

文章编号:1004-9037(2012)06-0646-06

一种基于TMS320C6713的语音信号频域盲分离算法

刘朝晨¹ 吴 强¹ 殷 超¹ 李清石¹ 许宏吉¹ 李玉军²

(1. 山东大学信息科学与工程学院, 济南, 250100; 2. 青岛海信数字多媒体技术国家重点实验室, 青岛, 266000)

摘要:提出了一种真实环境中语音信号频域盲分离算法的实现方案。基于TMS320C6713 DSK 平台,充分利用硬件资源,完成了语音盲分离处理算法的平台移植,并进行了编译环境、算法和程序3个级别上的软件优化,最终结果显示本系统具有较好的性能,在离线和在线情况下均取得了对真实环境中两说话人混合语音的有效分离。

关键词:频域盲分离;非平稳信号;Ping-Pong 缓存;软件优化

中图分类号:TN911.7

文献标识码:A

Blind Speech Separation in Frequency Domain Based on TMS320C6713 DSK

Liu Zhaochen¹, Wu Qiang¹, Yin Chao¹, Li Qingshi¹, Xu Hongji¹, Li Yujun²

(1. School of Information and Communication Engineering, Shandong University, Jinan, 250100, China;

2. Qingdao Hisense State Key Laboratory of Digital Multimedia Technology, Qingdao, 266000, China)

Abstract: An implementation for speech blind source separation of frequency-domain in a real environment is introduced. Based on the TMS320C6713 DSK platform, the hardware resources are used, and the blind speech separation on the DSP platform is implemented. Meanwhile, software optimizations on the levels of compiling environment, algorithm and program are conducted. Experimental results show good performance on two-speaker mixed speech separation in a real room on offline and online scenarios.

Key words: frequency-domain blind source separation; non-stationary signals; Ping-Pong buffering; software optimization

引 言

盲信号分离技术是一种在源信号、传输通道未知或所知先验知识很少的情况下,仅根据传感器获得的观测信号,利用信号处理算法实现对源信号进行有效估计的技术。由于语音信号自身的特性,语音盲分离成为盲分离领域中的一个难点与热点,它通常作为说话人定位、语音增强及语音识别等应用的前端处理,因而具有巨大的应用潜力和广阔的市场前景。

语音信号盲分离问题自20世纪80年代初提出以来,经过广大科研工作者20多年的努力,有了很

大的进步,出现了许多有价值 and 性能优秀的算法。但另一方面,这一研究领域仍然存在许多问题亟待进一步的研究和解决,其中一个重要的问题是目前的混合语音分离大多是建立在无噪环境中的瞬时混合^[1-2]情形下,只具有理论意义。真实环境下的语音信号盲分离是基于多通道卷积混合模型,针对卷积盲分离问题,目前主要存在两类算法:(1)时域分离算法,(2)频域分离算法。频域语音信号盲分离算法相对于时域盲源分离算法来说,因其较小的计算量和良好的分离性能而受到研究人员的关注。卷积型混合的语音分离由于模型本身的复杂性导致其计算复杂度远大于瞬时混合模型,当前绝大多

基金项目:国家重点基础研究发展计划(“九七三”计划)(2010CB735906)资助项目;中国博士后科学基金(2011M501092,2012M511508)资助项目;山东省博士后创新项目专项资金(201103003)资助项目;山东大学自主创新基金(2012ZD035,2011GN062)资助项目。

收稿日期:2012-06-26;**修订日期:**2012-10-08

数的算法均是在配置较高的PC 机上完成仿真与测试,在嵌入式平台上解决该问题的方案非常少,本文提出了一种语音信号频域盲分离算法的实现方案,基于TMS320C6713 DSK 平台设计了双通道语音采集电路,充分利用了有限的硬件资源,完成语音信号频域盲分离算法的移植,并进行了大量的软件优化,结果显示本系统具有较好的性能。

1 算法介绍

1.1 算法基本原理与实现过程

多通道卷积混合模型的语音盲分离问题可以表示如下

$$\mathbf{x}(t) = \sum_{\tau=0}^{P-1} \mathbf{A}(\tau) s(t - \tau) + \mathbf{n}(t) \tag{1}$$

