文章编号:1004-9037(2012)06-0703-07

分布式语音信号分离系统

杨志智 唐显锭 蔡 瑾 冯 辉

(复旦大学电子工程系,上海,200433)

摘要:传感器网络节点以OMAP3530处理器为核心,各主节点独立运行语音信号盲分离算法,通过节点间的自主协作和数据融合,分离出会场说话人的语音信号。语音分离算法为实时在线算法,利用声源信号间的相互独立性,在频域对混合信号进行分离。整个系统充分利用ARM核和DSP核的强大处理能力及传感器网络系统优势,经过合理的程序优化,有效地实现了多说话人场景下的语音分离。

关键词:无线传感器网络;分布式处理;实时盲信号分离;语音信号处理

中图分类号:TP391 **文献标识码**:A

Distributed Speech Signal Separation System

Yang Zhizhi, *Tang Xianding*, *Cai Jin*, *Feng Hui*

(Department of Electronics Engineering, Fudan University, Shanghai, 200433, China)

Abstract: Sensor network nodes use OMAP3530 for data acquisition, transmission and processing. Every main node runs blind speech signal separation algorithm individually. In addition to the source separation, proper collaboration and data fusion are introduced, and the system separates the speech signals of present persons in a meeting room. The blind source separation algorithm is real-time and on-line. It exploits the independencies of source signals, and executes the separation of mixed signals in frequency domain. The system makes the best use of the ARM core and DSP core's processing ability and the advantages of sensor network, thus achieving the multi-speaker speech signal separation efficiently after proper program optimization.

Key words: wireless sensor network; distributed processing; on-line blind source separation; speech signal processing

引 言

语音信号分离问题源于著名的"鸡尾酒会效应",即在复杂的混合声音中,人类能有效地选择并跟踪某一说话人的声音。近年来由于实际应用的需求使得语音信号分离成为了语音信号处理、移动通信和神经网络等领域的研究热点。

目前,语音信号分离的研究主要有计算声场景分析[1]和盲信号分离[2-3]两类方法。其中计算声场景分析方法的研究还处于初步阶段,且现有算法计算量普遍较大,很难实际应用,在多人同时说话情况下的分离效果通常比盲信号分离方法差[4]。盲信号分离是一种仅利用观测到的混合信号来估计源信号的方法,它是以独立成分分析(Independent component analysis,ICA)为理论基础,不要求有关源信号本身以及信号传输通道的知识。语音信号盲分离技术以其较好的性能在机器人语音识别、改进语音通信质量、提高语音可懂度以及信息抽取等方面发挥重要的作用。

独立成分分析以及相应发展起来的一些方法都是基于瞬时混合模型的,但是在实际的语音环境中,声音在传播过程中经过障碍物的反射和衍射,从麦克风或传感器接收到的信号不再是简单的瞬时混合信号,而是源信号在尺度和时延上经过缩放的卷积信号。对于卷积混合信号的盲分离,可以简单地分为时域和频域两种研究方法。时域方法往往需要巨大的计算量,收敛效果也不是很好,因此很

收稿日期:2012-06-16;**修订日期**:2012-09-29

难进行实际应用。频域算法运算时间短且易于在实际中实现，可以利用比较成熟的瞬时盲分离算法，但频域方法存在普遍难以解决并影响分离效果的问题，即幅值不确定性和次序不确定性。对于幅值不确定性，大部分频域算法都是采用信号归一化。而次序不确定性相对较难解决，目前已经提出了较多的算法解决该问题：利用源信号的平滑约束[5]，平均相邻频点的分离矩阵、限制时域滤波器长度、和考虑相邻分离矩阵的一致性；到达角方向估计，通过分析分离矩阵的方向性，可以估计出源信号的方向，从而对准次序[2]；将ICA方法推广至向量形式[4]。

