唤醒特性PRD

**产品需求说明书**

|  |  |
| --- | --- |
| 所属项目 | 思必驰\_天琴 |
| 所属部门 | 智能座舱中心 |
| 编辑 | 叶冬秀、王思 |
| 审核 | 柯学滨 |
| 版本 | 1.0 |
| 日期 | 2022-04-13 |

# 历史版本

|  |  |  |  |
| --- | --- | --- | --- |
| **日期** | **版本** | **变更描述** | **作者** |
| 2022/04/14 | 1.0 | 新建文档 | 叶冬秀、王思 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# 名词术语解释

|  |  |
| --- | --- |
| **术语** | **说明** |
| 阈值 | 唤醒阈值是用来衡量唤醒灵敏度的一个标准，阈值越高，唤醒越难，即灵敏度越低。  一般来说在设置唤醒阈值时，针对不同车企，不同车型，需要通过大量数据训练，不断进行测试，以选择最适合的一个数值。 |

目录

[历史版本 2](#_Toc104572254)

[名词术语解释 3](#_Toc104572255)

[1 产品概述 5](#_Toc104572256)

[1.1产品背景 5](#_Toc104572257)

[1.2 适配项目 6](#_Toc104572258)

[1.3 需求目的 6](#_Toc104572259)

[2 功能描述 6](#_Toc104572260)

[2.1 语音唤醒功能概述 6](#_Toc104572261)

[2.2 复数唤醒词 6](#_Toc104572262)

[2.3 自定义唤醒词 6](#_Toc104572263)

[2.3.1 自定义唤醒词的设置方式 7](#_Toc104572264)

[2.4 语音助理的唤醒方式 9](#_Toc104572265)

[2.4.1 语音唤醒 9](#_Toc104572266)

[2.4.2 方控和点击唤醒 10](#_Toc104572267)

[2.5 免唤醒设置 10](#_Toc104572268)

[2.5.1 全局免唤醒 10](#_Toc104572269)

[2.5.2 场景内免唤醒 11](#_Toc104572270)

[2.5.3 关键词免唤醒 12](#_Toc104572271)

[3 唤醒技术介绍 14](#_Toc104572272)

[3.1 语音唤醒的定义 14](#_Toc104572273)

[3.2 语音唤醒技术原理 14](#_Toc104572274)

[3.2.1 原理简介 14](#_Toc104572275)

[3.3 语音唤醒技术架构 16](#_Toc104572276)

[3.3.1 唤醒技术架构图 16](#_Toc104572277)

[3.4 唤醒技术特点 17](#_Toc104572278)

[3.5 语音唤醒功能说明 18](#_Toc104572279)

[3.5.1 主唤醒词 18](#_Toc104572280)

[3.5.2 快捷唤醒词 18](#_Toc104572281)

[3.5.3 唤醒词选择 19](#_Toc104572282)

[3.6 语音唤醒性能指标 19](#_Toc104572283)

[3.7 终端唤醒模式 21](#_Toc104572284)

[3.8 其他唤醒实现逻辑 21](#_Toc104572285)

# 1 产品概述

## 1.1产品背景

随着人工智能语音技术的发展，语音助理在智能汽车中得到广泛应用，使用唤醒词唤醒语音助理是用户使用语音功能时最高频场景之一。自定义唤醒词，即用户用户可以基于个人喜好，为语音助理取一个名字，成功后可以通过这个名字来唤醒语音助理，是最能够体现用户个性化的功能之一。然而，这个功能仍然面对着定制后唤醒词误唤醒率高、识别率低的问题。当技术瓶颈时,一个完整的唤醒词合法性判断系统、引导用户设置出更容易唤醒的词语是优化此问题的方法之一，本文档就是用于说明唤醒词及其相关功能设置。

## 1.2 适配项目

天琴所有项目

## 1.3 需求目的

通过语音功能完成对唤醒词自定义及相关功能设置的说明, 并从产品设计引导上, 优化关于唤醒词误唤醒率高、识别率低等问题，通过唤醒语音助手，帮助用户执行语音指令，为用户提供更方便，更人性化的唤醒方式。

# 2 功能描述

## 2.1 语音唤醒功能概述

语音唤醒功能主要是通过特定词语或者操作唤醒语音进入监听状态，以达到对用户说话进行收音的目的。自定义唤醒词功能主要提供语音指令输入和手动输入两种设置方式。在自定义唤醒词过程中，我们会对新唤醒词进行合法性判断，如敏感词过滤、叠词判断等等，帮助用户设置一个唤醒率更高的词语， 满足用户在车内通过语音完成对设备的控制和操作的需求。与此同时，车主可以通过按键和固定的词语来激活语音助理进行对话。

