

07 CarPlay Enhanced Siri

已关联2个iCafe卡片

福特phase6项目语音CarPlay Enhanced Siri软件需求说明书

文档编号：

保密等级：机密

百度ASD工程交付中心
(版权所有，违法必究)

0 修订历史

1	版本号	修订日期	修订记录	修订人	审核人
2	0.00.01	2023/01/11	初稿，定义carplay enhanced siri降噪功能	周倩雯	
	0.00.02	2023/03/09	因821为林肯车型，删除小度小度唤醒词（在林肯、电马车型不再支持）。结合项目情况标注TBD。输出客户确认	周倩雯	
3	0.00.03	2023/03/13	根据review意见，修改。 1.4删除参考文档；修改3.1.4.2 具体进入KWorVAD模式由Apple确认；补充3.4.2.1说明carplay只支持主驾唤醒。	周倩雯	
4	1.00.00	2023/03/22	根据福特review意见，修改。 应福特意见：修改3.1.4.2表格/3.1.4.3，自定义唤醒词和场景化命令词在Carplay页面在前台时屏蔽。补充YF的硬按键交互表格。	周倩雯	

1. 概述

1.1. 背景价值

CarPlay是苹果公司为车载系统开发的一项功能，该功能将iOS设备的界面通过USB或wireless(蓝牙+WiFi)投影显示在车载屏幕上，可以让驾驶员在车内通过车载屏幕直接方便地使用iOS设备上的功能。例如，在车载屏幕上直接使用iPhone或iPad上的地图、导航、Siri语音助手、音乐、电话、短信等各项iOS上的服务及功能。

换句话说，就相当于是在自己的车载系统上安装了一个“iOS操作系统”。该功能在很大程度上减少了驾驶员在车内对iPhone或iPad的直接使用及操作，从而降低了对驾驶员在车辆驾驶过程中的干扰。

苹果在2021年底更新了carplay验收标准，要求支持本地VR的车辆，需要支持siri相关的语音唤醒。对于后续需要接入carplay的车型，均需要通过苹果指定实验室的测试验收，验收通过后才可安装carplay的资格。

合作模块

1	企业	Carplay负责内容	进展&任务排期
2	福特	车企（需求来源方）	
3	延峰YF	普通siri	
	百度	Enhanced siri 降噪	1、整体计划以项目侧计划为准

2023/3/22 14:45	07 CarPlay Enhanced Siri		
4	提供Enhanced siri 语音基础能力：负责信号处理、唤醒、识别、语音合成等模块的实现，满足carplay模型语音能力。		
5	未定 依赖福特确认	carplay认证实验室： 1、carplay认证机构。 2、澄清并输出carplay场景需求 3、协助百度通过carplay认证。	
6	苹果	carplay接入需求来源方：对carplay功能做澄清，为接入carplay提供详细说明	澄清文档见附件

1.2. 名词解释

Ø

1.3. 阅读对象

百度：福特项目组CPM、PM、RD、UX、QA等团队
福特：对应产品团队

1.4. 参考文档

/

2. 需求列表

1	No.	Level1	功能描述
2	1	音频输入	满足苹果对音频输入要求
3	2	前端信号处理	满足苹果对音频信号处理要求
4	3	唤醒	在KW模式，支持标记对应事件及时间戳 并发送给carplay
5	4	语音识别	在VAD和KW模式，支持标记对应事件及时间戳 并发送给carplay

