1. **Overview**
   1. **Introduction**

With the rapid development of AI, the connected vehicle has become a trend, and there is an increasing need for smart vehicle voice system. Currently the most important terminal of Internet is the smart phone, while in the future, the automobile would be the most crucial terminal of Internet of Things and Internet of Energy. They would integrate with advanced technologies of energy, telecom and Internet to achieve the vehicle’s information sharing with people, other vehicles, road and cloud, etc., which is also referred to as V2X. To follow the trend and lead the industry, Baidu has launched the word’s first AI Interaction System between Human and Vehicle—Apollo DuerOS. Apollo DuerOS consists of technologies of Baidu Car Audio, Baidu Map, AR Navigation, Face Payment and so on, among which Car Audio is an important component. This article would mainly cover the DSP hardware solution.

The solution includes two parts: front-end signal acquisition and voice signal processing. Front-end signal acquisition is completed through a set of MIC array (2~3 MIC), while the voice signal processing is achieved by DSP. The DSP chips embedded into the motherboard of the vehicle would work on echo’s noise cancellation and directional voice pick-up, and then convey the processed signal to the main CPU layer, and finally the signals would flow into the Baidu Voice ASR Engine.

* 1. Why the combination of MIC series and DSP is chosen?

Firstly, there exists a lot of different noise in vehicle, e.g. noise from wind, running tiers, engine, air-conditions and so on. Therefore, if only one MIC is used, it’s impossible to cope with the complicated vehicle environment.

Secondly, with only one MIC, core functions like Beamforming, Directional Voice Pick-up would not be able to be completed.

Thirdly, to guarantee a comforting voice experience, in addition to handling the noise in-side the car, the impact of in-car music must also be addressed, that is to say, real-time echo cancellation capabilities is required. However, the computation power of vehicles-class CPU is far lower as of chips of other consumption goods, resulting in a bottleneck for echo noise cancellation. Therefore, an onboard DSP is needed to handle echo cancellation in real time.

* 1. **The main features**
* Fully clearing echo and mixing audio

Two-Channel Stereo Echo Cancellation (AEC) could remove the influence of echo suppression and ambiguous mixing performance.

* High quality removal of noise

It is able to eliminate noise, strengthen the microphone signal, optimize telephone function, and strengthen voice recognition without affecting the voice quality.

* Filter tier noise, wind noise, engine noise and so on

It can eliminate the impact of wind noise (from air conditioning and windows), engine and tire noise on the signal and improve vehicle performance.

* Sound source positioning and restriction

It can detect the sound source position within an angle of 180 ° formed by two microphones, and can improve the direction identification performance in special scenes. The voice recognition can be set within a specific range of angle, and the noise influence outside range could be blocked.

1. **Capabilities**
   1. **Echo cancellation**

The hardware echo cancellation technique in the Apollo DuerOS is designed based on the DSP platform. As shown in the figure below, in the vehicle scene, the speech is transmitted from the MIC and the sound of the music played from the speaker would also spread into the MIC as echo, which forms the speech + Echo mixing signal. Without echo cancellation processing, the speech recognition will be greatly adversely affected. However, existing speech recognition technologies can not automatically recognize that the number of sound sources from one signal input includes. In order to solve this problem, Speaker data is needed. The following is a detailed explanation on acoustic echo cancellation:

Acoustic Echo Cancellation (AEC) is a speech model of a far-end signal that is based on the correlation of the speaker signal with the multipath echo (Echo) produced by it. It is used to estimate the echo and is constantly modified from the coefficients of the self-adapted filter to make the estimation closer to the true echo. The echo estimate is then subtracted from the microphone's input signal to eliminate the echo, and the AEC also compares the microphone's input with the speaker's past value, then eliminates the delayed and prolonged multiple reflected echo.

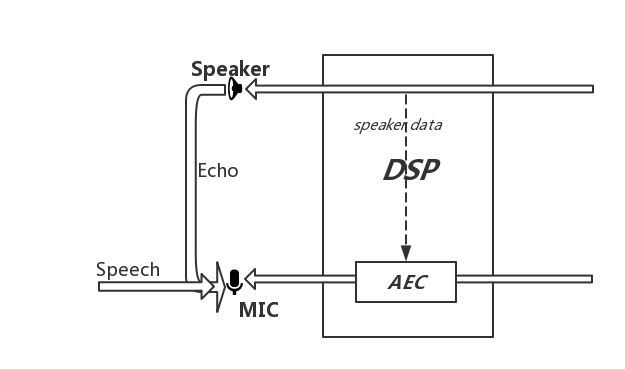


Figure 1 - AEC example diagram

* 1. **Performance display**

As is shown in the following three figures, Figure 2 is the MIC signal before AEC processing, Figure 3 is the reference signal input for the DSP, and Figure 4 is the resulted signal after AEC processing. It can be seen that the signal-to-noise ratio has improved a lot.

