

基于嵌入式以太网的语音通讯系统设计

吴平^{1,2} 曹晓琳^{1,2} 丁铁夫¹

¹(中国科学院长春光学精密机械与物理研究所, 长春 130021)

²(中国科学院研究生院, 北京 100039)

E-mail: pwwp@tom.com

摘要 设计了基于 ARM、DSP 和 uClinux 的嵌入式以太网来解决强背景噪声下的实时语音和指令数据通信问题, 开发了外部设备的驱动程序和网络应用程序, 并采用组播技术实现多个成员间的电话会议功能, 满足了语音实时传输要求。

关键词 嵌入式以太网 语音 ARM DSP uClinux

文章编号 1002-8331-(2006)16-0163-04 文献标识码 A 中图分类号 TP393

Design of Speech Communication System Based on Embedded Ethernet

Wu Ping^{1,2} Cao Xiaolin^{1,2} Ding Tiefu¹

¹(Changchun Institute of Optics, Fine Mechanics and Physics,
Chinese Academy of Sciences, Changchun 130021)

²(Graduate School of the Chinese Academy of Sciences, Beijing 100039)

Abstract: A embedded ethernet that base on ARM, DSP and uClinux is designed to resolve the real time speech and instrument datum communication in the high noise environment. Device driver and network application program is developed in uClinux embedded operating systems, telephone meeting among members is realized by the multicast technique, has met the real time demands of speech transmit.

Keywords: Embedded Ethernet, speech, ARM, DSP, uClinux

1 引言

在一个车载语音通讯系统中, 通话的背景噪声非常大, 需要借助电子设备才能解决成员间语音通话问题。原有的通话设备为模拟工作方式, 仅能传送语音, 不能传送数据, 且抗干扰能力差, 环境噪声对车内通话的影响非常严重, 可靠性低, 不便维修, 不具备连接新系统、新设备的能力。为此, 作为替代方案设计了一个基于 ARM 核处理器、DSP 数字信号处理器和 uClinux 操作系统的嵌入式以太网实现语音数据的采集、降噪和传输。

整个系统为基于 802.3 协议的嵌入式以太网, 网络拓扑结构为总线型连接方式, 如图 1 所示, 连接到网络总线的有语音终端(车内人员通话用)、数据终端(与车外数据通讯)、电台终端(与车外无线通话)、扩展终端等, 所有终端均以网络设备形式接入系统, 通过 TCP(UDP)/IP 协议进行语音和数据的传输。车载系统的各终端拥有不同的优先级和通话权限, 管理员终端(也是语音终端)拥有最高权限, 负责网络管理和对内对外的语音通话及数据收发控制。所有终端可接受网络提供的点对点的语音和数据服务, 也可以实现分组语音通信。各终端还可按需通过 RS232 与外部设备进行数据通讯。由于各终端的硬件结构基本相同, 故以下重点介绍语音终端的硬件和软件设计, 其它终端只需进行少量的硬件改动和软件设计即可实现。

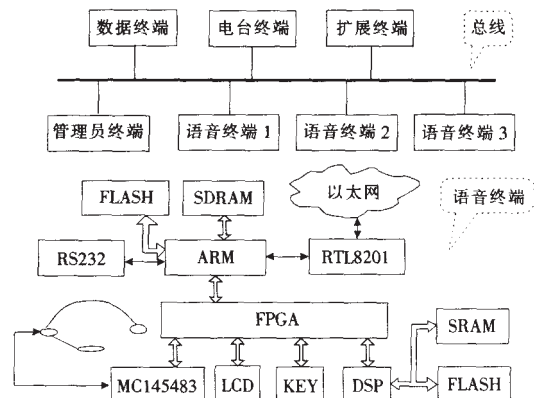


图1 系统网络拓扑结构和语音终端框图

2 语音终端硬件设计

终端采用 Samsung 公司的 ARM 核处理器 S3C4510B(下简称 ARM)为控制核心, 利用 ARM 内置的以太网控制器实现各终端的网络连接, 将 FPGA 内的语音模块、LCD 模块和 KEY 模块映射到 ARM 的外部 I/O Bank 区, 实现 ARM 与 DSP 的语音包交换以及 ARM 对 LCD 字符显示模块和键盘的控制。因为 FPGA 的三个模块不满足 Cache 的使用要求, 所以需将 ARM 的外部 I/O Bank 地址映射到非 Cache 空间, 在初始化时定义