语音信号盲分离算法大多基于麦克风阵列[2]，因为麦克风阵列能较好地估计出源信号的到达角方向。再结合语音信号频谱的平滑特性和谐波特性，能很好地解决频域算法中的次序不确定性问题。但基于麦克风阵列算法对于到达角相近的声源分离效果较差，且声源需位于麦克风阵列附近。为了弥补麦克风阵列空间中有效范围的不足和提升大房间或大会场中语音分离的效果，提出了分布式麦克风阵列[6]。然而分布式麦克风阵列在实际应用中成本较高，且系统搭建困难，灵活性较差。因而分布式麦克风网络[7]成为了前几年的研究热点，它具有易搭建、高容错、易扩展、低功耗等优势。分布式麦克风网络场景下可以采用一些集中式的信号处理算法分离语音，但该方法需要较大的网络传输开销，特别是当网络扩张时，网络开销急剧上升。

本文提出了一种基于无线传感器网络的分布式语音分离算法，并在OMAP3530平台上予以实现。该算法采用频域瞬时盲分离方法分离混叠语音数据并提取分离语音特征，各传感器节点通过交互分离语音特征完成全局信息融合。信息的分布式处理可以将复杂算法分摊到多个处理器节点处理，减小算法对硬件系统的需求。分布式网络中传输的数据是经过处理后的数据，网络传输负载大大减小。该分布式算法处理特别适用于源信号在空间中稀疏分布的情况，在这种场景下，传感器节点数一般多于声源数，因而在声源附近通常会存在一些采集节点，这些节点能得到该声源相对较高质量的采样。实验结果表明，该算法能有效地实现大房间或会场中的语音信号分离。

1 算法详细介绍

系统实现目标：(1)分离大型会场中一定数目说话人语音；(2)实时播放指定说话人语音。本系统采用“实时独立向量分析算法”[8]分离混叠语音信号，并利用基于分离信号Mel倒谱系数[9]的语音信号有效性评估算法确认或融合有效输出信号。

1.1 实时独立向量分析

在实际的语音环境中，麦克风或传感器接收到的信号是源信号在尺度和时延上经过缩放的卷积信号，系统中源信号与观测信号之间的关系为

$$x_i(t)=\sum_{j=1}^{L}\sum_{\tau=0}^{T-1}h_{ij}(\tau)s_j(t-\tau) \tag{1}$$

式中：$x_i(t)$为t时刻第i个观测信号；$h_{ij}(\tau)$为时域第j个源至第i个观测的传递函数；卷积时长为T；$s_j(t)$为t时刻第j个源信号；L为源信号数目。通过短时傅里叶变换，可以将时域信号$x_i(t)$变换到频域信号$x_i^{(k)}[n]$

$$x_i^{(k)}[n]=\sum_{t=0}^{K-1}\omega(t)x_i(nJ+t)\mathrm{e}^{-\mathrm{j}\omega_k t} \tag{2}$$

式中：$\omega_k=2\pi(k-1)/K$，$k=1,2,\cdots,K$；J为帧移长度；$\omega(t)$为窗函数。如果窗函数长度比混合滤波器长度长，则时域卷积可以转换为频域相乘，表达式为

$$x_i^{(k)}[n]\approx\sum_{j=1}^{L}h_{ij}^{(k)}\,s_j^{(k)}[n] \tag{3}$$

如果分离矩阵存在，则第i个分离信号由式(4)给出

$$y_i^{(k)}[n]=\sum_{j=1}^{M}g_{ij}^{(k)}\,x_j^{(k)}[n]\approx s_i^{(k)}[n] \tag{4}$$

为了求取分离矩阵，参照实时独立向量分析(Independent vector analysis, IVA)方法，使用Kullback-Leibler散度来衡量两个函数间的独立性：(1)联合概率密度函数$p(y_1,\cdots,y_L)$；(2)非线性函数，单个源信号的概率密度分布函数的乘积$\prod_{i=1}^{L}q(y_i)$。表达式如下

$$C=KL\Big(p(\hat{s}_1,\cdots,\hat{s}_L)\,\|\prod_{i=1}^{L}q(\hat{\boldsymbol{s}}_i)\Big)=\text{const.}-\sum_{k=1}^{K}\log|\det\boldsymbol{G}^{(k)}|-\sum_{i=1}^{L}E\log q(\hat{\boldsymbol{s}}_i) \tag{5}$$

