

# 基于 FPGA 的实时声源定位

李俊杰, 何友, 宋杰

(海军航空工程学院 信息融合技术研究所, 山东 烟台 264001)

**摘要:** 提出了利用 2 个麦克风基于 FPGA 的声源定位的方法。具体通过基于相位变换改进的互相关方法成功在低信噪比(10 dB)的噪声环境下完成声源定位。利用同样的算法和硬件结构, 可以在 1 片 FPGA 芯片上实现 5 组并行的时域处理的系统, 而且每个麦克风的功耗只有 77 mW~108 mW。

**关键词:** 声源定位; 时延估计; FPGA

中图分类号: TN912.34

文献标识码: A

## Real-time sound sources localization based on FPGA

LI Jun Jie, HE You, SONG Jie

(Research Institute of Information Fusion, Naval Aeronautical Engineering Institute, Yantai 264001, China)

**Abstract:** This paper presents a single FPGA implementation of a realtime sound localization, utilizing a cross correlation technique based on a modified version of the phase transform, successfully localizes sound sources in noisy environments with as low an SNR as 10 dB. Using the same algorithm and similar hardware architecture, it is shown that up to 5 parallel systems, all real time, can be implemented on a single FPGA while only utilizing an estimated 77 mW~108 mW per microphone.

**Key words:** sound source localization; time delay estimation; FPGA

实时声源定位在许多方面得到了应用, 例如声音的识别和电话会议, 可以利用阵列麦克风来实现对多个声源信号的获取和并行处理<sup>[1-3]</sup>。由于处理多路语音信号需要多个处理器, 使得其实现费用昂贵, 即便是使用 DSP, 系统也会带来很大的功耗, 因而限制了其在许多实际中的应用。例如 Brown 大学发展的大规模麦克风阵列系统利用多个 DSP 处理器和缓冲器来实现声源的定位, 每个麦克风的功耗达到了 400 mW。这大大超过了一些便携式设备(PDA 和手机)的功耗, 因此最好的解决办法是设计专用芯片。

本文将阐述声源定位系统在 FPGA 中的实现, 为专用芯片提供一个可行性参考, 具有很好的商业应用价值。以前采用 DSP<sup>[4]</sup>或是 DSP+FPGA<sup>[5]</sup>实现多路声源信号的定位, 而本设计的整个定位系统除了前端的模拟部分外其余部分均在 FPGA 中实现。采取有效的算法后, 整个硬件实现的功耗可以控制在 77 mW~108 mW 之间。

### 1 声源定位的算法

现有许多算法<sup>[1-4]</sup>实现声源定位, 包括基于信号子

空间的方法(例如 MUSIC 算法)和空间似然方法<sup>[2,4]</sup>等, 最为常用的方法是估计信号的对应的麦克对到达延时(TDOA)<sup>[3]</sup>估计方法。该方法的每一组麦克对将声源定位在 3 维空间的一个双曲面上, 这样通过多个麦克对确定的双曲面的交点能有效地实现声源的定位。TDOA 估计方法已进行了很多研究<sup>[3,6]</sup>, 最为普通的是广义互相关 GCC(Generalized Cross Correlation)方法<sup>[6]</sup>。与其他的方法相比, 基于 GCC 的方法计算量小、计算效率高。

假设 2 个麦克各自接收的信号分别为  $m_1(t)$  和  $m_2(t)$  (包括噪声、回响和声音的延时信号)。常用的估计延时的方法是互相关方法:

$$\tilde{\tau} = \arg \max_{\beta} \int_{\omega} W(\omega) M_1(\omega) \overline{M_2(\omega)} e^{-j\omega\beta} d\omega \quad (1)$$

式中,  $\tilde{\tau}$  为  $\tau$  的估计,  $M_1(\omega)$ 、 $M_2(\omega)$  分别是 2 路信号的傅里叶变换,  $W(\omega)$  为互相关的权函数。常用的  $W(\omega)$  有 2 种选择:

$$W_{\text{PHAT}}(\omega) = \frac{1}{|M_1(\omega)| |M_2(\omega)|} \quad (2)$$

$$W_{UCC}(\omega)=1 \quad (3)$$

PHAT 权系数对应的是相位变换，在回响环境中效果明显<sup>[1,3,6]</sup>。UCC 权系数为 1，对应的是单纯的互相关而没有经过滤波处理。

离散信号的 GCC 表示为：

$$\tilde{\tau}=\arg \max _{\beta} \sum_{k=0}^{N / 2} W(k)\left|M_1(k)\right|\left|M_2(k)\right| \cos (\theta(k)) \quad (4)$$