基金项目: 中科院长春光学精密机械与物理研究所青年创新基金资助项目(编号: Q03S08Z)

作者简介: 吴平(1970-), 男, 中科院长春光学精密机械与物理研究所博士生, 研究方向为嵌入式系统、语音信号处理。曹晓琳(1978-), 女, 中科院长春光学精密机械与物理研究所博士生, 研究方向为语音信号处理。丁铁夫(1946-), 男, 研究员, 博士生导师, 研究方向为嵌入式系统、图像处理与计算机数据通讯。

ARM 的外部 IO Bank0 基址为 0x7fd0000。

车载系统要求语音传输频段为 300Hz~3 400Hz, 在本设计中语音编码采用 PCM 格式, PCM 编码抗干扰能力强、语音质量好、容易叠加、便于进行数字信号处理。对语音数据的 PCM 编解码工作由 MC145483 芯片完成, 语音采样频率为 8kHz, 16 位编码, 传输带宽为 128kB/s。本文规定 DSP 与 MC145483 间交换的语音称“语音数据”, DSP 与 ARM 间交换的语音称“语音包”, 各终端间传送的语音称“语音 IP 包”。

语音数字处理系统以 TMS320VC5402(下简称 DSP) 为核心, DSP 通过 FPGA 内 PCM 模块实现对语音数据的采集和回送, 对采集的 PCM 带噪语音数据 DSP 首先进行端点检测和语音增强, 然后将处理后的纯净语音经 FPGA 语音模块交给 ARM, ARM 将收到的语音包处理为语音 IP 包发送给局域网上相关的通话人员。DSP 接收 ARM 发送的语音包后无需进行处理直接交给 MC145483 进行播放。

FPGA(EP1K30) 用来实现 ARM、DSP 与各外设的接口, 提供终端需要的控制信号和数据交换通道。FPGA 内部逻辑电路结构如图 2 所示, 图中未连接的输入输出线直接接 EP1K30 的 I/O 引脚, 并与处理器、外设对应连接。ARM 和 DSP 的数据总线、地址总线、读写信号线、需要的中断线和片选线等都接入 FPGA。图中 ARM 模块由地址线、片选线和读写线经译码产生 LCD 模块、KEY 模块及与 DSP 进行语音包交换的控制信号。DSP 模块经译码产生与 ARM 交换语音包及与 PCM 模块交换语音数据的控制信号。PCM 模块为 MC145483 提供编解码所需

的时钟信号, 并完成 MC145483 和 DSP 之间的语音数据交换。语音模块实现 ARM 与 DSP 之间的语音包交换。LCD 模块实现 ARM 对 LCD 的显示控制。KEY 模块为键盘提供行列扫描信号和键盘消抖电路。

ARM 和 DSP 之间的语音交换采用打包的办法, 通过 FPGA 内的语音模块实现, 在语音模块中使用 FPGA 的两个内建 RAM 作为语音缓存区, 分别存放发送和接收的语音数据包, 定义一个语音包为 16ms、256 字节。ARM 向 DSP 发送语音包的过程是: ARM 首先发送一个/ARM_CLR 信号来复位 RAM 地址(置 0), 然后写入语音数据, 写一个字节数据 RAM 地址加 1, 写满 256 字节后发送/DSP_INT1 中断信号给 DSP, 并同时 RAM 地址复位, DSP 在收到中断信号后即读取 256 个字节的语音数据, 完成一次语音包的发送。DSP 向 ARM 发送语音包的过程与此类似。