式中: P 是滤波器的长度,麦克风的观测信号向量 $\mathbf{x}(t)=[x_1(t),\cdots,x_{d_x}(t)]^T$, d_x 是麦克风的个数,源信号向量 $\mathbf{s}(t)=[s_1(t),\cdots,s_{d_s}(t)]^T$, d_s 是源信号的个数, $\mathbf{n}(t)$ 为加性噪声向量,混合通道 \mathbf{A} 具有 $d_x d_s P$ 个未知的参数。解混模型如下

$$\mathbf{u}(t) = \sum_{\tau=0}^{Q-1} \mathbf{W}(\tau) \mathbf{x}(t - \tau) + \mathbf{n}(t) \tag{2}$$

目标是估计出混合逆系统 \mathbf{W} ,从而就可以得到对源信号的估计 $\mathbf{u}(t)$ 。本系统中使用的算法^[3]利用了语音信号的非平稳性,通过改变观测信号在多个时间点上的互相关性来提供对未知的混合通道足够的限制,解决了语音盲分离这一问题,主要思路如下:将时域中的卷积混合模型转换到频域,在每一个频点上进行瞬时盲分离,同时解决每个频点存在的排序不确定性问题,从而避免了每个频点上得到的解混矩阵的行向量的顺序不一致,最后从频域再转换到时域得到解混矩阵 \mathbf{W} ,进行多通道有限长单位冲激响应滤波器(Finite impulse response, FIR)滤波得到估计的源信号 $\mathbf{u}(t)$ 。

1.2 频域盲分离算法的硬件实现分析

卷积混合模型的语音盲分离问题通常有两种解决方法:时域方法^[4-6]与频域方法^[7-10]。本系统采用的盲分离算法是一种频域算法,选择其在硬件实现的原因为:

- (1)盲分离的迭代算法基于二阶统计量,相比其他大多数的频域算法多采用高阶统计量(如联合近似对角化(Joint approximative diagonalization of eigenmatrix, JADE)算法是利用四阶累积量)的算法,在很大程度上简化了计算量;
- (2)该算法避免了复数矩阵求逆、复数矩阵特

征值分解等实现困难计算量大的数学运算;

(3)由于语音信号本身的复杂性,目前大多数频域方法仍是针对人工混合的语音信号进行盲分离,本算法是对真实有噪声环境下的语音信号进行盲分离,具有更重要的现实意义;

(4)该算法的数学运算主要基于大规模的矩阵运算及迭代操作,这样允许更加有效地在 TI 的 DSP 处理器上实现和优化。

该算法同时也存在着局限:本算法中提出的排序不确定性解决方法巧妙而简单,但与频域方法中解决排序问题使用的波达方向(Direction of arrival, DOA)估计及相邻频点间能量包络的相关性进行排序相比,在简化程序计算量的同时也带来了分离性能的下降。

2 算法的实现及其优化

2.1 系统概述

语音盲分离系统在充分利用 TMS320C6713 DSK 原有硬件电路的前提下,向前向后拓展了部分硬件电路,实现了双通道语音信号麦克风采集、预放大、语音盲分离、语音回放和存储等。系统逻辑如图 1 所示。

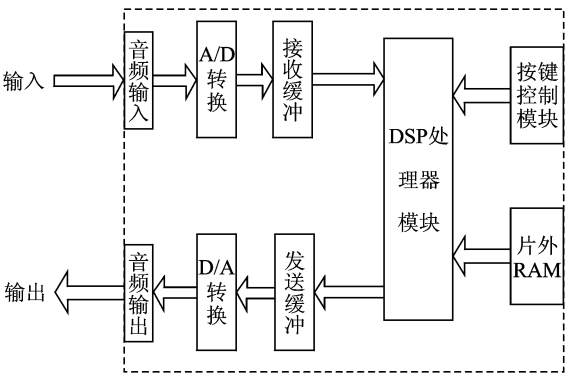


图1 系统逻辑框图

2.2 硬件系统

系统硬件电路主要包括如下模块:

- (1)麦克风接入电路:两个单声道麦克风采集得到的语音信号分别作为线性输入音频输入接口的左右声道信号输入系统,实现了双通道语音信号的同时采集。
- (2)信号预放大电路:隔离直流信号,完成对双通道语音信号电平增益的有效同等放大。
- (3)TMS320C6713 DSK:作为系统核心部分,完成 A/D 转换、语音盲分离算法处理、D/A 转换、双通道语音信号模拟输出等。硬件电路如图 2 所示。

