

江西省研究生创新专项资金项目

申 请 表

项 目 名 称: 智能家居环境下基于麦克风阵列的语音增强方法研究

申 请 人: 周静

指 导 教 师: 许春冬

培养单位(签章): 江西理工大学

填 报 时 间: 2018 年 5 月 1 日

江西省教育厅制

一、项目申报人基本情况					
姓 名	周 静	性 别	男		
出 生 年 月	1993 年 8 月	籍 贯	湖北荆州		
在读学历层次	硕士	入学日期	2017 年 9 月		
在 读 专 业	电子与通信工程				
身 份 证 号 码	511622199308241939				
指 导 师 姓 名	许春冬	研究方向	语音信号处理		
本科(硕士)毕业学 校	江西理工大学		专 业	电子信息工程	
所 在 院 系	信息工程学院		E-mail	1458311853@qq.com	
联 系 电 话	13125372142		手机	13125372142	
二、项目基本情况					
项目名称：智能家居环境下基于麦克风阵列的语音增强方法研究					
1、项目研究背景					
<p>(1) 智能家居行业发展现状</p> <p>国际智能家居行业近些年发展迅速，尤其是以美国、英国、日本、德国等国家为首的发达国家，智能家居产业发展相对较早，目前已经将智能语音平台、智能控制系统、智能家电、智能机器人、物联网等人工智能技术应用到智能家居环境下，且取得了不错的效果。中国报告大厅公开的《2016-2021 年中国智能家居产业市场运行暨产业发展趋势研究报告》数据显示^[1]，国内智能家居市场规模正逐年扩大，且随人工智能技术的发展，行业内出现较大范围的技术融合，使得许多前沿技术应用到智能家居</p>					

领域。但国内智能家居产业现状也存在许多不足，如缺乏人性设计、需求低价格高、标准不统一等。

智能家居产业在未来有着广阔前景。智研咨询发布的《2017-2022 年中国养老行业现状分析及投资战略研究报告》显示,2015 年中国 60 岁以上的老龄人口达 2.22 亿,预计到 2025 年达 3 亿以上;我国加速进入老龄化社会,提高老年人生活质量的问题随即而来,这就使得智能家居需求势必增长。同时随着智能产业的发展,**智能家居的概念也将越来越普及,人们对智能家居的信赖感也在逐步增强^[2]**,如人们在购买家居产品的时候越来越看重其融入的“智能元素”。到 2020 年我国全面建成小康社会,中国人均收入将大幅增强,同时随着技术的发展也逐步降低智能家居产品成本,正逐步的向人们消费水平靠近,且智能家居产业发展也被写进政府工作报告,备受重视。

(2) 麦克风阵列语音增强技术的发展

近些年来,利用麦克风阵列对语音信号进行噪声抑制甚至消除的方法已经成为语音增强领域的一个热门研究。迄今为止,出现了大量基于麦克风阵列的语音增强算法,可大致分为波束形成(Beamforming, BF)、后置滤波算法和盲源分离算法等^[3,4]。

波束形成可以分为固定波束形成(Fixed Beamforming, FBF)和自适应波束形成。FBF 算法最具代表性的为“延迟-相加”波束形成(Delay and Sum Beamforming, DSBF)算法,该算法将多个麦克风采集的语音信号做延迟处理,补偿各信号的时间差异以达到同步,然后加权求和;这类算法能够有效抑制非相干噪声(或散射噪声),但对相干噪声干扰(主要表现为方向性噪声和混响噪声干扰)抵抗力较低,因此这类算法较少单独使用。在 1972 年, Frost 提出了基于线性约束最小方差(Linearly Constrained Minimum Variance), LCMV 的自适应波束形成算法,其原理是对采集到的语音信号进行延迟补偿,然后根据输出信号中残留噪声对语音信号作约束,以输出噪声功率最小来达到语音增强的目的;1982 年, Griffith 和 Jim 基于自适应波束形成算法提出了基于广义旁瓣抵消降噪算法(Generalized Sidelobe Canceller, GSC), GSC 算法包括三个模块:固定波束形成器(Beamforming, BF)模块,阻塞矩阵(Blocking Matrix, BM)模块和自适应噪声消除(Adaptive Noise Canceller, ANC)模块。GSC 算法比较简单,很容易被实现,对方向性噪声的抑制效果也很好。但是由于在阻塞矩阵模块中阻塞矩阵不可以完全滤除目标语音信号导致自适应噪声消除模块的性能下降,从而影响 GSC 整个系统的降噪效果^[6]。随后许多研究人员根据经典的 GSC 算法提出了一些有效的改进的算法。其中, Hoshuyam 等人提出采用自适应阻塞矩阵的基于稳健系数自适应滤波器约束的语音增强算法(Coefficients Constrained Adaptive Filter, CCAF); Nordholm 提出的基于空间高通滤波算法的阻塞矩阵大幅度的提升了阻塞效果; Chu 和 Fang 等人提出了通过小波变换算法实现阻塞矩阵的设计; Widrow 和 Steams 提出了一个双重结构的波束形成算法;根据最小输出功率来调整系数,并且保证了波束方向为期望语

音的方向；Jan 和 Flanagan 通过估计出的声学传输路径函数来表示实际的传输路径提出了匹配滤波波束成型算法 (Matched Filter Beamforming, MFBF)；Grenier 利用声源和麦克风阵列之间存在的特性来估计传输函数 (Transfer Function, TF) 提出了许多基于 GSC 的语音增强算法；李芳兰和周跃海等人提出在 GSC 中引入可调波束次级改成器估计声源方位以抑制背景噪声影响的算法；Sharon Cannot 提出了基于传递函数的广义旁瓣抵消算法 (Transfer Function Generalized Sidelobe Canceller, TF-GSC)，该算法通过估计出的声学传输比例而非传输系数来对固定波束形成器和阻塞矩阵的进行重新的构建，利用在频域的近似相乘形式实现了阵列接收信号的传递函数模型，并阐述了 LCMV 和传统 GSC 算法的频域实现算法，该算法在方向性噪声的抑制中取得了很好的效果，但在散射噪声方面的性能却不是很好^[3]；经过对 TF-GSC 算法的进一步改进，Sharon Cannot 提出使用卷积代替 TF-GSC 中的近似相乘形式更加精确地模拟了阵列接收不同路径信号的传递模型 (Convolutional Transfer Function Generalized Sidelobe Canceller, CTF-GSC) 算法，该算法进一步提高了在嘈杂的室内环境下的语音增强系统的降噪能力。以上这些算法大多都是基于 GSC 算法的改进，但这些算法大多都依赖于阻塞矩阵的性能，在信源角度估计有偏差的情况下很难完全滤除期望语音信号，进而降低了系统的噪声抑制能力并且带来了不稳定性。

后置滤波算法是将波束形成算法处理后的信号做进一步处理，它能够较好的抑制散射噪声。现阶段其主要算法有基于维纳滤波的算法、基于最小均方误差 (MMSE) 的算法、基于卡尔曼滤波的算法、算法子空间、空时预测算法、基于自适应消噪 (Adaptive Noise Canceller, ANC) 算法以及多种基于短时谱的算法等。

最早在 19 世纪 70 年代，Allen 利用波束形成算法的方向性增益和维纳滤波器对非方向性噪声的降噪能力，将维纳滤波算法作为后置滤波用于语音增强算法中。1990 年，Zelinski 提出的通过采用维纳滤波来对固定波束形成输出的信号进行进一步的降噪处理，该算法的优点是在滤波处理过程中滤波器的系数是通过多通道的信号信息来估计，极大的提高了噪声的抑制能力。Simmer 引入维纳滤波和谱减法作为后置滤波来解决阵列之间低频信号相干性高的问题。Dal-Degan 和 Prati 提出在低频段使用谱减法，在高频段使用维纳滤波法，从而形成一个新的算法。后来，有研究者提出将 GSC 算法与最小均方误差的短时幅度谱算法 (Minimum Mean Square Error, MMSE) 结合的后置滤波，该算法在室外环境中对非平稳噪声有较好的噪声消除效果。邓俊杰和孙超提出将子带 GSC 与改进的维纳后置滤波相结合的算法，该算法具有更高的输出分段信噪比。但在室内环境下，语音信号经过反射和折射后出现多径传输，从而导致语音增强系统的噪声抑制能力和噪声消除能力严重下降。由于 TF-GSC 算法可以解决多径传输导致的系统下降问题，Cohen 和 Berdugo 首次提出了针对室内环境中的非平稳噪声信号的抑制算法。该算法利用 TF-GSC 算法的输出信号和阻塞矩阵输出的噪声信号