3 系统软件设计

uClinux 是针对控制领域的嵌入式 Linux 操作系统, 适合不具备 MMU 的微处理器, 所以 uClinux 采用实存储器策略, 所有程序访问的地址都是实际的物理地址。本系统选用 uClinux 嵌入式操作系统经过移植运行在以 ARM 为核心的硬件平台上, 将 FPGA 内的语音模块、KEY 模块和 LCD 模块作为 uClinux 的外设开发相应的设备驱动程序, 由操作系统进行管理和控制, 并利用 uClinux 集成的 TCP/IP 协议栈进行网络编程, 实现语音和数据的实时、可靠传输。

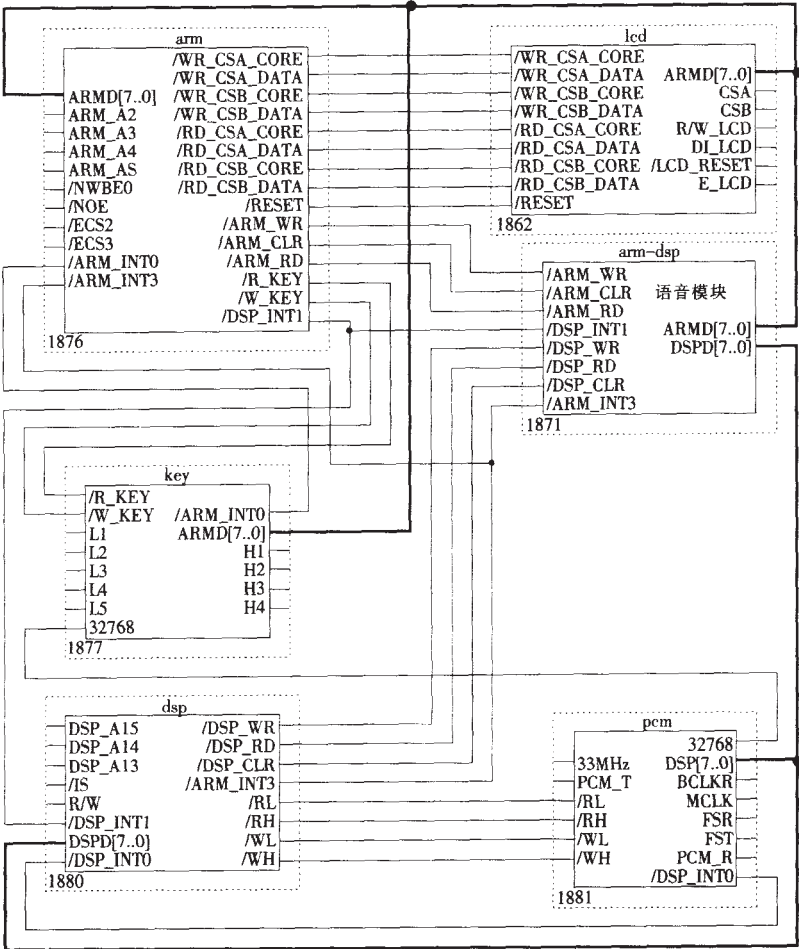


图 2 FPGA 内逻辑电路结构

3.1 设备驱动程序开发

uClinux 设备驱动程序可分为用户空间驱动程序和内核空间驱动程序, 由于 uClinux 嵌入式操作系统不具备 MMU, 用户程序可以直接访问物理地址, 对于不需要中断处理器的外部设备如本系统的 LCD 字符显示模块, 可以采用对其编写用户空间驱动程序的方法, 即可满足系统设计要求。相对内核空间驱动程序, 编写用户空间驱动程序可以与标准 C 库链接, 无需更多了解内核, 在实际应用中不必进行用户空间和内核空间的切换, 减少操作系统开销, 提高效率。

对于键盘、串口、语音模块等需要中断处理的设备要编写内核设备驱动程序, 内核设备驱动程序是操作系统内核和硬件设备之间的接口, 它为应用程序屏蔽了硬件的细节, 使得应用程序可以象操作普通文件一样对硬件设备进行操作。