式(5)对$g_{ij}^{(k)}$求导并利用自然梯度法即可得到分离矩阵梯度

$$\Delta g_{ij}^{(k)}=\sum_{l=1}^{L}(I_{il}-E[\varphi^{(k)}(y_i^{(1)},\cdots,y_i^{(K)})y_i^{*(k)}])g_{lj}^{(k)} \tag{6}$$

其中当$i=l$时，I_{il}等于1，否则等于0。非线性函数$\varphi^{(k)}(\cdot)$形式如下

$$\varphi^{(k)}(y_i^{(1)},\cdots,y_i^{(K)})=\frac{y_i^{(k)}}{\sqrt{\sum_{k=1}^{K}|y_i^{(k)}|^2}} \tag{7}$$

为了使该算法应用于在线情况，式(6)中求解期望部分需要做一定的近似。令

$$R_{ij}^{(k)}[n]=\varphi^{(k)}(y_i^{(1)}[n],\cdots,y_i^{(K)}[n])y_l^{*(k)}[n] \tag{8}$$

此处假设 $R_{ij}^{(k)}$ 只与当前帧有关，因而直接去掉式(6)中的期望计算。在线自然梯度为

$$\Delta g_{ij}^{(k)}[n]=\sum_{l=1}^{L}(I_{il}-R_{ij}^{(k)}[n])g_{lj}^{(k)}[n] \tag{9}$$

由于当源信号突然变小或消失时，式(8)所得的值会突然减小，从而使得式(9)所得的值突然增加，从而在更新过程中使得分离滤波器发散。为了缓解这一情况，引入不完全约束，对式(9)做一定的修正

$$\Delta g_{ij}^{(k)}[n]=\sum_{l=1}^{L}(\Lambda_{il}^{(k)}[n]-R_{ij}^{(k)}[n])g_{lj}^{(k)}[n] \tag{10}$$

式中 $\Lambda^{(k)}$ 为对角阵，对角元素等于 $R^{(k)}$。

由于在线算法中输入数据一般没有经过白化，所以该算法中很难使用Newton 梯度法。为了提升该算法的收敛性，引入归一化的自然梯度法，滤波器系数迭代算法修正为

$$g_{ij}^{(k)}[n+1]=g_{ij}^{(k)}[n]+\eta\sqrt{\xi^{-1(k)}[n]}\Delta g_{ij}^{(k)}[n] \tag{11}$$

式中：$\xi^{-1(k)}[n]$ 为归一化因子。

$$\xi^{(k)}[n]=\beta\xi^{(k)}[n-1]+(1-\beta)\sum_{i=0}^{L}|x_i^{(k)}[n]|^2/L \tag{12}$$

式中 β 为平滑因子，一般取 0.5。

为了解决频域盲信号分离中的幅值不确定性，使用最小扭曲原则，即让输出信号乘以幅度调整因子 $\mathrm{diag}(G^{-1(k)})$。

IVA 在实现语音频域盲分离过程中，无需估计语音信号的到达角，因而可以应用于分布式传感器网络中，因为在分布式传感器网络中无法估计语音信号的到达角；相比非负矩阵分解方法和计算声场景分析的语音分离方法，实时IVA 计算复杂度远远低于这些方法，能方便地应用于实际DSP 处理中，完成实时语音分离处理，而非一般的批处理；相比基于传统频域ICA 分离算法，IVA 方法更具鲁棒性，能适应更大时延和混响的场景，分离性能也相对较好。

在实时IVA 算法中，算法的收敛需要一定时间，即随着时间的推移，语音信号分离的性能会呈现逐渐提升的情况；语音分离算法并非完全消除其他干扰语音而得到纯净语音，而是在一定程度内提升语音信号的信号干扰比，即对干扰语音进行一定限度的抑制。