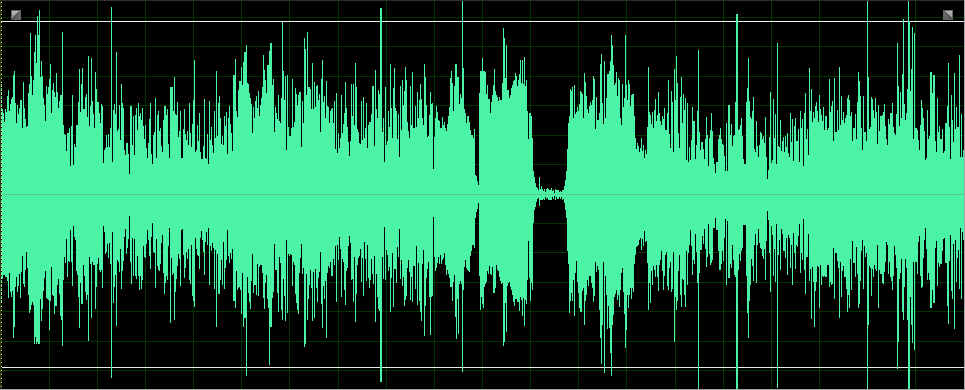


Figure 2 - MIC signal

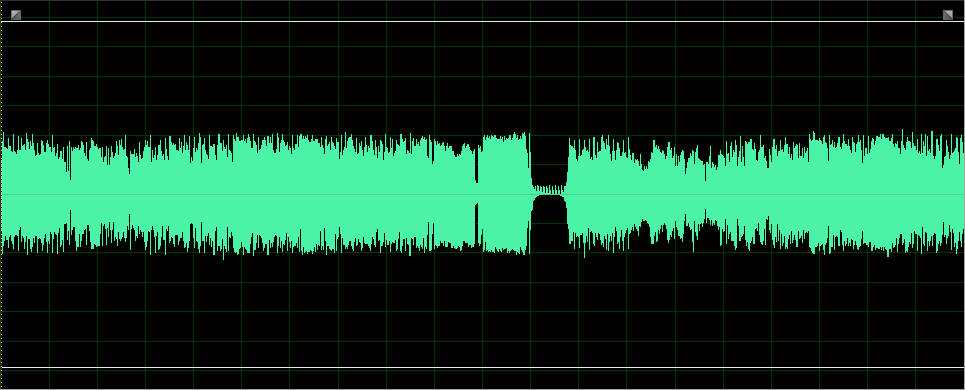


Figure 3 - Speaker signal

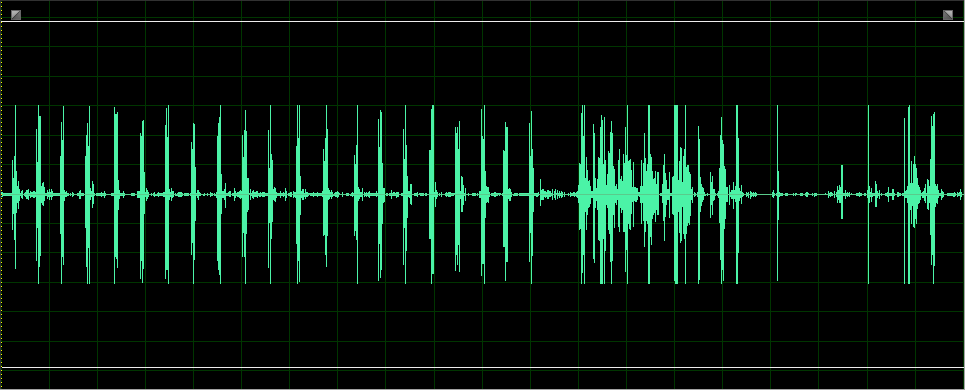


Figure 4 - Resulted Signal

* 1. **Reference solution for hardware design**

We recommend the following design solutions:

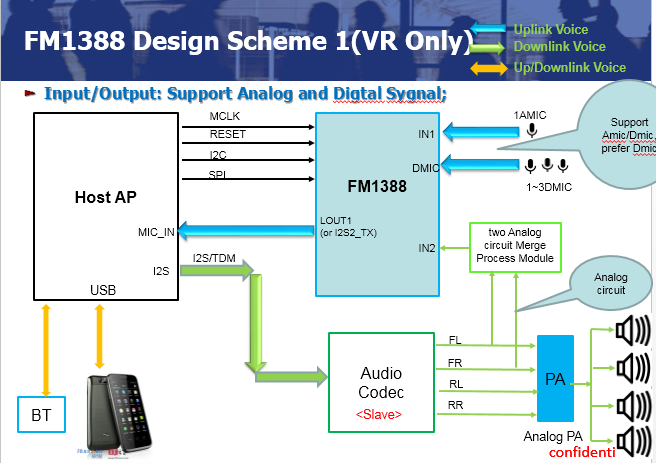


Figure 5 - Design Scheme 1

As is shown in Figure 5, on the uplink, the signal input supports an analog MIC or 1-3 digital MIC, and the signal output supports LOUT analog signal output or I2S digital signal output; while on the downlink, signal is transmitted by the host through the I2S to the AudioCodec, and the AudioCodec would select the left and right signals to simulate and merger them into an analog signal, and finally the signal is stuffed into the IN2 port.

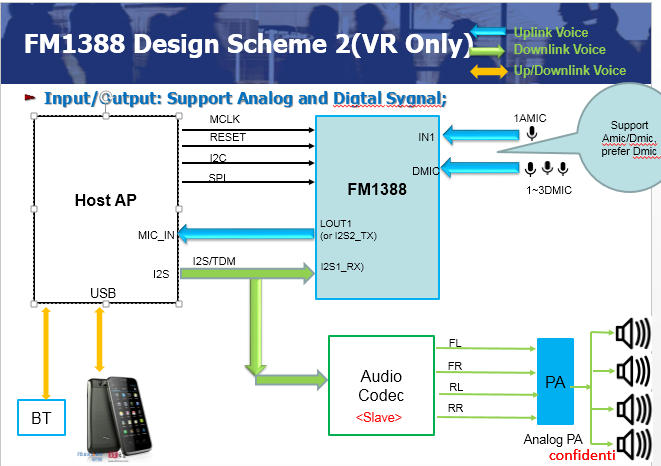


Figure 6 - Design Scheme 2

As shown in Figure 6, on the uplink, the signal input supports an analog MIC or 1-3 digital MIC, and the signal output supports the LOUT analog signal output or I2S digital signal output; while on the downlink, the signal is transmitted by the host by I2S to the AudioCodec, and at the same time divide one way of signal to DSP. There is no need to synthesize the loop signal to input DSP again for this scheme, but the disadvantage is that AudioCodec may deal with the sound effect, so the difference between the delivered reference signal and Echo signal will increase.

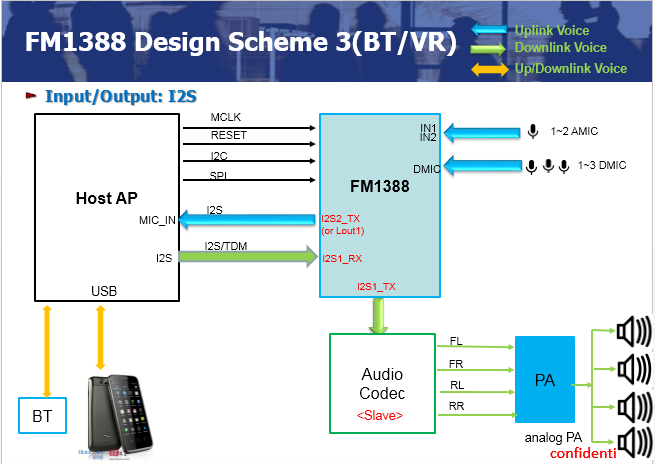


Figure 7 - Design Scheme 3

As shown in Figure 7, on the uplink, the signal input supports an analog MIC or 1-3 digital MIC, and the signal output supports LOUT analog signal output or I2S digital signal output; while on the downlink, the signal is transmitted by the host through I2S to the DSP, and after processed by DSP it would be output to the AudioCodec. The solution neither needs the synthesis loop signal to be input to the DSP, nor needs to separate the signal to the AudioCodec. However, like the second scheme, the AudioCodec may process the sound effect, so the difference between the reference signal and the Echo signal would Increase.