来组成多通道后验滤波器。随后，栗晓丽和傅丰林提出基于子带 TF-GSC 的算法，该算法在子带 TF-GSC 算法中的固定波束形成支路上引入维纳滤波算法，从而有效的抑制了麦克风拾取的语音信号中混有的相干性和非相干性的噪声。**和单一的 TF-GSC 算法相比，后置滤波算法极大的提高的系统的噪声抑制能力^[4]**。因为单一的 TF-GSC 算法只能有效的抑制方向性干扰信号，输出的增强信号中残留有较大成分的非相干性噪声，与此相反，后置滤波器算法对非相干性噪声具有很强的抑制作用。因此研究者们提出了将传统的阵列信号处理算法与基于短时谱的滤波算法结合的算法来进行语音增强系统的设计。

盲源分离（Blind Source Separate，BSS）方法也是基于波束形成技术的一种方法。BSS 算法能够将多个未知信号源混叠的混合信号分离出各声源成分，其在处理过程中不需要信源的先验信息。盲源分离技术研究早期主要是集中于基于高阶（HOS，高于 2 阶）统计量的方法，由于 HOS 收敛速度慢，研究重点就转向二阶统计量（SOS）。应用于盲源分离和辨识的模型主要有 SIMO 模型和 MIMO 模型。

(3) 智能家居环境下麦克风阵列语音增强技术的应用

智能家居环境下，采用麦克风作为音频信号的采集设备，且常为线性麦克风阵列，如图 1 示。语音增强技术作为智能声控系统识别和感知的一个重要前端处理，在整个智能语音控制系统中起着关键性作用。室内家居环境下，除了多种非目标声源干扰外，还会存在严重混响，且方向性噪声和散射噪声较为“顽固”，这严重影响着系统识别率，其模型如图 2 所示。现阶段，麦克风阵列在智能家居中的应用的技术主要有声源采集、声源定位、声源分离、信号增强等，如智能声控电视、声控窗帘、声控空调、智能视频电话、智能机器人等。许多智能器件上采用了单麦克风或双麦克风阵列作识别，在低信噪比环境下往往难以有效判决语音指令，给用户带来的体验感不强，采用多麦克又加大了成本。随着物联网技术的发展，未来的智能家居将是体系互联的，只需要在房间固定位置安装麦克风阵列即可做到对所有家居设备的声控管理，但现阶段国内各企业标准并不统一，因此这还是一个需要时间的过程。

智能家居环境下，有效识别往往是核心问题，那么应如何在多声源、强混响环境下做到有效识别呢？从现有研究来看，**基于麦克风阵列的语音增强技术无疑是最有效的方法**。应答于以上研究与发展的大环境，本项目提出了针对智能家居的麦克风阵列语音信号增强处理研究。

去混响是麦克阵列风语音增强算法研究的一个重要方向，尤其是针对室内环境而言，如何做到去除混响干扰尤为重要。适用于混响环境下的麦克风阵列语音增强技术主要有三类^[3]：1) **基于盲语音增强的方法(Blind Signal Enhancement Approach)**；2) **基于波束形成的方法(Beamforming Based Approach)**；3) **基于逆滤波的方法(An Inverse Filtering Approach)**。

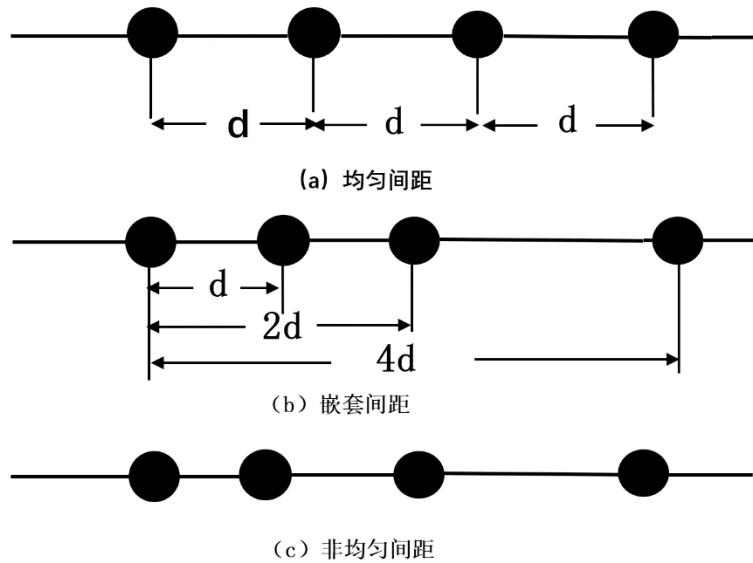


图 1 常用一维线性阵列

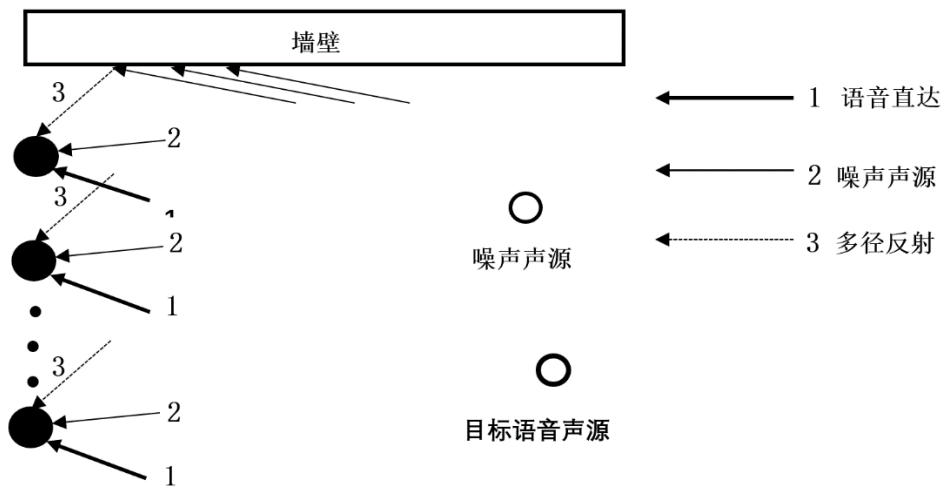


图 2 麦克风阵列采集语音信号模型

(4) 小结

本项目拟在国内外对麦克风阵列语音增强研究的基础上，重点研究智能家居声学环境下麦克风阵列的语音增强。拟通过解决**衰减补偿**提高信号增益，采用 MIMO 分级模型实现**源分离及去混响**，采用声源定位技术**判别远场和近场**，采用“**子带 CTF-GSC+维纳滤波算法**”抑制远场方向性噪声和散射噪声，采用**自适应波束形成技术**实现近场增强，并实现家居环境远场、近场增强算法模型自适应切换，建立智能家居环境复杂声学环境下目标语音信号的增强算法系统模型，以提高麦克风采集信号的“可懂度”，提升家居智能语音控制的体验感。研究成果应为智能家居声控前端处理提供重要理论依据，并引导现阶段智能家居声控行业的发展。

2、项目研究的主要内容

语音增强是语音识别中前段处理的一个重要环节，好的增强算法能够有效的提高语音信号的可懂度。现阶段，在语音增强方面已有一些相对成熟的算法，如基于深度学习（DNN）的算法等。在智能家居环境下，由于前沿的高精度算法复杂度高、运算量过大等因素，使得目前的硬件环境往往因成本过高、室内布局设计复杂、难以维护等因素不能给予较好的支持，从而理论可实现的算法物理实现难度过大。而一些传统的算法，虽然算法复杂度相对低很多，在设计、成本、维护等方面要求低很多，但在低信噪比环境下往往又很难有效的起到增强效果。智能家居环境下噪声声源主要包括其多人语音声源、家庭影视声源、厨房炊具声源、流水声源、饭桌餐具声源、室内敲打声、空调风扇等电器声源、脚步声源等类型，且多数情况存在严重混响，复杂的噪声环境给系统的识别带来了严重干扰。因此根据智能家居发展的现状，本项目将针对以下三个内容进行深入研究与探讨：

（1）现阶段多数基于麦克风阵列的语音增强方法都是通过抑制噪声来取得更好效果，而忽略了补偿信号自身（提高信号增益）来达到更高信噪比。因此本项目将根据声信号在介质中传播的模型，拟采用基于麦克风阵列的自适应波束形成等技术模拟出声信号的传播衰减模型，求取相关衰减系统函数，并设计衰减补偿滤波器，要求处理后达到较好突出目标信号、明显提升信噪比的效果；

（2）在众多麦克风阵列增强算法中，基于传递函数卷积混合的广义旁瓣抵消算法（Convolutional Transfer Function Generalized Sidelobe Canceller, CTF-GSC）对方向性噪声的抑制起到了较好的抑制效果，但对散射噪声却显得束手无策。而后置维纳滤波算法却可以在较小失真情况下有效的抑制散射噪声，因此本项目针对室内环境下的噪声和混响，将子带 CTF-GSC 算法与后置维纳滤波算法相结合，研究如何做到有效去除方向性噪声和散射噪声，并用于远场增强模型；