以下简要介绍语音设备的字符型设备驱动程序, 操作系统将 FPGA 内的语音模块抽象为语音硬件设备, 定义该设备驱动程序的 file_operations 结构为:

```
struct file_operations dpsound_fops
{
    read: dpsound_read, /* 读取语音模块的语音包 */
    write: dpsound_write, /* 将语音包写入语音模块 */
    ioctl: dpsound_ioctl, /* 对语音设备进行控制操作 */
    open: dpsound_open, /* 打开设备并初始化 */
    release: dpsound_release, /* 关闭设备 */
};

void init_dpsound(void) /* 语音设备初始化代码 */
{
    /* 注册新设备: 254 为拟申请的主设备号, "dpsound" 为设备名字, &dpsound_fops 为 file_operations 数据结构指针 */
    register_chrdev(254, "dpsound", &dpsound_fops);
    /* 申请中断: 1 为拟申请的中断号, dpsound_read_service 是中断处理函数指针, SA_INTERRUPT 表示申请快速中断, "dpsound_read" 为设备名称 */
    request_irq(1, dpsound_read_service, SA_INTERRUPT, "dpsound_read", NULL);
}

void dpsound_read_service(int irq, void* dev_id, struct pt_regs* regs) /* 中断处理程序 */
{
    CLEAR_PEND_INT(1); /* 清中断标志 */
    read_ram_sound(); /* 读取语音模块内的语音包 */
    wake_up_interruptible(&read_queue); /* 唤醒读进程 */
}
```

系统外设使用内核函数 register_chrdev() 向操作系统注册新设备驱动程序, 通过 file_operations 结构向操作系统内核进行说明, 该结构的每一个成员都对应着一个系统调用, 当用户进程对设备文件进行操作时, 系统调用通过设备文件的主设备号找到相应的设备驱动程序, 然后读取这个数据结构的函数指针。编写设备驱动程序的主要工作就是编写子函数来填充 file_operations 的各个域。

uClinux 对中断的处理属于系统的核心部分, 如果外部设备与内核以中断方式进行数据交换, 就必须把该设备的驱动程序作为系统内核的一部分, 把相关的中断号路由到正确的中断处理程序中。设备驱动程序通过调用 request_irq() 函数向内核申请中断, 并将一个中断服务函数挂到相应的处理队列中。

3.2 网络应用程序

uClinux 嵌入式操作系统集成有 TCP/IP 协议栈, 本系统对非语音数据采用可靠的 TCP/IP 协议实现, 以保证可靠通讯。而语音数据传送最重要的是实时, 且语音自身有一定的容错能力, 偶而一个包丢失或传送错误, 并不会严重影响语音的保真度和对语义的理解, 特别对组播和广播情况, 只能使用 UDP/IP 协议才能够实现系统要求的语音分组和电话会议。

为使网络中的多个主机同时接受到相同的报文, 可以采用单播、广播和组播方式实现。如采用单播方式, 源主机必须多次发送相同的报文, 效率较低。采用广播方式, 则源主机会把报文传送到局域网的所有主机上, 不能对接收主机进行动态管理。而采用组播方式, 源主机可只发送一次报文就能到达每个需要接收该报文的主机。组播能使一个或多个组播源只把数据包发送给特定的组播组, 而只有加入该组播组的主机才能接收到数据包。IP 组播通信依赖于 IP 组播地址, 在 IPv4 中第四类 IP 地址即 D 类地址被保留用作组播地址, 范围从 224.0.0.0 到 239.255.255.255。

系统对语音通讯采用基于 UDP/IP 协议的多线程编程方法, 接收 IP 组播的成员集合构成一个群组, 每个群组拥有唯一的组播地址来指定发送语音报文的目的组。一个组播组的成员是动态的, 某成员可以随时加入或离开组播组, 不属于组播组的成员也可以向该组播组发送数据包。需指出的是, IP 组播地址仅用作目的地址, 不能出现在任何数据报的源地址字段中。