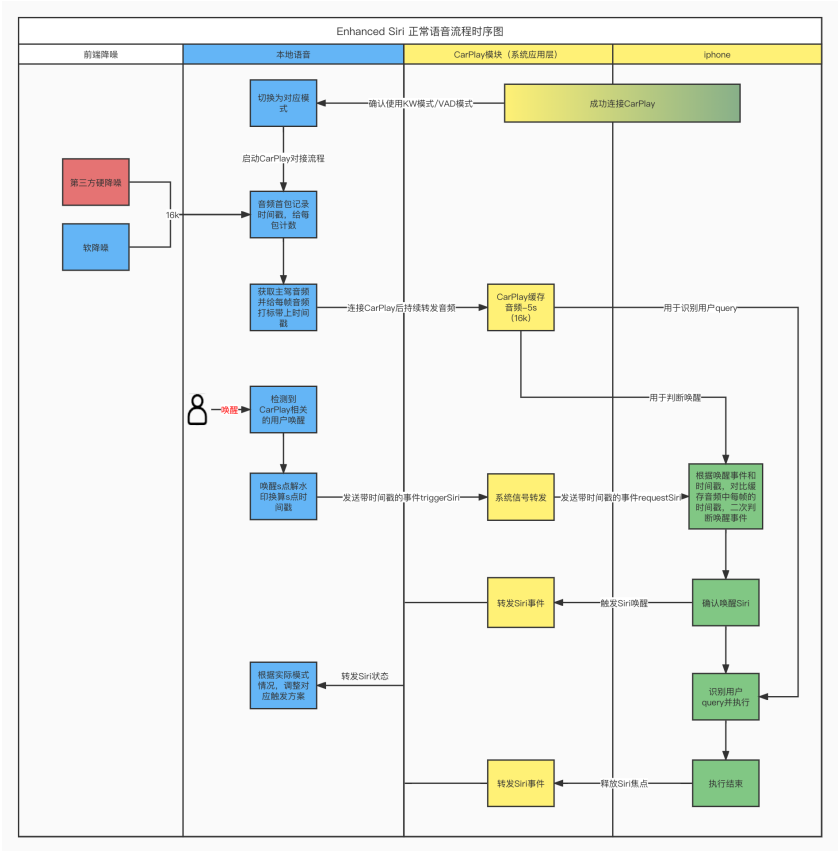
3. 需求描述

3.1. 总体产品设计

3.1.1. 概述

百度负责处理Carplay Enhaced Siri部分的音频输入与信号处理，普通siri部分由其他供应商负责。

3.1.2. Carplay唤醒原理图



该时序图仅表示正常运转流程，重复唤醒/硬按键/软按键的打断情况根据对应功能描述确定

3.1.3. 整体需求描述

百度提供Enhanced siri 语音基础能力：负责信号处理、唤醒、识别、语音合成等模块的实现，满足carplay模型语音能力。

3.1.4. 用户场景

3.1.4.1交互模式介绍

- 1. 普通模式，使用百度信号处理，语音支持唤醒、识别等需求，使用项目的定制化唤醒词“你好林肯” 场景化指令词可唤醒。
- 2. Carplay的关键词/KW模式，使用百度信号处理，音频流持续传输，语音支持“hey siri”的唤醒和“s”的时间戳透出。使用hey siri可唤醒，由apple手机端进行识别。
- 3. Carplay的vad模式，使用百度信号处理，音频流持续传输，语音支持vad的切分，给出起尾点。语音不支持唤醒，全部由apple手机端进行识别。

3.1.4.2 模型共存/互斥状态说明

连接Carplay后从普通VR模式进入Carplay模式，由iOS操作系统检测，具体执行KW 还是VAD模式，由Apple回传的信号决定。手机语言环境和车机语言环境是否一致，若手机的使用语言和车机的使用语言一致，则进入Carplay的关键词模式；若不同，则进入Carplay的vad模式。

详细描述见下图。

1	百度唤醒模型支持功能	静默态	本地VR活跃态	CarPlay活跃态
2	未连接CarPlay	小度小度 定制唤醒词“你好林肯” 免唤醒词 自定义唤醒词	小度小度 定制唤醒词“你好林肯” 免唤醒词 自定义唤醒词	/
	已连接CarPlay-关键词KW模式	小度小度 定制唤醒词“你好林肯” 免唤醒词（在Carplay页面在前台时屏蔽） Hey Siri/Siri	小度小度 定制唤醒词“你好林肯” 免唤醒词（在Carplay页面在前台时屏蔽）	重新触发唤醒： Hey Siri/Siri
4	已连接CarPlay-VAD模式	VAD模型	小度小度 定制唤醒词“你好林肯”	重新触发唤醒： VAD模型

- 1、静默态表示当前车机正常启动，无用户说话，无唤醒事件发生。
- 2、本地VR活跃态：表示本地VR正在识别或者播报tts。
- 3、CarPlay活跃态：表示CarPlay已被唤醒，由CarPlay模块处理相关逻辑。