（3）智能家居环境下，麦克风阵列所处环境易变为近场环境模型，所接收到的声信号波前弯曲率不可忽略，阵列所录入信号为球面波，而远场模型接收到的为平面波，此时如果再采用远场模型来处理，将导致麦克风阵列语音增强效果下降。因此本项目提出一种基于近场波束形成的麦克风阵列语音增强算法，拟通过阵列声源定位判断远场、近场，实现语音增强模型的自适应切换。

3、项目研究的关键技术及解决方法

项目研究内容及研究方法主要涉及以下几个关键技术。

3.1 信号衰减增益补偿

(1) **声学基础模型建立。**本项目是基于人声在空气中传播的衰减模型展开的。从时域角度来看，2m 左右的距离就可以给语音信号带来较大的幅值（能量）衰减；从频域角度来看，各频段信号的衰减与频率平方成反比，且随着距离的增加，高频部分衰减尺度显著扩大。具体的声学基础模型本文就不再给大家论述，具体内容可参考杜功焕等所著《声学基础》^[5]一书进行相关了解。

(2) 语音幅值衰减与高频率衰减幅值补偿。

幅值衰减系统拟合。根据上述语音信号的声学特点，项目将在消音室环境下通过在不同距离点上布置麦克风采集实验语音信号数据。对数据处理后，配对不同麦克风的相同帧，根据线性约束自适应波束成形器原理，计算信号帧间的自相关和互相关函数等参数，拟合出幅值衰减系统。

线性约束自适应波束成形器原理图：

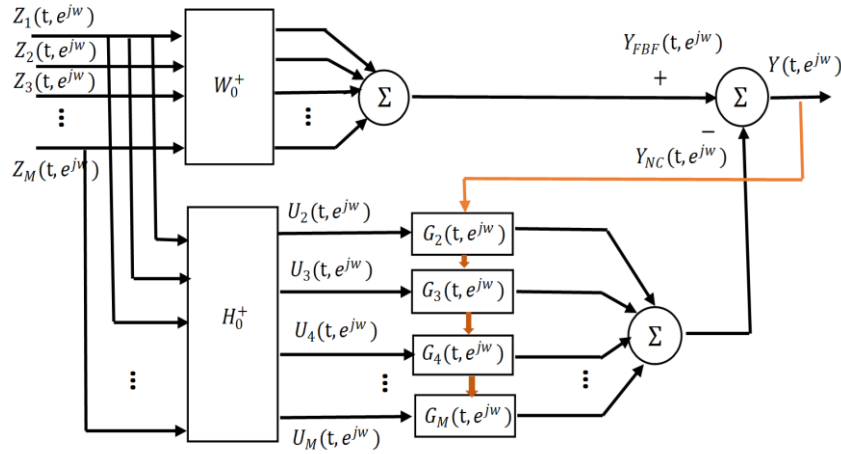


图 3 线性约束自适应波束成形器原理图

衰减系统拟合方程^[3]：

$$\begin{bmatrix} \hat{\varphi}^{(1)}_{z_m z_1}(e^{j\omega}) \\ \hat{\varphi}^{(2)}_{z_m z_1}(e^{j\omega}) \\ \vdots \\ \hat{\varphi}^{(K)}_{z_m z_1}(e^{j\omega}) \end{bmatrix} = \begin{bmatrix} \hat{\varphi}^{(1)}_{z_1 z_1}(e^{j\omega}) & 1 \\ \hat{\varphi}^{(2)}_{z_1 z_1}(e^{j\omega}) & 1 \\ \vdots & \vdots \\ \hat{\varphi}^{(K)}_{z_1 z_1}(e^{j\omega}) & 1 \end{bmatrix} \begin{bmatrix} H_m(e^{j\omega}) \\ \phi_{u_m z_1}(e^{j\omega}) \end{bmatrix} + \begin{bmatrix} \varepsilon_m^{(1)}(e^{j\omega}) \\ \varepsilon_m^{(2)}(e^{j\omega}) \\ \vdots \\ \varepsilon_m^{(K)}(e^{j\omega}) \end{bmatrix} \quad (1)$$

其中 $H_m(e^{j\omega})$ 即为所求衰减系统的系统函数。

(3) 滤波器计补

根据频率越高衰减越严重的特点，在 0~0.5kHz 频段作较小幅值固定补偿，在

0.5kHz~2.5kHz 频段作幅值递增的补偿, 在 2.5kHz~3.5kHz 作幅值递减的补偿, 在 3.5kHz~4 kHz 做较小幅值固定补偿, 其余频段不做补偿^[4-6]。根据拟合的幅值衰减系统确定补偿幅值相关参数。理论上, 补偿处理后所得信号频率特性将变得相对平坦, 且有用语音段明显凸起。预设计滤波器如图 4 所示。

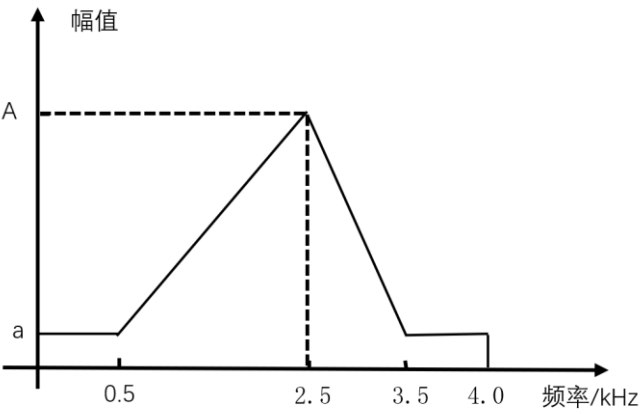


图 4 预设计补偿滤波器

3.2 “子带 CTF-GSC 算法+维纳滤波算法”抑制方向性噪声和散射噪声

3.2.1 CTF-GSC 算法结构、原理及子带 CTF-GSC 算法

CTF-GSC 算法由三部分构成^[4, 6, 7]：固定波速形成器（Fixed Beamforming, FBF）、阻塞矩阵（Blocking Matrix, BM）和自适应噪声抵消器（Adaptive Noise Canceller, ANC）。CTF-GSC 算法是以 GSC 算法为基础, 利用不同麦克风间传递函数（Relative Transfer Function, RTF）的比例系数对阻塞矩阵重构。传递函数的比例系数有两种估计方法, 一是忽略不同频率成分间的相互干扰, 通过传递函数相乘的形式得到 CTF-GSC 算法；二是将不同频率成分相互干扰考虑在内, 通过各传递函数卷积混合的形式得到 CTF-GSC 算法。实验证明第二种更为精确^[5, 17], 更适用于复杂的室内声学环境。

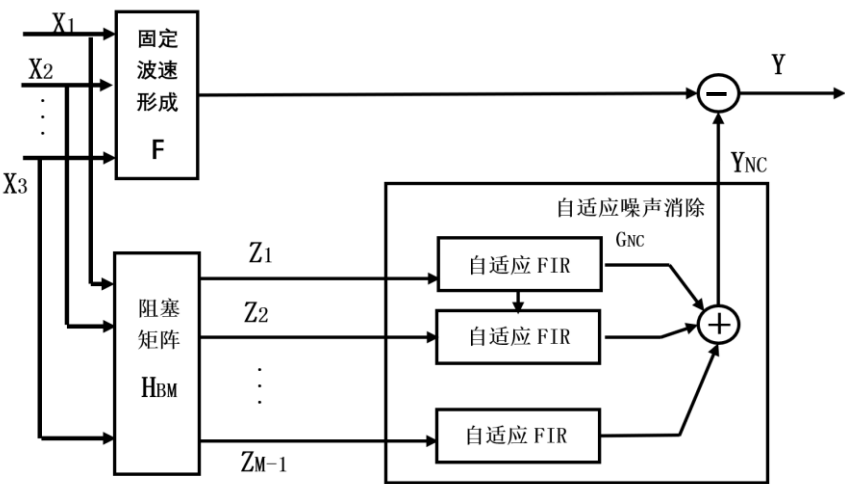


图 5 CTF-GSC 算法原理框图

求得卷积阵列 $H_m(k', k)$ 后，可计算得到混响情况下的阻塞矩阵：

$$H = \begin{bmatrix} -\hat{H}_2^H & -\hat{H}_3^H & \dots & -\hat{H}_M^H \\ I & 0 & \dots & 0 \\ 0 & I & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & 0 & I \end{bmatrix} \quad (2)$$

其中

$$\hat{H}_m = \begin{bmatrix} H_m(0,0) & 0 & \dots & 0 \\ 0 & H_m(1,1) & \dots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & 0 & H_m(N-1, N-1) \end{bmatrix} \quad (3)$$

