

麦克风阵列时延估计的 Cram   Rao 界限

王冬霞^{1,2}, 殷福亮², 陈 哲²

(1. 辽宁工业大学电子与信息工程学院, 辽宁锦州 121001; 2. 大连理工大学电子与信息工程学院, 辽宁大连 116023)

摘 要: 麦克风阵列拓扑结构对麦克风阵列语音处理系统具有重要影响. 该文针对麦克风阵列语音处理系统的时延估计性能, 通过对声源和麦克风空间位置的几何描述, 推导出了时延估计的 Cram   Rao 界限, 分析了麦克风阵列拓扑结构对其时延估计性能的影响, 并给出了常见的麦克风阵列拓扑结构时延估计性能的仿真实验. 实验结果有效地验证了理论分析所得出的结论.

关键词: 麦克风阵列; 阵列拓扑结构; 时延估计; Cram   Rao 界限

中图分类号: TN912. 3 **文献标识码:** A **文章编号:** 0372-2112 (2008) 12-2473-05

Cram   Rao Bound of Time Delay Estimation for Microphone Array

WANG Dong-xia^{1,2}, YIN Fu-liang², CHEN Zhe²

(1. Electronic and Information Engineering College, Liaoning University of Technology, Jinzhou, Liaoning 121001, China;

2. School of Electronic and Information Engineering, Dalian University of Technology, Dalian, Liaoning 116023, China)

Abstract: Microphone array placement has an important influence on the performance of microphone array speech processing system. For the performance of its time delay estimation (TDE), Cram   Rao Bound (CRB) of TDE is derived by use of geometrical description of source and microphones, and the effect of microphone array configurations on its TDE performance is analyzed in this paper. Simulation experiments are done on the TDE performance of conventional microphone array configurations. Simulation results effectively verify the conclusion.

Key words: microphone array; array configuration; time delay estimation; Cram   Rao bound

1 引言

麦克风阵列是由一组在空间不同位置的麦克风按一定规则布置所形成的阵列. 与单麦克风相比, 麦克风阵列能够有效地利用语音信号的时空信息, 进行房间混响和噪声环境下的语音增强和声源定位. 近年来, 麦克风阵列语音处理已经得到国内外学者的广泛关注, 并在车载系统、视频会议系统、机器人导航等领域有着重要的应用前景.

目前, 有关麦克风阵列语音增强和声源定位算法的研究已取得重要进展^[1~7], 并在实际中得到应用. 但在麦克风阵列语音处理中, 有关麦克风阵列拓扑结构对其语音处理系统性能影响的研究却相对较少. 实际上, 麦克风阵列拓扑结构对其性能具有重要的影响^[8]. 因此, 在阵列空间采样定理的基础上, 对麦克风阵列拓扑结构进行优化, 以进一步改善麦克风阵列语音处理系统的性能, 是一项很有意义的工作.

1995 年, Gazor 和 Genier 从阵列方向图距离、MVDR

估计器最小方差等方面, 分别给出了麦克风阵列拓扑结构对其性能影响的非解析表达式^[9]. 1998 年, Rabinkin 在其博士论文中给出一种麦克风阵列拓扑结构对时延估计误差和声源定位精度影响的非解析表达式^[10], 但该式是通过拟合实验数据而得到. 文献[11]和[12]分别研究了阵列拓扑结构对声源定位精度和阵列波束形成性能的影响.

考虑到包括语音增强和声源定位在内的大多数麦克风阵列语音处理系统都是基于时延估计技术, 因此, 本文针对麦克风阵列语音处理系统的时延估计性能, 推导了时延估计的 Cram   Rao 界限, 给出了麦克风空间摆放位置对麦克风阵列时延估计性能影响的一种解析表达式.

本文的具体结构安排如下: 在所选声学模型的基础上, 从理论上推导出麦克风阵列拓扑结构对其时延估计性能影响的一种解析表达式; 然后在实验部分, 详细验证了麦克风阵列拓扑结构对麦克风阵列系统时延估计性能的影响, 并得出相应结论.