1.2　语音信号有效性评估

一般在实际应用中无法预先知道源信号的个数，在处理过程中假定源信号数目等于观测信号数目，当实际源信号数目少于观测信号数目时，IVA 分离输出可以得到源信号估计和残余噪声。在本系统设计中假设源信号数目小于观测信号数目。本系统中主节点根据该节点分离语音特征和其他主节点分离语音特征选取有效输出，并将有效输出语音的特征进行广播。

Mel 倒谱系数(Mel frequency cepstrum coefficient，MFCC)及其一阶或二阶差分是说话人识别中常用的语音特征。这里使用MFCC 与其一阶差分作为语音特征，利用语音特征空间中两路语音间的欧几里得距离和相关系数作为这两路语音间的相似度度量，当相似度量大于阈值时，判定两路语音属于同一说话人，则选择其中能量较大的一路作为输出。算法步骤如下：

(1)利用前一语音分离算法得到的分离数据，求出其频谱平方，即能量谱，并用 M 个 Mel 带通滤波器进行滤波；由于每一个频带中分量的作用在人耳中是叠加的，因此将每个滤波器频带内的能量进行叠加，这时第 k 个滤波器输出功率谱 $x'(k)$。

(2)将每个滤波器的输出去对数，得到相应频带 的对数功率谱；并进行反离散余弦变换，得到 L 个MFCC 系数，一般 L 取12～16 个。MFCC 系数为

$$C_n=\sum_{k=1}^{M}\log x'(k)\cos[\pi(k-0.5)n/M]\qquad n=1,2,\cdots,L \tag{13}$$

(3)将这种直接得到的 MFCC 特征作为静态特征，再将这种静态特征做一阶差分，得到相应的一阶差分 MFCC 动态特征。两者组合得到特征向量 $\boldsymbol{v}$。

(4)计算多路数据语音特征两两间的欧式距离和相关系数，分别表示为 dist_{ij} 和 corr_{ij}，则第 i 路数据与第 j 路数据的相似度度量表示为

$$P_{ij}=\mathrm{corr}_{ij}/\mathrm{dist}_{ij} \tag{14}$$

(5)设定两个相似度判定门限，若相似度高于第一门限，则判两者为同一说话人；若相似度低于第二门限，则判两者为不同说话人；若处于两者之

间，则不做判断，由下一帧数据判断。

(6)根据(5)中判定结果，若某一路输出不与其他路相似或与某些路相似但该路能量值最大，则将该路有效值置1，否则置0。

以上即为主节点选择输出的判决算法，其中判决门限由实验测试给出。该算法直接利用语音特征的相似度作为语音分类的判断，性能可能差于基于隐马尔科夫模型(Hidden Markov model，HMM)的方法、高斯混合模型方法或基于人工神经网络的方法，但该算法无需学习模型参数或神经元权矢量，能对输入信号做出快速处理，从而满足实时的需求。

2 算法在DSP上的实现

2.1 硬件系统框架

本设计在OMAP3530处理器平台上实现分布式语音信号分离。分布式传感器网络系统由9个节点组成，每个节点包括1个麦克风、1个WIFI无线网卡和1个基于OMAP3530的处理器模块。硬件系统的实现框图如图1所示。