## 2.2 复数唤醒词

功能描述：复数唤醒词也叫自定义主唤醒词，企业用户可根据企业文化等需要定制主唤醒词，用户可以通过该唤醒词来唤醒语音助理，为了降低误唤醒概率，目前支持企业用户最多可定制三个唤醒词，如“你好，小驰/小驰，你好/嘿 小驰”等，需要付费定制。

## 2.3 自定义唤醒词

* 功能描述：自定义唤醒词是指在主唤醒词的基础上，车主用户可以基于个人喜好等，为语音助理取一个别称，成功后可以通过该别称来唤醒语音助理。在自定义唤醒词过程中，我们会对新唤醒词进行合法性判断如敏感词过滤、叠词判断等，通过严格且合法的判断来过滤掉辨识度低的唤醒词，提升对语音助理的唤醒率。自定义唤醒词的设置方式主要分为手动设置和语音设置两种，同时支持用户通过语音或手动的方式对语音助理的的唤醒词进行修改和删除等操作。
* 原理描述:针对不同车企，不同车型，通过大量数据训练（需定制）可以得到相应的唤醒阈值，输入语音信号经前端信号处理，提取声学特征，输入到唤醒声学模型，计算唤醒词的得分，并与唤醒词对应的阈值进行比较，若某唤醒词得分大于或等于阈值，则该唤醒词被检测到，并执行相应的操作；反之，则不唤醒。阈值越高，唤醒的灵敏度越低，唤醒越难，反之亦然。

### 2.3.1 自定义唤醒词的设置方式

#### 2.3.1.1 手动设置自定义唤醒词

1、功能描述：用户通过语音形象界面->语音设置->自定义唤醒词的路径，找到手动修改唤醒词的设置项，手动输入名称并保存，即可设置。

2、手动设置唤醒词的要求：

* 自定义唤醒名称不可为空，且最多设置一个。
* 支持任意的中文汉字组合，暂时不支持英文唤醒词设置。唤醒词的长度在3到5个汉字之间，4个汉字为最佳（手动输入自定义唤醒词名称时，车机页面提示支持输入的字数为3~5个中文字）。
* 不支持输入的唤醒词名称包含敏感词（脏话、涉政、涉黄、伟人名字等）,自定义唤醒词会进行敏感词过滤，当自定义唤醒词的拼音中包含敏感词拼音时或音调，都会返回“不能包含敏感词汇哦”的反馈提示。
* 唤醒词应尽量避免生活化、口语化以及日常生活中使用频率很高的词汇，且避免使用多音字，唤醒词之间的音节覆盖应尽量多，差异大，避免使用叠字以及连续零声母的词汇。
* 唤醒词的发音应尽量响亮，且易开口，发音清晰分辨的词。
* 手动输入自定义唤醒词名称时，车机系统自动检测是否符合要求，不符合的话要提示用户修改，并给出默认示例，自定义唤醒词引导。
* 自定义唤醒词名称录入后，点击保存后才生效。

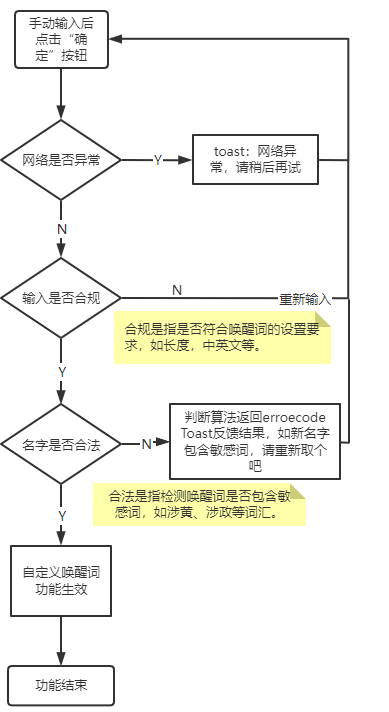
3、其他情况说明

为了提高唤醒成功率和提升唤醒效果，自定义唤醒词最多只支持用户设置一个，且自定义唤醒词只记忆用户最新定义的唤醒词，之前定义的唤醒词失效，且原先出厂时自带的主唤醒词会永久保留。

