**Spec Tracking Number: PIS2030**

**Voice Assistant语音助手**

**Document Owner**

This specification is owned by:

Zhang jiajia

Infotainment Group, ESS, PATAC

+86 21 50165016 -(ext)515425

[Jiajia2\_zhang@patac.com.cn](mailto:Jiajia2_zhang@patac.com.cn)

目录/Contents

[1 需求概述 Requirement Overview 4](#_Toc65248361)

[1.1 背景价值 Background Information 4](#_Toc65248362)

[1.2 名词解释Terminology 4](#_Toc65248363)

[1.3 功能全景 Panorama of Functions 7](#_Toc65248364)

[2 需求列表Requirement List 7](#_Toc65248365)

[3 需求描述Requirement Description 9](#_Toc65248366)

[3.1 语音识别Voice Recognition 9](#_Toc65248367)

[3.1.1 VR技能列表 List of VR Features 9](#_Toc65248368)

[3.1.2 总体要求 General Requirements 9](#_Toc65248369)

[3.1.3 性能要求Performance Requirements 10](#_Toc65248370)

[3.1.4 语音唤醒Voice Trigger 14](#_Toc65248371)

[3.1.5 麦克风降噪Microphone Noise Reduction 15](#_Toc65248372)

[3.1.6 声源定位Sound Source Localization 15](#_Toc65248373)

[3.1.7 语音端点检测Voice Endpoint Detection 16](#_Toc65248374)

[3.1.8 语音纠错Voice-in Error Correction 16](#_Toc65248375)

[3.1.9 硬按键要求Hard Key Requirements 16](#_Toc65248376)

[3.1.10 对话管理器Dialog Manager 18](#_Toc65248377)

[3.1.11 Barge-In 21](#_Toc65248378)

[3.1.12主动交互 Active Interaction 21](#_Toc65248379)

[3.1.13 二次交互 Secondary Interaction 23](#_Toc65248380)

[3.1.14 所见即可说 Seeing Is Saying 24](#_Toc65248381)

[3.2 TTS 24](#_Toc65248382)

[3.2.1 总体要求General Requirements 24](#_Toc65248383)

[3.2.2 性能要求Performance Requirements 25](#_Toc65248384)

[3.2.3 多音字的发音Pronunciation of Polyphones 25](#_Toc65248385)

[3.2.4 数字的发音Pronunciation of Numbers 25](#_Toc65248386)

[3.2.5 常见符号的发音Pronunciation of Common Symbols 26](#_Toc65248387)

[4 系统需求System Requirements 27](#_Toc65248388)

[4.1 账号相关Account Related 27](#_Toc65248389)

[4.2 应用内设置In-application Setting 27](#_Toc65248390)

[4.3 外部调用External Call 29](#_Toc65248391)

[4.3.1 拟人形象助手/Personified Assistant 29](#_Toc65248392)

[4.3.2 氛围灯随动/Ambient light 30](#_Toc65248393)

[4.3.3 车辆故障查询/ Vehicle fault inquiry 30](#_Toc65248394)

[4.4 版本升级Version Upgrade 30](#_Toc65248395)

[4.5 埋点需求Event Tracking Requirements 30](#_Toc65248396)

[4.6 相关CAN信号Related CAN Signals 30](#_Toc65248397)

[4.7 相关标定项 Related Calibration Items 30](#_Toc65248398)

[5 附录Appendix 31](#_Toc65248399)

[**Revision Log** 31](#_Toc65248400)

# 需求概述 Requirement Overview

## 背景价值 Background Information

语音作为人类最自然的交互方式之一，非常适合汽车驾驶场景，不仅能解放双手，避免视线分散，同时在打通车内功能、对接网络服务后，更能快速满足用户多样化的需求。

As one of the most natural interaction modes of human beings, voice is very suitable for being applied in the scenarios of vehicle driving. It frees both hands of drivers and avoids distraction of their sight. In the meantime, after the in-vehicle functions are connected to each other and the network services are accessed, it is capable of satisfying, in a faster manner, the diversified needs of users.

车载语音助手定位于一个能满足车主驾车过程中的安全、效率、娱乐等方面的贴心助手。

The in-vehicle Voice Assistant is positioned as a considerate assistant that can satisfy the owner’s requirements in safety, efficiency, entertainment and other aspects during driving.

## 名词解释Terminology

**语音激活检测(VAD)：**指系统判断外界什么时候存在“有效语音”，什么时候不存在。当判断外界存在有效的语音时，机器才会对其进行后续的识别操作。

**Voice Activation Detection (VAD):** it involves the process in which the system judges when there is an “effective voice” outside and when there is not. Only when the system judges that there is an effective voice outside, shall the machine perform the subsequent recognition operation accordingly.

**语音唤醒（VT）：**指系统通过识别“有效语音”中是否存在某个关键词（唤醒词），来判断用户接下来是否要发出指令，是否需要进行响应。

**Voice Trigger (VT):** it involves the process in which, by recognizing whether the “effective voice” contains any keyword (trigger word), the system judges whether the user is going to give a command and whether a response is required.

**N-Best列表：**指系统识别的结果列表，包括每种结果的可能性，即置信度。

**N-Best list:** it refers to the list of results recognized by the system, covering the possibility of each result, namely the confidence.

**语音打断（Barge-In）**：指在设备说话时，用户可以直接通过语音进行打断，而不需要通过用手按下停止键或是强行等设备把话说完。

**Barge-In**: when the device is speaking, the user is allowed to barge in on it directly by voice, without having to manually press the Stop key or passively wait until the device finishes the speaking.

**麦克风阵列：**由一定数目的声学传感器（一般是麦克风）组成，用来对声场的空间特性进行采样并处理的系统。

**Microphone array:** it refers to the system consisting of a certain number of acoustic sensors (typically microphones) that are used to sample and process the spatial characteristics of sound field.

**语音技能**：指可识别的某类语音指令集合。

**Voice features:** it refers to the voice command set of certain type which is recognizable.

**SNR：**用户语音与驾驶环境噪音的平均信噪比，语音识别频率范围： 200 Hz to 7.5kHz。

**SNR:** it refers to the average signal-to-noise ratio between user voice and driving environment; the range of voice recognition frequency falls within 200Hz to 7.5kHz.

**字错误率 WER：**

**Word error rate (WER):**

* S：替换的字数
* S: the number of words replaced
* D：删除的字数
* D: the number of words deleted
* I：增加的字数
* I: the number of words added
* N：一段语音对应文本的总字数
* N: the total number of words in a phonetic text

**句错误率 SER：**

**Sentence error rate (SER):**

* L：参与识别的句子总数
* L: the total number of sentences involved in recognition
* H：识别正确完全一致的句子数
* H: the number of sentences recognized exactly as they are

**MOS评分(Mean Opinion Score)**：平均主观意见分

**MOS:** Mean Opinion Score

按照5分制给出，5分为最高分，1分为最低分，一般以3分为可接受标准，精确度为0.5，用于评价自然度、音质、可懂度和总体感觉。

Measured by the 5-point system, the score of 5 points is highest and that of 1 point is lowest; generally, the minimum acceptable score is 3 points; with an accuracy of 0.5, MOS is used to evaluate the naturalness, sound quality, intelligibility and overall sense.

