# 百度Apollo车载语音DSP硬件设计方案

# 一、概述

## 1.1简介

随着人工智能的快速发展，网联车已经是大势所趋，车载语音需求与日俱增。现在互联网最重要的终端，是智能手机。**未来能源网、物联网最重要的终端，则将是汽车！**它们将融合现代能源技术、通信与网络技术，实现车与人、车、路、云等（V2X）能源和智能信息交换、共享。由此，百度推出了全球首款人车AI交互系统----Apollo小度车载系统，该系统集成了百度车载语音、百度地图、AR导航、人脸支付等核心能力，其中车载语音是Apollo小度系统的重要组成部分，本文主要介绍车载语音DSP硬件设计方案。

这套方案分为前端信号采集、语音信号处理，前端信号采集是通过一组MIC阵列（2-3MIC）来实现的，语音信号处理则是将DSP芯片嵌入到车机主板上，由DSP进行回声消噪、定向拾音等处理，最后将DSP将结果信号输送到主CPU层，继而灌入百度语音ASR引擎。

## 1.2为什么一定要选用MIC阵列与DSP结合的方式呢？

其一，车载噪声众多，包括风噪、胎噪、路噪、引擎声、空调风扇等，如果只是选择单一MIC，是无法应对车载复杂环境的；

其二，如果只选用单一MIC，无法实现Beamforming，定向拾音等核心功能；

其三，除了要应付车内噪音外，要实现好的语音体验，还需要解决车内音乐的影响，即需要实时的回声消除能力，而车规级CPU的运算能力比消费品类芯片是要低很多的，此时要进行回声消噪，CPU的性能往往会成为瓶颈。因此，需要一款车载DSP来实时处理回声消除。

## 1.3主要特点

* 回声、混音全面清除

*两通道立体声回声消除（AEC），同时清除回声抑制和含糊混音性能之间的影响*

* 高质量噪音清除

*在不影响语音质量的前提下将噪声进行消除，加强麦克风信号，优化语音电话沟通，强化语音识别。*

* 过滤胎噪、风噪、引擎噪声等

*清除风声噪音（开空调、车窗和敞篷引来的风声）、发动机和轮胎噪音对信号的影响，改进车用性能。*

* 声源定位及限制

*能检测双麦克风形成的180°夹角内的声源位置角度，特殊场景下可提高方向辨识性能。也可以限制对某个角度范围的语音进行识别，抑制设定范围外的噪音影响。*

# 二、性能

## 2.1回声消噪

Apollo小度系统中的硬件回声消除技术是基于DSP平台来设计的，如下图所示，车载场景下，人声(Speech)从麦克风(MIC)传入，而从扬声器(Speaker)播放的音乐声，此时也会以回声(Echo)的形式传入麦克风(MIC)，形成了Speech+Echo的混音信号。如果不作回声消除处理，将会给语音识别带来很大的不利影响，现有语音识别技术无法自动从一路信号输入识别出这一路信号包含几种声源，为了解决这种问题，需要引入参考信号(speaker data)。以下是声学回声消除的详细解释：

*声学回声消除(AEC)是对扬声器信号(Speaker)与由它产生的多路径回声(Echo)的相关性为基础，建立远端信号的语音模型，利用它对回声进行估计，并不断地修改自适应滤波器的系数，使得估计值更加逼近真实的回声。然后，将回声估计值从话筒的输入信号中减去，从而达到消除回声的目的，AEC还将话筒的输入与扬声器过去的值相比较，从而消除延长延迟的多次反射的声学回声。*