语音设置自定义唤醒词成功后，语音设置项的名称也需要同步修改。

自定义唤醒词设置失败，即保留上次设置结果。

4、手动设置自定义唤醒词逻辑流程图



#### 2.3.1.2 语音设置自定义唤醒词

1、功能描述:车主用户可以通过输入语音指令来设置唤醒词，如使用“给你起个名字

#名字#”等指令来设置唤醒词，生效后语音助理会提示相关信息，如“好的，以后可以说#名字#呼叫我”等;同时，车主用户也可以使用语音指令多轮交互实现唤醒词的语音设置，如“我想给你取个名字”—>“你想叫我什么呢”—>“就叫#名字#”—>“好的，以后可以说#名字#来呼叫我

”。

语音设置唤醒词的要求：

* 支持任意的中文汉字组合，暂时不支持英文唤醒词设置。唤醒词的长度在3到5个汉字之间，4个汉字为最佳。
* 不支持输入的唤醒词名称包含敏感词（脏话、涉政、涉黄、伟人名字等）,自定义唤醒词会进行敏感词过滤，当自定义唤醒词的拼音中包含敏感词拼音时或音调，都会返回“不能包含敏感词汇哦”的反馈提示。
* 唤醒词应尽量避免生活化、口语化以及日常生活中使用频率很高的词汇，且避免使用多音字，唤醒词之间的音节覆盖应尽量多，差异大，避免使用叠字以及连续零声母的词汇。
* 唤醒词的发音应尽量响亮，且易开口，发音清晰分辨的词。

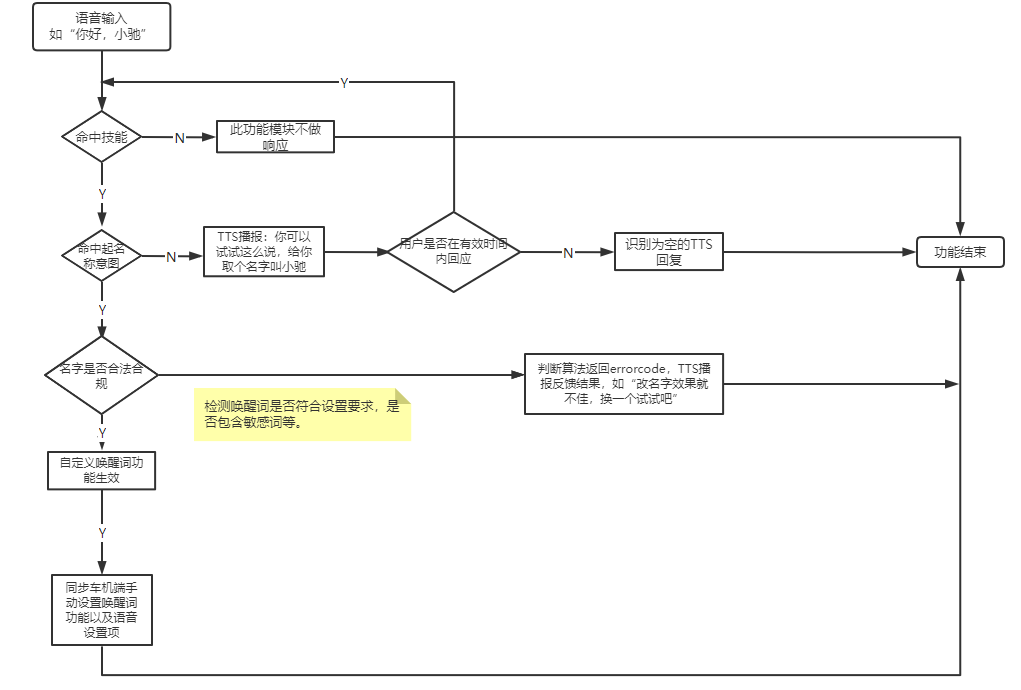
3、其他情况说明

为了提高唤醒成功率和提升唤醒效果，自定义唤醒词最多只支持用户设置一个，且自定义唤醒词只记忆用户最新定义的唤醒词，之前定义的唤醒词失效，且原先出厂时自带的主唤醒词会永久保留。

语音设置自定义唤醒词成功后，语音设置项的名称也需要同步修改。

自定义唤醒词设置失败，即保留上次设置结果。

4、语音设置自定义唤醒词逻辑流程图



## 2.4 语音助理的唤醒方式

### 2.4.1 语音唤醒

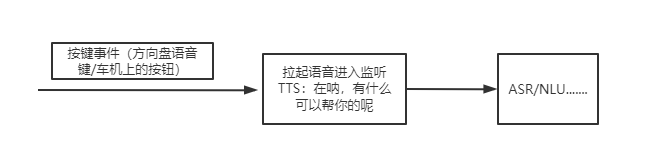
1、功能描述：用户通过输入的唤醒词指令直接唤醒语音助理。如当用户说“你好小驰/小驰你好/嘿小驰”时，语音助理被唤醒，进入新一轮的语音交互。