|  |  |  |
| --- | --- | --- |
| **级别**  **Level** | **MOS值**  **MOS value** | **用户满意度**  **User satisfaction** |
| 优  Excellent | 5.0 | 非常自然，语音达到了广播级别水平，很难区分合成语音和广播语音的区别，听起来非常相似。从整体上来说语音清晰流畅，声音悦耳动听，非常容易理解，听音人非常乐意接受。  Being highly natural, the voice reaches the broadcast level and it is hard to tell it is a synthetic voice or a broadcast voice, because they sound very close. Overall, the voice is clear and smooth, bearing a pleasant sound and a readily comprehensible content and thus easily acceptable to the listeners. |
| 4.5 | 自然，听起来完整没有明显不正常的韵律起伏，比较清晰流畅，比较容易理解，达到了人们普通对话的质量，听音人愿意接受。  Being natural, the voice sounds complete and is free of obvious abnormal rhythmic fluctuations; in the meantime, the voice is relatively clear and smooth, bearing a readily comprehensible content, reaching the quality of people’s daily conversation, and thus acceptable to the listeners. |
| 4.0 | 还可以，没有出现明显的分词错误和严重的语言韵律错误，有很少的一两个音节不太清除，听音人可以没有困难的理解语言的内容，听音人多数认为可以接受。  Being not bad, the voice involves no obvious word segmentation errors and serious language rhythm errors; although it contains few, one or two, syllables that are not clear, the listeners are capable of effortlessly understanding the content and thus most listeners are willing to accept it. |
| 良  Good | 3.5 | 不太自然，语音还算流畅，语音中的错误比较少，偶尔有几个音节不太清楚，韵律起伏比较正常，错误比较少，多数听音人勉强可以接受。  Being not natural, the voice is passably smooth; featured by only few errors, occasionally several unclear syllables, and relatively normal rhythm fluctuations that contain insignificant errors, the voice is barely acceptable to most listeners. |
| 中  Medium | 3.0 | 可接受，语音不太流畅，有比较容易察觉的语言错误，有一些不太正常韵律起伏，一般情况下可以努力理解语音的内容，不太愿意接受。  Being acceptable, the voice is not that smooth and involves easily perceivable language errors and some irregular rhythm fluctuations; therefore, the listeners, under normal circumstances, should try hard to figure out the voice and are unwilling to accept it. |
| 差  Poor | 2.0 | 比较差，语音不流畅，听起来只是把单独的音节简单的堆砌到一起，没有正常的韵律起伏，有些词不是太清晰，难于理解，整体上听音人可以听懂一些内容，但是不能接受。  Being poor, the voice is not smooth and sounds like that it is just a product from the simple stacking of individual syllables; without normal rhythmic fluctuations and containing some ambiguous words, the voice is hard to be understood; although the listeners are able to get part of it, they are unwilling to accept it. |
| 劣  Inferior | 1.0 | 明显的是机器音，很不清楚，语音无流畅可言，只能听懂只言片语，基本上无法理解，完全无法接受。  Obviously, it is a machine-made voice and very unclear, involving no such thing as smoothness; the listeners are able to understand only a few words and basically get nothing, thus completely unwilling to accept it. |

**唤醒率：**说出唤醒词时，语音助手被成功唤醒的比率。

**Trigger rate:** it refers to the rate at which the Voice Assistant is successfully triggered when a trigger word is spoken out.

**误唤醒次数**：噪声环境下，没说唤醒词时，语音系统在规定时间内误触发唤醒的次数。

**Number of false triggers:** it refers to the number of times the voice system is falsely triggered, in a noisy environment and with no any trigger word brought up, within a prescribed time.

## 功能全景 Panorama of Functions

# 需求列表Requirement List

（细分并描述需求的主要功能模块，列出功能模块的优先级：1，2，3，1最低，3最高。）

(Subdivide and describe main functional modules of the requirements, and then list out priority of the functional modules, namely 1, 2, 3: 1 represents the lowest; 3 represents the highest.)

|  |  |  |
| --- | --- | --- |
| Subfeature | Description | Priority |
| 技能列表  List of features | 语音技能分类  Classification of voice features | 3 |
| 自然语音识别  Natural voice recognition | 支持方言口音的普通话  Support the mandarin with dialect accent | 3 |
| 中英文混合语音识别  Recognition of voice mixed with Chinese and English | 语音转文本，支持中英文姓名、中英文歌曲、中英文POI等。  The voice-to-text supports Chinese/English name, Chinese/English song, Chinese/English POI, etc. | 3 |
| 自然语义理解  Natural semantic comprehension | 支持识别任意语句中的语义  Support the recognition of semantics in arbitrary statement | 3 |
| ~~自定义~~语音技能云端更新  Self-defined voice features | 后台支持动态增加说法，修改对话逻辑  Dynamically add statements/modify dialog logic in the background | 3 |
| 自定义唤醒词  Self-defined trigger word | 支持用户自定义语音唤醒词  Support the user-defined voice trigger word | 3 |
| One-shot语音唤醒  One-shot wakeup | 支持用户唤醒词和语音命令连起来说  Support the user to string together any trigger words and voice commands | 3 |
| 免唤醒  Trigger-free | 常用说法唤醒命令词、限时免唤醒、快捷命令词  Common expression used as trigger command word, trigger-free within a limited time, and quick command word | 3 |
| 麦克风降噪  Microphone noise reduction | 支持麦克风阵列、ECNR  Support microphone array and ECNR | 3 |
| 声源定位  Sound source localization | 支持识别声源方位  Support the recognition of sound source direction | 3 |
| 语音端点检测  Voice endpoint detection | 识别有效人声  Recognize effective human voice | 3 |
| 语音纠错  Voice-in error correction | 基于上下文，修改特定场景下的关键信息  Modify the key information in a specific scenario based on the context | 3 |
| 硬按键要求  Hard key requirements | VR过程中支持并响应方向盘按键  Support and respond to the keys of steering wheel in the process of VR | 3 |
| 文本回显  Text echo | 支持实时将用户语音转成文本显示  Support the real-time conversion of user voice into text display | 3 |
| 语音交互对话流  Voice interactive dialog flow | 展示对话历史，支持触摸操作  Display dialog history and support touch operation | 3 |
| 对话模式  Dialog mode | 支持详尽与简洁的对话模式  Support both exhaustive and concise dialog modes | 2 |
| 对话风格/发音人  Dialog style | 支持不同对话风格如：中规中矩、轻松活泼  Support both serious and breezy dialog styles | 2 |
| 语速调节  Speech rate adjustment | 支持较慢、标准、较快3档语速  Support 3 scales of speech rate, namely slow, standard and fast | 3 |
| 提示语设置  Prompt setting | 支持用户设置提示语  Support the user’s prompt setting | 2 |
| 话题切换  Topic switching | 支持用户切换话题  Support the user’s topic switching | 3 |
| 跨场景上下文理解  Cross-scenario context understanding | 支持理解和关联用户上下文对话内容  Support the understanding and correlating of the user’s dialog context | 3 |
| 不支持指令的回复  Reply for unsupported command | 友好回复系统不支持或暂不支持的功能  Give a friendly reply for the function that is not supported or temporarily not supported by the system | 2 |
| Barge-In | 支持用户~~随时~~输入新的语音指令打断当前对话  Support the user to enter a new voice command at any time | 3 |
| 主动交互  Active interaction | 支持特殊场景语音交互的主动触发  Support the active triggering of voice interaction under a special scenario | 3 |
| 二次交互  Secondary interaction | 支持二次交互  Support secondary interaction | 3 |
| 所见即可说  Seeing is saying | 支持识别HMI界面看得见的文字  Support the recognition of visible words on the HMI interface | 3 |
| TTS发音人  TTS speaker | 多个发音人、男女声、特色方言  Multiple speakers, male & female voices, and featured dialects | 3 |
| TTS特殊符号处理  TTS special symbol processing | 支持正确朗读号码、时间、日期、标点、常见符号等  Support correct reading of numbers, time, date, punctuation, common symbols, etc. | 3 |

# 需求描述Requirement Description

## 语音识别Voice Recognition

### VR技能列表 List of VR Features

语音识别模块应当能够将所有可识别的语音指令放在同一个Level实现。即系统应当能够自动地分析用户的语句，并理解用户语句的分类。

The voice recognition module shall be able to place all recognizable voice commands at the same Level for execution. That is, the system shall be capable of automatically analyzing the user’s statements and understanding classification of the user’s statements.

该语音识别系统需要支持下面的语音技能分类：

The voice recognition system shall support the classification of following voice features:



关于OnStar技能，系统应根据标定来判断车上是否配备了OnStar模块，对于配备OnStar的车型，VR应支持OnStar技能，对于不配备OnStar模块的车型，VR可以不支持OnStar技能。

With regard to the features of OnStar, the system shall judge whether the vehicle is equipped with OnStar module based on the calibration; for the model equipped with OnStar, VR shall support the features of OnStar; for that not equipped with Onstar, VR is allowed not to support the features of OnStar.

### 总体要求 General Requirements

1. 系统应支持自然语音识别。

1. The system shall support natural voice recognition.

* + 语音识别的过程中，支持粤语方言识别，支持四川口音普通话识别（~~例如上海口音，东北口音）~~、语句顺序无严格规定。
  + In the process of voice recognition, the dialect accents (such as Shanghai accent and northeast accent) shall be supported; however, no strict rules are set to the sequence of statement.

1. 系统应支持中英文混合语音识别。

2. The system shall support the recognition of voice mixed with Chinese and English.

* 当系统语言为非中文时，不影响语音识别功能、语音设置，在用户切换语言后，系统应提示当前只支持中文语音识别。具体提示方式不限，由交互定义。
* When the system language is non-Chinese, the system shall prompt the user that only Chinese voice recognition is supported.
* 系统应支持识别中英文姓名、中英文歌曲、中英文POI等
* The system shall support the recognition of Chinese/English name, Chinese/English song, Chinese/English POI, etc.
* 系统应支持识别用户通讯录中用英文存储的联系人名、媒体资源中以英文命名的歌曲等
* The system shall support the recognition of the contact name saved in English in the user’s address book, the song titled in English in the media resources, etc.