- 4、在连接CarPlay之后需要持续不断送音频给到CarPlay模块，断开CarPlay连接之后停止。
- 5、如遇Carplay意外退出情况，则默认回到未连接Carplay-静默态
- 6、当本地VR唤醒检测打开时，可语音输入，**小度小度**、你好林肯重新进入本地VR识别态；本地VR识别态当本地VR识别检测打开时，可语音输入，**小度小度**/你好林肯重新进入本地VR识别态；当Carplay的关键词检测打开时，使用hey siri可唤醒，由apple手机端进行识别，进入Carplay活跃态；当Carplay的VAD检测打开时，语音不支持唤醒，全部由apple手机端进行识别，进入Carplay活跃态。
- 7、百度处理前端音频，在连接CarPlay之后需要持续不断送音频给到CarPlay模块，断开CarPlay连接之后停止。
- 8、**CarPlay活跃态时，需要正常上报相关唤醒事件，如果不满足唤醒要求苹果会自行拦截。（1114更新）**

3.1.4.3 语音唤醒特殊相关要求（由车企或tier1输入）

- 1、自定义唤醒词：**在Carplay页面在前台时屏蔽。建议全部屏蔽。避免产生自定义唤醒词和siri相近造成误唤醒**
- 2、场景化命令词：**在Carplay页面在前台时屏蔽。建议全部屏蔽。避免造成可能存在的one-shot系统时延问题，避免潜在bug。例如query「Hey Siri，导航到火车站」**
- 3、硬/软按键唤醒：连接carplay后仍然可以使用方向盘硬/软按键唤醒语音（系统侧YF负责的硬按键交互表格如下供参考）。

1	CarPlay Siri与硬/软语音按键交互	车机当前状态	短按语音硬按键 (<0.6s)	长按语音硬按键 (>=0.6s)	点击本地语音软按键	长按Carplay内Home button
2	未连接Carplay	本地VR未激活	唤醒本地VR，开启VR识别	提示用户进行蓝牙配对，Carplay连接	唤醒本地VR，开启VR识别	不存在
3		本地VR已激活	会退出当前VR，不会重新唤醒	退出当前VR，提示用户进行蓝牙配对，Carplay连接	会退出当前VR，不会重新唤醒	不存在
4	已连接Carplay	本地VR、Siri未激活	唤醒本地VR，开启VR识别	唤醒Carplay Siri	唤醒本地VR，开启VR识别	唤醒CarPlay Siri
5						
6		本地VR已激活	会退出当前VR，不会重新唤醒	退出本地VR，不唤醒Carplay Siri	会退出当前VR，不会重新唤醒	1.VPA结果页时CarPlay会被退到后台，该场景不存在
7			--按键信息以及本地VR状态传输给手机,是否唤醒Siri取决于手机端		2. TTS触发时，可操作Home button唤醒Siri，打断本地VR	
8		CarplaySiri已激活	退出CarPlay语音，不会激活本地语音			
9			-取决于手机侧策略	退出CarPlay	打断Siri，激活本地VR --本地VR状态传输给手机，是否激活取决于手机侧	忽略该长按操作，维持当前Siri -取决于手机侧策略
				，不会重新唤醒CarPlay Siri		
				-取决于手机侧策略		

3.2. 音频输入要求

3.2.1. 概述

满足苹果对音频输入的要求

3.2.2. 音频输入说明

Carplay需要支持以下相关功能。

To support an enhanced Siri experience in CarPlay, the accessory must implement the following.

1. 两个新的音频流，支持音频输入和输出。

Two new audio streams for audio input and audio output.

2. 「始终打开」一个麦克风输入流存储在一个循环的缓冲区中。

“Always on” microphone input stream stored in a circular buffer.

3. 回声消除和降噪（EC/NR）模块，用于取消通过汽车扬声器播放的任何音频，并降低座舱噪音，如道路或风扇噪音。

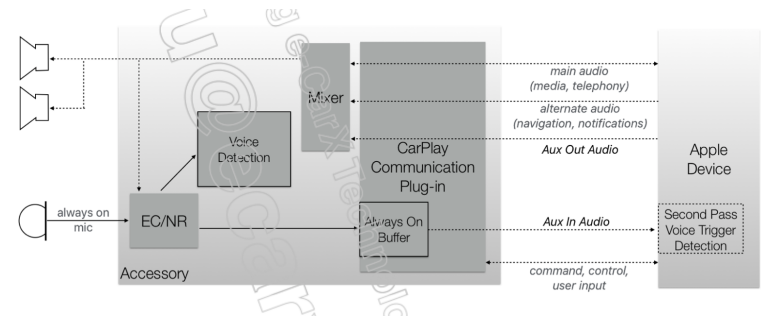
Echo cancelation and noise reduction(EC/NR) module to cancel out any audio playing through the car speakers, and to reduce cabin noise such as road or fan noise.