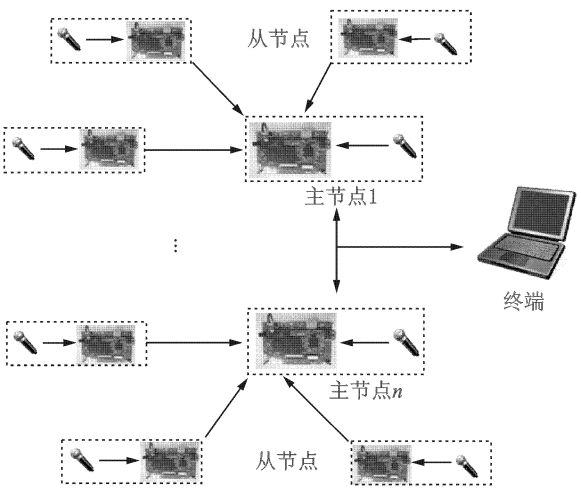

图1 硬件系统框图

以上系统框图显示了各部分之间的连接关系。其中麦克风连接处理器模块中的声音采集模块，实现语音采集；WIFI无线网卡负责传感器网络中节点间的信息交互。整个网络拓扑分为两级，执行语音信号分离运算的节点称为主节点；只负责语音采集的节点称为从节点。从节点采集其所在位置的语音信号，通过无线WIFI网络传输至其相邻主节点，主节点利用相邻从节点上传的数据和自身采集的数据进行语音信号盲分离，同时主节点间通过无线网络交互分离语音特征，主节点利用所有分离语音特征确认有效输出，以免与其他主节点输出同一说话人语音，终端通过网络接收所有主节点分离语音并实现本地播放。

2.2 软件系统框架

本文采用TI公司提供的Digital Video SDK开发环境进行软件开发，该开发环境包括ARM和DSP的交叉编译工具链及其他软件开发的工具。软件系统的框图如图2所示，图中只列举了一个主节点和从节点，其余节点软件系统一致。

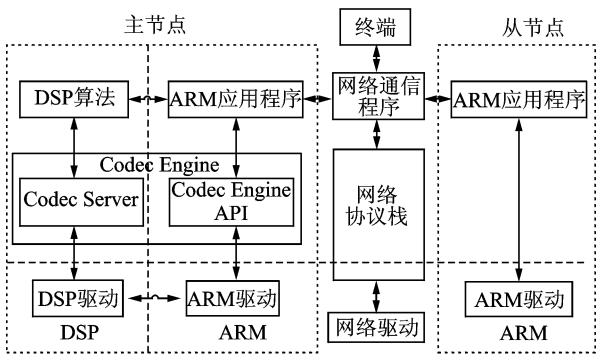

图2 软件系统框图

OMAP3530的ARM核运行基于Linux操作系统的应用程序，所用的外围设备都由ARM负责控制。ARM负责传感器网络中节点调度，时钟同步，采集语音数据，语音数据对齐和网络传输等任务。主节点ARM需把语音数据传送到ARM和DSP的共享缓冲内存区，交由DSP处理。DSP处理结束后，主节点ARM负责把语音数据传送至终端。DSP核上运行DSP/BIOS实时操作系统和语音信号盲分离与语音有效性评估算法。所有的算法接口都符合TI xDAIS标准，由Codec Engine调用。除了算法，DSP核上还集成了管理内存和DMA的Framework Component。

ARM核和DSP核的通信由TI提供的Codec Engine软件框架负责。Codec Engine是介于应用程序和具体算法之间的软件模块，其中的VISA API通过stub和skeleton访问Engine SPI最终调用算法。ARM和DSP的所用共享缓冲内存都是通过CMEM模块在DDR中分配的，缓冲内存地址连续且与DSP核Cache对齐。

实际操作中，所有节点的应用程序完全一致，主从节点的区分只需用一个标志加以区分，从而执行不同功能。这样的设计有利于扩充或缩减传感器网络的大小，而且能够灵活切换主从节点。

2.3　算法流程图

本系统算法主要是由独立向量分析和语音有效性评估两个算法组成，分别完成语音信号分离和有效语音提取。算法流程图如图 3 所示，该流程图只描述单个主节点中 DSP 处理算法流程。其中一帧语音数据包括多路语音数据，分别由主节点及其相邻从节点采集，长度为512。流程图中语音有效性评估之前数据为多路数据，有效性评估算法分别应用于各路分离输出数据，当某路输出评估值大于阈值，则该路数据保留并执行短时傅里叶反变换，得到时域语音信号，否则丢弃该路输出数据。