1. 系统应支持自然语义理解。

3. The system shall support natural semantic comprehension.

* 系统应当能够自动识别用户说出的任意语句，并能够判断出用户说出的语句属于哪一个分类，并正确地提取用户语句中的关键信息，从而根据用户的请求完成相应的动作。
* The system shall be able to automatically recognize the arbitrary statement spoken out by the user, judge which category the statement spoken out by the user belongs to, and then correctly extract the key information in the user's statement, thereby completing the corresponding action according to the user’s request.

1. 系统应支持本地及云端的混合语音识别。

4. The system shall support the recognition of both local and Cloud mixed voices.

1. 系统应系统应支持本地及云端的并行识别，并有能力为用户选择更优的识别结果。

5. The system shall support the local and Cloud parallel recognition and be capable of selecting a better recognition result for the user.

* 当网络连接异常导致无法联网时，系统应支持离线语音识别，具体支持的离线语音技能请参照3.1.1语音识别技能列表。
* When the Internet access fails because the network connection goes abnormal, the system shall support offline voice recognition; see 3.1.1 List of VR Features for the supported offline voice features.

1. 系统应支持在后台自定义语音技能，动态增加说法、修改对话逻辑等。

6. The system shall support the self-defined voice features in the background to dynamically add statements, modify dialog logic, etc.

### 性能要求Performance Requirements

系统应满足在不同噪音环境下语音识别率、唤醒率等要求，参考噪音环境要求如下：

The system shall meet the requirements on voice recognition rate, trigger rate, etc under different noisy environments. The requirements on noisy environments for reference are shown as follows:

* Idle停车环境：车辆静止，发动机开着，Average SNR>=14（平均信噪比）
* Idle: the vehicle stands still and the engine is idling, Average SNR>=14 (average signal-to-noise ratio)
* Noise1噪声环境1：城市道路，车速40-80，14> Average SNR>=10
* Noise 1: urban road, vehicle speed 40-80, 14> Average SNR>=10
* Noise2噪声环境2：高速道路，车速超过80，周围声音非常嘈杂，10> Average SNR
* Noise 2: express way, vehicle speed above 80, very noisy surrounding, 10> Average SNR

#### 语音识别率要求Requirements on Voice Recognition Rate

|  |  |
| --- | --- |
| **噪音环境**  **Noisy environment** | **识别率 (句正确率)**  **Recognition rate (sentence accuracy)** |
| Idle | 95%以上  Above 95% |
| Noise1 | 90%以上  Above 90% |
| Noise2 | 85%以上  Above 85% |

#### 分类语音识别率要求 Requirements on Classified Voice Recognition Rate

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 分类条目  Classified entry | 说明  Description | 说法备注  Remark on statement | 噪声环境  Noisy environment | 识别率（句正确率）  Recognition rate (sentence accuracy) |
| 基本命令识别率  Recognition rate of basic command | 对离线基本命令的识别率（离线基本命令指打开App、车辆控制、音乐播放控制、帮助之类的命令）  Recognition rate of offline basic commands (offline basic commands refers to commands for App activation,  vehicle control, music play control, help, etc) | 必须包含音频播放操作，车辆控制操作的说法  The statement on audio play operation and vehicle control operation shall be contained | idle | 95%以上  Above 95% |
| noise1 | 95%以上  Above 95% |
| noise2 | 90%以上  Above 90% |
| 电话号码数字识别率  Recognition rate of telephone number | 对1-11位电话号码的识别率  Recognition rate of telephone number with 1-11 digits | 必须包含特例：2/5/8连着的号码、11位分段拨号、11位连续拨号  The special cases shall be contained: 2/5/8 connected number, 11-digit segmented dialing, and 11-digit continuous dialing | idle | 95%以上  Above 95% |
| noise1 | 90%以上  Above 90% |
| noise2 | 85%以上  Above 85% |
| 电话簿人名识别率  Recognition rate of name in phone book | 对动态生成的蓝牙电话簿人名的识别率  Recognition rate of name in the dynamically generated Bluetooth phone book | 必须包含英文谐音人名的说法  The expression for homonym of English name shall be contained | idle | 95%以上  Above 95% |
| noise1 | 90%以上  Above 90% |
| noise2 | 90%以上  Above 90% |
| 电台搜索识别率  Recognition rate of radio station search | 收音机频率输入要支持逐位按数字读法和自然数读法  Support digit-by-digit number reading and natural number reading for radio frequency input | 必须包含中文+字母+数字的说法  The pronunciation for combination of Chinese + letter + number shall be contained | idle | 95%以上  Above 95% |
| noise1 | 90%以上  Above 90% |
| noise2 | 88%以上  Above 88% |
| 歌曲搜索识别率  Recognition rate of song search | 最大2000首歌曲  2,000 songs at most | 必须包含歌曲名大于10，英文歌曲的说法  The pronunciation for song name greater than 10 and English song shall be contained | idle | 90%以上  Above 90% |
| noise1 | 88%以上  Above 88% |
| noise2 | 85%以上  Above 85% |
| 在线播客节目识别率  Recognition rate of online podcast | 最大2000个播客节目  2,000 podcasts at most | 必须包含热门播客节目的说法  The pronunciation for popular podcast shall be contained | idle | 92%以上  Above 92% |
| noise1 | 88%以上  Above 88% |
| noise2 | 85%以上  Above 85% |
| 兴趣点/街道名识别率  Recognition rate of point of interest/street name | 某个城市兴趣点/街道名  Point of interest/street name of a city | 必须包含POI，街道名的说法  The pronunciation for POI and street name shall be contained | idle | 90%以上  Above 90% |
| noise1 | 88%以上  Above 88% |
| noise2 | 85%以上  Above 85% |

#### 语音交互成功率要求Requirements on Success Rate of Voice Interaction

对于车载场景下各语音技能的给定语料，其交互成功率应在语音识别无误的情况下，保证100%的成功率。

With regard to the given corpus for each voice feature under the in-vehicle scenario, the interaction success rate shall reach be 100% on the condition that the voice recognition is correct.

交互成功率=交互成功的次数/总数据量。

Interaction success rate = number of successful interactions/total data volume.

#### 唤醒率要求 Requirements on Trigger Rate

|  |  |  |
| --- | --- | --- |
| **噪音环境**  **Noisy environment** | **唤醒率要求(主唤醒词)**  **Requirements on trigger rate (key trigger word)** | **误唤醒次数要求（4小时内）**  **Requirements on number of false triggers (within 4 hours)** |
| Idle | 95%以上  Above 95% | 最多1次  Once at most |
| Noise1 | 90%以上  Above 90% | 最多1次  Once at most |
| Noise2 | 90%以上  Above 90% | 最多1次  Once at most |

#### 响应时间要求Requirements on Response Time

用户通过按键触发VR到VR开始响应之间的时间应小于1s。

The time taken between the user’s trigger of VR by pressing the key and the VR’s start of response shall be less than 1s.

用户通过语音唤醒的方式触发VR到VR开始响应离线指令之间的时间应小于1.5s。

The time taken between the user’s trigger of VR by voice and the VR’s start of response shall be less than 1.5s.

网络良好的情况下，用户说完语音命令到VR给出语音反馈或执行相关动作的时间应小于1.5s。

On the condition that the network connection is sound, the time taken between the user’s completion of voice command to the VR’s voice feedback or execution of related action shall be less than 1.5s.

### 语音唤醒Voice Trigger

用户可以通过语音唤醒语音助手，也可以通过点击本地语音助手拟人形象、VR 应用图标、方向盘硬按键唤醒语音。

系统在免打扰模式下应禁止语音唤醒~~等相关~~功能，但是允许用户通过以上其他方式手动唤醒语音助手。

在安吉星E call、APA、camera 使用过程中，禁止语音功能，如果用户点击常驻的语音助手拟人形象时，需要友好提示用户，具体提示参考交互。安吉星E call请参考PIS-2079，Camera使用状态请参考PIS-2019，APA 请参考PIS-2048.

Under the Do-Not-Disturb Mode (DND Mode), the system shall disable such functions as voice trigger.

#### 唤醒词Trigger Word

系统应支持用户开启和关闭语音唤醒功能，语音唤醒的出厂默认状态为开启。

The system shall support the user to activate/deactivate the function of voice trigger, which is activated by the factory default setting.

系统应支持用户自定义语音唤醒词，系统默认的主唤醒词应读取标定。用户修改唤醒词后，系统默认的主唤醒词仍需要支持唤醒功能。

The system shall support the user-defined voice trigger words, but the system default key trigger words shall be read for calibration.