### 2.4.2 方控和点击唤醒

1、功能描述：用户通过方向盘上的语音按键及车机上按钮，实现唤醒语音交互的功能，当按键唤醒事件触发时，语音助手状态立即切换为唤醒状态，语音进入新的一轮交互。除了在特殊情况倒车情况下按键不响应，其他场景均支持。

2、当车主用户第一次按下语音按键的时候，语音助理被唤醒，唤醒状态下按下第二次的时候，语音助理退出。

#### 2.4.2.1 方控和点击唤醒的逻辑流程图



## 2.5 免唤醒设置

### 2.5.1 全局免唤醒

1、功能描述：全局免唤醒功能是为用户提供常用高频操作的快捷语音指令操作，不需要唤醒语音助理直接说，固定说法，既有唤醒功能，也同时具有指令执行功能。全局免唤醒的快捷指令需要用户完整按照特定预置的词语进行表达，命令词生效范围也根据该命令词的功能而定。如当音乐播放器播放时，用户可以直接说“上一首/下一首/暂停播放”等命令实现切歌和暂停操作。

2、全局免唤醒指令清单

|  |  |  |
| --- | --- | --- |
| **唤醒场景** | **模块** | **快捷唤醒命令** |
| 免唤醒（全局） | 车机系统 | 增大音量 |
| 减小音量 |
| 关闭声音 |
| 打开屏幕 |
| 关闭屏幕 |
| 打开声音 |
| 关闭声音 |
| 调亮屏幕 |
| 调暗屏幕 |
| 返回主页/桌面 |
| 音乐/电台 | 暂停播放 |
| 继续播放 |
| 上一首 |
| 下一首 |

### 2.5.2 场景内免唤醒

1、功能描述：场景内免唤醒是指在某些特定的页面下生效的命令词，它是以词组的方式出现，以帮助用户在该页面和流程下进行快速的选择和意向表达操作，仅在目标功能场景下支持的唤醒词，其他场景使用无效。用户需要按照预置的特定的词语进行完整表达，才能触发生效。如在来电时，用户可以通过直接说“接听/挂断”来进行来电操作，而不需要唤醒，在非来电时直接说接听/挂断语音指令不响应。

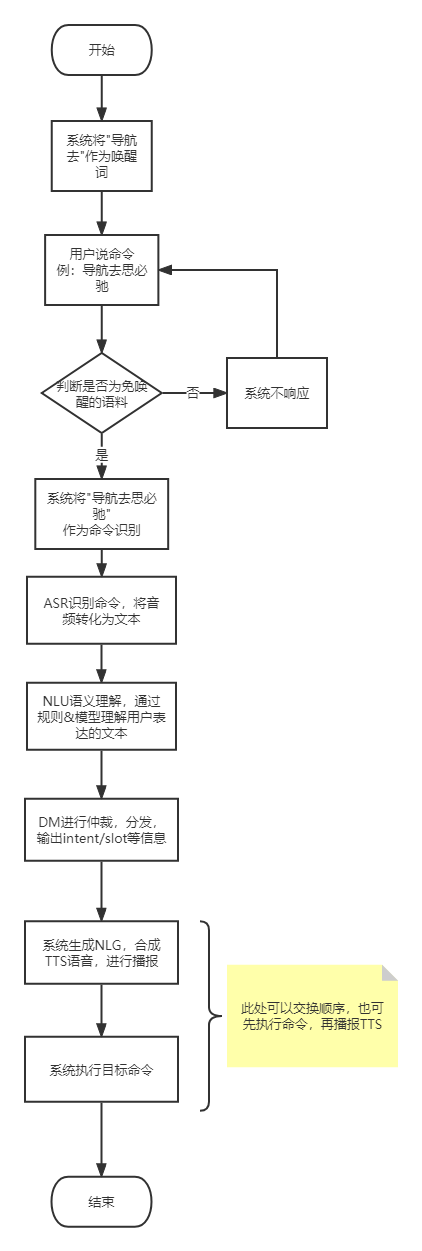
2、场景内免唤醒指令清单

|  |  |  |
| --- | --- | --- |
| **唤醒场景** | **模块** | **快捷唤醒命令** |
| 免唤醒（场景） | 地图/导航（地图和导航状态都生效） | 打开路况 |
| 关闭路况 |
| 放大地图 |
| 缩小地图 |
| 白天模式 |
| 夜间模式 |
| 车头朝上 |
| 北朝上 |
| 2D模式 |
| 3D模式 |
| 导航（只有在导航状态下才生效） | 查看全程 |
| 躲避拥堵 |
| 高速优先 |
| 不走高速 |
| 避免收费 |
| 往哪儿走 |
| 还要多久 |
| 返回导航 |
| 关闭导航 |
| 来电触发 | 接听 |
| 挂断 |