4. 语音检测模块，用于检测任何活动语音，以及特定关键字被说出时的语音。

Voice detection module to detect any active speech, and also when a specific keyword is spoken.

5. 混音器模块，用于组合通过主音频播放的媒体、通过辅助输出音频播放的Siri提示以及通过备用音频播放的任何音频。

Mixer module to combine media playing through MainAudio, Siri prompts playing through Aux Out Audio, and any audio played over Alternate Audio



3.3. 前端信号处理

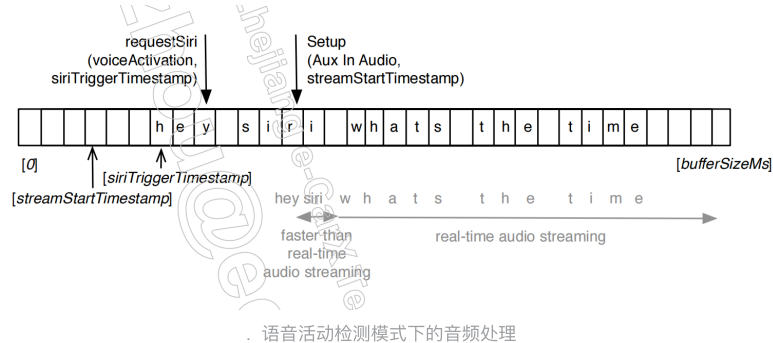
3.3.1. 概述

满足苹果对音频信号处理要求

3.3.2. 说明

3.3.2.1.语音检测模式（VAD）

- i. 当VAD模式开启时，在一段时间没有说话后，当检测到有人声说话时。标记为语音开始的SoS事件，保留SoS事件的SiriTrigger时间戳。如果随后出现300ms的连续语音，则需要发送requestSiri（语音激活）命令，并提供检测到SoS事件时的SiriTrigger时间戳。
- ii. 当存在至少300毫秒的连续人声语音时，视为一个有效的语音区域。每两个有效的语音区域之间，必须存在大于400ms的非语音区域。否则将视为一个语音区域。
- iii. 扬声器发出的声音（收音机、音乐、导航提示、语音交互系统）和车内或车周围的环境声音（转向信号、喇叭、交通、环境）不得触发SoS事件。
- iv. 从麦克风接收语音区域开始，到检测到语音开始（SoS事件）的延迟不得超过140毫秒。
- v. 在确认检测到语音单词后的50ms内，发送带有参数voiceActivation和SiriTriggerTimestamp的requestSiri命令到Carplay相关模块。（苹果要求，测试环境不明，最终指标待确认）
- vi. 对话丢失率（SMR-Speech Miss Rate）-对于典型的人声对话，每小时的对话丢失率不得超过0.1%。（苹果要求，测试环境不明，最终指标待确认）
- vii. 错误唤醒率（FWR-False Wake Rate）-错误的语音活动报告，不得超过每小时15次。（苹果要求，测试环境不明，最终指标待确认）



3.3.2.2.相关音频输入需求详细解释

音频输入。Siri的麦克风音频使用新的辅助输入音频流类型传输到设备。Aux-In-Audio流是单向的，将音频从附件传输到设备。请参阅辅助音频语音识别。

Audio Streams

Siri interactions use two new audio streams for audio input and output.

Audio input. Microphone audio for Siri is transferred to the device using a new Aux In Audio stream type. The Aux In Audio stream is unidirectional and transfers audio from the accessory to the device. See Aux In Audio—Speech Recognition.

Audio output. Siri prompts are played through a new Aux Out Audio stream type. The Aux Out Audio stream is unidirectional and transfers audio from the device to the accessory. See Aux Out Audio—Speech Recognition.

The Aux In Audio and Aux Out Audio streams are set up by the device as needed, when Siri is listening or speaking back to the user. The accessory must be capable of setting up the Aux In Audio stream or Audio Out Audio stream at any time, upon request. The device may request to set up one, or both streams.