#### One-shot语音唤醒One-shot Wakeup

系统应支持One-shot Wakeup语音唤醒功能，即支持用户可以不通过按键来触发VR，而是直接通过语音触发VR，并且用户可以把触发VR的唤醒词和语音命令连起来说。

The system shall support the function of One-shot Wakeup, that is, the user is allowed to trigger VR without pressing the key but directly through the voice, as well as allowed to string together any trigger words and voice commands that trigger VR.

#### 免唤醒Trigger-free

系统应支持常用说法唤醒命令词，即对于常见的业务范围支持用户使用唤醒命令词（而非唤醒词）直接操作常用功能，支持的唤醒命令词如“我想导航到xxx、打电话给xxx、我想播放xxx、我想查询xxx”， 具体参见语音技能列表。该功能应支持开关，默认为开启。

The system shall support the common expression used as trigger command word, that is, for the common business scope, the system shall support the user to apply the trigger command word (instead of trigger word) to operate common functions directly; the supported trigger command words include “I want to navigate to xxx/call xxx/I want to play xxx/I want to query xxx”: see the List of Voice Features for details. This function, activated by default, shall be able to be activated and deactivated.

系统应支持限时免唤醒，即在唤醒后的一段激活时间内，无需再次唤醒即可直接使用语音指令。激活时间默认为15秒。

The system shall support the trigger-free function within a limited time, that is, the voice command is able to be used directly within a period of activation time after its being activated with no need to be triggered once again. The activation time is set to 15s by default.

系统应支持快捷命令词，即用户可免唤醒使用固定命令词操作系统常用功能，如播放/暂停播放，上一首/下一首等，具体参见语音技能列表。

The system shall support the quick command word, that is, when the trigger-free function is enabled, the user is allowed to use the fixed command word to operate the system’s common functions, such as Play/Pause, Previous/Next, etc, etc: see the List of Voice Features for details.

### 麦克风降噪Microphone Noise Reduction

系统应支持多路麦克风阵列。

The system shall support multiple microphone arrays.

系统应支持Beam Forming波束成形技术，以实现抑制目标方向的干扰源，增强目标方向的声音强度的目的。

The system shall support the Beam Forming technology for the purpose of suppressing the interference source in the target direction and enhancing the sound intensity in the target direction.

系统应支持麦克风主动降噪，回声消除、噪声抑制、去除混响，以保证语音质量和识别率。

The system shall support active noise reduction, echo cancellation, noise suppression, and reverberation removal for the microphone to ensure voice quality and recognition rate.

当Local VR、Carplay VR、Carlife VR被唤醒后，系统应通知空调模块使其降低风量，待VR结束后，系统应通知空调模块使其恢复风量。

When Local VR/Carplay VR/Carlife VR is activated, the system shall notify the air-conditioning module to reduce the air volume; after VR is deactivated, it shall notify the air-conditioning module to restore the air volume.

### 声源定位Sound Source Localization

根据车型麦克风布置，对于硬件支持的车型，系统应支持多音区声源定位，包括主驾、副驾、第二排左、第二排右、~~第三排左、第三排右的~~声源定位。车型标定请参考4.7章节。

According to the model’s microphone arrangement, the system shall support multi-range sound source localization, including the sound source localization for driver side, front passenger side, left/right side of second row.

声源定位应支持可配置，包括~~自动响应所有录音方向、~~仅响应主驾方向、仅响应唤醒人方向。默认为仅响应唤醒人方向~~响应所有录音方向~~。具体文言以交互为主。

The sound source localization shall be capable of being configured, including automatic response to the recording direction, which is the default setting, and response only to the voice from driver side.

~~系统应支持~~仅响应唤醒人方向~~谁唤醒谁说~~，即谁唤醒成功后则只有他/她可以执行本次对话流程。

The system shall support the principle that the one triggering the system shall be the one speaking, that is, the one managing to trigger the system shall be the only one to complete the dialog process.

唤醒成功时，系统应支持显示当前声源方向。

After being triggered, the system shall support the display of current sound source direction.

对于车控指令，系统应支持匹配该声源方向的定向控制，例如主驾通过语音打开车窗，系统应支持开启主驾车窗。

With regard to the vehicle control command, the system shall support the directional control that matches the sound source direction, for example, when the driver gives a voice command to open the window, the system shall support the opening of driver side window.

### 语音端点检测Voice Endpoint Detection

系统应支持用户启动或唤醒语音识别后，立即可以说话，而非必须等待提示语/提示音播放完毕。

The system shall support the user to speak immediately after activating or triggering the voice recognition, rather than having to wait for the prompt/prompt tone to finish playing.

系统应支持识别有效人声，屏蔽无效声音信号。

The system shall support the recognition of effective human voice and the shielding of invalid sound signal.

### 语音纠错Voice-in Error Correction

基于上下文理解算法，系统应支持对电话、票务、天气等场景下关键信息进行修改，如：电话联系人，电话号码段，天气城市等，进而提升语音交互的效率和体验。

Based on the context understanding algorithm, the system shall support the modification of key information under such scenarios as telephone, ticketing and weather, involving telephone contact, telephone number segment, weather & city, etc, thereby improving the efficiency and experience of voice interaction.

### 硬按键要求Hard Key Requirements

VR过程中应能支持并响应~~方向盘~~硬按键。一般情况下包括如下按键：SWC\_VR、SWC\_ MUTE&HANGUP~~挂断~~、KEY\_CODE\_POWER ~~Power~~、SWC\_VOL\_UP/ SWC\_VOL\_DOWN~~音量~~、~~KEY\_CODE\_HOME Home、KEY\_CODE\_MEDIA~~ ~~Media~~。按键定义请参考PIS-2002.

In the process of VR, the system shall support and respond to the keys of steering wheel. In general, the following keys are contained: VR, Hang-up, Power, Volume, Home, and Media.

* SWC\_VR （TBD）
  + 短按SWC\_VR唤醒本地VR; 长按SWC\_VR系统无需响应本地VR
  + 本地VR过程中，并且TTS播放时，如果短按SWC\_VR键，则终止TTS的播放；
  + In the process of local VR, when TTS is playing, provided that the VR key is pressed, the playing of TTS shall be terminated;
  + 本地VR过程中，并且TTS没有播放时，如果短按SWC\_VR键，则退出本地VR；
  + In the process of local VR, when TTS is not playing, provided that the VR key is pressed, the local VR shall be exited;
  + 如果本地VR没有启动，IPhone也没有连接车机，并且也没有incoming call的情况下，长按SWC\_VR 的响应请参考PIS-2035. ~~系统无需响应~~；
  + VR过程中，如果默认语音助手是本地VR，~~并且IPhone没有连接车机，没有incoming call的情况下~~，长按SWC\_VR系统无需响应本地VR；如果默认语音助手不是本地VR，长按SWC\_VR响应对应的默认语音助手。默认语音助手的设置请参考4.2章节。
  + In the process of VR, on the condition that neither the iPhone is connected to the Infotainment System, nor there is an incoming call, provided that the VR/telephone key is long pressed, the system shall not make a response;
  + ~~如果用户在免打扰模式下操作SWC\_VR按键，请参考PIS-2057.系统应退出免打扰模式并响应VR按键~~。
  + 通话状态下，不需要响应VR按键的长按、短按。
* SWC\_MUTE&HANGUP~~挂断~~键 (仅Buick车型支持)
  + VR过程中，按下SWC\_ MUTE&HANGUP~~挂断~~键，则退出VR并执行相应动作；
  + TTS播放中，按下SWC\_ MUTE&HANGUP，中断TTS并执行相应动作.
* KEY\_CODE\_POWER ~~Power~~键
  + VR过程中，如果长按KEY\_CODE\_POWER ~~Power~~键，应退出VR并进入TOD画面
  + In the process of VR, provided that the Power key is long pressed, VR shall be exited while TOD screen shall be accessed
  + VR过程中，如果短按KEY\_CODE\_POWER ~~Power~~键，应退出VR并执行相应动作
  + In the process of VR, provided that the Power key is short pressed, VR shall be exited while corresponding action shall be performed.
* SWC\_VOL\_UP/ SWC\_VOL\_DOWN~~音量~~键
  + VR过程中，非TTS播报时，按下SWC\_VOL\_UP/ SWC\_VOL\_DOWN~~音量+/-~~键，调节媒体~~语音播报~~音量
  + ~~VR过程中，~~TTS播报时，按下SWC\_VOL\_UP/ SWC\_VOL\_DOWN键，调节语音播报音量
  + ~~非VR过程，TTS播报时，SWC\_VOL\_UP/ SWC\_VOL\_DOWN音量+/-键，可以调节语音播报音量~~
  + ~~非VR过程，非TTS播报时，按下SWC\_VOL\_UP/ SWC\_VOL\_DOWN音量+/-键，调节媒体音量；~~
* ~~KEY\_CODE\_HOME键~~
* ~~HOME key~~
  + ~~VR过程中，如果按下KEY\_CODE\_HOME Home键，应退出VR并执行相应的按键动作~~
  + ~~In the process of VR, provided that the HOME key is pressed, VR shall be exited while the corresponding key action shall be performed~~
* ~~KEY\_CODE\_MEDIA Media键~~
* ~~Media key~~
  + ~~VR过程中，如果按下KEY\_CODE\_MEDIA Media键，应退出VR并执行相应的按键动作~~
  + ~~In the process of VR, provided that the Media key is pressed, VR shall be exited while the corresponding key action shall be performed~~
* ~~Onstar按键~~
* ~~OnStar key~~
* ~~VR过程中，按Onstar按键安吉星白键需要停止语音识别并立即切到安吉星画面. 请参考PIS-2079.~~
  + ~~In the process of VR, press the white key of OnStar to stop the voice recognition and immediately switch to the screen of OnStar~~