### 2.5.3 关键词免唤醒

1、功能描述：关键词免唤醒是指在导航、音乐和电话等核心场景下，用户可以不用说出唤醒词，直接合并关键词说出自己的意图，实现更自然的人机对话，即【关键词+可配置的内容】组成的指令，支持说法“导航去xxx/我想听xxx/打电话给xxx”,如“导航去世界之窗/我想听红豆/打电话给张三”，为了降低误唤醒率，暂时不支持对该功能语义做泛化。

#### 2.5.3.1 关键词唤醒词逻辑流程图



不同的唤醒词概念，主要是从产品功能的来区分的。主副唤醒词就是目前唤醒业务主要支持的，由客户定制的唤醒词，比如"你好xx"。快捷唤醒词、免唤醒词，属于一种拓展的功能，基于唤醒模型的通用能力，设定一些车机交互的指令词，命令词。

# 3 唤醒技术介绍

## 3.1 语音唤醒的定义

随着便携式移动设备和智能硬件设备的普及，通过语音对话的方式实现人机交互已大势所趋，并且语音交互大大增强了产品的用户体验。语音唤醒作为一种重要的语音交互的入口，已经在众多产品中得到应用和普及，比如Google 的“Okay Google”，Apple的“Hey Siri”，Amazon的“Alexa”，小米的“小爱同学”，阿里的“天猫精灵”，OPPO的“小布小布”，海信电视的“海信小聚”，车萝卜的“你好萝卜”等等，语音唤醒已经广泛应用在智能音箱，智能手机，智能车载以及智能IOT设备中。

语音唤醒在学术上被称为keyword spotting(简称KWS)，在连续语音流中实时检测出说话人特定片段(keyword)。与语音识别的区别是，语音识别是一个识别任务，主要目的是将语音转化为文字，预先不知道语音中所包含的文字内容，通过语音识别系统，将语音对应的文字转化出来。语音唤醒是一个确认任务，已经预先定义了关键字或唤醒词，语音唤醒系统检测语音流中是否包含该关键字或唤醒词，并给出特定的响应。识别和确认是两种不同类型的任务，类似的还有说话人识别和说话人确认。

## 3.2 语音唤醒技术原理

### 3.2.1 原理简介

汽车内部声学环境比较恶劣，存在各种噪声的干扰人机交互，需要对采集到的麦克风信号做处理后，后续的唤醒识别模块才能有性能保证。车载用到的语音增强算法主要包含：1、回声消除2、降噪算法，可以分为单麦克风降噪和双麦克风降噪。

回声消除全名是声学回声消除(Acoustic Echo cancellation, AEC)。声学回声指的是设备自身扬声器播放的声音经不同的路径一次或多次反射后进入麦克风所产生的回声集合,也可称作设备自噪声。用户通过语音同设备进行交互的时候，回声信号和干净的语音信号混合，这会恶化采集到的语音信号的信噪比，严重干扰后续的信号处理算法和唤醒识别模块的性能。所以原始麦克风信号要先通过回声消除算法模块，消除设备自噪声，以达到提升信噪比的目的。回声消除主要的原理是采用自适应滤波技术，动态的实时跟踪车内的声学信道，参考音经过这个信道的滤波，模拟出传到麦克风处的回声，最后原始的麦克风信号减掉这个回声信号，达到消除车载设备自噪声的目的。

降噪算法主要处理的噪声是指车载的环境背景噪声，比如空调噪声，发动机噪声、胎噪、风噪，这类噪声通常不具有空间指向性，噪声能量随时间的变化相对比较平稳。降噪算法主要的原理是，采用专门为车载噪声特点设计的噪声跟踪算法，去实时动态的估计目前车内噪声特征，然后利用语音增强算法把估计出的噪声抑制掉，达到增强用户语音的目的最终提升在车载噪声环境下的语音系统性能。