因为 UDP 不提供大块数据的分割功能, 为保证实时传输, 需要合理设计语音发送缓冲区的大小, 使得一个语音缓冲区内的数据可以通过一个 UDP 发送。每个以太网帧的数据段最大长度为 1500 字节, 除去 IP 头和 UDP 报头, 每个 UDP 可以承载的最大语音数据量为 1472 字节。本系统实际语音发送缓冲区设为 256 字节, 与 ARM 和 DSP 语音交换的数据包大小一致, 即 ARM 收到 DSP 一个语音包后即处理为 UDP 报文进行发送, 这样既保证实时传输, 又不使网络负载过大, 同时如果网络中偶而丢失一个语音 IP 包, 因其中语音信息较少, 对语音的质量影响也不大。

为使接收端实时连续播放语音, 采用预存储技术解决抖动问题, 设置一个预存储量, 语音 IP 包数目达到此数量时才进行播放。通过实验, 本系统在实际应用中接收循环缓存区设为 1024 字节, 预存储值为 1 包(256 字节), 即 DSP 在收到第 2 包语音后才开始将第 1 包及以后收到的语音送 MC145483 进行播放, 语音延时为 16ms, 这样处理后的语音听起来十分流畅。

4 语音信号处理

由 DSP 对采集到的 PCM 带噪语音进行处理, 算法设计包括端点检测和语音增强两部分。统计表明, 实际通话中有 50% 左右的静音期, 因此对采集到的语音数据进行端点检测, 在没有语音时停止语音 IP 包的发送, 减少网络负荷, 改善语音传送的实时性。在低信噪比情况下, 短时能量与过零率等参数都不能很好地区分语音段与非语音段, 因此采用倒谱系数来作为端点检测的参数, 即假定前几帧信号是背景噪声, 计算这些帧的倒谱系数, 并将其平均值作为背景噪声的倒谱系数的估计值, 接下来计算每帧信号的倒谱系数, 计算每帧信号的倒谱系数与噪声倒谱系数估计值的倒谱距离, 根据倒谱距离的轨迹采用门限判决的方法进行语音端点检测。

车载通讯系统的背景噪声主要有低频周期性噪声(发动机周期性运转)、脉冲噪声和宽带噪声,对低频周期性噪声进行IIR数字滤波即有明显抑制。对宽带噪声采用基于短时谱估计的谱相减法进行处理,该算法利用加性噪声与语音不相关的特点,在假设噪声是统计平稳的前提下,用无语音间隙测算得到的噪声频谱估计值取代有语音期间噪声的频谱,与含噪语音频谱相减,从而获得语音频谱的估计值。

假设纯净语音 $s(n)$ 被加性噪声 $d(n)$ 所污染,形成带噪语音 $y(n)$, 即 $y(n)=s(n)+d(n)$ 。通常需要加窗分帧处理,所以以上式可表示为:

$$y_m(n)=s_m(n)+d_m(n) \quad (1)$$

式中, m 为语音帧编号, n 为每帧语音中采样点的编号。

设 $y_m(n)$ 的傅里叶系数为 Y_k , $s_m(n)$ 的傅里叶系数为 S_k , $d_m(n)$ 的傅里叶系数为 N_k 。直接对 (1) 式进行傅里叶变换,并用无话时的 $|N_k|$ 统计平均 $\lambda_n(k)$ 作为噪声谱估计值,则有幅值谱相减法的经典形式:

$$|Y_k|=|S_k|+|N_k|=|Y_k|-E[|N_k|]=|Y_k|-\lambda_n(k) \quad (2)$$

图 3 为基于幅值谱相减法的实现框图,对输入带噪语音分

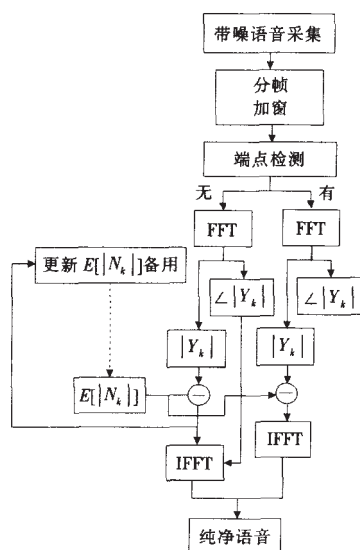


图 3 幅值谱相减法实现框图

帧加窗后,对有语音帧,提取出带噪语音相位,并用之前最新的无语音帧得到的噪声频谱幅度估计值替代此帧噪声的频谱幅度,从带噪语音频谱幅度中将其扣除作为此帧语音频谱幅度估计值。对语音频谱幅度估计值插入原带噪语音相位作为语音频谱的估计值,对其进行反傅里叶变换即可得到增强后的语音信号。