收稿日期: 2007-05-10; 修回日期: 2008-07-05

基金项目: 国家自然科学基金 (No. 60372082)

$$\begin{aligned}
&= \frac{1}{9} \sum_{m=0}^{N-1} \left[\frac{ds((n-m) - ij)}{dt} \cdot \frac{dx_s}{dij} \cdot \frac{dx_s}{dij} \right. \\
&\quad + \frac{ds((n-m) - ij)}{dt} \cdot \frac{dy_s}{dij} \cdot \frac{dy_s}{dij} \\
&\quad \left. + \frac{ds((n-m) - ij)}{dt} \cdot \frac{dz_s}{dij} \cdot \frac{dz_s}{dij} \right]^2 \\
&= \frac{1}{9} \sum_{m=0}^{N-1} \left[\frac{ds((n-m) - ij)}{dt} \cdot \frac{1}{dij} \cdot \frac{dx_s}{dij} \right. \\
&\quad + \frac{ds((n-m) - ij)}{dt} \cdot \frac{1}{dij} \cdot \frac{dy_s}{dij} \\
&\quad \left. + \frac{ds((n-m) - ij)}{dt} \cdot \frac{1}{dij} \cdot \frac{dz_s}{dij} \right]^2 \quad (11)
\end{aligned}$$

$$\begin{aligned}
- E \left[\frac{d^2 \ln p(x, ij)}{d^2 ij} \right] &= \frac{1}{9c^2} \sum_{m=0}^{N-1} \left\{ \left[\frac{ds((n-m) - ij)}{dt} \right]^2 \left[\frac{|r_{js}| |r_{is}|}{\cos \phi_s (|r_{js}| (x_i - x_s) - |r_{is}| (x_j - x_s))} \right]^2 \right. \\
&\quad + \left[\frac{ds((n-m) - ij)}{dt} \right]^2 \left[\frac{|r_{js}| |r_{is}|}{\sin \phi_s (|r_{js}| (y_i - y_s) - |r_{is}| (y_j - y_s))} \right]^2 \\
&\quad \left. + \left[\frac{ds((n-m) - ij)}{dt} \right]^2 \left[\frac{|r_{js}| |r_{is}|}{\sin \phi_s (|r_{js}| (z_i - z_s) - |r_{is}| (z_j - z_s))} \right]^2 \right\} \\
&= \frac{1}{9c^2} \sum_{m=0}^{N-1} \left\{ \left[\frac{ds((n-m) - ij)}{dt} \right]^2 \cdot \frac{1}{c^2} f(r_{is}, r_s, r_{js}) \right\} \quad (16)
\end{aligned}$$

其中

$$\begin{aligned}
f(r_{is}, r_s, r_{js}) &= \left[\frac{|r_{js}| |r_{is}|}{\cos \phi_s (|r_{js}| (x_i - x_s) - |r_{is}| (x_j - x_s))} \right]^2 + \left[\frac{|r_{js}| |r_{is}|}{\sin \phi_s (|r_{js}| (y_i - y_s) - |r_{is}| (y_j - y_s))} \right]^2 \\
&\quad + \left[\frac{|r_{js}| |r_{is}|}{\sin \phi_s (|r_{js}| (z_i - z_s) - |r_{is}| (z_j - z_s))} \right]^2 \quad (17)
\end{aligned}$$

为了进一步简化表达式并不失一般性,假设采样间隔足够小,式(16)的求和运算可以用积分形式来表示,即

$$- E \left[\frac{d^2 \ln p(x, ij)}{d^2 ij} \right] = \frac{1}{9c^2} \cdot \frac{1}{T_s} \int_0^{T_s} \left(\frac{ds(t - ij)}{dt} \right)^2 dt \cdot \frac{1}{c^2} f(r_{is}, r_s, r_{js}) \quad (18)$$

式中 T_s 为 N 个采样点对应的的时间. 考虑到语音信号能量可以表示为

$$= \int_0^{T_s} s^2(t) dt \quad (19)$$

并注意到, $\frac{1}{c^2} = 1/(2B)$, $\frac{1}{c^2} = N_0 B (N_0/2$ 为噪声功率), 则式(18)可以进一步表示为

$$- E \left[\frac{d^2 \ln p(x, ij)}{d^2 ij} \right] = \frac{1}{9c^2} \cdot \int_0^{T_s} \left(\frac{ds(t - ij)}{dt} \right)^2 dt \cdot \frac{1}{c^2} f(r_{is}, r_s, r_{js})$$

如图1所示,设声源 r_s 的水平角和仰角分别为 θ_s 和 ϕ_s , 则该声源单位时间内传播的位移在 x 坐标轴上的投影为

$$\begin{aligned}
dx_s &= c dt \cdot \frac{x_s}{\sqrt{x_s^2 + y_s^2}} \cdot \frac{\pm \sqrt{x_s^2 + y_s^2}}{\sqrt{x_s^2 + y_s^2 + z_s^2}} \\
&= c dt \cos \theta_s \cos \phi_s \quad (12)
\end{aligned}$$