固定波速形成模块矢量矩阵为 $F = [F_1, F_2, \dots, F_m]$ ：

$$F_m = \begin{bmatrix} F_m(0) & 0 & \dots & 0 \\ 0 & F_m(1) & \dots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & 0 & F_m(N-1) \end{bmatrix} \quad (4)$$

其中

$$F_m(k) = H_m(k, k)K(k) \quad (5)$$

$$K(k) = \left[\sum_{m=1}^M H_m^H(k, k)H_m(k, k) \right]^{-1} \quad (6)$$

则可得到固定波束形成模块输出参考信号：

$$\{\hat{d}_1^{FBF}\}_{CTF} = K(k) \sum_{m=1}^M F_m^H(k)X_m(k) \quad (7)$$

阻塞矩阵输出噪声信号：

$$Z_m(k) = X_m(k) - H_m(k, k)X_1(k), \forall m = 2, \dots, M \quad (8)$$

基于上述原理，提出改进的 CTF-GSC 算法——**子带 CTF-GSC 算法**，其原理框图如图 6 所示。

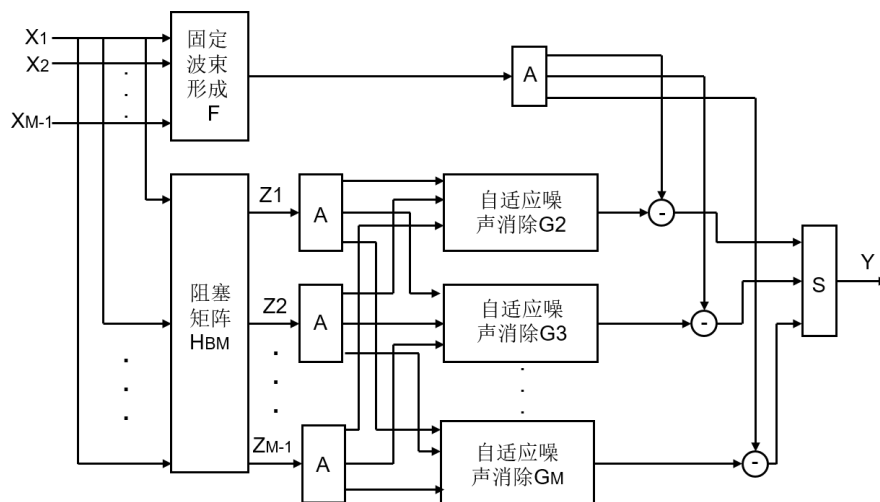
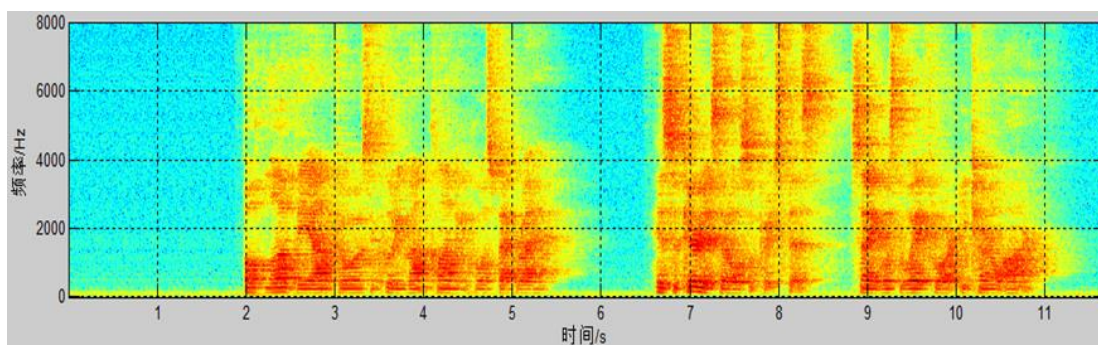
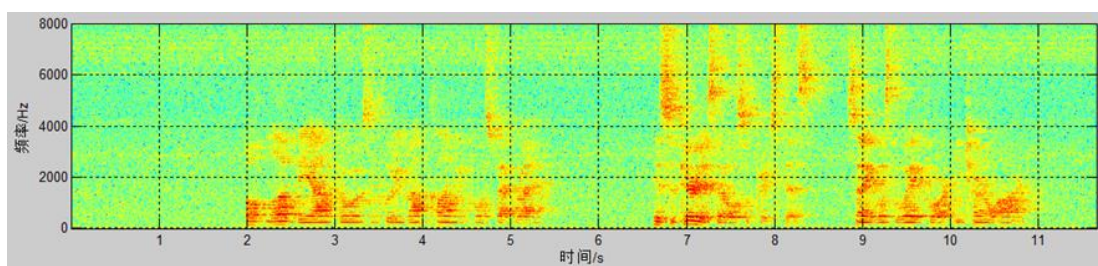


图 6 子带 CTF-GSC 算法原理框图

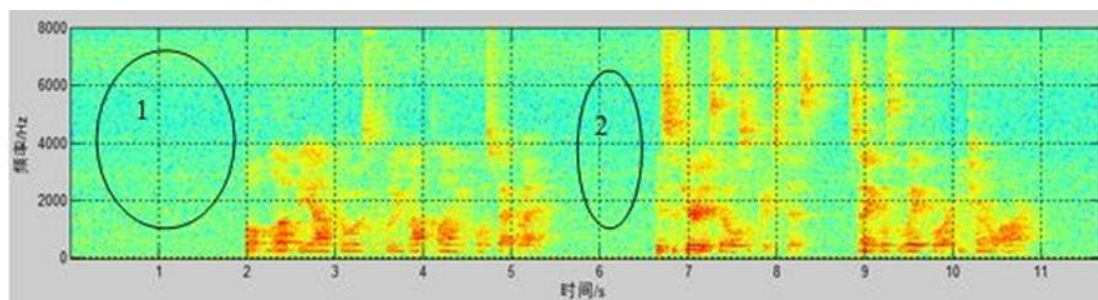
CTF-GSC 算法与子带 CTF-GSC 算法混响情况下降噪效果对比：



(a) 原始混响语音信号语谱图



(b) CTF-GSC 算法去混响语谱图



(c) 子带 CTF-GSC 算法去混响语谱图

图 7 CTF-GSC 与子带 CTF-GSC 算法去混响效果对比

从语谱图分析,子带 CTF-GSC 算法去混响效果好于 CTF-GSC 算法,更适用于室内环境,因此本项目将采用子带 CTF-GSC 算法与后置维纳滤波算法相结合。

3.2.2 后置滤波——维纳滤波算法^[4, 8, 9]

智能家居环境下,语音信号采集时会受到方向性噪声、散射噪声、混响等噪声的影响。目前混响、方向性噪声都可以通过一些基于麦克风阵列的语音增强方法抑制,并能起到不错的效果,如波束形成技术。但基于波束成形的技术对散射噪声却无法起到好的抑制效果,而维纳滤波算法可较好的抑制散射噪声,因此采用维纳滤波算法来做后置滤波。

维纳滤波的基本原理:

$$Y(k) = H(k) \cdot X(k) \quad (9)$$

其中 $X(k)$ 为带噪信号的傅里叶变换, $H(k)$ 为增益函数, $Y(k)$ 是对纯语音 $S(k)$ 的最优估计。假设语音信号与噪声不相关,则得最小均方误差:

$$D(k) = (H^2(k) - 2H(k) - 1) \cdot E[S^2(k)] + H^2(k) \cdot E[N^2(k)] \quad (10)$$

当 $D(k)=0$ 时可求得增益函数

$$H(k) = \frac{E(S^2(k))}{E[S^2(k)] + E[N^2(k)]} \quad (11)$$

基于先验信噪比的维纳滤波,先验信噪比由后验信噪比直接判决:

$$SNR_{prio}(k) = \alpha \frac{E(\hat{S}^2(k-1))}{E[N^2(k)]} + (1 - \alpha) \cdot \max[SNR_{post}(k) - 1, 0] \quad (12)$$

α 为常数,通常取 0.95。

传统维纳滤波和基于先验信噪比的维纳滤波会因估计出的噪声谱存在误差,易造成语音信号失真和降噪效果不明显等结果,因此可考虑通过语音活动检测(VAD)及小波估计算法对噪声作进一步精确估计,从而改进增益函数达到更好效果。

$$H(k) = \frac{SNR_{prio}(k)}{SNR_{prio}(k) + \mu} \cdot \beta \quad (13)$$

其中 μ 为先验信噪比补偿因子, β 为增益修正因子。

3.3 近场环境下自适应波束形成方法及消噪^[11, 19]

相比于远场，近场麦克风阵列采集到得声信号为球面波，如图 8 所示。近场与远场信号处理的区别在于信号源与噪声的描述，如传播矢量，噪声相关矩阵等。相对于原始声源信号，反射波对应的镜像声源往往位于较远的位置，因此可以通过距离差来减少混响和噪声的影响。项目提出基于近场优化的自适应波束成形语音增强，用于智能家居环境近场情况下降噪。