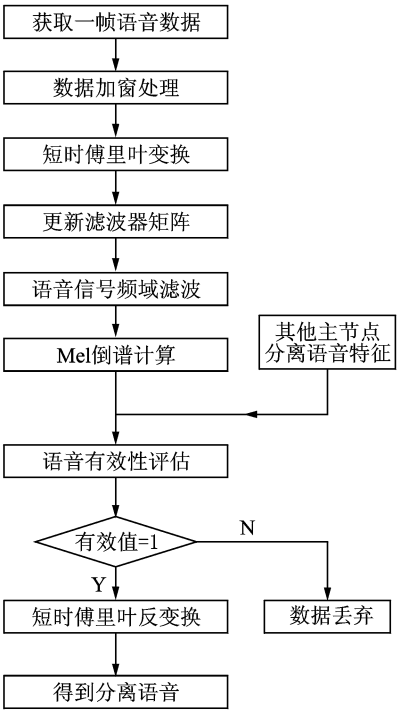

图 3　算法流程图

2.4　算法在 DSP 上的实现难点

无线传感器网络节点硬件设计和时钟同步。在该算法中，要求节点间的同步精度为毫秒级，为了减小节点硬件的复杂度，该系统直接采用 WIFI 无线网络对传感器节点进行时钟同步，由于无线网络固有的延迟较大，要达到毫秒级的同步精度需要有较好的算法。同步算法面临的最大问题在于传输延迟的不确定性，而传输延迟主要从以下几个方面产生：

(1)发送时间：发送端在构造消息的过程中所花费的时间。这包括操作系统的内核协议处理中带来的延迟、应用层的同步程序被操作系统进行进程切换所带来的延迟等。

(2)获取时间：获取时间主要指网络设备获取传输信道的时间，因为传输网络往往是多用户的，往往不得不等待一段时间才能进行发送。

(3)传输时间：消息从发送端发送出去后达到接收端所花费的时间。

(4)接收时间：接收端网络设备从信道中接收下消息并提醒主机消息到达所花费的时间。

在以上产生误差的时间中，获取时间应该说是对同步误差贡献最大的一部分。因为这部分时间随网络负载的变化会产生非常大的波动。本系统采用参考广播同步机制(Reference broadcast synchronization，RBS)同步协议[10]，此协议最大的特点在于利用通信协议的物理层广播消息进行同步，这样尽管广播的发送时间和获取时间都是未知的，但是由于广播出去的仅仅是一个物理信号，所以对所有接收端来说，这个时间是完全相同的，这就相当于去除掉了发送时间和获取时间所带来的误差，从而提高同步精度。RBS 同步协议的过程如下：

(1)一个广播节点广播出 m 个消息包(广播节点本身不参与同步)；

(2)n 个待同步节点中的每一个都在收到消息包的时候记录下当时的系统时间；

(3)待同步节点交换记录的时间戳；

(4)第 i 个节点可以计算出与任意一个其他节点 j 的时间差，计算的公式如下

$$\forall\ i\in n,\ \&\ j\in n:\ \mathrm{Offset}[i,j]=\frac{1}{m}\sum_{k=1}^{m}(T_{j,k}-T_{i,k}) \tag{15}$$

式中：n 为待同步节点的个数；m 为广播消息包的个数；$T_{r,b}$ 为第 r 个节点收到第 b 个广播包的时间。

实验结果表明，该算法保证了传感器节点间毫秒级的同步精度。

2.4.1　音频播放器的缓存设计

由于整个系统音频数据的传输是数据流形式的，且节点语音采集的速率与终端语音播放的速率不会完全一致，这是由硬件时钟不一致决定的。因而播放过程中可能会出现播放缓存塞满或不足的情况，使得语音播放无法正常运行，这就需要完善语音播放模块。