语音唤醒实现的技术方案有很多种，主要有基于语音识别的语音唤醒，基于音素搜索的语音唤醒，以及基于kws-filler解码的语音唤醒等。每种语音唤醒技术都有其优缺点。综合各种技术方案的优缺点，我们实现了一种基于深度神经网络的语音唤醒技术，利用深度神经网络强大的学习能力对声学单元进行建模，并通过一系列的技术优化，使得我们的语音唤醒技术达到国内领先水平。

## 3.3 语音唤醒技术架构

### 3.3.1 唤醒技术架构图

**前端**

**信号处理**

***输入语音***

**声学**

**特征提取**

**声学**

**模型**

**唤醒词列表**

**唤醒结果**

输入语音信号经前端信号处理，提取声学特征，输入到唤醒声学模型，计算唤醒词的得分，并与唤醒词对应的阈值进行比较，若某唤醒词得分大于或等于阈值，则该唤醒词被检测到，并执行相应的操作；反之，则不唤醒。

双MIC增强技术架构

****

车载双麦克风增强技术从功能上分为主驾模式与全局模式，在全局模式下，利用空间信息与噪声特性，主副驾的声源语音都会得到增强，使其获得相比单麦更高的唤醒率与识别率；在主驾模式下，通过波束形成技术，同时结合DOA估计，对主驾声源进行增强，而对副驾的语音进行抑制，实现主驾位置更高的唤醒率与整车更低的误唤醒率。

双麦克风增强的特点：

1）高唤醒率与识别率：在全局与主驾模式下，语音交互的唤醒率与识别率都较单麦有很大的提升；

2）低误唤醒率：在主驾模式下，除主驾驶方向的语音得到增强外，来自其他方向的语音被抑制了，在很大程度上降低了误唤醒；

3) 抗干扰：在主驾模式下，能较好地抑制来自副驾驶方向的语音，使副驾驶唤不醒且不能进行语音交互，让车载设备成为主驾驶的语音助手，让主驾驶的语音交互不受副驾驶或其他方向的语音的影响；

4）抗噪声：在车载噪声环境(风噪、胎噪、发动机噪声与空调噪声等)下，仍具有较高的唤醒率与识别率。

## 3.4 唤醒技术特点

我们的基于深度神经网络的语音唤醒技术，经过一系列的优化方法，使得我们的技术在业内达到领先水平。我们的技术有以下特点：

1. 唤醒性能优

唤醒性能优主要表现在较高的唤醒率与极低的误唤醒率。在相对安静的场景，我们的唤醒率接近100%，在车载噪声复杂场景，我们的唤醒率也能达到85%以上，综合唤醒率达到90%以上，处于业内领先水平。在较高唤醒率的同时，我们能保持极低的误唤醒率，每个唤醒词的误唤醒率小于48小时一次，远低于行业水平。

2）唤醒词任意设置

我们的方法支持在使用过程中任意的设置唤醒词，用户可以随意选择喜欢的唤醒词（唤醒词的不同，性能会有差异），为用户的使用提供了极大的便利。

3）功耗低，资源占用少

我们的方法计算量小，运算简单，资源占用少，功耗低。“零”延时语音唤醒要求快速响应，我们的方法做到延时在200ms以内，用户体验实现真正的“零”延时。

## 3.5 语音唤醒功能说明

### 3.5.1 主唤醒词

思必驰语音唤醒技术可通过在设备或软件中预置主唤醒词，并设定相对应操作，当用户发出该语音指令时，设备便从休眠状态中被“唤醒”，并作出指定响应，开启人机交互。

### 3.5.2 快捷唤醒词

思必驰语音唤醒技术可通过在设备或软件中预置快捷唤醒词，并设定相应的操作，当用户发出该语音指令时，设备直接快速执行相应的操作，而无需进行多余的语音交互，大大增加了人机交互的效率。

### 3.5.3 唤醒词选择

思必驰语音唤醒技术支持任意的中文汉字组合，但并不代表任意的汉字的组合都适合用来做唤醒词。不同的唤醒词选择，性能也是不同的。为了达到最佳的唤醒性能，我们对唤醒词的选择有如下建议：

1. 唤醒词3到6个汉字之间，4个字为最佳；
2. 唤醒词的音节尽量差异大，避免重复的发音；
3. 唤醒词的发音响亮，易开口；
4. 尽量避免生活化、口语化的词汇。