对于Carplay、Carlife等手机互联方案中的语音识别功能，其按键要求请参考手机互联Spec PIS-2004。

For the voice recognition functions in the mobile phone interconnection programs such as Carplay and Carlife, please see Spec PIS-2004, the mobile phone interconnection, for their key requirements.

### 对话管理器Dialog Manager

#### 文本回显Text Echo

系统应支持文本回显，实时将用户语音转成文本显示。

The system shall support the text echo to convert user voice into text display in real time.

当网络连接异常导致无法联网时，系统应仍支持文本回显。

When the Internet access fails because the network connection goes abnormal, the system shall still support the text echo.

#### 语音交互界面Voice Interaction Interface

语音交互界面应支持展示会话过程中的所有对话内容，包括用户的语音输入转写结果和系统反馈的内容，具体表现请参照交互设计。

The voice interaction interface shall support the display of all contents generated by the dialog, including the user’s voice-to-text result and system feedback: see the interaction design for details.

语音交互界面应支持用户手动操作并及时更新操作状态，具体表现请参照交互设计。

The voice interaction interface shall support the user’s manual operation and update the operation status in time: see the interaction design for details.

对于某些反复多轮的相似对话内容，如电话号码纠错，系统应支持将多条相似对话内容的最终结果合并为一条，具体表现请参照交互设计。

For some repeated rounds of similar dialogs, such as telephone number error correction, the system shall support the integration of final results of multiple similar dialogs into one: see the interaction design for details.

#### 对话模式Dialog Mode

为更好地，有针对性地服务于各类用户，自然语言识别系统应能分别支持详尽与简洁的对话模式。并且，系统应提供用户设置对话模式详尽与简洁的能力。

To better serve all types of users in a targeted manner, the natural language recognition system shall be able to support both exhaustive and concise dialog modes. Also, the system shall provide the user with the capability of setting the dialog mode to exhaustive or concise.

对话模式的出厂默认值应为详尽。

The factory default setting for dialog mode shall be exhaustive.

#### 对话风格（发音人）Dialog Style

为了提供各年龄段用户更好的用户体验，自然语言识别系统应支持多种~~两种~~对话风格，例如中规中矩、轻松活泼。并且，系统应提供用户~~设置~~切换对话风格的能力。

To provide a better user experience for users of all ages, the natural language recognition system shall be able to support both serious and breezy dialog styles. Also, the system shall provide the user with the capability of setting the dialog style.

需要特别注意的是，系统应使用不同~~2种~~TTS人声以分别对应不同~~中规中矩与轻松活泼两种~~对话风格（发音人），TTS发音人要求请参考3.2.1，TTS人声的选择应通过泛亚Review。

What calls for special attention is that the system shall adopt two TTS voices to correspond with the two dialog styles, namely serious and breezy, respectively and the selected TTS voices shall pass the Review of Pan Asia.

对话风格的出厂默认值应为中规中矩。

The factory default setting for dialog style shall be serious.

#### 语速Speech Rate

自然语言识别系统的对话系统应有能力支持较慢、标准、较快3档语速供用户选择和设置。并且对应档位的语速需通过泛亚Review。

The dialog system of the natural language recognition system shall support 3 scales of speech rate, namely Slow, Standard and Fast, for user’s selection and setting. Also, the speech rate of corresponding scale shall pass the Review of Pan Asia.

语速的出厂默认值应为标准。

The factory default setting for speech rate shall be standard.

#### 提示语/提示音Prompt

提示语/提示音即用户使用语音识别功能时听到的欢迎语/提示音。系统应支持用户设置提示语功能的开启和关闭，并且支持用户选择提示音或自定义提示语，提示语的长度应限制在2-8个汉字范围内，其中不支持标点符号。提示音需要通过泛亚Review。

The prompt is the welcome greeting that the user hears when activating the function of voice recognition. The system shall support the user to enable/disable the prompt function and to self-define the prompt; the length of prompt shall be confined to 2-8 Chinese characters, in which no punctuation is allowed.

提示功能的出厂默认状态应为~~关闭~~开启。

The factory default setting for prompt function is disabled.

如果用户在设置中开启提示功能，系统初始的默认提示语应为“~~您好~~在呢”，~~可以增加动态的提示语，~~每次唤醒后的提示语需要支持随机、动态的变化，其中动态的提示语包含但不仅限于“你好”、“哈喽”、“来啦”、“在”、“来咯”、“hi”等。

如果用户设置自定义提示语后，提示语固定，不做动态变化。

When the user enables the prompt function in the setting, the initial default prompt given by the system shall be “Hello”.

#### 话题切换 Topic Switching

系统应支持在任何时候响应用户切换话题，并正确地切换到新的意图，继续与用户进行对话。例如用户说完“播放周杰伦的歌”后，系统反馈用户找到多个结果，此时若用户回复“播放王菲的歌”或“打电话给周杰伦”，系统应切换到用户新的意图，去播放新的歌曲或打电话。

The system shall support responding to the user’s topic switching at any time and correctly switching to the new intent to carry on the conversation with the user. For example, after the user finishes speaking “play Jay Chou’s song” and the system feeds back the user that multiple results are found, provided that the user replies with “play Faye Wong’s song” or “call Jay Chou” at that time, the system shall switch to the user’s new intent, and play the new songs or make the phone call.

#### 上下文理解Context Understanding

在语音交互过程中，系统应记住用户每轮的对话内容，支持指代消除、成分补全，与用户进行多轮对话，直到完成用户意图。例如用户询问”北京的天气怎么样?”, 再说”导航去那”， 系统应可以自动利用上下文信息进行指代消解, 分析出用户是想”导航去北京”。具体支持的上下文理解场景，请参考3.1.1语音技能列表。

In the process of voice interaction, the system shall remember the content of each round of conversation with the user, support the anaphora elimination and element completion, and conduct multiple rounds of conversations with the user until the user’s intents are satisfied. For example, when the user asks “How is the weather in Beijing?” and then commands that “navigate to there”, the system shall be able to automatically apply the context information to complete the anaphora elimination and figure out that the user wants to “navigate to Beijing”. See 3.1.1 List of Voice Features for the supported context understanding scenarios.

#### 不支持指令的回复Reply for Unsupported Command

对于系统无法支持或暂时无法实现的语音控制功能，系统应给出友好的应答话术。

For the voice control function that is not supported or is temporarily unrealizable by the system, the system shall give a friendly response.

目前能识别的系统不支持或暂不支持的功能包含但不限于：对安全气囊的控制、巡航控制、车机开关机、后视镜折叠/打开、车辆点/熄火等。针对这些系统不支持或暂不支持的功能，用户通过语音助手进行控制时应给予友好的回应，如 “这个我暂时还没学会”，具体回复由交互设计确认。

Currently, the functions that are recognizable but not supported or temporarily not supported by the system include, but are not limited to, airbag control, cruise control, on/off of infotainment system, folding/unfolding of rearview mirror, vehicle start/stop, etc. For these functions that are not supported or temporarily not supported by the system, when the user tries to control them through the Voice Assistant, the system shall give a friendly response like “I have not yet mastered this skill”: the specific response shall be subject to the interaction design.

### Barge-In

语音识别系统应支持用户在VR过程中~~随时用语音~~通过唤醒方式（不仅限于唤醒词）打断对话流程，输入新的语音指令。

The voice recognition system shall allow the user to barge in the dialog process with voice at any time during the VR process and input a new voice command.

系统应支持允许用户在有歌曲播放/收音机播放等车机系统播放过程中，随时用语音打断对话流程，输入新的语音指令。

The system shall allow the user to barge in the dialog process with voice at any time and input a new voice command during the song play/radio play, etc through the infotainment system.