式中， $\theta(k)=\angle M_1(k)-\angle M_2(k)-2 \pi F_s k \beta / N$  表示相位误差， $N$  表示总的采样点数， $F_s$  表示采样频率。(4)式可以看成是在不同频率下的相位误差的补偿函数。理想情况下，相位误差将会接近零，此时(4)式取得最大值，余弦相位误差选择函数将会产生大的相位补偿，使相位误差降到最低。

如果采用方波作为补偿函数，(4)式可以写成：

$$\tilde{\tau}=\arg \max _{\beta} \sum_{k=0}^{N / 2} W(k)\left|M_1(k)\right|\left|M_2(k)\right| \operatorname{rect}\left(\frac{\theta(k)}{\varepsilon}\right) \quad (5)$$

式中， $\varepsilon$  表示方波的脉冲宽度， $\operatorname{rect}(t)=\begin{cases} 1 & |t| < 1 \\ 0 & \text{else} \end{cases}$

与(4)式相比，(5)式用来估计 TDOA 的好处在于计算量的减少，而且在低的信噪比下具有更好的性能，所以本文采用的是(5)式在 FPGA 中的实现。

## 2 FPGA 的实现

设计中只考虑最简单的一对麦克风的 TDOA 的估计，每个麦克风接收的信号经过放大、带通滤波后以 20 kHz 的频率进行采样，每个采样的数据宽度为 24 位。取量化后的高 8 位送往 FPGA (Xilinx Virtex II 2000) 中进行运算。尽管只讨论了 1 对麦克风的情况，多个麦克风对也能在同一个 FPGA 中实现 (本文将会在后面具体介绍)。

输入的采样信号分别存储在 2 个 Buffer 里(Buffer 的大小为 256~1 024 个采样点)，然后将输出信号经过汉宁窗滤波，将数据转化为 16 位的浮点数存储在 FFT 的 Buffer 中，FFT 模块将各自的 Buffer 中的数据进行 FFT 运算，如图 1 所示。

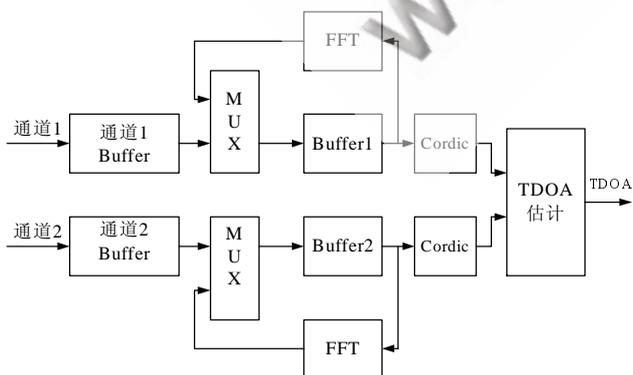


图 1 系统框图

在 FPGA 的实现过程中采用 CORDIC 算法，它将傅

里叶变换的复数表达形式转换成幅度和相位表达形式，在能够减少计算量的同时不增加硬件的资源。2 路信号的幅度和相位计算出来后利用(5)式得到 TDOA 的估计，如图 2 所示。估计的过程涉及到根据(5)式搜索最大的  $\tau$ ，搜索的范围从  $-30 T_s \sim 30 T_s$  (步进为采样周期  $T_s$ )， $\varepsilon=0.5$ 。为了实现 TDOA 的实时估计，硬件部分主要由 3 部分组成：输入信号采集部分、FFT 的计算及幅相转换部分、TDOA 的估计部分。2 个 Buffer 已经能够满足 GCC 前端数据缓存的要求，但考虑到 CORDIC 算法的误差，设计中多用了 1 个 Buffer 储存数据的复数形式，为系统的误差分析提供数据，确保定位精度。