5 系统验证与结论

目前,由三个语音终端组成的样机已调试完成,样机终端上有两个进程运行,一个是数据传输进程,采用基于 TCP/IP 协议的服务器/客户端模式,管理员终端为服务器,其它为客户端,由管理员控制客户端是否加入语音通话组。另一个是基于 UDP/IP 协议的语音传输进程,该进程又有收语音和发语音两个线程,收语音线程可以接收任意数目语音终端发来的语音并对语音进行叠加处理。语音进程与数据进程通过 FIFO(有名管道)进行通信,语音进程受数据进程控制以确定该终端是否加入组播组,语音进程的运行优先级高于数据进程。样机系统在初始化时,为每个终端分配不同的 IP 地址和编号,管理员可直接利用终端编号进行相应的通话操作和控制。样机在实际工作中,语音传输非常流畅、无失真,语音传输延迟约几十 ms,语音和数据传输的实时性和可靠性达到系统的设计要求。本系统为 10M 嵌入式以太网,语音传输带宽为 128Kb/s,系统可满足 10 个以上语音终端的语音通话功能。

本文介绍了一种基于嵌入式以太网的语音通讯系统,相比于传统的因特网 IP 电话,本系统主要用于语音实时通讯,网络数据主要是语音,其它数据较少,而且限于在局域网范围内传送,网络流量和负载都较低,语音数据传送路径和时延相对固定。实践证明,采用嵌入式局域网的设计方案能够满足特殊环境下各成员间的实时语音通讯要求。(收稿日期:2005 年 8 月)

参考文献

- 1.张辉译.GNU/Linux 编程指南[M].第 2 版,清华大学出版社,2002
- 2.魏永明,骆刚,姜君译.Linux 设备驱动程序[M].中国电力出版社,2002
- 3.张雄伟等.现代语音处理技术及应用[M].机械工业出版社,2003
- 4.朱民雄等.计算机语音技术[M].北京航空航天大学出版社,2002
- 5.13.A P Dempster, N M Laird, D B Rubin. Maximum likelihood from incomplete data via the em algorithm[J]. Journal of the Royal Statistical Society, 1977; 39(1): 1-38
- 6.14.L R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition[C]. In: Proceedings of the IEEE, 1989; 77(2): 257-285
- 7.15. <http://www.sinobiometrics.com>
- 8.16. Benadbelkader C, Cutler R, Davis L. Motion-based Recognition of People in EigenGait Space[C]. In: Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, 2002
- 9.17. Foster J P, Nixon M S, Prugel-Bennet A. Automatic Gait Recognition Using Area Based Metrics[J]. Pattern Recognition Letters, 2003; 24: 2489
- 10.18. Lee L, Grimson W E L. Gait Analysis for Recognition and Classification[C]. In: Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, 2002
- 11.6. 晓雯等. 基于步态的人体身份检测与识别[J]. 计算机工程与应用, 2004; 40(5): 82-83
12. Jang-Hee Yoo, Mark S Nixon, Chris J Harris. Extraction and description of moving human body by periodic motion analysis. <http://www.bib.ecs.soton.ac.uk/data/7422/ps/CATA-O2-R29.ps>
13. Jang-Hee Yoo, Mark S Nixon, Chris J Harris. Extraction human gait signature by body segment properties. <http://www.bmva.ac.uk/meetings/meetings/02/6March02/soton2.pdf>
14. 王亮, 胡卫明, 谭铁牛. 人运动的视觉分析综述[J]. 计算机学报, 2002; 25(3)
15. 朱志刚等译. 数字图像处理[M]. 电子工业出版社, 1998
16. 章毓晋著. 图像分割[M]
17. Amit Kale. Algorithms for gait-based human identification from a monocular video sequence[D]. Doctor of Philosophy, 2003