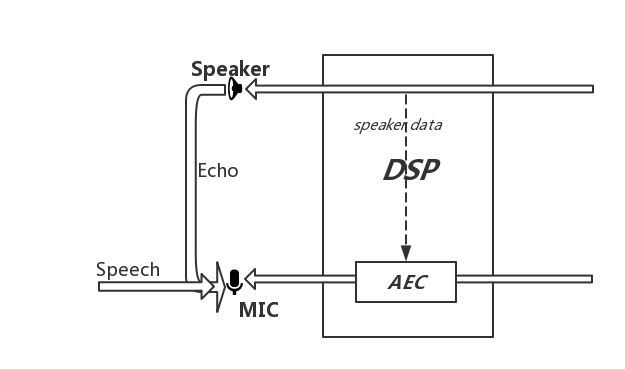


Figure 1 - AEC示例图

## 2.2效果展示

如下面三幅图所示，Figure2是AEC处理前的MIC信号，Figure3是输入给DSP处理的参考信号，Figure4是AEC处理后的结果信号，可以看出，经过回声消噪之后的信号信噪比提高了很多。

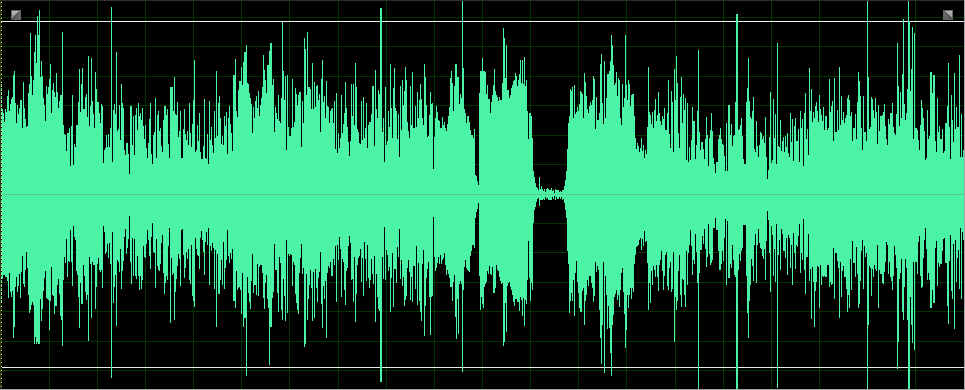


Figure 2 - MIC信号

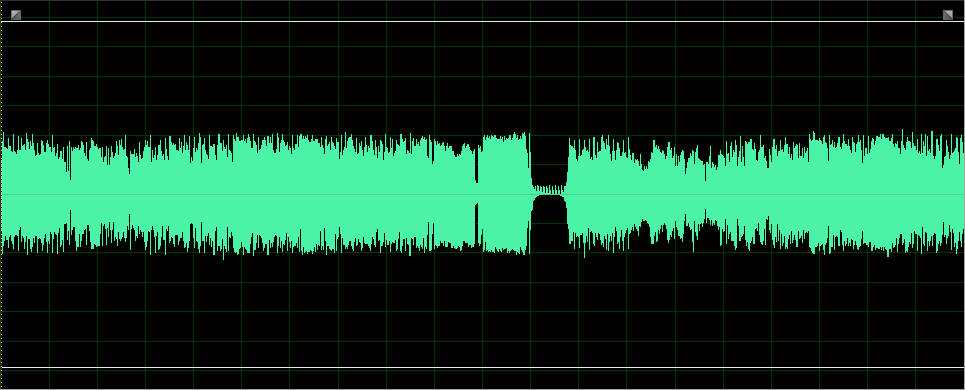


Figure 3 - Speaker信号

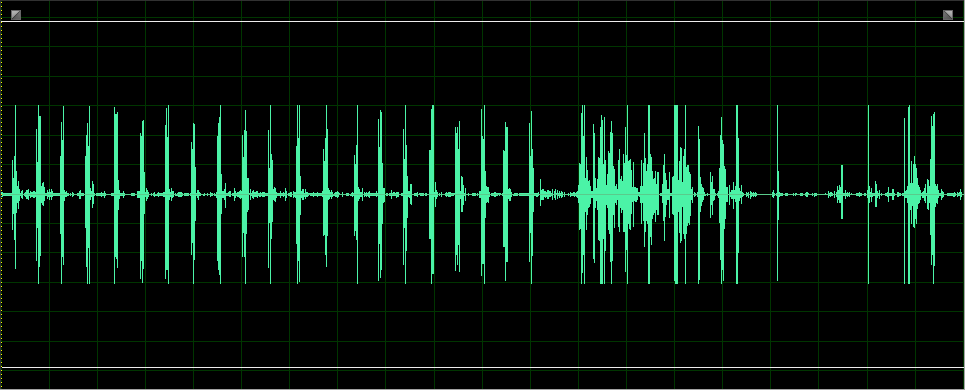


Figure 4 - 结果信号

## 2.3硬件设计参考方案

我们推荐以下几种设计方案：

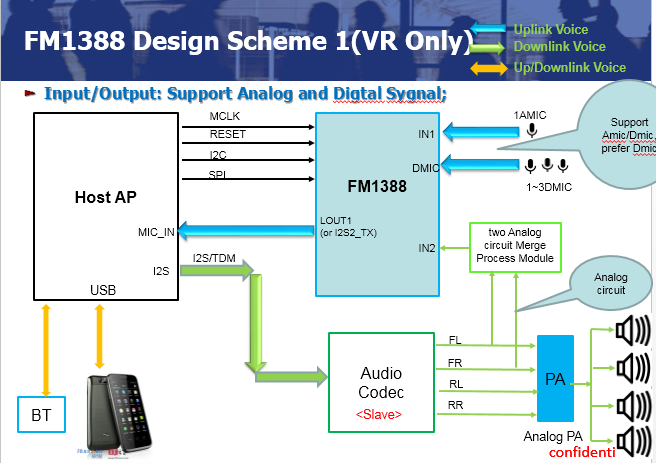