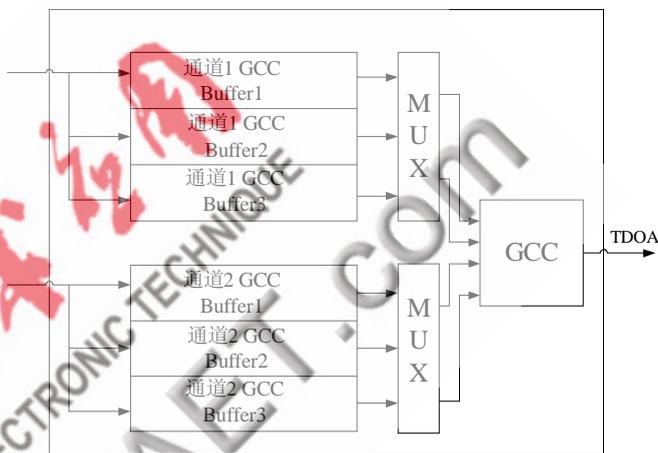


图 2 TDOA 估计模块

## 3 实验结果

在实验中 Buffer 的大小为 1 024 位，对应的时间缓存时间为 50 ms。麦克风对按前面介绍的方法安置。第一个实验是对固定声源的定位，讲话人在房间内固定的地方说话，如图 3 所示。

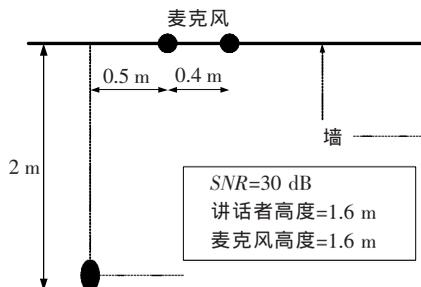


图 3 固定声源定位的示意图

要注意保持麦克风对与讲话者嘴的高度一致。将麦克风得到的数据经过放大、滤波、采样和 FPGA 处理。系统噪声由麦克风、放大器或滤波器等器件引入，信噪比为 30 dB。利用麦克风获取的 50 ms 的信号帧，分别用 GCC 和 PHAT 的权值来估计讲话者的 TDOA。将每帧的到达时间时延  $\tau$  转换成波达方向(DOA) $\phi=\arcsin\left(\frac{v \tau}{d}\right)$ ，其

中,  $v$  表示声音的传播速度 (取 345 m/s),  $d$  表示 2 个麦克风之间的距离 ( $d=0.4$  m)。

通过 DOA 的估计和实际的 DOA 来计算 DOA 的误差, 实际的 DOA 可以通过讲话者在环境的实际位置得到。DOA 误差如图 4 所示。利用 PHAT 权值的 TDOA 定位精度要比 UCC 的好, 因此试验中利用的是 PHAT 权值。为了得到基于 FPGA 的声源定位系统在运动的声源和不同背景噪声下的性能, 本文利用 PHAT 权系数分别进行实验。如图 5 所示, 讲话者从一个地方移动到另外一个地方, 整个移动持续 1 min, 讲话者始终面对着麦克风对。

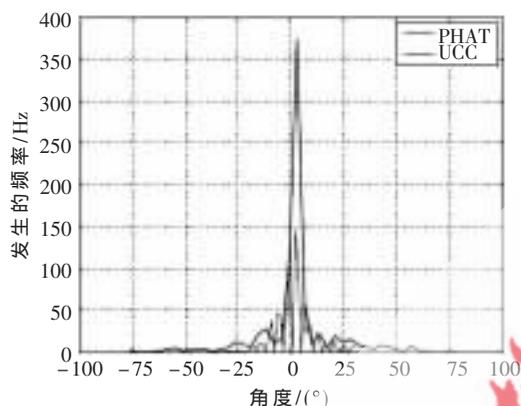


图 4 DOA 误差分布图

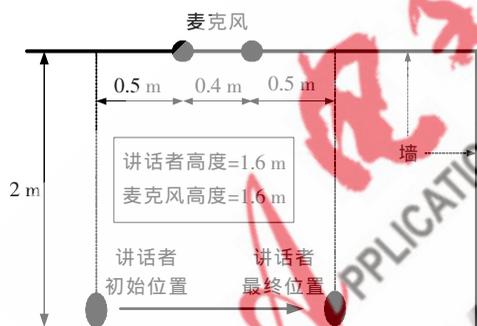


图 5 运动声源定位的示意图

分别在信噪比为 30 dB、20 dB、10 dB、0 dB 的条件下对 2 个不同的讲话者做试验, 当信噪比为 30 dB 时, 噪声只由传感器和信号处理系统自身引入; 当信噪比为 20 dB、10 dB、0 dB 时, 则通过提供一个高斯噪声源来提供, 通过调整噪声的强度可以实现信噪比的变化。信噪比为 30 dB 的 DOA 误差如图 6(a) 所示, 在不同的信噪比条件下,  $0^\circ$  的位置具有共同的峰值, 但是随着信噪比的降低 (20 dB 和 10 dB 分别对应图 6(b) 和图 6(c)), 误差越来越大, 当降到 0 dB 时起不到定位的作用, 如图 6(d) 所示。