Aux-In-Audio语音识别（增强Siri体验的新功能）

除非另有规定，附件必须符合ITU-T P.1110中规定的建议和测试标准。

采样率：表1描述了音频输入流的要求。附件必须只接受并提供具有相同位深的以下采样率。

Sample Rate	Encoding	Bit Depth	Channels	Duplexing	Requirement
16kHz	LPCM	16 bit	mono	n/a	Mandatory for CarPlay over USB
16kHz	OPUS	16 bit	mono	n/a	Mandatory for CarPlay over wireless

Aux In Audio—Speech Recognition (new for enhanced Siri experience)

Unless otherwise specified, the accessory must meet the recommendations and test criteria specified in ITU-T P.1110.

Sample Rate: Table 1 describes requirements for the audio input stream. The accessory must only accept and provide the following sample rates with equal bit depth.

麦克风

建议使用定向麦克风和线性波束形成麦克风阵列，以提高信噪比。

Microphones: Directional microphones and linear beam-forming microphone arrays are recommended for improved SNR.

上行输出电平

在标称测试布置中，附件必须为设备提供恒定的上行输出电平。不建议使用AGC。如果配件调整信号增益，则在每个口语中增益必须保持不变。附件上行链路输出测得的标称电平必须为A加权 -30 dB ± 2 dB均方根（RMS），以相对于满标度的单位表示（dBFS（A））。或者，如果使用ITU测量程序，标称电平可能为13 dB ± 2 dB SLR。

Uplink output level: In the nominal test arrangement, the accessory must provide a constant uplink output level to the device. The use of an AGC is not recommended. If the accessory adjusts signal gain, the gain must be held constant across each spoken utterance. The nominal level measured at the uplink output of the accessory must be A-weighted -30 dB ±2 dB root-mean-square (RMS), expressed in units relative to full-scale (dBFS(A)). Alternatively, the nominal level may be 13 dB ±2 dB SLR if using the ITU measurement procedure.

延迟

从配件麦克风接收声音到CarPlay Communication Plug-in接收代表该声音的LPCM音频缓冲区的延迟不得超过60ms。

Latency: Latency from the receipt of sound at the accessory’s microphone to the receipt of a buffer of LPCM audio by the CarPlay Communication Plug-in representing that sound must not exceed 60ms.

3.3.2.3.始终打开缓冲区

为了通过语音启动Siri并在按下按钮时立即激活，配件必须始终收听并将麦克风输入记录到循环缓冲器中。

苹果会缓存5s的音频。并且要求每10ms发送一个包。

Always On Buffer

In order to implement launching Siri by voice and instant activation on button press, the accessory must always listen and record microphone input into a circular buffer.

3.3.2.4.回声消除和降噪

必须对录制的麦克风输入应用回音消除，以补偿通过附件扬声器播放的任何音频。这包括所有使用主音频或备用音频播放的CarPlay音频、本地媒体播放和通知。

必须对录制的麦克风输入进行降噪处理，以减少驾驶室内驾驶噪音的干扰。

Echo Cancellation and Noise Reduction

Echo cancellation must be applied to the recorded microphone input to compensate for any audio played through the accessory speakers. This includes all CarPlay audio played using Main Audio or Alternate Audio, native media playback and notifications.

Noise reduction must be applied to the recorded microphone input to reduce interference from driving noise in the cabin.

3.3.2.5.信噪比

为了进行语音识别，附件必须在上行输出中保持足够的信噪比（SNR）。为了保持较高的语音识别精度，绝对信噪比应超过20 dB，并应尽一切合理努力在ITU-T P.1110附录D中规定的所有典型驾驶条件下保持较高的信噪比。建议使用高于20 dB的值，在静止车辆和慢速驾驶速度下，信噪比至少应为25 dB。

Signal-to-noise ratio: The accessory must maintain adequate signal-to-noise ratio (SNR) in its uplink output for speech recognition purposes. To maintain high speech recognition accuracy, the absolute SNR should be over 20 dB and all reasonable efforts should be made to maintain high SNR in all typical driving conditions as specified in Annex D of ITU-T P.1110. Values above 20 dB are recommended, and the SNR should be at least 25 dB in a stationary vehicle and at slow driving speeds.