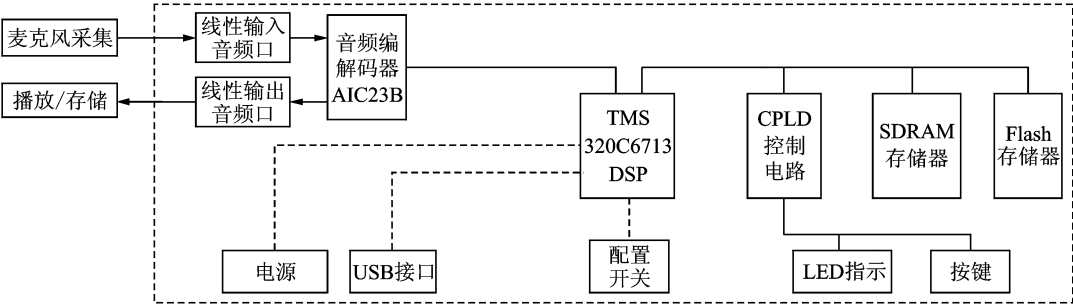


图 2 硬件电路

2.3 软件系统

系统设计了离线和在线两种情况真实环境下两说话人的混合语音盲分离。离线情况时,手动使用开关控制语音信号的采集、盲分离算法处理、播放/存储的启动与停止,可以获得良好的盲分离性能;在线情况时,在假设说话人位置不动的前提下,系统自动对前端输入缓冲区的语音数据进行盲分离算法处理,分离之后的语音信号连续输出。本系统的软件系统开发基于 DSP/BIOS 嵌入式实时内核,它是编译器 CCS 提供的一个尺寸可裁剪的工具,有效解决多任务之间的实时调度。离线、在线情况下软件系统的主流程图如图 3,4 所示。