实时声源定位系统在 Xilinx 公司的 xc3s1000 FPGA 中实现, 按文中提出的算法和实现方法可以实现 10 dB 信噪比以上的声源定位, 整个系统具有很好的鲁棒性。通常的处理多对麦克风对的信号算法常常利用 DSP, 然

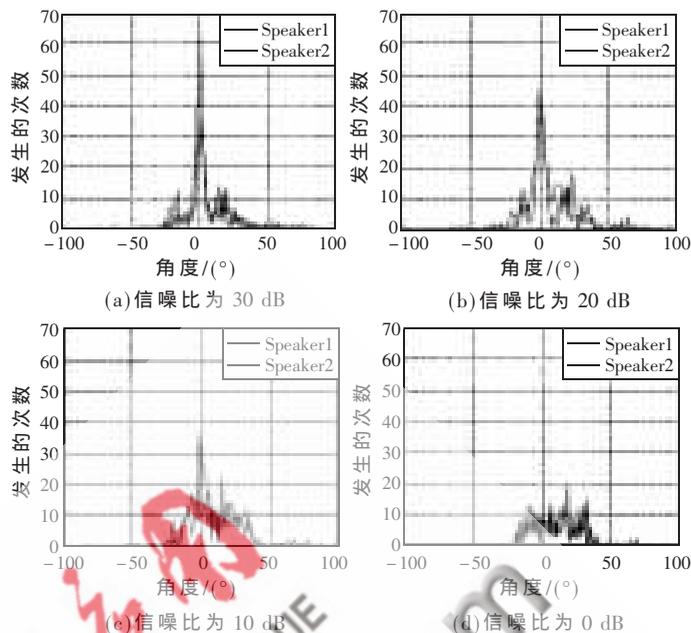


图 6 不同信噪比条件下的 DOA 误差分布图

而功耗的需求以及外围设备的复杂使得 DSP 在许多场合受到限制。利用本文介绍的算法, 在 FPGA Virtex II Pro-70 中并行处理 6 个 TDOA 估计模块, 时钟选取为 10 MHz, 功耗可以控制在 0.776 W~1.074 W 之间, 试验消耗的逻辑门为 1 192 793。通过流水线处理, 如果系统的时钟为 100 MHz, 在 1 块 FPGA 中能并行处理 50 个 TDOA 估计模块, 不过功耗要增大到 7.76~10.74 W 之间。由于 Virtex II Pro-70 有足够的存储空间, 因此输入 Buffer 部分无需额外的逻辑门。与其他的方法相比, 如大型麦克风阵列每个麦克的功耗为 400 mW<sup>[5]</sup>, 本文的方法的平均功耗要小得多 (每个麦克 77~108 mW), 为减少 VLSI 电路的功耗提供了一个切实可行的途径。

#### 参考文献

- [1] AARABI P, MAHDAVI A. The relation between speech segment selectivity and time delay estimation accuracy. In Proceedings of ICASSP, May 2002.
- [2] AARABI P, ZAKY S. Robust sound localization using multi-source audiovisual information fusion. Information Fusion, 2001, 3(2): 209-223.
- [3] BRANDSTEIN M S, SILVERMAN H. A robust method for speech signal time-delay estimation in reverberant rooms. In Proceedings of ICASSP, May 1997.
- [4] PONCA M, SCHAUER C. FPGA implementation of a spike-based sound localization system. In 5th International Conference on Artificial Neural Networks and genetic Algorithms, 2001.
- [5] SILVERMAN H F, PATTERSON W R, FLANAGAN J L. The huge microphone array. Brown University Technical Report, May 1996.
- [6] KNAPP C H, CARTER G. The generalized correlation method for estimation of time delay. IEEE Transactions on ASSP, 1976, 24(4): 320-327.

(收稿日期: 2009-03-06)