### 3.1.12主动交互 Active Interaction

系统应支持如下场景的主动交互。该功能应支持开关，~~默认为开启，~~具体开关项参考场景引擎PIS-2060。

The system shall support the active interaction under the following scenarios. This function, activated by default, shall be able to be activated and deactivated.

|  |  |  |
| --- | --- | --- |
| **场景**  **Scenario** | | **需求描述**  **Requirement description** |
| ~~导航场景~~  ~~Scenario of navigation~~ | ~~导航路线有更新~~  ~~Navigation route updated~~ | ~~系统导航过程中，如果检测到当前路况差且有更快的路线时，主动询问用户是否更换到更快的路线。~~  ~~During the navigation process, when it is detected that the current road condition is poor and there is a faster route, the system shall actively ask the user whether to switch to such faster route.~~ |
| ~~堵车时~~ | ~~主动提醒用户预计拥堵时间，并讲个笑话~~ |
| ~~超速时~~ | ~~当车速超过120km/h时，主动提醒用户安全驾驶，并播放舒缓的音乐~~ |
| 车辆异常场景  Scenario of vehicle abnormality | 油量过低  Low fuel level | 当油量过低时，主动提示用户"当前油量太少了，需要加油了，需要帮您找最近的加油站吗？”油量报警值参考PIS-2069 /PIS-2085 Warning number 139/332  在安吉星E call、APA、camera 使用过程中只需提示用户“油量过低”，不需要继续询问。  When the fuel level is too low, the system shall actively remind the user that “current fuel level is too low and the vehicle shall be refueled; shall I help you to locate the nearest gas station?” |
| 电量过低  Low battery | 电动车型，当电量过低时，主动提示用户"当前电量太少了，需要充电了，需要帮您找最近的充电站吗？”电量报警值参考PIS- TBD  在安吉星E call、APA、camera 使用过程中只需提示用户“电量过低”，不需要继续询问。  For the electric vehicle, when the battery is too low, the system shall actively remind the user that “Battery is too low and shall be charged; shall I help you to locate the nearest charging station?” |
| 后窗未关  Rear window left open | 熄火时如果用户的后窗没有关闭，主动提示用户记得关闭后窗。车窗信号请参考PIS-2045第3.3.1章节  Provided that the rear window is left open when the engine is turned off, the system shall actively remind the user to close the rear window. |
| 胎压异常提醒  Alert for abnormal tire pressure | 当胎压异常时应主动提示用户，询问用户是否要去附近的修理厂，如果用户说是则触发导航去修理厂，胎压异常信号请参考PIS-2069附表.  When the tire pressure goes abnormal, the system shall actively prompt the user and ask the user whether to head for the nearby repair shop; provided that the user replies Yes, the system shall trigger the navigation to the repair shop |
| 节假日场景  Scenario of festivals and holidays | 生日主动触发  Active trigger for birthday | 开机当日若为用户生日时，祝福用户生日快乐！系统应支持读取车辆账户中的生日字段，作为默认生日日期。若用户切换账户，系统应支持重置生日提醒的时机。  Provided that the day when the system gets started up falls on the user’s birthday, the system shall bless the user a happy birthday. The system shall support the reading of the birthday field in the vehicle account and then take it as the default birthday date. Where the user switches to another account, the system support the resetting of the timing for birthday reminder. |
| 重要节假日的主动触发  Active trigger for important festivals and holidays | 重要节假日需包含：中秋、端午、元宵、情人节、除夕、春节假日期间、圣诞节。  The important festivals and holidays shall include: Mid-Autumn Festival, Dragon Boat Festival, Lantern Festival, Valentine’s Day, New Year’s Eve, Spring Festival Holiday, and Christmas.  重要节假日的主动触发，既可以通过语音触发的方式，比如“主人，今天是冬至记得吃饺子哦”，也可以只通过改变语音助手呈现形象的方式，具体由交互设计确认。  The active trigger for important festivals and holidays is allowed to be realized through voice (for example, “Master, today is the Winter Solstice; please remember to eat dumplings”) or just by changing the way the Voice Assistant presents the image, which, however, shall be subject to the interaction design. |

当系统处于通话中、免打扰模式、语音交互、TTS播报、语音输入法、语音对讲过程中时，若有主动交互待触发，系统应将其放入缓存序列，待上述事件结束后，若仍满足触发条件，则系统应发起主动交互。

When the system is in a call, DND mode, voice interaction, TTS broadcast, phonetic entry method or voice intercom, provided that there is any active interaction to be triggered, the system shall put it into the cache sequence; then, after the aforesaid event ends, provided that trigger condition is still met, the system shall initiate an active interaction.

### 二次交互 Secondary Interaction

二次交互场景下，当系统反馈出多个内容结果时，系统应支持用户进行二次选择和确认，如“第x个/页”、“ 上一页”、“ 下一页”等。

Under the scenario of secondary interaction, when the system feeds back multiple content results, the system shall allow the user to make secondary selection and confirmation, such as “the xth piece/page”, “previous page”, “next page”, etc.

二次交互场景下，当系统反馈出多个内容结果时，系统应支持用户说出结果中的关键词进行快捷选择。关键词应在2个字或以上，且是连续的文字。若用户所说的关键词不是唯一的，系统应筛选出结果后，需要用户进行再次选择确认。二次交互支持范围具体由交互设计确认。具体支持的二次交互指令，请参考3.2语音技能列表。

Under the scenario of secondary interaction, when the system feeds back multiple content results, the system shall allow the user to speak out the keywords in the results for quick selection. The keyword shall contain 2 or more words that are consecutive. Provided that the keyword spoken out by the user is not unique, after filtering out the results, the system shall ask the user to make a second selection and confirmation. The supported range of secondary interaction shall be subject to the interaction design. See 3.2 List of Voice Features for the supported secondary interaction commands.

### 所见即可说 Seeing Is Saying

在语音助手使用中，系统应支持识别当前HMI界面可以点击的~~所有能够被看得见的~~文字部分，支持用户通过语音~~免唤醒，~~直接操作当前界面。~~例如在Popup界面，若有重试和取消按钮，系统应支持识别重试和取消指令，并执行对应重试和取消动作。~~需要支持的页面控件类型有：包含文字的List/ Tab/ Tag/ button，控件的具体定义请参考交互设计PIS-3000。对于同音或者完全相同的文字，需要让用户进行二次选择。如检索到同名的联系人时，当用户读出该联系人姓名后，需要询问用户选择第几个。

The system shall support the recognition of visible words on the current HMI interface, thereby enabling the user to directly operate the current interface through voice in a trigger-free manner. The supported “Seeing Is Saying” interface shall be subject to the interaction design.

## TTS

### 总体要求General Requirements

系统应支持将文字段转换为语音进行播放。

The system shall support the conversion of text field to speech for playing.

对于TTS转换的语音长度不应有任何限制。

There shall be no restriction on the length of speech converted by TTS.

系统应支持中文TTS、英文TTS，包括中英文混读。

The system shall support Chinese TTS and English TTS, including mixed reading in Chinese and English.

系统应支持多个发音人、男女声、特色方言的语音播报，对应语音播报特色需由供应商提供给泛亚Review。

需要支持5个标准普通话发音、1个四川口音普通话、1个粤语发音。

The system shall support the voice broadcast by multiple speakers, male & female voices, and featured dialects; the supplier shall provide the corresponding voice broadcast features to Pan Asia for Review.