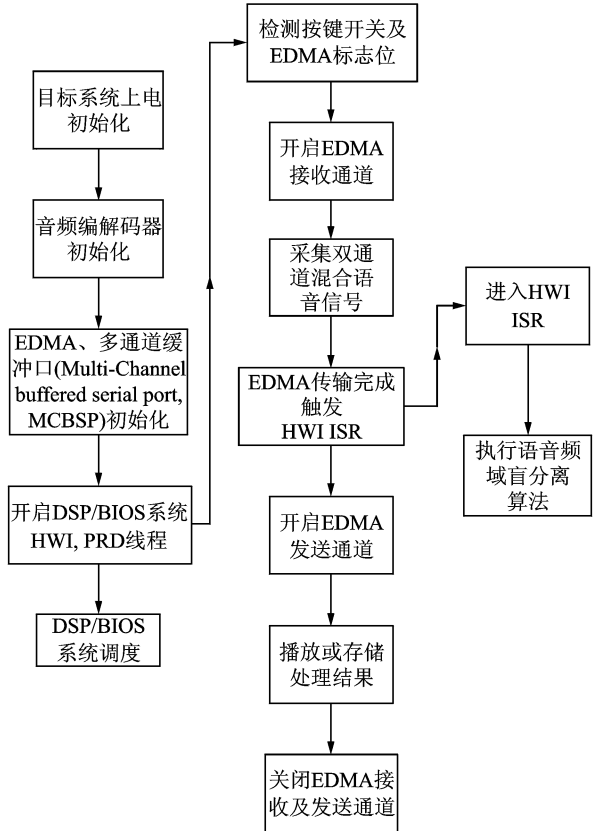


图 3 离线情况下软件系统的主流程图

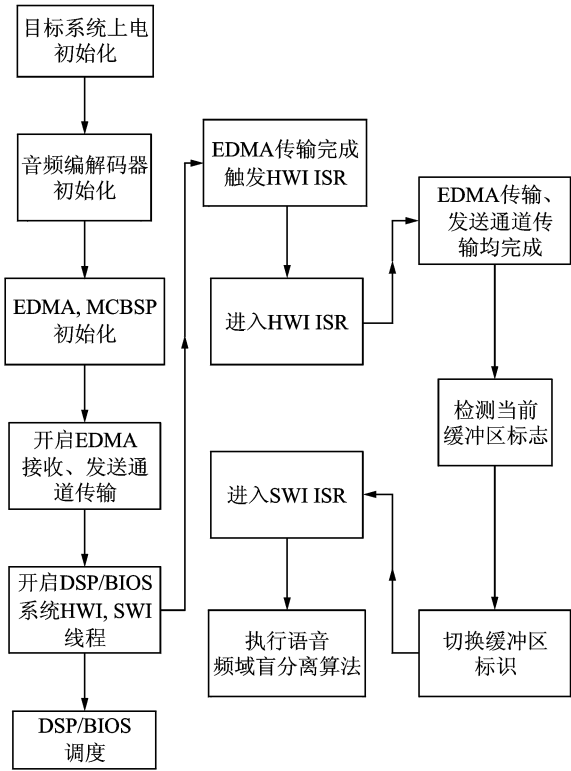


图 4 在线情况下软件系统的主流程图

2.4 算法在 DSP 上的实现难点与优化方案

2.4.1 实现难点分析

(1)该算法属于基于浮点操作语音信号数据处理范畴,虽然 TMS320C6713 是一款浮点型 DSP 处理器,但在本系统应用背景下:存储空间有限;处理器主频有限;算法基于二阶统计量要求采集足够多的语音信号进行批处理,有效实现语音盲分离仍是一项值得挑战的工作。

(2)算法所需进行的数学运算复杂、所操作的数据量较大,其中涉及三角函数运算、浮点数据与整型数据的相互转换、时频域运算的相互转换、复数矩阵求内积及多维(二维、三维、四维)矩阵的乘法运算等,要求确定合理的存储策略及算法的数据结构。

(3)在处理器性能有限及可利用的存储空间有限的情况下,系统实现的时间复杂度与空间复杂度是互补的,从代码级考虑功能实现,需要采取必要的优化措施,反复调试,以达到在空间复杂度满足的条件下尽可能降低时间复杂度。

2.4.2 优化方案选择

搭建最小系统并完成语音信号频域盲分离算法的平台移植,为了进一步提高系统的性能,实现系统的实时运行,提出了一种基于三阶段的优化方案,分别从编译环境级优化、算法级优化及程序级优化三方面对算法进行优化。

2.4.2.1 编译环境级优化

(1)CCS 编译器设置

TI C6000 的 CCS 编译器可以通过对代码进行各种级别的优化,以减小可执行目标文件尺寸提高代码的执行速度。编译器所做的优化包括进行软件流水,进行循环优化,删除全局共有的子表达式,删除全局未使用的赋值语句,把循环中的对数组的引用转变为递增的指针形式进行循环展开。

(2)存储器分配策略

TMS320C6713 片内采用基于二级缓存的存储架构,同时 DSK 在片外扩展 16 MB 的 SDRAM。由存储器的存储技术可知存储器的层次结构服从金字塔型,更快更小同时成本更高的存储器在更高层。利用基于 SRAM 的 L_1 , L_2 高速 Cache 存储器特别重要,这是因为从高速 Cache 存取数据的性能比从片外 SDRAM 存取数据的程序运行速度快过一个数量级。但是在算法的代码中,特别是数据所占空间远远大于片内存储空间,不可能将其全部配置于片内 RAM,因此需要合理地配置片内与片外存储空间,将程序所使用的代码和数据有效地分配在存储空间中,最大限度地发挥 DSP 的处理性能。

由于算法的数据结构复杂多样,数据量大,经过测试,将段 .bss,.far,.data 段分配到片外 RAM,其余段 .text,.stack,.cinit,.switch 及与 DSP/BIOS 系统相关的段均分配在片内 RAM。同时考虑到程序的空间复杂度,将堆分配在片外 RAM,有选择地指定时间局限性不高的数据结构动态的创建与释放。此外将栈的大小设置为最大值 0Xffff,以最大限度地提高数据的访问速度。

(3)使用 TI 提供的优化函数库

由于 TI 提供的函数库是已经高度优化后的运算函数,所以只需在链接时调用库中相关的链接目标文件(.o)。经测试,使用这类库函数对算法进行优化可以大大提升算法的性能,为算法的实现带来极大的便利。在本算法中主要用到以下两个函数库:

①TMS320C67X FastRTS Library^[11]

这是 TI 提供的适用于 C67X 平台的针对浮点型数据数学函数运行时的快速支持库,算法中用到的函数及用途如表 1 所示。

②TMS320C67X DSPLib^[12]

这是 TI 提供的针对于 C67X 平台的常用的数字信号处理运算库,主要包括以下几个类别:自适应滤波、基本数学运算、快速傅里叶变换、矩阵运算、相关、卷积及其他运算。经过测试,本系统算法中用到的函数及用途如表 2 所示。

(4)保证实时连续性所采取的处理措施

①采用乒乓缓存。若要达到连续实时处理,那么当接收缓冲区存满之后,DSP 必须在下一个新的音频数据到来之前处理完整个接收缓冲区,否则新数据就将覆盖在接收缓冲区。为了解决这一问题,引入了 Ping-Pong 缓存传输机制,即在发送端和接收端各使用两个缓冲区,一个 Ping 缓冲区和一个

表 1 TMS320C67X FastRTS Library 的运用

| 函数名 | divsp() | sqrtp() | atansp() | cosp() | sinsp() | recips() |
|-----|-------------------------------------|-----------|----------|--------|---------|---------------------------------|
| 功能 | 浮点除 | 求平方根 | 求反正切 | 求余弦 | 求正弦 | 求倒数 |
| 使用 | 计算信号向量标准差; 生成旋转因子; 调整 IFFT 幅度 | 计算信号向量标准差 | 生成旋转因子 | 生成旋转因子 | 生成旋转因子 | 计算互能量谱矩阵; 信号归一化; 计算 m 值 |

表 2 TMS320C67X DSPLib 的运用

| 函数名 | DSPF_sp_vecsum_sq() | DSPF_sp_cfft2_dit() | DSPF_sp_icfft2_dif() |
|-----|---------------------|-----------------------------|----------------------|
| 功能 | 计算向量均方和 | 快速傅里叶变换核心运算 | 逆快速傅里叶变换核心运算 |
| 使用 | 计算信号向量标准差 | 短时傅里叶变换; 迭代运算解混矩阵时域变换到频域 | 迭代运算解混矩阵频域变换到时域 |

Pong 缓冲区。这种传输机制的基本思想是:当系统在填充 Ping 缓冲区时,CPU 处理 Pong 缓冲区中的数据,确保当前的数据传输不会对 Pong 缓冲区进行重写;当 Ping 缓冲区填充完毕,CPU 转向处理 Ping 缓冲区中的数据,同时新的数据传输对 Pong 缓冲区进行填充,这一过程以此反复。因此只要再通过其他的优化方案保证 CPU 处理数据的时间少于采集满一个缓冲区的时间就可以保证系统运行的连续性。

②使用 EDMA 通道链接传输技术。TI 提供的片上外设增强型直接存储器访问(Enhanced direct memory access, EDMA)允许使数据的传输独立于 CPU 进行,CPU 能够更专注于软件的逻辑控制与 DSP 算法处理。在使用乒乓缓存传输技术时,EDMA 控制器需要为发送端和接收端各维护两张传输参数 RAM 表。其中一张参数 RAM 表用于 Ping 缓冲区读写,另一张表用于 Pong 缓冲区读写,这存在一个问题:当缓冲区填充完毕后,EDMA 控制器会产生中断,中断处理必须在下一次采样到来之前为下一个缓冲区重新分配传输参数 RAM 表,这显然对时间上要求较高。因此通过使用 EDMA 通道链接传输技术来解决这一问题,一旦缓冲区填充完毕时,EDMA 控制器将自动装载另一张参数 RAM 表开始下一个缓冲区的填充。

2.4.2.2 算法级优化

(1)在线场景下语音盲分离的优化。在线情况下,为了达到系统实时连续地运行,必须创建大小合适的缓冲区。如果缓冲区长度太小,允许 CPU 对另一缓冲区中的数据处理的时间就越短,同时也会由于数据量太少而不会获得良好的分离效果;如果缓冲区太长,虽然允许 CPU 对另一缓冲区中的数据处理的时间变长了,但也会使 CPU 由于需要处理的数据量过大而造成计算复杂度加大,同时也会造成输出的严重延迟。