进一步得到

$$\frac{dx_s}{dt} = c \cos \theta_s \cos \phi_s \quad (13)$$

同理

$$\frac{dy_s}{dt} = c \sin \theta_s \cos \phi_s \quad (14)$$

$$\frac{dz_s}{dt} = c \sin \phi_s \quad (15)$$

对式(4)中的 x_s 、 y_s 和 z_s 求导,并将其结果与式(13)、(14)和(15)的结果均代入式(11)中,得到

$$\begin{aligned}
&= \frac{1}{9c^2} \cdot \int_0^{T_s} \left(\frac{ds(t - ij)}{dt} \right)^2 dt \cdot \frac{1}{c^2} f(r_{is}, r_s, r_{js}) \quad (20)
\end{aligned}$$

设麦克风接收信号即麦克风阵列输入信噪比为

$SNR_{in} = \frac{1}{N_0/2}$, 语音信号 $s(t)$ 的傅里叶变换为 $S(\omega)$, 根据傅里叶变换的微分特性和 Parseval 定理, 式(20)可以进一步表示为

$$- E \left[\frac{d^2 \ln p(x, ij)}{d^2 ij} \right] = \frac{SNR_{in}}{9c^2} \cdot \frac{(2\pi)^2 / S(\omega)^2 d}{\int S(\omega)^2 d\omega} \cdot \frac{1}{c^2} f(r_{is}, r_s, r_{js}) \quad (21)$$

设 $Q = \frac{(2\pi)^2 / S(\omega)^2 d}{\int S(\omega)^2 d\omega}$ 为语音信号带宽

的度量,则式(21)可以简化为

$$-E\left[\frac{d^2 \ln p(\mathbf{x}, \mathbf{y})}{d^2 \mathbf{y}}\right] = \frac{\text{SNR}_{in}}{9c^2} \cdot Q \cdot \frac{2}{is} \cdot f(r_{is}, r_s, r_{js}) \quad (22)$$

由式(22)构造 Fisher 信息 $J(ij)$ 为

$$J(ij) = -E\left[\frac{d^2 \ln p(\mathbf{x}, \mathbf{y})}{d^2 \mathbf{y}}\right] = \frac{\text{SNR}_{in}}{9c^2} \cdot Q \cdot \frac{2}{is} \cdot f(r_{is}, r_s, r_{js}) \quad (23)$$

于是,时延估计的 Cramér-Rao 界限 $\text{CRB}(ij)$ 为

$$\begin{aligned} \text{CRB}(ij) &= J^{-1}(ij) = \frac{1}{-E\left[\frac{d^2 \ln p(\mathbf{x}, \mathbf{y})}{d^2 \mathbf{y}}\right]} \\ &= \frac{9c^2}{\text{SNR}_{in} \cdot Q \cdot \frac{2}{is} \cdot f(r_{is}, r_s, r_{js})} \end{aligned} \quad (24)$$

将式(17)代入式(24)中并整理,得到

$$\begin{aligned} \text{CRB}(ij) &= \frac{9c^2}{Q} \cdot \frac{1}{\text{SNR}_{in}} \cdot \frac{1}{2} \left[\left(\frac{\mathbf{r}_{is} \odot \mathbf{r}_s}{r_{is} r_s} \right)^2 + \left(\frac{\mathbf{r}_{js} \odot \mathbf{r}_s}{r_{js} r_s} \right)^2 \right. \\ &\quad - 2 \left[\left(\frac{x_s}{r_s} \right)^2 \frac{(x_i - x_s)(x_j - x_s)}{|r_{is}| |r_{js}|} \right. \\ &\quad + \left. \left(\frac{y_s}{r_s} \right)^2 \frac{(y_i - y_s)(y_j - y_s)}{|r_{is}| |r_{js}|} \right. \\ &\quad \left. \left. + \left(\frac{z_s}{r_s} \right)^2 \frac{(z_i - z_s)(z_j - z_s)}{|r_{is}| |r_{js}|} \right] \right] \quad (25) \end{aligned}$$

式中“ \odot ”表示矢量内积。

如图1所示,设矢量 \mathbf{r}_s 与矢量 \mathbf{r}_{is} 之间的夹角为 θ_{is} , 矢量 \mathbf{r}_s 与矢量 \mathbf{r}_{js} 之间的夹角为 θ_{js} , 声源与麦克风之间的矢量 \mathbf{r}_{is} 和 \mathbf{r}_{js} 的水平角分别为 ϕ_{is} 和 ϕ_{js} , 它们的仰角分别为 ψ_{is} 和 ψ_{js} , 根据矢量分析理论^[14], 式(25)可以进一步表示为

$$\begin{aligned} \text{CRB}(ij) &= \frac{9c^2}{Q} \cdot \frac{1}{\text{SNR}_{in}} \cdot \frac{1}{2} \left[\cos^2 \theta_{is} + \cos^2 \theta_{js} \right. \\ &\quad - 2 \left(\cos^2 \theta_s \cos \theta_{is} \cos \theta_{js} + \sin^2 \theta_s \sin \theta_{is} \sin \theta_{js} \right. \\ &\quad \left. \left. + \cos^2 \phi_s \cos \phi_{is} \cos \phi_{js} \right) \right], 1 \leq i < j \leq M \quad (26) \end{aligned}$$