### 性能要求Performance Requirements

TTS引擎从收到文本到开始播放的时间应小于1s；

The time taken from the receipt of text by TTS engine to its start of broadcast shall be less than 1s;

TTS语音质量的MOS评分应大于4.0分；

The MOS of TTS voice quality shall be greater than 4.0 points;

### 多音字的发音Pronunciation of Polyphones

系统应支持根据上下文和语境自动判断多音字的发音，包括人名、地名为多音字的情况。举例如下：

The system shall support automatic judgment of the pronunciation of polyphones based on the context and scenario, including the polyphones in people and place names. Some examples are given as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **测试语料**  **Corpus for test** | **词组**  **Word group** | **字**  **Word** | **期望读音**  **Expect pronunciation** | **错误读音**  **Wrong pronunciation** |
| 普通人参无此效果  Pǔ tōng rén shēn wú cǐ xiào guǒ  (Ordinary ginseng has no such effect) | 人参  Rén shēn (ginseng) | 参  Shēn/cān | Shen(第一声)  Shen (high-level tone/first tone) | Can(第一声)  Can (high-level tone/first tone) |
| 银行行长行走在人行道上  Yín háng hang zhǎng zǒu zài rén xíng dào shàng  (The bank president is walking on the sidewalk) | 银行  Yín háng (bank) | 行  Xíng/háng | Hang(第二声)  Hang (rising tone/second tone) | Xing(第二声)  Xing (rising tone/second tone) |
| 行长  Hang zhǎng (president) | 行  Xíng/háng | Hang(第二声)  Hang (rising tone/second tone) | Xing(第二声)  Xing (rising tone/second tone) |
| 人行道  Rén xíng dào (sidewalk) | 行  Xíng/háng | Xing(第二声)  Xing (rising tone/second tone) | Hang(第二声)  Hang (rising tone/second tone) |
| 打电话给仇老五  Dǎ diàn huà gěi qiú lǎo wǔ  (Call Qiu Laojiu) | 仇老五  Qiú lǎo wǔ  (Qiu Laojiu) | 仇  Qiú/chóu | Qiu(第二声)  Qiu (rising tone/second tone) | Chou(第二声)  Chou (rising tone/second tone) |

### 数字的发音Pronunciation of Numbers

系统应按照汉语习惯对出现数字的语句采用正确的发音。举例如下：

The system shall support the correct pronunciation of the number-contained statement according to Chinese habits. Some examples are given as follows:

|  |  |  |
| --- | --- | --- |
| **测试语料**  **Corpus for test** | | **期望的TTS结果**  **Expected TTS result** |
| 数字进制  Digital base | 新车优惠896,500元  Xīn chē yōu huì 896,500 yuán  (RBM 896,500 off for new car) | 新车优惠八十九万六千五百元  Xīn chē yōu huì bā shíjiǔ wàn liù qiān wǔ bǎi yuán |
| 年代  Year | 公元4000年前  Gōng yuán 4,000 nián qián  (4,000 BC) | 公元四千年前  Gōng yuán sì qiān nián qián |
| 秦朝建立于公元前221年  Qín cháo jiàn lì yǔ gong yuán qián 221 nián  (The Qin Dynasty was founded in 221 BC) | 秦朝建立于公元前两百二十一年  Qín cháo jiàn lì yǔ gong yuán qián liǎng bǎi èr shíyī nián |
| 时间  Time | 2017/10/27  (October 27, 2017) | 二零一七年十月二七号  Èr líng yī qī nián shí yuè èr qī hào |
| 23:11 | 二十三点十一分  Èr shísān diǎn shíyī fēn |
| 5pm | 下午五点  Xiàwǔ wǔ diǎn |
| 电话  Phone number | 182-3123-3213 | 幺八二三幺二三三二幺三  Yāo bā èr sān yāo èr sān sān èr yāo sān |
| 分数/小数/百分数  Fraction/decimal/percentage | 1/2  (One second) | 二分之一  Èr fēn zhī yī |
| 65%~70% | 百分之六十五至百分之七十  Bǎi fēn zhī liù shíwǔ zhì bǎi fēn zhī qīshí |
| 0.9882 | 零点九八八二  Líng diǎn jiǔ bā bā èr |

### 常见符号的发音Pronunciation of Common Symbols

系统应支持包含标点符号、单位、英文夹杂等情况的语句的正确发音。举例如下：

The system shall support correct pronunciation of the statement containing punctuation, unit, English inclusion, etc. Some examples are given as follows:

|  |  |  |
| --- | --- | --- |
| **测试预料**  **Corpus for test** | | **期望的TTS结果**  **Expected TTS result** |
| 标点符号  Punctuation | 视频时长60’5’’  Shì pín shí cháng 60’5’’  (Video duration 60’5’’) | 视频时长为60 分5秒  Shì pín shícháng wéi 60 fēn5 miǎo |
| 屏幕比例为10:9  Píng mù bǐ lì wèi 10:9  (The screen ratio is 10:9) | 屏幕比例为10比九  Píng mù bǐ lì wèi 10 bǐ jiǔ |
| 单位  Unit | $0.50 | 零点五零美元  Líng diǎn wǔ líng měi yuán |
| 1μＭ | 一微米  Yī wéi mǐ |
| 40.3°C | 四十点三摄氏度  Sì shí diǎn sān shè shì dù |
| 英文夹杂  English inclusion | 我真的不是很care  Wǒ zhēn de bù shì hěn care  (I really don’t care) | 我真的不是很care  Wǒ zhēn de bù shì hěn care |
| 4G的最高下载速度  4G de zuì gāo xià zài sù dù  (The maximum download speed of 4G) | 4G的最高下载速度  4G de zuì gāo xià zài sù dù |

# 系统需求System Requirements

## 账号相关Account Related

系统应支持基于帐户提供个性化的语音服务；

The system shall support personalized voice services based on the account.

需要有能力保存用户添加reminder的录音，reminder功能请参考PIS-2031.

## 应用内设置In-application Setting

1. 语音唤醒词

1. Voice trigger word

* 1. 系统应支持开启或关闭语音唤醒功能。

a) The system shall support activation/deactivation of the voice trigger function.

* 1. 系统应支持用户开启、关闭自定义唤醒词，开启后可以修改语音唤醒词，唤醒词长度在~~2~~ 3~6个字之间。需要对用户修改的词汇审核，若因为敏感词汇、叠音词等导致的审核不通过，需要友好的告知用户设置失败的原因。具体参考交互文档。

b) The system shall support the user to modify the voice trigger word, which shall have a length of 3~6 characters.

1. ~~语音启动~~ 提示音/提示语

2. Prompt for VR activation

* 1. 系统应支持开启或关闭~~语音启动~~提示语。

a) The system shall support activation/deactivation of the prompt for VR activation.

* 1. 系统应支持编辑或修改~~语音启动~~提示语。支持用户选择提示音或提示语，提示语长度范围在2~8个字之间。需要对用户修改的词汇审核，若因为敏感词汇、叠音词等导致的审核不通过，需要友好的告知用户设置失败的原因。具体参考交互文档。

b) The system shall support the user to edit or modify the prompt for VR activation, which shall have a length of 2~8 characters.

1. 语音播报

3. Voice broadcast

* 1. 系统应支持设置语音交互的语速，分为较慢、标准、较快三档。默认值为标准。

a) The system shall support setting of the speech rate for voice interaction, which is divided into three scales, namely Slow, Standard and Fast, and set to Standard by default.

1. 限时免唤醒

4. Trigger-free function within a limited time

* 1. 系统应支持开启/关闭以及切换限时免唤醒时长功能，默认为15秒。该功能应支持用户选择3秒、15秒、30秒、60秒~~，默认为开启~~

a) The system shall support activation/deactivation of the trigger-free function within a limited time, which is 15 s default.

1. 常用说法免唤醒

5. Trigger-free function realized upon common expression

* 1. 系统应支持开启或关闭常用说法免唤醒，默认为开启。

a) The system shall support activation/deactivation of the trigger-free function realized upon common expression, which is activated by default.

1. 对话模式

6. Dialog mode

* 1. 系统应支持用户设置对话模式，包括详尽与简洁。默认为详尽。

a) The system shall allow the user to set the dialog to an exhaustive or concise mode, which is set to the exhaustive mode by default.

1. 发音人

针对不支持彩蛋功能的车型，系统应支持用户~~选择~~切换语音播报发音人，支持切换5个标准普通话发音人、1个四川口音的普通话发音人。

针对支持彩蛋功能的车型，系统应支持用户~~选择~~切换语音播报发音人，需要预留一个普通话发音人给彩蛋功能，在彩蛋激活之前，支持切换~~5~~ 4个标准普通话发音人、1个四川口音的普通话发音人；在彩蛋激活后，支持切换5个发音人、1个四川口音的普通话发音人。彩蛋功能中语音相关需求及适用车型请参考PIS-2100第3章节。

当用户选择粤语识别后，仅支持1个粤语发音人。

语音发音人的样音包请参考RTC域TTS样音包。默认为第一个发音人。

发音人具体文言显示参考交互。当用户选择一个发音人后，需要播放简短的语音样音向用户展示。

支持用户选择定制化的TTS发音。该TTS发音由用户在手机端百度账号中训练定制，车机端登陆百度账号后同步用户定制化的TTS。最多支持5个定制化的TTS语音包。

如果用户没有登录百度账号，可以引导用户登录后，同步定制化的TTS语音包。具体参考交互设计。

~~对话风格Dialog style~~

* 1. ~~系统应支持用户设置对话风格，包括中规中矩与轻松活泼。默认值为中规中矩。~~

~~a) The system shall allow the user to set the dialog to a serious or breezy dialog style, which is set to the serious style by default.~~

1. ~~主动交互~~

~~8. Active interaction~~

* 1. ~~系统应支持开启或关闭主动交互，默认为开启。~~

~~a) The system shall support activation/deactivation of the active interaction, which is activated by default.~~

1. 声源定位

9. Sound source localization

a)系统应支持设置声源定位模式，包括~~自动响应所有录音方向、~~仅响应主驾方向、仅响应唤醒人方向。默认为~~自动响应所有录音方向~~仅响应唤醒人方向。具体文言以交互为主。

The system shall support setting of the sound source localization mode, including automatic response to the recording direction, which is the default setting, and response only to the voice from driver side.