2.4.2　算法实现及代码优化

尽管该语音分离算法是实时在线算法，但还是具有运算量大的特点，这增加了算法在 DSP 上的实现难度。算法中的运算大多是复数矩阵运算，且含有很多复数矩阵求逆的情况，为了让算法能够实时地在 DSP 上运行，需要对该算法中运算量较大

的部分进行优化。虽然C64x+DSP善于计算定点乘加运算,但编译器不能有效地对C语言进行优化,必须调用TI提供的函数库或者进行汇编优化才能充分地发挥DSP的乘加运算能力。C64x+DSP不支持浮点运算,当算法中需要高精度计算时只能把浮点运算转化为定点运算。

3 性能与结果

在系统测试中,采用信号干扰比(Signal to interference,SIR)作为语音信号分离性能的评价,SIR的计算参考文献[11]。仿真场景设定为8 m×6 m×3 m的房间中放置9个传感器节点,位置如图4所示,主节点周围3个节点为该主节点相邻从节点。所有麦克风节点和声源离地面的高度都为1.5 m。3个声源分别为1个男声、2个女声,语音长度都为100 s,随机选择其中的30 s作为一次仿真源信号,共做100次实验。利用源镜像模型技术[12]仿真得到每个麦克风所对应的房间冲击响应,房间混响时间设定为500 ms。在所有仿真实验中,语音数据采样率为8 kHz,设定短时傅里叶变换长度为512,帧移设定为128,窗函数采用Hanning窗。式(12)中平滑因子β取0.5,式(11)中学习率η设定为0.5。

仿真场景如图4所示,其中仿真场景1包含源信号1,2,3;仿真场景2包含源信号1,2,3。两种场景下语音分离结果分别如图5,6所示。其中分离输出信号干扰比SIR_{out}是由100次实验求期望所得,且每秒计算1次SIR。从仿真场景1结果可以看出,系统经过约15 s之后各个源信号估计的SIR值达到最大值,干扰小于信号的1/10。对比场景1和2结果可以看出,当系统中源信号相隔较远时,系统收敛更快,且输出性能也更好,该性能的提升主要是由于源信号1和2的距离加大,致使源信号的输入SIR提高,从而使得整体性能得到提升。传感器网络系统的引入即是针对源信号在空间中稀疏分布的场景,而且该场景下很有可能有一个节点靠近某个源信号,这就大大提升了语音信号分离的性能。