式(26)的结果表明,麦克风阵列空间摆放即麦克风阵列的拓扑结构以一种解析方式影响着麦克风阵列语音处理系统的时延估计性能:如果声音传播信道是理想信道,没有损耗即 $\theta_{is} = 1$, 则麦克风阵列系统时延估计性能依赖于声源矢量 \mathbf{r}_s 的方向以及矢量 \mathbf{r}_{is} 和 \mathbf{r}_{js} 的方向,而与声源距离 $|r_s|$ 、声源与两麦克风的间距 $|r_{is}|$ 和 $|r_{js}|$ 均无关(实际信道不是理想信道,故与距离 $|r_{is}|$ 有关);此外,输入信噪比越高,麦克风阵列系统时延估计性能越好;反之,性能越差。

4 仿真实验与结果分析

4.1 实验环境

选择房间大小为 7m × 5m × 3m, 房间冲激响应采用

Image 模型产生^[15]。混响时间 T_{60} 可以表示为墙壁反射系数的函数,即

$$T_{60} = -\frac{0.163V}{S \log(1 - \alpha)} \quad (27)$$

式中 V 、 S 与 α 分别表示房间体积、总面积与墙壁反射系数。本实验选取各面墙的反系数 $\alpha = 0.75$, 此时混响时间约为 300ms。

在实验中,麦克风阵列由 6 个麦克风组成,其拓扑结构分别为间距为 d 的均匀线阵、间距为 d 的倍数的对数线阵和间距为 d 的均匀圆阵(一种特殊的面阵,圆阵半径为 r)。这些阵列结构均为麦克风阵列语音处理系统(语音增强和声源定位)中常用的麦克风阵列拓扑结构。语音源距麦克风阵列中心 3m, 阵列所接收的语音信号为平面波。语音信号入射方向与阵列水平面夹角成 30° 、 60° 、 90° 、 120° 和 150° 方式而改变。语音信号采样率为 8kHz。

4.2 实验结果

采用互功率谱相位时延估计算法(CSP)进行时延估计^[16]。由于麦克风数目为 6 个,因此共产生 $C_6^2 = 15$ 个时延估计结果。对真实值和估计值之间误差的方差进行 Monte Carlo 仿真,得出不同情况下的时延估计方差曲线。

当麦克风间距 $d = 0.15\text{m}$ 、混响时间为 300ms 时,在不同输入信噪比情况下,时延估计性能如图2所示。从图2可以看出,随着输入信噪比的增加,麦克风阵列系统时延估计精度提高。当信噪比到达 15dB 以上时,时延估计精度的提高趋势相对平缓。在三种典型的阵列拓扑结构中,均匀圆阵的时延估计性能略好于对数线阵,对数线阵的时延估计性能略优于均匀线阵。

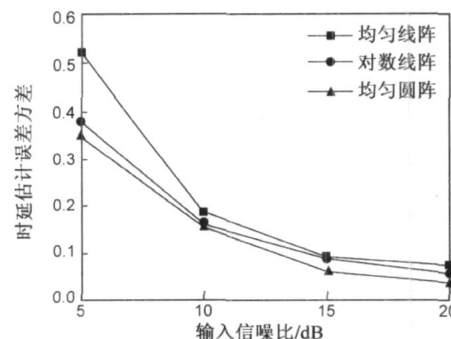


图2 麦克风阵列时延估计性能与输入信噪比之间的关系

当输入信噪比为 5dB、混响时间为 300ms 时,在不同麦克风间距情况下,时延估计性能如图3所示。从图3可以看出,随着麦克风间距的增加,麦克风阵列系统时延估计精度有所提高。在所选的声学环境下,均匀圆阵的时延估计性能好于对数线阵;对数线阵的时延估计性能略优于均匀线阵。