在语音源位置固定情况下,短时间内解混滤波器 W 可以看作是不变的。利用这一特点,对长度为 $BUFSIZE = 128\ 000$ 的缓冲区,取 $BUFSIZE/2$ 大小进行算法处理首先获得解混滤波器 W ,再对缓冲区的全部数据进行 FIR 滤波最终获得估计的源信号。经此优化后,时间复杂度可节省 $BUFSIZE/2$ 次浮点乘法操作。

(2)多通道 FIR 滤波优化。估计的源信号是由观测信号经解混矩阵 W 进行 FIR 滤波得到的, W 包括 T 个数据单元(本系统中为 2×2 的实数矩阵),经仿真发现 W 中的最后 $(T-Q)$ 个数据单元

元素值为 0 或近似为 0(数量级在 10^{-16} 或 10^{-17}),故只需要 W 中的前 Q 个数据单元即可。经仿真验证了该想法的有效性,时间复杂度可节省 $(T-Q) \times BUFSIZE$ 次浮点乘法操作。

2.4.2.3 程序级优化

虽然 CCS 已经进行了 -O2 级别的优化,但是没有任何编译器能用一个好的算法或数据结构来替代低效率的算法或数据结构,因此在对程序的优化上仍有很多方法来协助编译器生成更高效的代码,采取的措施包括减少连续的函数调用,循环展开除了大量使用以上的两种程序级优化手段,在程序中也采用了其他一些更为普遍的优化措施,包括消除不必要的片外 RAM 的引用,引入临时变量来保存结果、使用 restrict 关键字,避免存取变量别名使用问题及将数组代码的查找、读取等操作形式转换为指针代码,这样编译器就能产生高效的程序。

3 性能测试与优化结果

为了验证算法的有效性,首先在 PC 上对算法进行仿真,房间的混音时间为 500 ms 左右,以 12 kHz 采样率采集两组语音信号,观测信号的信噪比为 11 dB。仿真结果如图 5 所示,其中图 5(a-b)是观测信号,图 5(c-d)是估计的源信号,作为对比图 5(e-f)是真实的源信号。通过比较估计信号和真实源信号的波形,观测信号已成功实现盲分离,

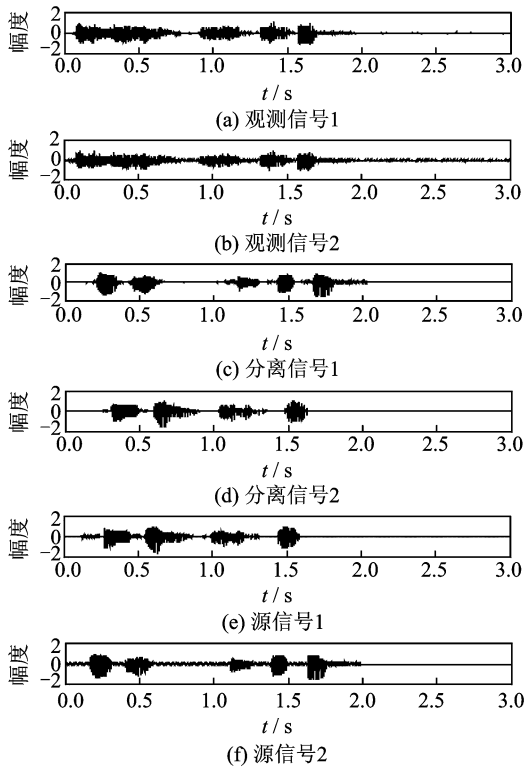


图 5 算法仿真结果

说明算法在真实环境中是适用的。

通过 CCS 提供的内存读取工具,将在目标系统运行得到的解混滤波器 W 数据结果与 PC 平台下 Matlab 环境算法运行得到的 W 相比较,计算精度在 10^{-4} 之上均保持一致,同时人在听觉上的主观感受也验证了算法的性能。

算法在目标平台完成移植后,采用 2.4.2 节所介绍的方法对算法的相关模块进行各种优化。在离线情况下对长度 $\text{BUFFSIZE}=64\ 000$ 的缓冲区进行语音盲分离,选取 3 个关键部分的代码,对优化前后的运算性能进行了对比,如表 3 所示。表 4 展示了离线情况和在线情况下系统的整体优化效果。综合表 3 和表 4,可以看出优化之后算法的性能明显提升。