Figure 5 - Design Scheme 1

如Figure5所示，上行链路，信号输入支持1个模拟MIC或者1-3数字MIC，信号输出支持LOUT模拟信号输出或者I2S数字信号输出；下行链路，由主机端通过I2S输出到AudioCodec，再由AudioCodec选择前左、前右两路信号进行模拟合并成一路模拟信号，最后塞入IN2端口。

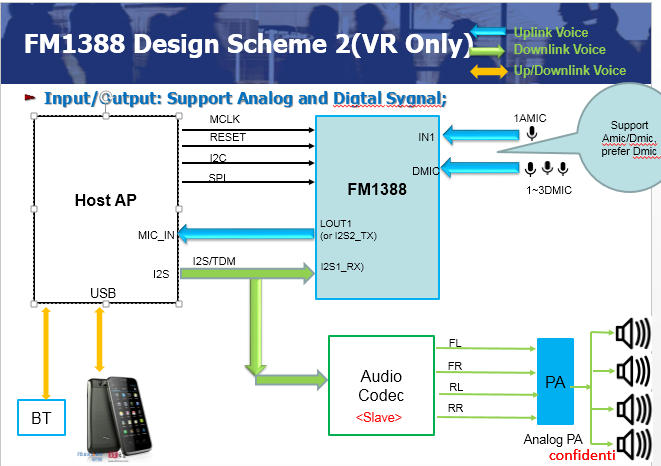


Figure 6 - Design Scheme 2

如Figure6所示，上行链路，信号输入支持1个模拟MIC或者1-3数字MIC，信号输出支持LOUT模拟信号输出或者I2S数字信号输出；下行链路，由主机端通过I2S输出到AudioCodec，同时分出一路信号输送给DSP，此方案就无需再合成回路信号输入DSP，但缺点是AudioCodec可能会进行音效处理，因此输送的参考信号跟Echo信号相差会增大。