在实际声学环境下,图2和图3所示的时延估计

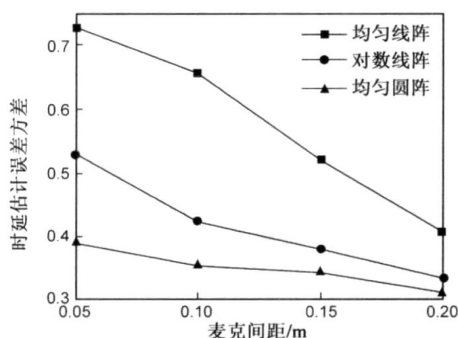


图3 麦克风阵列拓扑结构与时延估计性能之间的关系

Cramér-Rao 界限均大于时延估计 Cramér-Rao 界限的理论值。但是,实测值与理论值趋势基本一致,即它在一定程度上验证了式(26)所得到的结论。

5 结束语

针对麦克风阵列语音处理系统的时延估计问题,在时延估计 Cramér-Rao 界限推导的基础上,得到了一种描述麦克风阵列拓扑结构对其系统时延估计性能影响的解析表达式。这些结论对于麦克风阵列语音处理系统的应用设计具有重要的指导意义。

参考文献:

- [1] Omologo M, Svaizer P. Acoustic source location in noisy and reverberant environment using CSP analysis [A]. IEEE International Conference on Acoustics, Speech, and Signal Processing [C]. Adelaide, Australia: IEEE Signal Processing Society Press, 1996. 921 - 924.
- [2] Brandstein S M. A Framework for Speech Source Localization Using Sensor Arrays [D]. USA: Brown University, 1995.
- [3] Fischer S, Simmer U K. Beamforming microphone array for speech acquisition in noisy environments [J]. Speech Communication, 1996, 20 (2): 215 - 227.
- [4] Doclo S, Moonen M. GSVD-based optimal filtering for single and multimicrophone speech enhancement [J]. IEEE Transactions on Signal Processing, 2002, 50 (9): 2230 - 2244.
- [5] MCCOWAN A L, BOURLARD H. Microphone array post-filtering for diffuse noise field [A]. IEEE International Conference on Acoustics, Speech, and Signal Processing [C]. Florida, USA: IEEE Signal Processing Society Press, 2002. 1 - 905 - 1 - 908.
- [6] Wang Dongxia, Yin Fuliang. A subband adaptive learning algorithm for microphone array based speech enhancement [A]. Lecture Notes in Computer Science [C]. Berlin: Springer, 2005. 592 - 597.
- [7] 王冬霞, 殷福亮. 联合波束形成与谱减法的麦克风阵列语音增强方法 [J]. 大连理工大学学报, 2006, 46 (1): 121 - 126.
- [8] Johnson D, Dudgeon D. Array Signal Processing: Concepts and

Techniques [M]. New Jersey: Prentice-Hall, 1993.

- [9] Gazor S, Grenier Y. Criterion for position of sensors for microphone array [J]. IEEE Transactions on Speech and Audio Processing, 1995, 3 (4): 294 - 303.
- [10] Rabinkin D V. Optimum Sensor Placement for Microphone Arrays [D]. New Brunswick Rutgers: The State University of New Jersey, 1998.
- [11] Yang B, Scheuing J. Cramer-Rao bound and optimum sensor array for source localization from time differences of arrival [A]. IEEE International Conference on Acoustics, Speech and Signal Processing [C]. Philadelphia, USA: IEEE Signal Processing Society Press, 2005. 961 - 964.
- [12] Wang D X, Yin F L. Experiment evaluation of microphone array placement for speech enhancement [A]. 6th International Symposium on Test and Measurement [C]. Dalian: Dalian University of Technology Press, 2005. 1583 - 1586.
- [13] Harry L Van Trees. Detection, Estimation, and Modulation Theory [M]. New York: John Wiley and Sons Inc, 2003.
- [14] 蔡蓓蓓, 许履瑚, 梁在中, 等. 实用数学手册 [M]. 北京: 科学出版社, 2002.
- [15] Allen J B, Berkley A D. Image method for efficiently simulating small room acoustics [J]. Journal of the Acoustic Society of America, 1979, 65 (4): 943 - 950.
- [16] Omologo M, Svaizer P. Use of the cross-power-spectrum phase in acoustic event localization [J]. IEEE Transactions on Speech and Audio Processing, 1997, 5 (3): 288 - 292.

作者简介:



王冬霞 女, 1975 年 12 月出生于辽宁省海城市。博士, 辽宁工业大学副教授。研究方向为语音信号处理和阵列信号处理的理论与应用。
E-mail: dxwanglucky@yahoo.com.cn



殷福亮 男, 1962 年 9 月出生于辽宁省抚顺市。大连理工大学教授, 博士生导师。研究方向为语音处理、图像处理和现代通信技术。
E-mail: flyin@dlut.edu.cn