表 3 局部优化结果

| 结果分析 | FIR 滤波 | 计算信号 能量 | 512 点 FFT |
|--------|-------------|------------|--------------|
| 优化之前 | 381 299 191 | 4 275 278 | 1 583 280 |
| 优化之后 | 226 250 207 | 1 527 400 | 563 456 |
| 性能提升/% | 68.5 | 179.9 | 190.0 |

表 4 系统整体优化结果

| 结果分析 | 离线情况 | 在线情况 |
|----------|-------|-------|
| 分离语音时长/s | 4 | 8 |
| 优化之前/s | 21.8 | 16.0 |
| 优化之后/s | 7.3 | 7.6 |
| 性能提升/% | 198.6 | 110.5 |

4 结束语

基于 TI TMS320C6713 DSK 平台,通过设计前端双声道麦克风采集电路,搭建硬件平台,并在该平台上完成语音信号频域盲分离算法的移植与优化,最终实现了真实环境下对两说话人的混合语音离线与在线两种情况下的有效分离。本系统的创新点在于:(1)充分利用了有限的硬件资源,完成真实环境下语音的盲分离;(2)合理使用算法的数据结构,同时结合芯片及编译器特点运用编译器级、算法级和程序级等多种优化方案,使优化后的语音盲分离算法具有良好的性能。从市场需求来看,语音盲分离技术通常作为说话人定位、语音增强及语音识别等应用的前端处理,本系统作为对该领域的初步探索,具有很大的应用前景及市场潜力。

参考文献:

[1] Choi S, Cichocki A. Adaptive blind separation of speech signals: cocktail party problem[C]//International Conference on Speech Processing. Seoul, Korea: ICSP,1997:617-622.

[2] O’Grady P D, Pearlmutter B A, Richard S T. Survey of sparse and non-sparse methods in source separation[J]. International Journal of Imaging Systems and Technology, 2005,15(1):18-33.

[3] Lucas P, Clay S. Convolutional blind separation of non-stationary sources [J]. IEEE Transactions on Speech and Audio Processing, 2000,8(3):320-327.

[4] Amari S. Natural gradient works efficiently in learning[J]. Neural Computation, 1998,10(2):251-276.

[5] Douglas S C, Sawada H, Makino S. A spatio-temporal fastICA algorithm for separating convolutional mixtures[C]//IEEE International Conference on Acoustics, Speech, and Signal Processing. Dallas, TX, USA: IEEE,2005:165-168.

[6] Zhang L Q, Cichocki A, Amari S. Multichannel blind deconvolution of nonminimum phase systems using filter decomposition[J]. IEEE Transaction on Signal Processing, 2004,52(5):1430-1442.

[7] Nion D, Mokios K N, Sidiropoulos N D, et al. Batch and adaptive parafac-based[J]. IEEE Transactions on Blind Audio, Speech, and Language Processing, 2010,18(6):1193-1207.

[8] Sawada H, Mukai R, Araki S, et al. A robust and precise method for solving the permutation problem of frequency-domain blind source separation [J]. IEEE Trans on Speech and Audio Processing, 2004, 12(5):530-538.

[9] Murata N, Ikeda S, Ziehe A. An approach to blind source separation based on temporal structure of speech signals[J]. Neurocomputing, 2000,41(1/4): 1-24.

[10] Parra L C, Alvino C V. Geometric source separation: merging convolutional source separation with geometric beamforming [J]. IEEE Transactions on Speech and Audio Processing, 2002,10(6):352-362.

[11] Texas Instruments Incorporated. Signal processing examples using the TMS320C67X digital signal processing library[EB/OL]. 2002-10-06. <http://www.ti.com>.

作者简介:刘朝晨(1988-),男,硕士研究生,研究方向:盲信号理论与应用,E-mail:sdulzc@163.com;吴强(1978-),男,讲师,研究方向:仿脑计算、语音信号处理、生物医学工程、机器学习;殷超(1988-),男,硕士研究生,研究方向:网络多媒体技术;李清石(1987-),男,硕士研究生,研究方向:多媒体信息处理;许宏吉(1976-),男,副教授,研究方向:空时信号处理、智能天线;李玉军(1973-),男,博士后,网络多媒体技术。