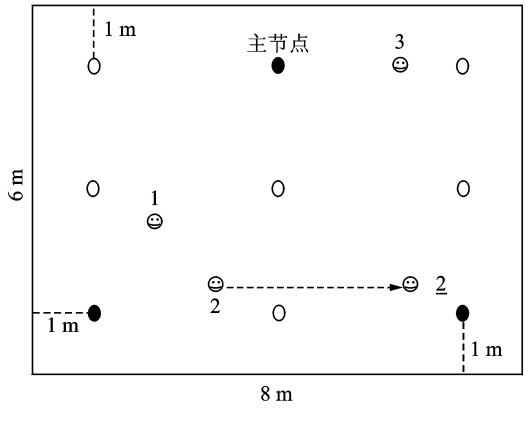

图4 仿真场景

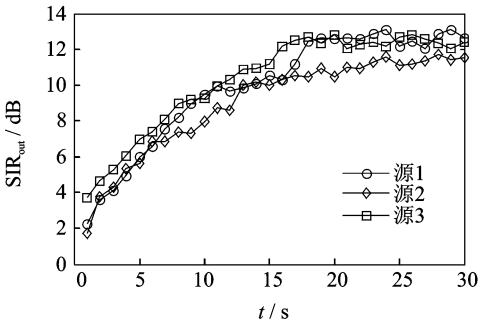

图5 仿真场景1中输出SIR曲线

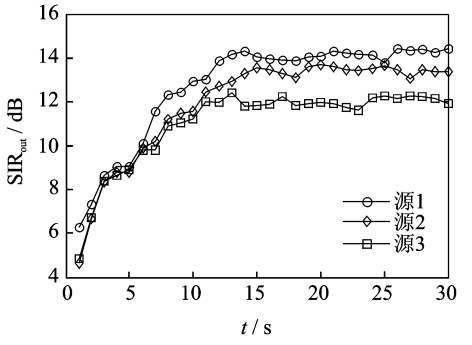

图6 仿真场景2中输出SIR曲线

系统在真实环境中的运行结果显示,该系统能有效地分离会场说话人语音,终端能获取较高SIR语音输出,达到了该系统的设计目的。

4 结束语

本文在OMAP3530平台上实现分布式语音信号分离系统,OMAP3530平台的强大运算性能为语音信号处理算法提供了实现的基础。在Codec Engine的软件架构下,本项目扩展了本文采用的语音处理算法的接口,在DSP核上实现语音分离和提取算法。同时,ARM核负责数据采集、网络通信、时钟同步、系统管理等工作。分布式语音分离算法利用传感器网络优势,针对大型会场场景,采用频域盲信号分离的算法解决著名的"鸡尾酒会问题"。该系统在计算机听觉、语音通信、助听器、远程网络会议、语音分离方面都有着广泛的应用前景。

参考文献:

[1] Wang D L, Brown G J. Computational auditory

scene analysis: principles, algorithms and applications[M]. Hoboken, NJ: Wiley-IEEE Press, 2006: 199.

[2] 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 Speech and Audio Processing, 2004, 12 (5): 530-538.

[3] Andr J W, van der K, Wang DeLiang, et al. A comparison of auditory and blind separation techniques for speech segregation[J]. IEEE Trans Audio Speech Lang Process, 2001, 9(3): 189-195.

[4] Kim T, Attias H T, Lee S Y. Blind source separation exploiting higher-order frequency dependencies [J]. IEEE Trans Audio Speech Lang Process, 2007, 15(1): 70-79.

[5] Huyen Dam H, Nordholm S, Yong Low S, et al. Blind signal separation using steepest descent method [J]. IEEE Trans Signal Process, 2007, 55(8): 4198-4207.

[6] Nesta F, Omologo M. Cooperative wiener-Ica for source localization and separation by distributed microphone arrays[C]//IEEE International Conference on Acoustics, Speech, and Signal Processing. [S. l.]: IEEE, 2010: 1-4.

[7] Dmochowski J P, Liu Zicheng, Chou P A. Blind source separation in a distributed microphone meeting environment for improved teleconferencing[C]// IEEE International Conference on Acoustics, Speech, and Signal Processing. [S. l.]: IEEE, 2008: 89-92.

[8] Taesu K. Real-time independent vector analysis for convolutive blind source separation[J]. IEEE Transactions on Circuits and Systems, 2010, 57(7): 1431-1438.

[9] 韩纪庆，张磊，郑铁然. 语音信号处理[M]. 北京：清华大学出版社，2004：84.
Han Jiqing, Zhang Lei, Zheng Tieran. The speech signal processing[M]. Beijing: Tsinghua University Press, 2004: 84

[10] Elson J, Girod L, Estrin D. Fine-grained network time synchronization using reference broadcasts [C]//Proceedings of the 5th Symposium on Operation System Design and Implementation. [S. l.]: OSDI, 2002: 147-163.

[11] Vincent E, Gribonval R, Pévotte C. Performance measurement in blind audio source separation [J]. IEEE Trans Audio Speech Lang Process, 2006, 14 (4): 1462-1469.

[12] Lehmann E A, Johansson A M. Prediction of energy decay in room impulse responses simulated with the image-source model[J]. Acoust Soc Am, 2008, 124(1): 269-277.

作者简介：杨志智(1988-)，男，硕士研究生，研究方向：语音信号处理，E-mail: 10210720044@fudan.edu.cn；唐显锭(1988-)，男，硕士研究生，研究方向：分布式信号处理；蔡瑾(1989-)，女，硕士研究生，研究方向：语音信号处理；冯辉(1980-)，男，讲师，研究方向：分布式信号处理理论及应用。