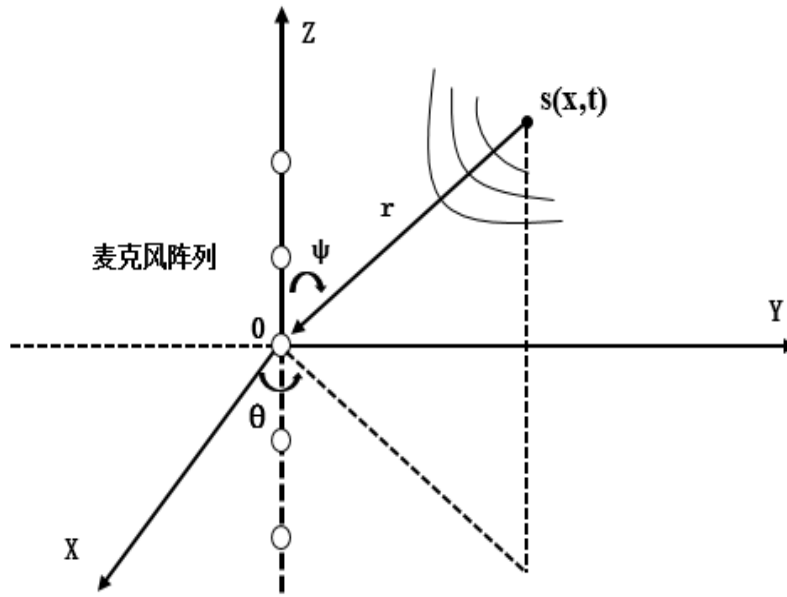


图 8 球面坐标系下的声波与阵列

声波 $s(r, t) = s(r, \theta, \psi, t)$ 的波动方程表示为：

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial s}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial s}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 s}{\partial \theta^2} = \frac{1}{c^2} \frac{\partial^2 s}{\partial t^2} \quad (14)$$

其中 c 为声速， θ 和 ψ 分别为声源的方位角和仰角。

$$\text{可得波动方程解为: } s(r, \omega, t) = \frac{A}{r} e^{j(\omega t - kr)}. \quad (15)$$

该方法系统框图如图 9 所示，主通道采用近场优化波束形成器，辅助通道采用双自适应波束形成，然后通过语音自动检测（VAD）对阻塞滤波器的输出信号进行语音检测，利用门限值来控制噪声消除滤波器的自适应波束形成过程。

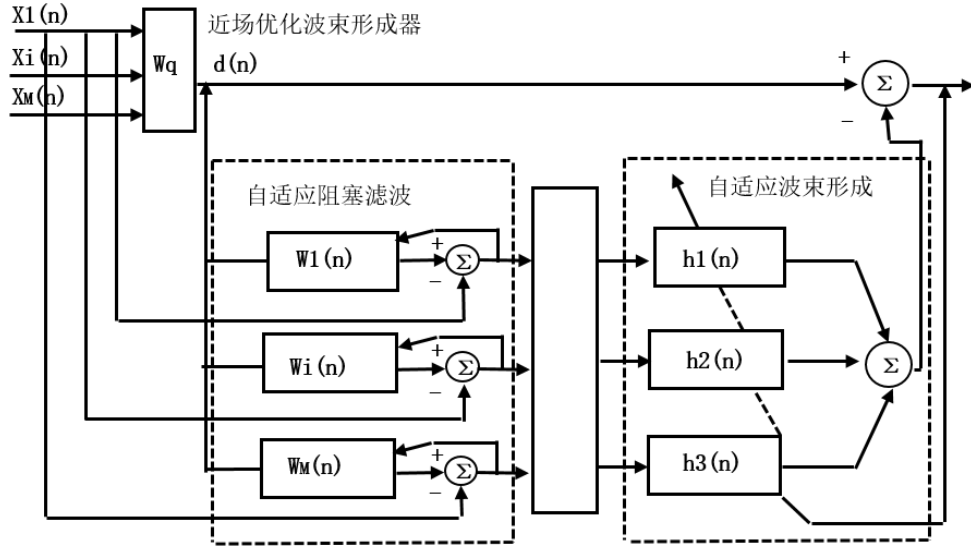


图9 近场麦克风阵列语音增强方法原理框图

近场优化滤波器输入数据 $x(n)$ 协方差矩阵:

$$\mathbf{R}_{xx}(n) = (1 - \beta)\mathbf{R}_{xx}(n-1) + \beta\mathbf{x}^T(n)\mathbf{x}(n) \quad (16)$$

自适应阻塞滤波器输出误差信号矩阵:

$$\mathbf{E}(n) = \begin{bmatrix} \mathbf{X}(n) - \mathbf{D}^T(n)\mathbf{W}(n) \\ (1 - \mu_1)\mathbf{E}(n-1) \end{bmatrix}_{1}^{P_1-1} \quad (17)$$

式中 μ_1 为步长。

多通道自适应波束形成的滤波器 h 权矢量更新方程^[17]:

$$\mathbf{h}_i(n+1) = \begin{cases} \mathbf{h}_i(n) + \mu\mathbf{Y}_i(n)(\mathbf{Y}_i^T(n)\mathbf{Y}_i(n) + \delta\mathbf{I})^{-1}\mathbf{z}(n), & Q(n) \geq Q_{ths} \\ \mathbf{h}_i(n), & Q(n) \leq Q_{ths} \end{cases} \quad (18)$$

3.4 麦克风阵列声源分离及去混响^[4, 13, 14, 20]

3.4.1 模型建立

前面已经提到智能家居环境下存在多声源、混响、背景噪声等干扰,图10描绘了家居环境下的声学场景。在 M 个声源, N 个麦克风采集系统下, $s_m(k)$ 表示不同声源, $v_n(k)$ 表示不同麦克风接收到的背景噪声, $g_{mn}(k)$ 表示不同麦克风接收到的不同声源的信号。

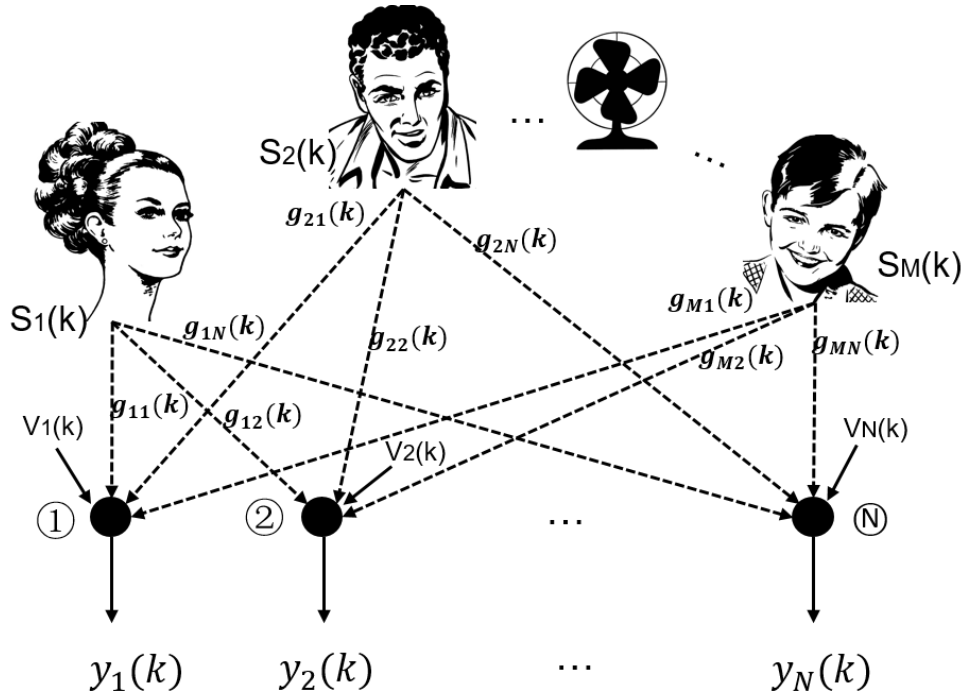


图 10 智能家居环境下麦克风阵列信号采集模型

第 N 个麦克风输出可表示为^[6]:

$$y_n(k) = \sum_{m=1}^M g_{nm}(k) * s_m(k) + v(k), \quad n = 1, 2, \dots, N \quad (19)$$