## 3.6 语音唤醒性能指标

语音唤醒系统评价标准通常有四个方面：唤醒率，误唤醒，响应时间，功耗。

1. 唤醒率

顾名思义，唤醒率就是指唤醒某个设备的概率。通用的术语也叫召回率(Recall)。通常测试的方式为，在某个场景，测试了M次，成功唤醒了N次，则唤醒率(Recall)=N/M\*100%。 注意：唤醒率是一种概率，按照大数定律，只有测试达到一定次数才有意义。举个极端的例子，测试一次，假如唤醒了，则唤醒率为100%，假如没有唤醒，则唤醒率为0。一般要求一定的测试人数，测试次数。

算力内存允许的情况下，可定制多个唤醒词，一般5个词以内，各个唤醒词性能基本无变化。再多唤醒词，尤其近似唤醒词，会存在互串的风险。且多个唤醒词，误唤醒的风险会线性增加。

1. 误唤醒

误唤醒是指没有人说唤醒词，但设备却反馈唤醒信息的概率。一般以多长时间误唤醒次数来表示，比如系统误唤醒24小时1次，48小时4次等。误唤醒与唤醒率一样是一种概率，需要一定的测试时间。一般至少24小时，正常48小时，更多72小时/96小时等等。

1. 响应时间

语音唤醒在用户说出唤醒词之后，设备需要在极短的时间内给出反馈。如果反馈延迟比较大（比如1s以上），用户会感到明显的卡顿，会影响用户体验。注：语音唤醒延迟是系统延迟的一部分，假如系统还有界面显示或语音合成播报，这些模块的延迟都会影响用户体验。

1. 功耗

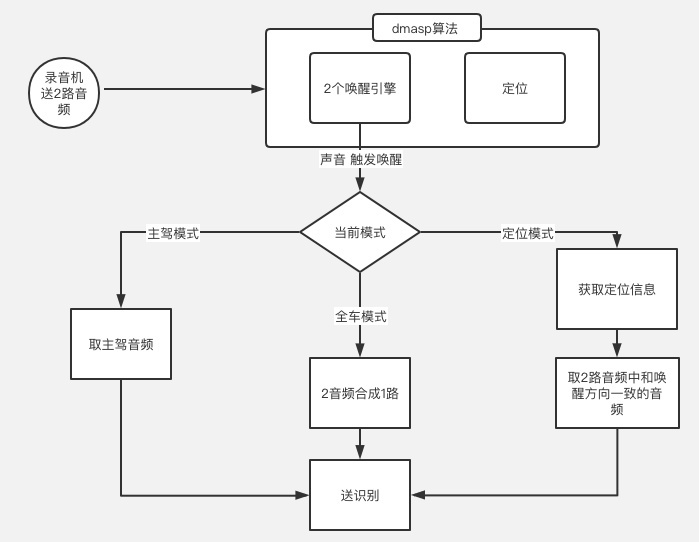
功耗是语音唤醒系统的一个重要指标和评价标准，尤其是对穿戴可移动设备。

一个语音唤醒系统，需要在以上评价标注中找到一种平衡。比如误唤醒和唤醒率是一对负相关的指标，误唤醒要求越高，唤醒率就越低，误唤醒要求越低，唤醒率会越高，要根据产品的应用场景结合用户体验，设置合理的阈值，使得误唤醒和唤醒率达到一种平衡。同样，相同算法或者模型的情况下，响应时间，功耗与唤醒性能（综合唤醒率和误唤醒）也是负相关，唤醒性能越高，响应时间和功耗会越大。语音唤醒系统，必然要在唤醒性能（唤醒率和误唤醒），响应时间，功耗间达到平衡。

