全双工PRD

**产品需求说明书**

|  |  |
| --- | --- |
| 所属项目 | 思必驰\_天琴 |
| 编辑 | 叶冬秀 |
| 审核 | 柯学滨 |
| 版本 | 1.0 |
| 日期 | 2022-05-10 |

# 历史版本

|  |  |  |  |
| --- | --- | --- | --- |
| **日期** | **版本** | **变更描述** | **作者** |
| 2022/05/10 | 1.0 | 初版 | 叶冬秀 |
| 2022/05/30 | 1.1 | 新增TTS异常情况 | 柯学滨 |

目录

[历史版本 2](#_Toc1925)

[一、产品概述 4](#_Toc4566)

[1.1 产品背景 5](#_Toc32104)

[1.2 适配项目 5](#_Toc4372)

[1.3 需求目的 5](#_Toc12569)

[二、功能描述 5](#_Toc25947)

[2.1 功能概述 5](#_Toc23380)

[三、 场景描述 5](#_Toc4636)

[3.1 全双工流程图 5](#_Toc28895)

[3.2 场景一：TTS播报时，打断播报 6](#_Toc9139)

[3.3 场景二：一次唤醒多次对话 6](#_Toc21646)

[3.4 拒识集 7](#_Toc489)

[3.5 异常情况 7](#_Toc9794)

[四、 测试指标 7](#_Toc12143)

**一、产品概述**

**1.1 产品背景**

语音助手的目标是智能+流畅，就像人与人面对面交流一样自然。但是目前市面上的语音助手，大多为半双工的语音交互，每次交互都需要叫唤醒词，而且用户无法随时打断对话。在实际车载场景中，很多用户反馈：语音唤醒比较麻烦、vad容易截断、tts播报造成干扰等问题。区别于半双工语音交互必须遵循一问一答的方式，全双工语音交互可以实现机器语音播报的同时，持续监听用户的语音输入并进行识别，用户可以随时打断或抢说，是一种更自然的、更接近人与人交流方式的语音交互方式。

同时，目前市面上的车机、车镜等产品，联网能力、性能配置逐渐升级，再加上5G网

络的即将到来，让全双工机制有了落地的可能。

在全双工延时倾听多轮对话的场景下，语音交互结束后，语音处于半退出状态，保持继续识别，除了闲聊、百科技能，其他技能都可以支持。DUI平台支持自定义延时聆听时长，推荐20秒。