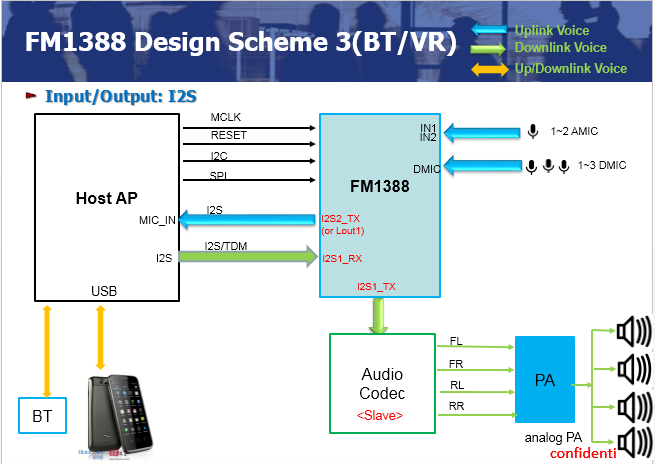


Figure 7 - Design Scheme 3

如Figure7所示，上行链路，信号输入支持1个模拟MIC或者1-3数字MIC，信号输出支持LOUT模拟信号输出或者I2S数字信号输出；下行链路，由主机端通过I2S输出到DSP，DSP处理后输出给AudioCodec，此方案既无需合成回路信号输入DSP，也无需分出一路信号给AudioCodec，但缺点同方案二一样，AudioCodec可能会进行音效处理，因此输送的参考信号跟Echo信号相差会增大。

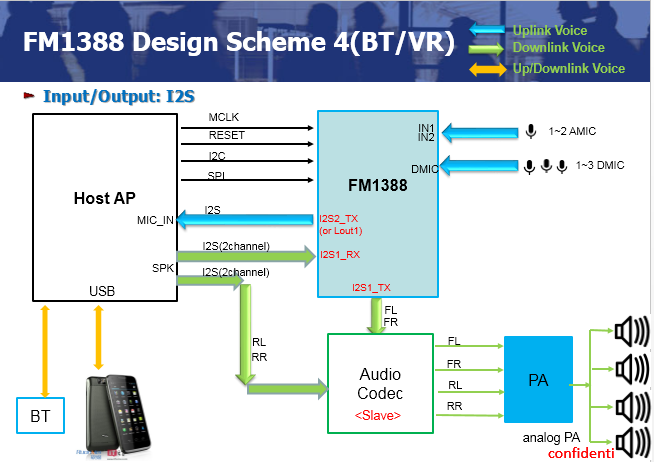


Figure 8 - Design Scheme 4

如Figure8所示，上行链路，信号输入支持1个模拟MIC或者1-3数字MIC，信号输出支持LOUT模拟信号输出或者I2S数字信号输出；下行链路，由主机端通过I2S输出两路通道（FL\FR）

两路通道（RL\RR）到AudioCodec，同时分出一路信号输送给DSP，此方案就无需再合成回路信号输入DSP，但缺点是AudioCodec可能会进行音效处理，因此输送的参考信号跟Echo信号相差会增大。

以上四种设计方案，可供合作OEM厂商根据自己的情况进行选择，或者依据实际情况再制定新的合作方案。

## 2.4我们所推荐的接入步骤

1、百度车联网与OEM厂商完成第一轮workshop，完成前期接洽，包括芯片应用领

域应用情况，芯片价格，供货情况等，以便确定合作方式。

2、百度提供相关DSP设计参考资料以及芯片DataSheet给OEM厂商作硬件设计参考。

3、OEM厂商要提供方案设计框架图，以及电路设计框图给百度确认；

双方确认没问题后，厂商根据电路图做出PCB Layout。同时厂商软件驱动工程师也需要

并行集成驱动程序。

4、厂商 PCBA 回来后，先自行测量各个电压供电是否OK，时钟是否OK，测量上电时

序是否符合要求等，确保硬件正常工作。软件驱动工程师编写驱动代码并初始化

DSP，确保DSP 能进入正常工作状态。

5、安装完整结构机壳，联系百度车联网，安排双方联调时间

# 三、硬件设计规格

## 3.1 MIC设计规格

百度DuerOS MIC阵列方案对MIC并无特别要求，只要是正规厂家生产，并合格出厂的产品，均能满足要求。但是为了获得理想的效果，我们还是推荐MIC尽量达到以下要求：