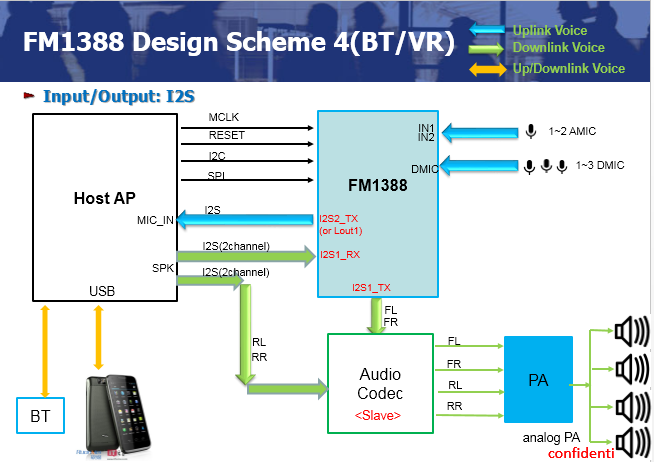


Figure 8 - Design Scheme 4

As is shown in Figure 8, on the uplink, the signal input supports an analog MIC or 1-3 digital MIC, and the signal output supports the LOUT analog signal output or I2S digital signal output; while on the downlink, the signal is transmitted by the host by I2S to two channels ((FL \ FR) and (RL \ RR)) to the AudioCodec, and at the same time one way of signal is separated and sent to the DSP, The scheme no longer needed to synthesize the loop signal input DSP, but the disadvantage is that AudioCodec may carry out sound processing, so the difference between reference signal sent with Echo signal will increase.

The above four design options are for OEM manufacturers to choose according to their own situation, or they can develop a new cooperation solution based on the real situation.

* 1. **Our recommended access procedure**

1. Baidu and OEM manufacturers cooperates to complete the first round of workshops and the pre-contact about the chip’s applications, prices, availability etc. so as to determine the way of cooperation.
2. Baidu provides relevant DSP design reference materials and chip DataSheet to OEM manufacturers for hardware design reference.
3. OEM manufacturers needs to provide the framework of scheme design and circuit diagram design to Baidu for confirmation;

After both sides’ confirmation, the manufacturer makes the PCB Layout according to the circuit diagram, while at the same time the manufacturer’s software engineers need to integrate the driver program.

1. After the PCBA, the manufacturer should make sure whether the voltage power supply and clock is available and the power ordering complies with the requirement on their own to guarantee that the hardware functions well. And the software engineers should code the driver program and initialize DSP to ensure that DSP is in a normal working condition.
2. After the installation of a complete structure, contact Baidu for arrangements for co-adjustment.
3. **The hardware design specifications**
   1. **MIC design specifications**

Baidu DuerOS MIC array solution has no special requirements on the MIC. As long as they are qualified products manufactured by legal companies, they are acceptable. However, in order to achieve the desired effect, we recommend that the MIC meet the following requirements as much as possible:

* SNR> = 60dB
* Sensitivity:
* Analog Microphone, -38dB (± 3dB)
* Digital Microphone, -26dB (± 3dB)
* in MIC array design:
* the distance between 2 MIC is within 4-20CM
* if there are 2 AMIC, AMIC’s phase consistency should be guaranteed.

Note: If the sensitivity is higher than the above reference value, the effect will be better.

* 1. **DSP hardware platform**

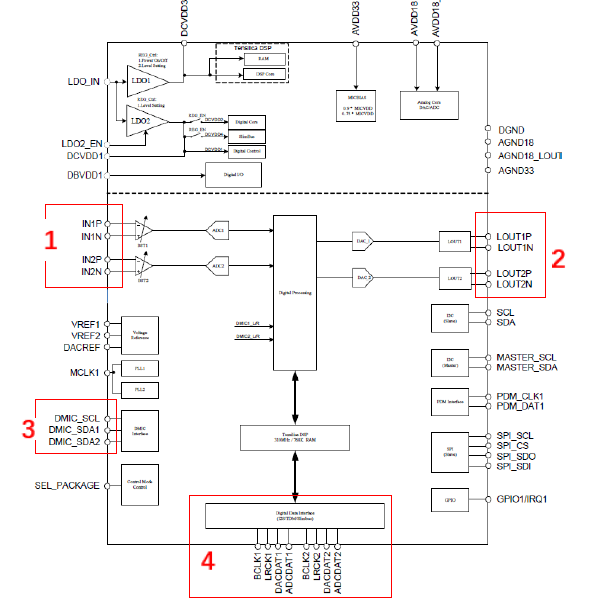


Figure 9 - DSP Interface Diagram

1. **Hardware specifications:**

* There are 2 ADC analog input interfaces which can be used as analog signal input, or to connect 2 analog MIC. Its THD + N can reach -92dB and SNR 98dB.
* There are 2 sets of DAC analog output interfaces. THD + N can reach -90dB and SNR 101dB.
* There are 2 groups of DMEG input interface which can 4 DMICs. Fudi FM1388 echo de-noising algorithm can support up to 3 MIC working at the same time.
* There are 2 sets of independent I2S / PCM / TDM digital interface, which can support up to 24-bit, 8KHZ ~ 192KHZ

1. **The main hardware interface:**

* 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

1. **The main performance indicators:**

* 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