在没有加性背景噪声的情况下，由分离和去混响所产生的信号可以表示为

$$z_a(k) = HG s_{ML}(k) \quad (20)$$

$$z_a(k) = [z_1(k) z_2(k) \dots z_M(k)]^T \quad (21)$$

$$H(k) = \begin{bmatrix} h_{11}^T & h_{12}^T & \dots & h_{1N}^T \\ h_{21}^T & h_{22}^T & \dots & h_{2N}^T \\ \vdots & \vdots & \ddots & \vdots \\ h_{M1}^T & h_{M2}^T & \dots & h_{MN}^T \end{bmatrix} \quad (22)$$

$$G = \begin{bmatrix} G_{11} & G_{12} & \dots & G_{1M} \\ G_{21} & G_{22} & \dots & G_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ G_{N1} & G_{N2} & \dots & G_{NM} \end{bmatrix} \quad (23)$$

要使 $z_m(k) = s_m(k - \tau_m)$, 这里 τ_m 为时延常数, 则分离与去混响条件表示为

$$HG = U = \begin{bmatrix} U_{11}^T & O_{L1}^T & \dots & O_{L1}^T \\ O_{L1}^T & U_{22}^T & \dots & O_{L1}^T \\ \vdots & \vdots & \ddots & \vdots \\ O_{L1}^T & O_{L1}^T & \dots & U_{MM}^T \end{bmatrix} \quad (24)$$

3.4.2 源分离^[13, 18]

以一个 2×3 的 MIMO 模型（两个声源三个麦克风）做简洁说明。对于一个 2×3 的 MIMO 系统, 可以利用两个麦克风输出信号一次性完成由两个相互干扰的声源同时存在所引起的同信道干扰（Co-Channel Interference, CCI）。如通过以下步骤去掉 $Y_1(z)$ 和 $Y_2(z)$ 中有 $S_2(z)$ 所引起的干扰成分^[11, 19]:

$$Y_1(z)G_{22}(z) - Y_2(z)G_{12}(z) = [G_{11}(z)G_{22}(z)]S_1(z) + [G_{22}(z)V_1(z) - G_{12}(z)V_2(z)] \quad (25)$$

类似的, 在这两个麦克风由 $S_1(z)$ 所导致的干扰成分（从第二个声源的视角来看）同样能够被去除。简化过程如图 11 所示。

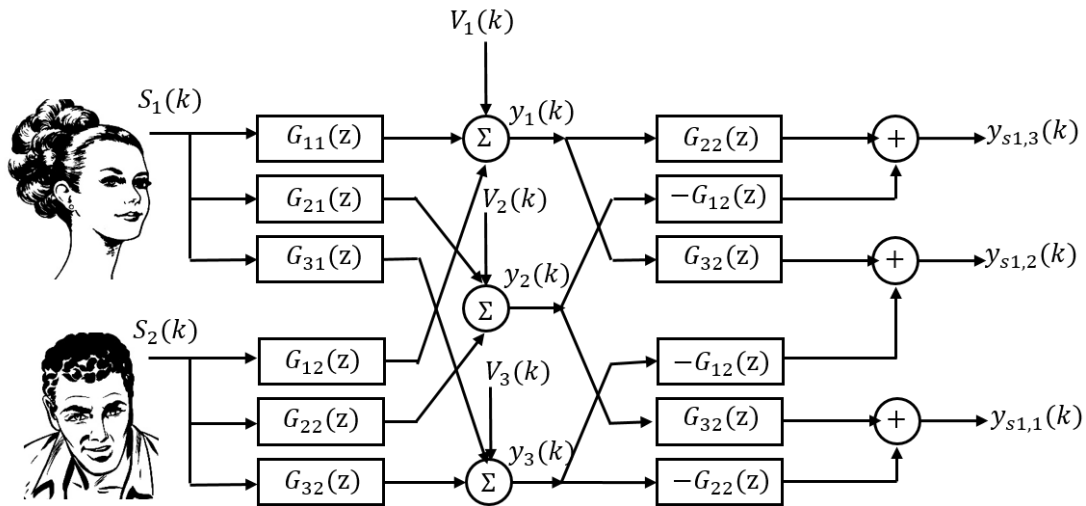


图 11 2×3 的 MIMO 模型源分离过程

3.4.3 去混响

除噪声污染以外, 混响是导致语音质量恶化的另一大因素^[3]。混响会导致时间和频率模糊, 从而导致语音信号谱包络和精细化结构失真^[13], 因此存在混响情况下, 语音可懂度会降低很多。这些问题极大地推动了语音处理和语音信号通信系统中对有效去混响方法的需求。

语音去混响可以分为三类^[15, 16]: 基于语音源模型的去混响, 采用同态处理的语音混响分离, 利用信道反转和均衡方法的语音去混响。其中, 信道反转和均衡方法在去混响中取得较好效果, 研究中拟采用该方法去混响。信道反转和均衡方法中常用的有

三种^[19]：直接反转法（也称迫零法），最小均方误差（MMSE）或最小二乘法（LS），多通道逆定理法（MINT）。相关文献表明^[4]，MINT 方法更适合于 SIMO 系统。

3.5 主要技术应用实施框架：

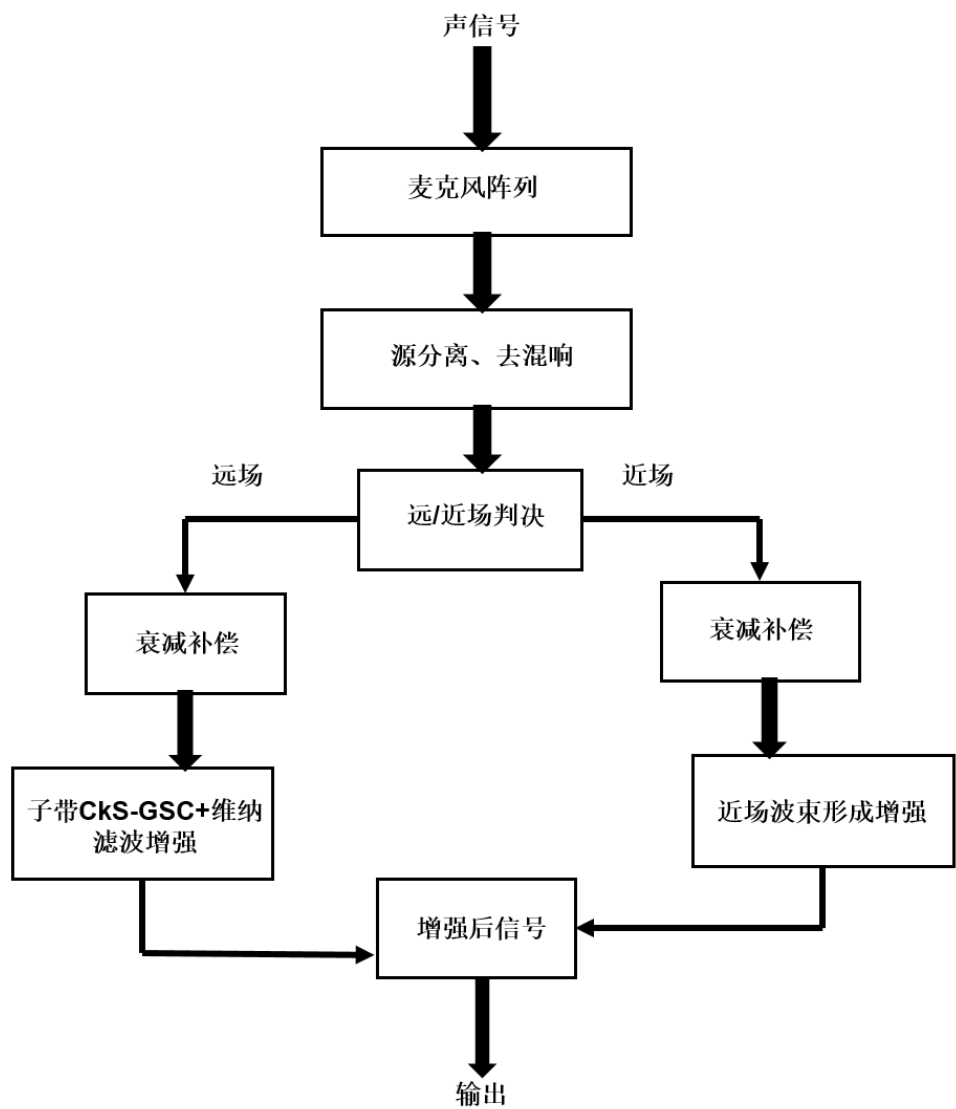


图 12 主要技术搭建实施框架

通过以上关键技术，对声源信号分理出目标声源并去除混响，在通过衰减补偿滤波器提升信噪比，并根据目标声源对远场、近场作判决。通过判决结果，确定目标声源适用远场增强模型还是近场增强模型，实现增强模型自适应切换。远场增强模型能够通过子带 CTF-GSC 算法结合维纳滤波算法有效去除方向及散射噪声，近场模型采用自适应波束形成有效做到增强。有效的寻求适应该场合的增强模型，可提高在智能家居复杂声学环境中增强系统的适应能力，从而取得更好的增强效果。