* SNR >= 60dB
* Sensitivity: Analog Microphone，-38dB(±3dB)；
* Digital Microphone, -26dB(±3dB)
* MIC阵列设计中，2个MIC之间的距离4-20CM。
* MIC阵列设计中，如果是2个AMIC，生产厂家需要确保AMIC的相位一致性。

备注：如果灵敏度比以上参考值还高，效果会更好。

## 3.2 DSP硬件平台

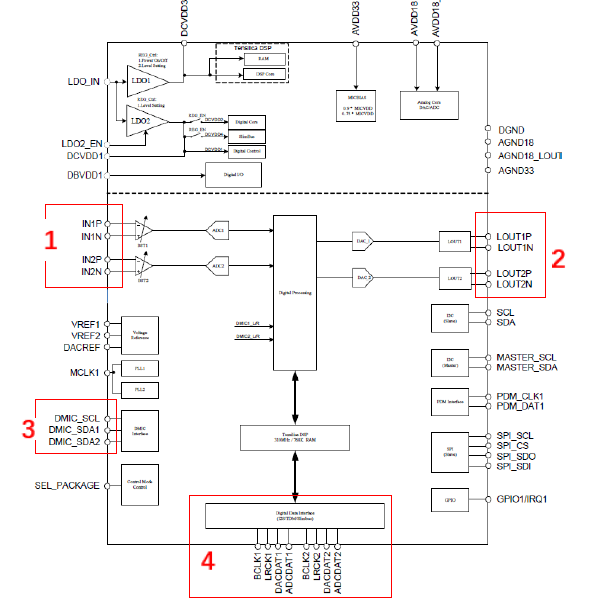


Figure 9 - DSP接口图

### 1、硬件规格：

* 1）具有2组ADC模拟输入接口，可以作为模拟信号输入，或者接2个模拟MIC。其THD+N可以到-92dB, SNR达到98dB.
* 2）具有2组DAC模拟输出接口. THD+N可以到-90dB, SNR达到101dB.
* 3）具有2组DMIC输入接口，可以接4个DMIC，富迪FM1388回声消噪算法目前可以支持到3个MIC同时工作。
* 4）具有2组独立的I2S/PCM/TDM数字接口. 可以支持到24-bit,8KHZ ~ 192KHZ

### 2、主要硬件接口

* Highly integrated mixed-signal IC for voice processing
* On-chip advanced DSP subsystem running advanced voice processing algorithms
* 2 ADCs for analog microphone inputs , -95dB THD+N, 102dBA SNR
* 2 DACs for analog line level outputs, -90 dB THD+N, 100 dBA SNR
* 2 analog pre-amp providing amplitude boosts of +20/24/30/35/40/44/50/52 dB
* Line-in and line-out digital signals via digital interfaces: Two 24-bit/8KHz~192kHz
* I2S/TDM digital interfaces.
* Supports up to 4 digital mic inputs, and 2 analog mic inputs
* Co-processor mode operation:
* Host boot from I2C interface
* Host interface flexibility:
* Control: master/slave I2C, SPI
* Data: TDM, I2S
* Automotive-grade, TQFP-48 pin
* Consumer-grade, TQFP-48 pin

### 3. 主要性能指标

* Telephone, wideband(HD), and super-wideband voice support
* 2/3/4 Microphone Array processing for noise reduction
* Voice Recognition enhance mode
* Acoustic echo cancellation with linear and proprietary non-linear filtering
* Fast convergence
* 60dB ERLE on acoustic echo cancellation
* 12 ~ 30 dB noise suppression
* Bright Voice Enhancement (BVE) to provide intelligent equalization on downlink

voice

* Automatic Gain Controls
* Dynamic Range Controls
* Parametric equalization for uplink and downlink voice signals
* VDA/P.1100/P.1110 tests
* Non-invasive tuning support