3.3.2.6.混响消除

3.3.2.7.声源定位

3.3.2.8.混响消除

3.4. 唤醒

3.4.1. 概述

在KW模式，支持标记对应事件及时间戳 并发送给carplay

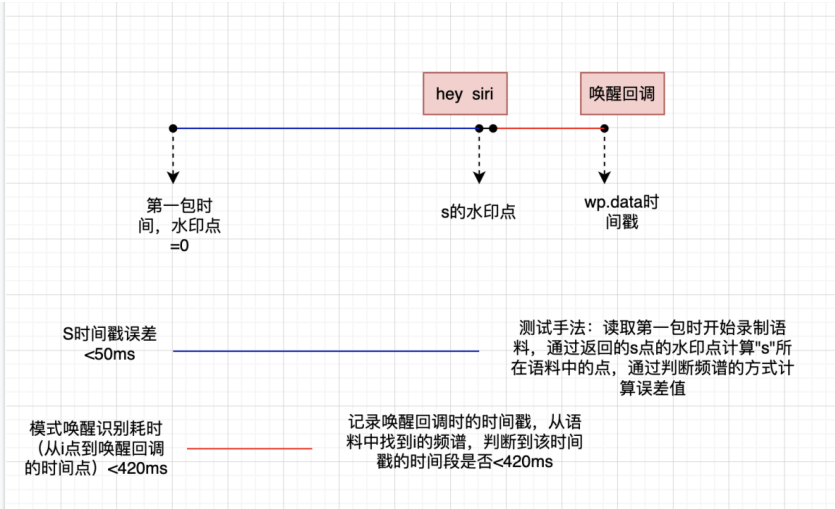
3.4.2. 说明

3.4.2.1唤醒词描述

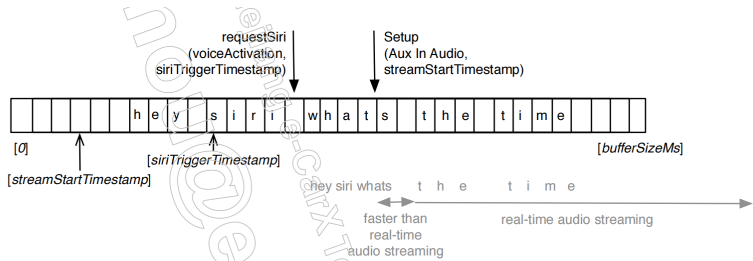
唤醒词：hey siri和siri。**Carplay**的认证只测试主驾（会屏蔽副驾等方位的唤醒事件，因为Apple只接收一路音频）。但是在用户场景下可以是车内任何位置都可使用earplay。且两种模式下均使用车机mic收音。

3.4.2.2关键词模式（KW模式）

- i. 当KW模式开启时，当检测到用户说出关键词“Hey Siri”时，需要发出requestSiri（voiceActivation）指令，并且附带siriTriggerTimestamp时间戳（检测到“Siri”中“S”的时间，siri的标注在s前），以表示关键词的开始时间。siriTriggerTimestamp时间戳相比较原始录音的误差需要在50ms以内。
- ii. 从麦克风接收到“Siri”关键字（第二个“i”音）时开始计算时长，直到关键字检测完成，发出唤醒指令，总的延迟不得超过420ms。The latency from receiving the end of the “Siri” keyword (second “i” sound) at the accessory’s microphone until keyword detection is completed must not exceed 420 ms.)
- iii. 在确认检测到唤醒关键词后的50ms内，发送带有参数voiceActivation和SiritiggerTimestamp的requestSiri命令到Carplay相关模块。（苹果要求，测试环境不明，最终指标待确认）
- iv. 错误拒绝率（FRR-False Reject Rate）：对于有效的“Hey Siri”/“Siri”唤醒语音，唤醒率需要达到99%以上。（苹果要求，测试环境不明，最终指标待确认）
- v. 错误接受率（FAR-False Acceptance Rate）：每小时有效外部语音音频的不得超过2。（苹果要求，测试环境不明，最终指标待确认）
- vi. ~~关键词唤醒性能问题——车机性能能不能达到相关要求标准——待YF确认能否提供足够的内存/算力，百度自测具体占用数值。~~