4、实验过程设计

(1) 搭建信号采集环境

主要器材：线性麦克风阵列（8 个及以上），前置放大器，NI-USB9234 采集卡，计算机^[21]。

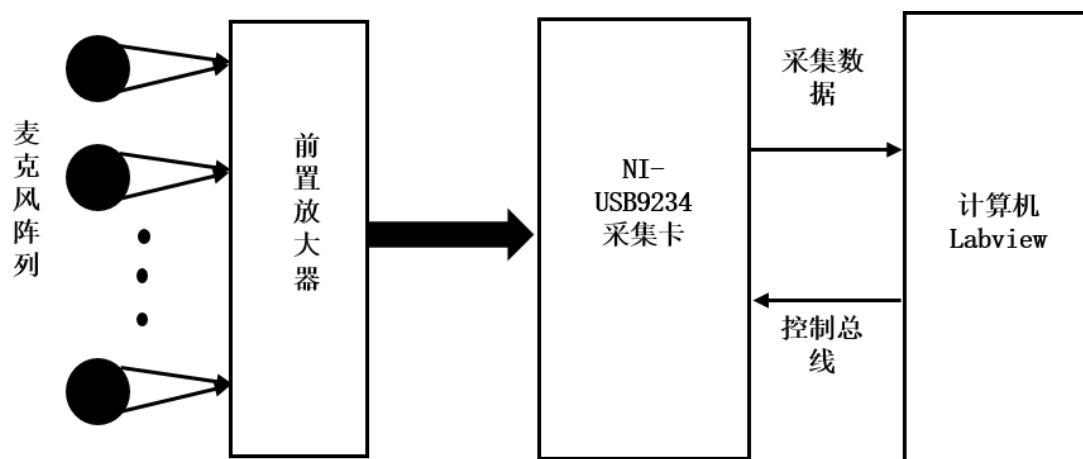


图 13 声信号数据采集系统示意图

(2) 实验仿真软件环境

所用仿真软件：MATLAB 2014a、ProteusPro、keil4、Visual C++。

(3) 研究设计步骤

a) 采用一组中国传媒大学消声室采集的声学数据，仿真模拟出衰减系统，设计出衰减补偿滤波器，并对实验效果做比较，再应用与室内环境并不断优化；

b) 使用一组室内混响下采集的数据，通过 MATLAB 建立仿真实验测试子带 CTF-GSC 算法与维纳滤波算法结合的散射噪声抑制效果，并针对家居环境根据实验数据和效果不断优化；

c) 根据近场语音增强模型建立仿真环境，采用一组近场环境下采集的数据，测试近场环境下自适应波束形成方法及消噪性能，通过声源定位判决测试远场、近场切换下的增强效果，并根据实验结果不断优化；

d) 在经费充足的境况下，结合已有技术，将以上三个实验取得较好效果的方法集成到一个针对家居环境的麦克风阵列语音增强系统上，经软件仿真实现后，再做硬件

设计实现。此过程涉及到软件系统设计和硬件系统设计，需要购买相关器材，如数字信号处理器（建议采用德州仪器生产）、模数转换器（ADC）、数模转换器（DAC）、SDRAM、FlashROM 等，需要一定量的经费支持开展^[20]，NI9234 数据采集卡较为昂贵，考虑采用其他数据采集卡，如 USB-6009 采集卡等。

5、项目研究创新点

（1）在智能家居环境下，提出一种衰减补偿滤波的方法以提高信噪比，达到语音增强的目的；

（2）分析智能家居环境下方向性噪声、散射噪声、混响的特点，结合现有方法，提出一种将子带 CTF-GSC 算法与维纳滤波算法相结合的方法，以达到对远场散射噪声更好的抑制效果；

（3）分析家居环境麦克风阵列近场的模型，提出一种适应智能家居环境的自适应波束近场语音增强方法，并通过声源定位实现近场、远场增强模型的自适应切换。

6. 附主要参考文献：

[1]安琪. 2017 物联网智能家居行业现状分析[N]. 中国建材报, 2017. 08. 24.

[2]陈功正, 王腾, 陆畅等.人工智能时代智能家居行业发展研究[J].合作经济与科技, 2018. 03.

[3] Sharon Gannot, David Burshtein, Ehud Weinstein, Signal Enhancement Using Beamforming and Nonstationarity with Application to Speech.IEEE Transactions on signal processing,VOL.49,NO.8,August 2001.

[4]cob Benesty, Jingdong Chen, Yiteng Huang. Microphone Array Signal Processing[M],2008.

[5]杜功焕, 朱哲民, 龚秀芬. 声学基础[M]. 江苏南京: 南京大学出版社, 2012.

[6]室内环境下基于麦克风阵列的语音增强算法研究[D].广东深圳: 深圳大学, 2016.

[7]雁萌.一种低信噪比条件下的语音增强方法[C].第六届全国人及语音通讯学术会议, 中国深圳, 2001,11,20:75-77.

- [8]积玉, 郑莹, 何衍庆.衰减频率特性及其应用[J].石油化工自动化, 2004,1:34.
- [9]ang H, Chu P. Voice source localization for automatic camera pointing system in video Speech and Signal Processing, Munich, Germany, 1997:187-190.
- [10]静然.基于麦克风阵列的语音增强算法研究[D].四川成都: 电子科技大学, 博士毕业论文, 2007,06.
- [11]荣. 后置维纳滤波和可调波束成形器的语音信号增强[J].火力与指挥控制, 2018,01:43(1).
- [12]abibian S, Akbari A. Noise reduction from speech signal based on wavelet transform and KullbackLeibler divergence[J].International Symposium on Telecommunications, 2008:787-791.
- [13]inwang Wan, Zhenyang Wu. Improved Speech Source Localization in Reverberant Environments Based On Correlation Dimension[C].IEEE Conference on Industrial Electronics and Applications, 2009.
- [14]霞, 周彬, 贾冲等译.麦克风阵列信号处理[M].北京: 国防工业出版社, 2016,01.
- [15] Chatlani N, Soraghan J J. MD-based Filtering (EMDF) of Low-Frequency Noise for Speech Enhancement [J]. IEEE Transactions on Audio Speech and Language Processing, 2012,20(4):1158-1166.
- [16]关晓彬. 基于空间域的麦克风阵列近场声源定位与增强研究[D].四川成都: 西南交通大学, 2012,06.
- [17]胡勇.麦克风阵列语音信号增强算法研究[D].四川成都: 电子科技大学, 2014,06.
- [18] 张小飞, 汪飞, 徐大专.阵列信号处理理论和应用[M].北京:国防工业出版社, 2010.
- [19] Junli Liang, Ding Liu. Passive Localization of Mixed Near Field and Far Field Sources Using Two stage MUSIC Algorithm[J],IEEE Trans Signal Processing, vol.58.2010.
- [20] 林静然.基于麦克风阵列的说话人跟踪及语音分离实现技术的研究[D].成都:电子科技大学, 硕士学位论文, 2005.
- [21]李幸远.麦克风阵列语音增强系统设计与实现[D].陕西西安: 西安电子科技大学, 2008,01.

项目预期研究结果：

- （1）通过声学衰减模型，揭示语音信号室内随距离远近的衰减的系统关系，并通过该关系设计出语音信号衰减补偿增益滤波器，多数室内环境达到 2dB 以上提升；
- （2）通过子带 CTF-GSC 算法与维纳滤波算法相结合的方法，有效的去除方向性噪声和散射噪声，为语音系统散射噪声去除提供理论依据；
- （3）以目标声源定位方法判决远场、近场增强模型，做到准确判决、迅速转换，达到强自适应性效果；
- （4）集成面向智能家居的麦克风阵列增强系统，对比现有方法，在室内环境下做到更好的语音增强效果；
- （5）走访相关智能家居企业，测试算法的前沿性，并做相关技术引荐；
- （6）在相关领域学术期刊上发表 2 篇以上文章，拟投 SCI、EI 类别期刊；
- （7）设计出该方案下面向智能家居的语音增强系统，并申请一份软件著作权。

项目时间进度安排和阶段目标		
	考核时间节点（年/月）	阶段目标（阶段考核指标）
1	2018 年 6 月 —— 2018 年 8 月	<ul style="list-style-type: none">● 项目成员开展项目研究方案及实施细则讨论；● 分配相关技术研究工作，要求查阅大量文献，研究成果需提供充足可信理论依据；● 邀请相关领域硕士导师给研究人员作学术报告；● 邀请相关专家给项目研究成员作学术研讨。
2	2018 年 8 月 —— 2019 年 1 月	<ul style="list-style-type: none">● 检验第一阶段成果，要求达到初步检测效果，提供相关研究步骤实施详细记录，并核查理论依据可信度；● 整理现阶段研究成果，对各小块工作进行整合，实施衰减增益补偿、子带 CTF-GSC 算法与维纳滤波算法相结合、远近场增强模型自适应切换研究

