5):809-814.

LI Chengzhi, Qu Tianshu, WU Xihong. A modified AEDA algorithm for sound source localization and tracking[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2005,41(5):809-814. (in chineses)

[2] 郭兰申. 面向反恐领域基于多源信息融合的机器人感知系统研究[D]. 天津:河北工业大学,2006.

GUO Lanshen. Research on robot perception system for anti-terrorism based on multi-source information fusion[D]. Tianjin:Hebei University of Technology, 2006. (in chineses)

[3] 金乃高,陈喆. 基于分层采样粒子滤波的麦克风阵列说话人跟踪法[J]. 电子学报,2008,36(1):194-198. JIN Nai-gao, CHEN Zhe. Microphone array speaker tracking using layered sampling particle filtering[J]. ACTA Electronica Sinica, 2008, 36(1):194-198. (in chineses)

[4] 殷福亮,陈喆. 基于加权子空间拟合的声源定位与跟踪方法[J]. 电子与信息学报,2008,30(9):2134-2137.

YIN Fu-liang, CHEN Zhe. Weighted subspace fitting sound source localization method based on particle filtering[J]. Journal of Electronics & Information Technology, 2008, 30(9):2134-2137. (in chineses)

[5] 蔡卫平. 基于麦克风阵列的声源定位与跟踪算法研究[D]. 南京:东南大学,2010.

CAI Weiping. Research and tracking algorithm for sound source localization based on microphone array [D]. Nanjing: Southeast University, 2010. (in chineses)

[6] 冯道宁,王浩. 基于麦克风阵列和虚拟仪器的声源定位[J]. 计算机工程与应用,2012,48(21):130-132. FENG Dao-ning, WANG Hao. Speech source localization based on microphone arrays and labview[J]. CEA, 2012, 48(21):130-132. (in chineses)

[7] 邵怀宗,林静然,彭启琮,等. 基于麦克风阵列的声源定位研究[J]. 云南民族大学学报,2004,13(4):256-259.

SHAO Huaizong, LIN jingrui, PENG Qicong, et al. Study on localization for speech based on microphone array[J]. Journal of Yunnan Nationalities University, 2004, 13(4):256-259. (in chineses)

[8] 陆晓燕. 基于麦克风阵列实现声源定位[D]. 大连:大连理工大学,2003.

LU Xiaoyan. Implementation of acoustic source localization based on microphone array [D]. Dalian: Dalian University of Technology, 2003. (in chineses)

[9] 周浩祥. 基于麦克风阵列的声源定位方法研究[D]. 大连:大连理工大学,2002.

ZHOU Haoxiang. Research on acoustic source local-

- ization methods based on microphone array [D]. Dalian: Dalian University of Technology, 2003. (in chineses)
- [10] 严素清, 黄冰. 传声器阵列的声源定位研究[J]. 电声技术, 2004(12): 27-30.
YAN Suqing, HUANG Bing. Study on acoustic source localization based on microphone arrays[J]. Loudspeaker And Microphone, 2004 (12): 27-30. (in chineses)
- [11] YOUN D H, AHMED N, CARTER G C. On using the LMS algorithm for time delay estimation [J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1992, 30(5): 798-801.
- [12] JIE Huang, Kume K, Saji A. et al. Robotic spatial sound socalization and its 3D sound human Interface [C]//First International Symposium on Cyber Worlds (CW2002), Tokyo: 2002: 191-197.
- [13] 王宏禹, 邱天爽. 自适应噪声抵消和时间延迟估计 [D]. 大连: 大连理工大学出版社, 1999.
WANG Hongyu, QIU Tianshuang. Adaptive noise cancellation and time delay estimation [D]. Dalian: Dalian University of Technology press, 1999. (in chineses)
- [14] JIE Huang, OHNISHI N, SUGIE N. A biomimetic system for localization and separation of multiple sound sources instrumentation and measurement [J]. IEEE Transactions on Volume, 1995, 44(3): 733-738.
- [15] SMITH J O, ABEL J S. Closed-form least-squares source location estimation from range difference measurements [J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1987, 35(12): 1661-1669.

Sound source localization algorithm based on time delay estimation with Matlab

CHEN Xiang-yang, LIU Song, LIU Pei, XU Ying, SHEN Chao

1. School of Computer Science and Engineering, Wuhan Institute of technology, Wuhan 430205, China;
2. The College of Post and Telecommunication of Wuhan Institute of Technology, Wuhan 430074, China;
3. Science and Technology Department, Hubei Provincial Archives Bureau, Wuhan 430071, China

Abstract: Methods of time delay estimation and sound source localization in robots, video conferences and meeting recordings were reviewed, including time delay estimation of generalized cross correlation function, minimum variance of adaptive time delay estimation, position angle distance algorithm, spherical difference localization algorithm and linear interpolation. According to the simulation of various algorithms with Matlab, considered the cross correlation feature of input signal and noise in simulation, an improved generalized time delay estimation of cross correlation was presented using the reformed phase transformation method. The simulation results show that using the improved generalized cross correlation algorithm can obtain more efficient and accurate positioning combined with spherical difference positioning technology.

Key words: sound source localization; time delay estimation; adaptive estimation; interpolation method

本文编辑:陈小平