【KW模式】性能指标判断标准



关键字检测模式下的音频处理


3.5. 语音识别

3.5.1. 概述

在VAD和KW模式，支持标记对应事件及时间戳 并发送给carplay

3.5.2. 说明

a.测试case

 carplaytestcase文案.xlsx
14KB

b.测试场景以及测试指标TBD

- vii. 每个场景都应通过相应的测试要求
- viii. 苹果测试场景需求（6个测试场景）
 - 1. 测试1车辆怠速（0公里/小时或0英里/小时）， 风扇转速低档位。
 - 2. 测试2车辆怠速（0公里/小时或0英里/小时）， 风扇转速高档位。
 - 3. 测试3汽车驾驶（60公里/小时或37英里/小时）， 风扇转速中档位。
 - 4. 测试4汽车驾驶（120公里/小时或75英里/小时）， 风扇转速中档位。
 - 5. 测试5汽车驾驶（120公里/小时或75英里/小时）， 风扇转速中档位。通过主机进行主动媒体播放。
 - 6. 测试6汽车驾驶（120公里/小时或75英里/小时）， 风扇转速中档位。通过Apple设备进行主动媒体播放。
- ix. 对于测试指标如果达不到相应标准但差距较小时，存在「豁免」流程。测试完成后，依赖主机厂、tier1和苹果方面进行沟通。如果数据差异过大，则无法豁免。
 - 1. 首次实车自测场景计划如下（需要在实车入实验室验证）：
 - a. 821低配（TBD具体需通过认证的配置依赖福特、YF确认）：空调档位0-7档，分三级档位，低档（1档），中档(4档)，高档(7档)。

功能点	测试场景	车速 (km/h)	空调档位 (TBD结合 实车确定具 体档位)	媒体播放	媒体播报 到达mic 音量档位	综合噪声到达mic音 量 (dBA) (TBD待 实车测试标定)	综合噪声到达mic音 量 (dBA) (实车测 试最终实际音量)	备注
821低配 carplay认证	怠速	0	低档	无	0	47		符合预期，空调低档(1档)最终达到48分贝
	怠速	0	高档	无	0	70		空调高档(7档)最终达到58分贝
	中速	60	中档	无	0	65		空调中档(4档)最终达到52分贝，空调开到最大7档才62
	高速	120	中档	无	0	70		空调中档(4档)最终达到56分贝，空调开到最大7档才67
	高速	120	中档	通过主机进行 主动媒体播放	20格	80		符合预期，空调中档(4档)最终达到83分贝
	高速	120	中档	通过Apple设备 进行主动媒体 播放	16格	80		符合预期，空调中档(4档)最终达到81分贝

3.6. 技术实现详情

暂无

3.7. 语音指令需求

如涉及语音指令需求，列在此处。*标注以语音指令集为准。*

3.5. 车控需求

涉及车控逻辑需求，列在此处。物理按键、音频焦点等

3.6. 激活需求

明确激活限制状态下具体限制功能。

3.7. 账号&支付需求

如涉及耦合账号、支付等相关需求，列在此处。

3.8. 国际化需求

如涉及国际化能力，请将HMI中文文本与对应英文翻译贴出或添加附件。

3.9. 隐私权限需求

如需申请系统隐私权限，请与此处列表描述。

3.10. 走行规制需求

如有客户走行安全需求，在此处描述。

3.11. 主题换肤需求

如需支持系统主题换肤等功能，在此处描述。

3.12. 非功能需求

如有性能、验收指标等要求，列在此处，如CPU、内存、存储、网络等性能要求。

4. 运营

项目冷启动的整体运营方案

5. 数据

5.1. 报表需求

5.1.1. 报表一名称

1.报表描述；

2.指标列表。

报表由指标构成，可描述报表中展现哪些指标，及各指标可视化形态，如柱状图、折线图、热力图等。

5.2. 指标需求

5.2.1. 指标一名称

1.指标描述

简要描述指标及分析目的。

2.指标公式（聚合方式）

指标可以直接依赖埋点生成公式，也可以依赖另一个指标进行二次、三次计算。

3.指标维度

5.3. 埋点需求

暂无

6. 外部依赖

系统carplay

7. 合规意见


#个人信息收集、处理合规性、隐私协议内容描述等

8. 安全意见

#软件安全需求补充


9. 附件


carplay认证相关附件如下：


 CarPlay Enhanced Siri Experience R3[1].pdf
669KB


ITU认证相关附件如下：


 T-REC-P.50-199303-S!!PDF-E.pdf
128KB

 T-REC-P.50-199802-!!App!!PDF-E.pdf
2MB

 T-REC-P.50-199909-!!!PDF-E.pdf
260KB

 T-REC-P.50-200005-!!Err1!PDF-E.pdf
84KB

 ITU-T_P_1100__01_2019_.pdf
4MB

 P1100-Reporting template for clause 12 testing.xlsx
21KB