10. 默认语音助手（TBD）

系统默认的语音助手是本地语音助手，同时也支持用户选择CarPlay Siri或CarLife 小度作为系统的默认语音助手。语音助手之间的仲裁逻辑（包含硬按键等操作）请参考PIS-2004.

* 当用户选择CarPlay Siri或CarLife 小度作为default语音助手时，点击本地语音助手拟人形象或VR 应用图标后，需要唤醒本地语音助手；如果此时default语音助手正在使用中，将会打断default语音，唤醒本地语音助手
* 当用户选择CarPlay Siri或CarLife 小度作为default语音助手时，通过方向盘硬按键会唤醒对应的语音助手，不会唤醒本地语音助手
* 当用户选择一种语音助手作为default语音时，只会响应当前默认语音助手的唤醒词
* 当前没有CarPlay、CarLife连接时，设置项对应的选项需要区别显示，具体参考交互

11. 帮助

支持在setting页面查看语音帮助，具体显示请参考交互。

12. 拟人形象助手

针对不支持彩蛋功能的车型，支持用户切换多个语音拟人形象助手。拟人形象助手请参考4.3.1章节。

针对支持彩蛋功能的车型，支持用户切换多个语音拟人形象助手。在彩蛋模式未激活时，需要隐藏该彩蛋模式的拟人形象助手这一项；彩蛋模式激活后，需要显示该拟人形象助手在列表中。

13. 方言识别

支持用户选择普通话识别或粤语识别~~，选择粤语后，发音人界面需要仅显示一个粤语发音人~~。

## 外部调用External Call

系统应对外提供语音识别和自然语言理解的调用能力。

The system shall support outward invocation of the voice recognition and natural language understanding.

系统应对外提供TTS播报、播报人选择、播放速度调节的能力。

The system shall support outward provision of the TTS broadcast, broadcaster selection, and broadcasting speed adjustment.

### 拟人形象助手/Personified Assistant

系统应支持对外提供语音助手使用中识别的状态(即VR 过程)，包括监听中、识别中、处理中、语音播报中。

系统需要通过拟人形象来展示上述不同状态，各品牌的默认拟人形象需要有差异，具体由交互设计定义。支持用户自定义拟人形象，并上传至云端保存。用户可以从云端下载可替换的形象、用户自定义的形象，~~下载方式TBD~~。

The system shall support outward provision of the voice recognition state, including under monitoring, under recognizing, under processing, and under voice broadcasting.

### 氛围灯随动/Ambient light

语音助手使用状态中，需要发送状态给氛围灯，氛围灯具体需求及支持的车型请参考PIS-2090.

### 车辆故障查询/ Vehicle fault inquiry

当车辆发生故障在仪表显示telltale时，在仪表显示期间，需要支持用户通过语音查询当前故障信息，并给出友好提示。需要显示的Telltale故障请参考附件，仪表Telltale具体描述请参考PIS-2062。

如果仪表显示多个故障时，VR需要按照触发时机逐一播报。



## 版本升级Version Upgrade

系统应支持语音助手核心能力的独立升级，即通过SGM App Shop（APP更新），如识别算法更新、唤醒模型更新、对话流程及对话能力的更新、对话话术的更新、识别率及性能的优化更新、新增在线垂类等。

The system shall support independent upgrade for the core capabilities of Voice Assistant, such as recognition algorithm, trigger model, etc.

## 埋点需求Event Tracking Requirements

请参考PIS-2007 埋点需求。

See PIS-2007 Event Tracking Requirements for details.

## 相关CAN信号Related CAN Signals

车控信号请参考PIS-2045; 空调相关信号请参考PIS-2014；氛围灯信号请参考PIS-2090。

## 相关标定项 Related Calibration Items

系统默认的语音唤醒词应读取车型品牌标定；标定名称为：GMBrand。

标定值为1对应别克，默认语音唤醒词为“你好，别克”；

标定值为2对应凯迪拉克，默认语音唤醒词为“你好，凯迪拉克”；

标定值为3对应雪佛兰，默认语音唤醒词为“你好，小雪”。

The system default voice trigger words shall be read for calibration;

声源定位的标定项：VR\_SpeechZone\_Microphone\_type，标定值为0代表没有麦克风、标定值为1代表仅一个模拟麦克风，这两个标定的车型不需要支持声源定位。

# 附录Appendix

N/A

**Revision Log**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Version** | **Date** | **Section** | **Description** | **Author** |
| 0.6 | 2021/1/11 | 3.1.1 | 更新VR 技能列表，澄清车辆状态参考文档 | Zhang Jiajia |
| 3.1.6; 4.2 | 删除声源定位全方位模式 |
| 4.2 | 补充自定义唤醒词开关 |
| 2021/1/18 | 4.3.3 | 新增车辆故障查询 |
| 2021/3/3 | 4.2 | 修改声源定位的默认项 |
| 4.7 | 补充声源定位的标定项 |
| 0.5 | 2020/10/14 | 3.1.1 | 更新VR 技能列表，补充新功能 | Zhang Jiajia |
| 3.1.9 | 经确认OnStar无此需求，删除VR交互中OnStar的需求 |
| 2020-11-11 | 4.2 | 根据北美VR策略，补充默认的VR设置；在setting中补充帮助入口 |
| 2020-11-12 | 4.3.2 | 根据CR 66784新增氛围灯随动功能 |
| 4.2;3.1.12 | 删除主动交互的开关项，具体参考场景引擎文档 |
| 3.1.6 | 根据CR 57949澄清声源定位需求 |
| 2020-11-19 | 3.1.2 | 补充语音功能不受其他语言环境的影响 |
| 4.2 | 根据CR 67374增加彩蛋功能发音人拟人形象助手需求 |
| 4.7 | 补充品牌标定项 |
| 2020-11-27 | 3.1.4 | 根据PIS-3000的画面优先级，补充禁止使用语音的场景 |
| 3.1.9 | 删除不存在的硬按键，澄清MUTE&HANGUP按键响应 |
| 2020-12-10 | 4.2 | 补充语音发人样音包的参考文档及默认选项 |
| 2020-12-16 | 3.1.6 | 由于性能损耗过大，删除声源定位的全方位响应模式 |
| 0.4 | 2020/6/17 | 3.1.2 | 澄清方言、口音的识别需求 | Zhang Jiajia |
| 3.1.3.5 | 澄清离线指令的响应时间 |
| 3.1.4.1 | 增加用户修改唤醒词后，系统默认的主唤醒词仍需要支持唤醒功能 |
| 3.1.5 | 澄清麦克风降噪的场景 |
| 3.1.6，4.2 | 修改声源定位文言描述 |
| 3.1.9 | 调整硬按键的名称 |
| 3.1.10.6 | 增加提示音的选择 |
| 3.1.12 | 删除堵车、高速等工况的主动交互 |
| 3.2.1 | 澄清语音播报的具体需求 |
| 4.1 | 澄清reminder保存录音的需求 |
| 4.2 | 删除对话风格设置，归类为语音播报人的选择设置项,修改提示语/提示音，根据CR 45462增加自定义TTS; 修改显示免唤醒选择项 |
| 4.3 | 补充VR 使用状态的描述 |
| 4.7 | 删除安吉星的标定需求 |
| 2020/7/10 | 3.1.10.6 | 修改默认提示音/提示语为开启状态，增加动态提示语 |
| 4.3.1 | 根据CR 45462补充拟人形象的功能描述 |
| 4.4 | 根据CR 45462补充版本升级需求 |
| 3.1.4.3 | 修改限时免唤醒的时间选择 |
| 3.1.12 | 删除导航变更路线的主动交互（与导航功能重复）补充主动交互相关的信号 |
|  | 2020/10/28 | 3.1.14 | 澄清所见即可说支持的页面 |  |
| 2020/11/4 | 3.1.14 | 澄清所见即可说支持的二次确认说法 |
| 0.3 | 2019/12/31 | All | 增加中英文对照翻译；  Add Chinese-English translation; | Jin Zhiyi |
| 0.2 | 2019/9/3 | 3.1.10.7 | 增加话题切换功能  Add the function of topic switching | Jin Zhiyi |
| 0.1 | 2019/8/8 | All | Initial | Jin Zhiyi |