我们一般是根据设备的算力来定制唤醒模型的大小，比如车机级别，唤醒模型大小1M左右；也有车载小系统，唤醒模型200K ; 也有低功耗芯片级，唤醒模型70k。

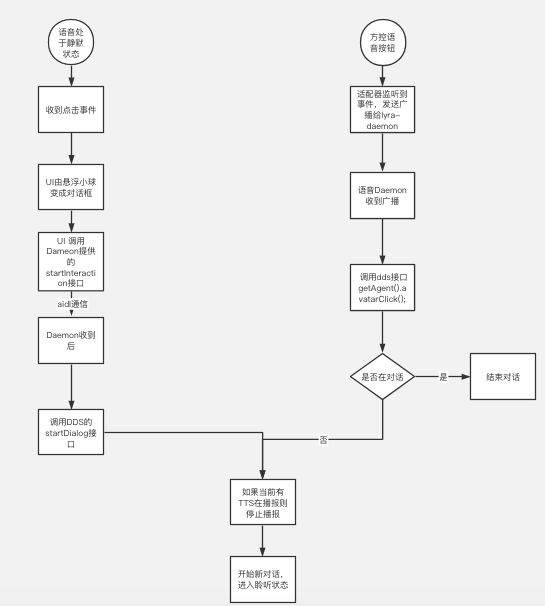
## 3.7 终端唤醒模式

终端有三种唤醒模式处理，如下图所示：

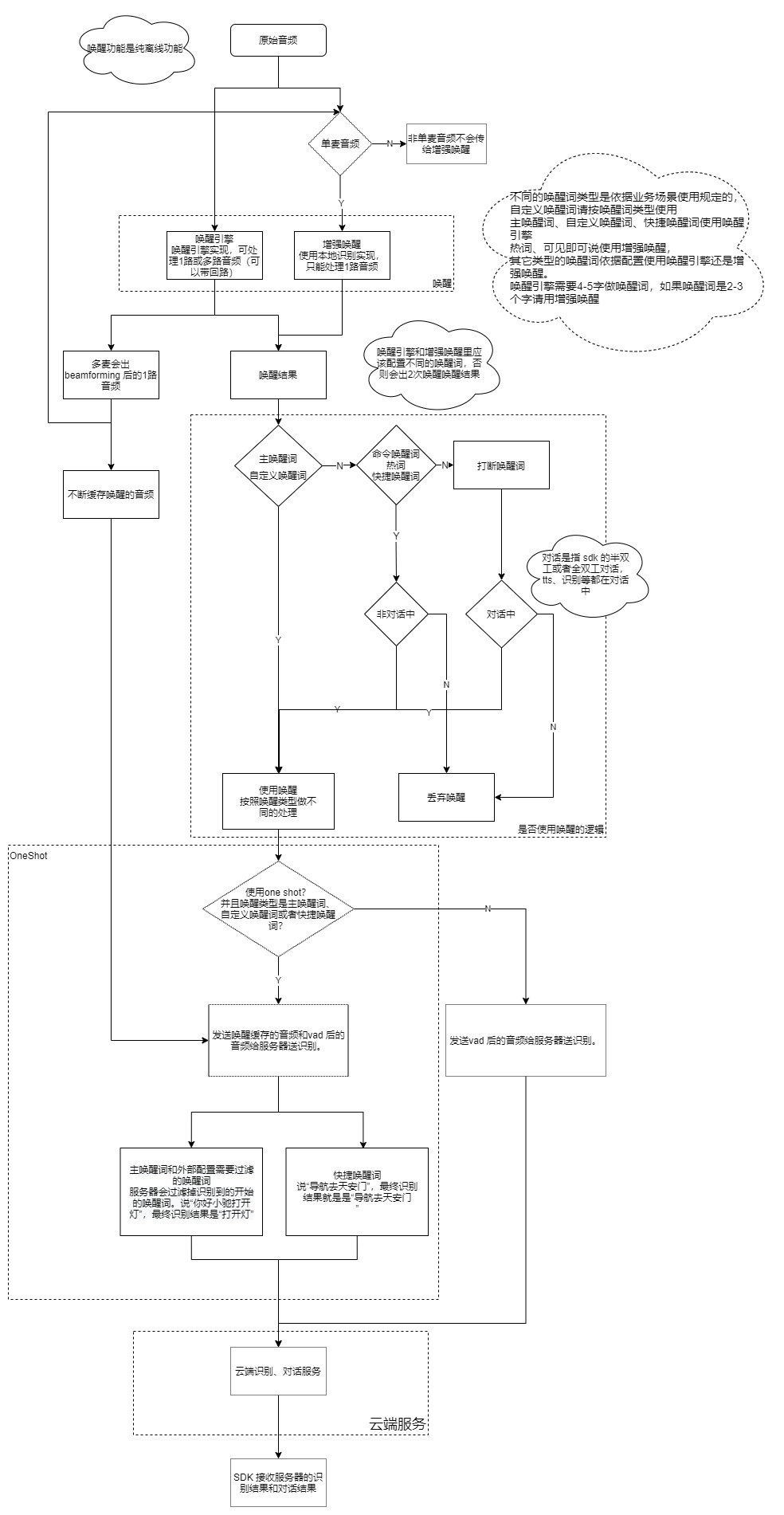
****

## 3.8 其他唤醒实现逻辑

方控和点击UI 触发对话，其逻辑实现如下流程图所示：

****

不同语音唤醒词的逻辑实现流程图，如下：

****