		<p>工作，并规定研究成果提交时间，要求与现有算法作详细对比，并达到更优效果；</p> <ul style="list-style-type: none"> ● 到相关企业参观学习，学习企业研究、开发及制作实施流程，了解企业在阵列增强方面面临的难题，如苏州思必驰公司等。
3	<p>2019 年 1 月 —— 2019 年 6 月</p>	<ul style="list-style-type: none"> ● 检验第二阶段成果，要求相关研究人员对每个模块研究过程做详细陈述，对比现有算法说明优点与不足，并由每位研究人员作点评，讨论相关解决优化方案； ● 核查每个实验环节，经反复经验证确认无误，邀请 3 名以上硕士生导师作审核点评，针对导师意见作修正； ● 验证算法整合后的实验效果，并开始从系统应用角度设计该语音增强软件，并撰写软件设计说明书（使用手册），填写著作申请登记表； ● 再次到企业做相关技术交流，验证算法方案的可行性； ● 分组开始撰写论文。
4	<p>2019 年 6 月 —— 2019 年 7 月</p>	<ul style="list-style-type: none"> ● 将研究过程及结果完整的记录成册，撰写完整的研究报告，并将撰写的论文投至相关期刊（目标期刊为 SCI、EI 类别）。 ● 将著作申请材料提交至相关机构（版权保护中心）审查。
预 期 主 要 成 果 形 式		<p>1、论文论著 ✓ 2、研究报告 ✓ 3、新产品（或农业新品种） 4、新装置 5、新材料 6、新技术（新方法、新工艺） ✓ 7、计算机软件 ✓ 8、其他</p>
三、项目人员情况		
<p>项目负责人主要科研成果：</p> <p>周静，男，1993 年生，中共党员，现为电子与通信工程系硕士在读学生。作为一名语音信号处理方向的硕士研究生，他对语音信号处理已有较为深入的研究，在语音增强、麦克风阵列信号处理、声信号端点检测等方面有一定的基础知识，能够较为熟练使用科大讯飞的六元环状麦克风阵列。硕士一年级期间，他阅读大量声学处理相关论文和书籍，有扎实的声学信号基础；它具备一定的电路基础，曾多次自主完成复杂电路的设计及焊接工作。他精通使用 MATLAB、Proteus Pro、keil4、Visual C++等软件，尤其在 MATLAB 和 C++程序设计上有较好的基础。且他有着较强的组织能力和管理能力，团队意识较强。</p>		

项目组织及管理：

- (1) 制定详细的项目研究进展计划，并严格执行；
- (2) 根据项目内成员相关情况做到明确分工，确保每位成员切身参与到该项目中；
- (3) 定期向相应的指导老师作项目进展汇报，并共同商讨研究过程中遇到的难以解决的问题，以此激励每位项目成员积极主动做好自己的模块工作；
- (4) 定期开展集体活动，增强项目凝聚力；
- (5) 确立以负责人为核心的团队模式，由负责人严格监管项目研究进度；
- (6) 勤于向导师咨询，并根据导师建议作相关调整；
- (7) 在学院的统一管理下，定期作进展汇报。

指导教师情况：

许春冬，男，博士、副教授、硕士生导师，现任江西理工大学信息工程学院电子与通信教研室副主任。主要业绩如下：

1. 发表论文 18 篇，其中 SCI 和 EI 收录论文 10 余篇，2009 年获学院多媒体教学竞赛第二名，教学设计和效果优秀。

2. 主持参与课题及经费：

已结题：

(1) 基于耳机重放的 3D 音频系统 (GJJ13426)，江西省教育厅科技课题一般项目，2 万元，主持。

(2) 以地域为依托的大学生创新创业教育研究与实践 (10YB139)，江西省教育规划课题，0.3 万元，主持。

(3) 时滞广义混沌同步机理及其在视频安全通信中的研究 (A11062002)，国家自然科学基金地区项目，26 万元，参加。

正承担：

(1) 基于麦克风阵列的语音信号缺失频谱修复研究 (GJJ150681)，江西省教育厅科技课题一般项目，3 万元，主持。

(2) 基于张量分析的空间音频信号压缩与重建技术研究 (61571044)，国家自然科学基金面上项目，60 万元，参加。

(3) 基于声学韵律特征的社交人格感知 (61473041)，国家自然科学基金面上项目，80 万元，参加。

(4) 新常态下矿业“或有环境负债”PPP 模式治理机理研究 (15BJY060)，国家社科基金一般项目，20 万元，参加。

(5)单通道语音信号分离与增强关键技术研究 (NSFJ2015-G21), 江西理工大学自然科学基金项目, 1 万元, 主持。

(6)《压缩编码理论》获批 2017 年江西理工大学研究生优质课程, 3 万元, 主持。

3. 江西省高校科技成果奖二等奖: 时滞系统的非线性动力学机理及其在安全通信中的应用研究, 江西省教育厅, 2013 年 9 月, 排名第五。

代表性论文:

1) 许春冬, 张震, 战鸽, 应冬文, 李军锋, 颜永红. 面向语音增强的约束序贯高斯混合模型噪声功率谱估计, 声学学报, 2017, 42 (5) :633-640.

2) **XU Chun-dong**, XIA Ri-sheng YING Dong-wen, LI Jun-feng, YAN Yong-hong. HMM-based noise estimator for speech enhancement[J]. Journal of Beijing Institute of Technology. 2014. 23(4): 549-556.

3) 许春冬, 夏日升, 应冬文, 李军锋, 颜永红. 面向语音增强的序贯隐马尔可夫模型时频语音存在概率估计[J]. 声学学报, 2014, 39 (5) :647-654.

4) 许春冬, 李军锋, 裘嫻, 夏日升, 颜永红. 两扬声器配置下的串声消除系统参数优化设置[J]. 计算机应用, 2014, 34 (5) :1503-1506.

5) 许春冬, 战鸽, 应冬文, 李军锋, 颜永红. 基于隐马尔可夫模型的非监督噪声功率谱估计[J]. 数据采集与处理. 2015, 30 (2) :359-364.

6) 许春冬, 王晶, 战鸽, 应冬文, 李军锋, 颜永红. 基于功率谱包络动态分割的鲁棒语音端点检测[J]. 北京理工大学学报 (自然科学版), 2015, 35 (11) :1189-1193.

7) **Chundong Xu**, Junfeng Li, Risheng Xia, Dongwen Ying, Yonghong Yan, An Analysis-Synthesis Approach for Enhanced Stereo Audio Reproduction, Proceedings of 3DSA2013, Osaka, Japan, 2013.6.26-2013.6.28.

8) **Chundong Xu**, Yiqing Liu. A Design of Embedded Video Server Based on H.264, International Conference on Communication Systems and Network Technologies, CSNT 2012, harbin, china, 2011.01.18-2011.01.20.

9) **Chun-Dong, Xu**, Mu-Gui, Zhou. A BP neural network activation function used in exchange rate forecasting, 2011 International Conference on Information Technology and Agricultural Engineering, Qiqihar china, 2010.12.25-2010.12.26.

10) **Chundong XU**, Jianbing Yi, Jiyuan Wang. Application of genetic algorithm on digital watermark against geometric attacks, 2011 3rd international conference on computer and automation engineering, Chongqing china, 2011.01.21-2011.01.23.

四、项目经费情况				
支出科目	金额（元）	计划根据及理由		
研究设备采购	6419	主要采购实验设备		
		物品名称	数量	单价（元） 该项总价（元）
		电容式驻极体	8	320 2560
		USB-6009 数据采集卡	2	800 1600
		前端放大器	2	400 800
		D/A 转换器+控制主板（高精度，4 通道以上）	2	150 300
		A/D 转换器+控制主板（高精度，4 通道以上）	2	120 240
		DSP 数字信号处理器 TMS320C6713B	2	185 370
		Flash ROM	4	30 120
		SD RAM	4	10 40
		电容（一套）	1	20 20
		电阻（一套）	1	20 20
		跳线（20cm,10cm）	2	20 20
		电烙铁	2	55 110
		焊锡丝（一卷）	1	19 19
		焊接板	20	40 80
		麦克风支架	1	120 120
论文发表版面费	6000	发表两篇论文		
软件著作权费用	2000	申请一份智能家居语音增强系统的软件著作权		
学术交流	800	邀请 3 位相关专家开展论坛、研讨等相关学术活动		
企业交流经费	600	用于与相关企业交流探讨活动的支出		
其他支出	400	研究室用于研究的其他用品		
合计	16219	立项若经费不足，可向学院或导师申请，自由调整。		