**1.2 适配项目**

天琴所有项目

**1.3 需求目的**

语音助手达到更加智能、流畅的交互体验。

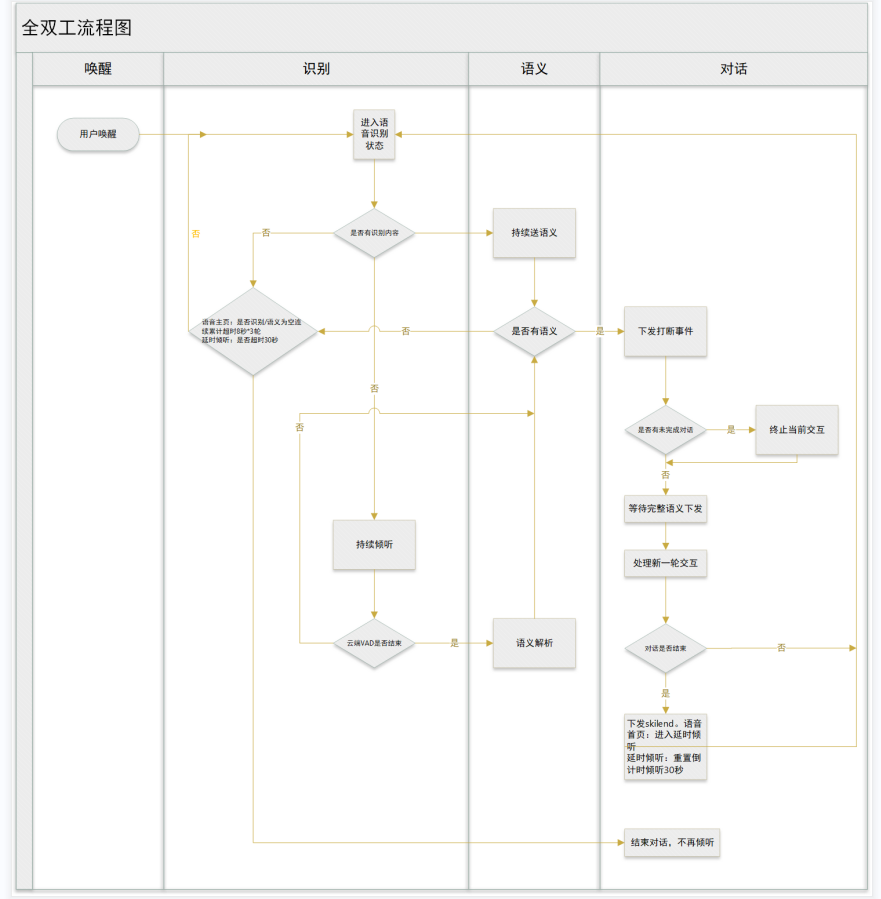
**二、功能描述**

**2.1 功能概述**

|  |  |  |
| --- | --- | --- |
| 序号 | 特性 | 描述 |
| 1 | 长交互 | 一次唤醒长交互，结束全双工对话之前，识别长开，不需要唤醒就可以识别输入内容（推荐20秒） |
| 2 | 主动打断 | 基于语义的主动打断，客户端在播放TTS的时候，识别是没有关闭的，用户可以继续说话，如果输入有效内容会打断当前正在播报的TTS（不包含闲聊和百科语料）同时支持唤醒打断 |
| 3 | 模式切换 | 全双工和半双工，交互模式可以切换（注：切换时，会强制停止当前轮对话，需要重新唤醒后，新模式才生效） |

1. **场景描述**

**3.1 全双工流程图**



**3.2 场景一：TTS播报时，打断播报**

A：客户端默认打开全双工模式（用户可以用语音指令，或者界面点击，关闭全双工模式）。

B：TTS播报时，语音图标保持识别动效，表示机器同时也在聆听。

C：播报时，输入非指定技能的说法，不响应。

D：播报时，输入指定技能的说法，显示识别文本，并执行指令。

A：若新指令执行结果有新的TTS播报，则打断原有TTS；若无，则不打断原有TTS。

**3.3场景二：一次唤醒多次对话**

1、结束对话20秒以内，输入任务型技能的说法，小窗仅显示识别文本，并执行指令。目前全双工技能范围（导航/音乐/车控/电话/电台/系统控制/多媒体控制/天气），非范围内技能（百科/闲聊）的说法不响应（拒识），可根据项目要求修改技能范围。全双工技能范围指的是延时聆听下的锁域范围，在首轮对话时仍是所有技能都有可能落域。

2、大于30秒时，退出继续倾听，语音形象聆听动效消失。

**3.4 拒识集**

1、用户输入无效语音，如：吧啦啦啦啦啦123456。

2、环境噪音/TTS播报语音。

3、语义识别后为非指定领域内的技能，如：闲聊和百科，也有可能是天气、股票等。

4、首轮单实体拒识。如：用户在唤醒语音助理，进入连续对话后的首轮输入说“世界之窗、稻香”等词语时拒识。

**3.5 异常情况**

1、无网络时不支持全双工功能，需要在连续对话开关处加上文字说明。

2、TTS设计上，当离线时，需从在线TTS切换为离线TTS；当使用复刻音色离线时，需要自动切换为离线音色使用。

3、在唤醒后的首轮对话中，对于不能进入任务型、通用型、流程型技能的说法，会绝大部分被知识型技能闲聊兜底，闲聊会根据实际输入给予回复，让语音系统不会显得过于呆板。当闲聊技能也无法兜底时，会反馈听不懂。在进入延时聆听后，说法不进入闲聊技能，只响应全双工下设置的锁域技能，保证拒识效果。

1. **测试指标**
2. 回声消除。全双工功能使用户能够打断TTS语音播报下发指令，这个时候语音形象在播放TTS时会把自己的声音也录音，所以需要消除这部分声音，更准确有效地识别到用户输入的内容。
3. 拒识。全双工开启期间，IP语音形象一直在聆听，在此期间会录取许多噪音，如非人声（背景噪音以及周围环境产生的声音）、不清晰的人声、无效人声（闲聊），拒识功能就是把这些无效的语音过滤掉。（实际指标以双方测试提供为准）