①

②

③

④

⑤

语音芯片

电控

Wifi

本地识别同时以音频包发送到云端处理，若本地识别小于阈值以云端处理为主

播放小苹果

云端

播放小苹果：

1. 根据相关协议打包上传音频包；
2. wifi端作相应的处理后将音频包发送到云端;
3. 云端做响应的识别、语音理解等动作
4. 云端下发语音播报的URL及歌曲URL
5. WIFI SDK接口分别处理语音播报的URL及歌曲URL，然后根据相关协议打包将音频包发送到本地端，然后本地端做相应的处理后播报语音及歌曲。

①

语音芯片

电控

Wifi

本地识别同时以音频包发送到云端处理，若本地识别大于阈值，则以本地为主

打开空调

云端

打开空调：

1. 本地做识别识别到相关的结果后大于阈值，就以本地识别为主，就根据本地相关协议传给电控，同时播放语音，忽略云端的识别（此场景同上）。

①

②

③

④

⑤

语音芯片

电控

Wifi

本地识别同时以音频包发送到云端处理，若本地识别小于阈值且云端又识别失败，云端下发语音播报

云端

若本地识别小于阈值，同时上传云端又未识别

1. 根据相关协议打包上传音频包；
2. wifi端作相应的处理后将音频包发送到云端;
3. 云端做响应的识别、语音理解等动作，但是未识别到
4. 云端下发未识别所对应的相关语音播报URL
5. WIFI SDK接口处理语音播报的URL，然后根据相关协议打包将音频包发送到本地端，然后本地端做相应的处理后播报语音

1、 你们模组是只能发送和接收PCM数据，不做编解码？

R:语音模组以特定UART协议压缩音频包上传PCM数据，播放只支持MP3解码。

2、空调示例部分：

你们的模组会做本地ASR识别，如果识别结果⼤于阈值，则直接本地转⽂本，匹配控制命令？

R:是的。如果本地识别结果大于阈值，就直接以本地识别的结果为主，云端过来的指令直接不响应。

WiFi模组也会做⽂本与指令的匹配，对应打包⾳频包？这个你们的意思是wifi模组做控制命令匹配？

R:我们模组以特定UART协议格式对数据进行发送接收的，模组接收的数据包必须是它能解析的特定协议格式。Wifi模组发送给我们模组的数据包必须是与我们模组解析的协议格式对应。

3、播放⼩苹果示例部分：

需要的AI能⼒是哪些？除了语⾳识别，语义理解，是还需要云端的⾳频内容点播、百科、问答等能⼒？

R：内容点播、百科、问答等能力暂时不需要。

4、⼯作流程：

1）家居控制命令流程：拾⾳--⾳频编码--⾳频上传--ASR--命令⽂本--控制命令匹配--指令下发--控制设备

2) AI交互流程（点播⾳频、百科等）：拾⾳--⾳频编码--⾳频上传--ASR+NLU--AI交互匹配--⾳频下⾏--⾳频解码-播放

R：判断机制：

如果本地识别到了就以本地识别为主，忽略云端；

如果本地未识别到，云端识别到了，则就以云端识别为主；

如果本地未识别到，云端也未识别到，则云端需要下发对应识别不到的语音播报；

上述三种机制都需要相对应的语音播报。

工作流程：

1)家居控制命令流程：

--uart协议打包上传到wifi（比如Dueros）--wifi端相应的解包通过相关算法压缩后经http协议上传云端--云端识别解析下发URL及CMD--WIFI端SDK解析URL返回MP3数据及cmd下发—本地端解包按照命令进行设备控制

拾⾳（模拟音频转换成PCM）--本地识别（若本地识别到了，以本地识别为主，不响应云端云端识别结果） --ASR--命令⽂本--控制命令匹配--控制设备

2）AI交互：拾音-- uart协议打包上传到wifi（比如Dueros）--wifi端相应的解包通过相关算法压缩后经http协议上传云端--云端识别解析下发URL—WIFI端SDK解析发挥MP3数据下发—本地解码mp3进行播放

5、拾⾳/播放：歌尔模组；⻨克⻛、喇叭是挂在歌尔模组上的吗？

R：是的。

1、 你们模组是只能发送和接收PCM数据，不做编解码？

R:语音模组以特定UART协议压缩音频包上传PCM数据，播放只支持MP3解码。

2、空调示例部分：

你们的模组会做本地ASR识别，如果识别结果⼤于阈值，则直接本地转⽂本，匹配控制命令？

R:是的。如果本地识别结果大于阈值，就直接以本地识别的结果为主，云端过来的指令直接不响应。

WiFi模组也会做⽂本与指令的匹配，对应打包⾳频包？这个你们的意思是wifi模组做控制命令匹配？

R:我们模组以特定UART协议格式对数据进行发送接收的，模组接收的数据包必须是它能解析的特定协议格式。Wifi模组发送给我们模组的数据包必须是与我们模组解析的协议格式对应。

3、播放⼩苹果示例部分：

需要的AI能⼒是哪些？除了语⾳识别，语义理解，是还需要云端的⾳频内容点播、百科、问答等能⼒？

R：内容点播、百科、问答等能力暂时不需要。

4、⼯作流程：

1）家居控制命令流程：拾⾳--⾳频编码--⾳频上传--ASR--命令⽂本--控制命令匹配--指令下发--控制设备

2) AI交互流程（点播⾳频、百科等）：拾⾳--⾳频编码--⾳频上传--ASR+NLU--AI交互匹配--⾳频下⾏--⾳频解码-播放

R：判断机制：

如果本地识别到了就以本地识别为主，忽略云端；

如果本地未识别到，云端识别到了，则就以云端识别为主；

如果本地未识别到，云端也未识别到，则云端需要下发对应识别不到的语音播报；

上述三种机制都需要相对应的语音播报。

工作流程：

1)家居控制命令流程：

--uart协议打包上传到wifi（比如Dueros）--wifi端相应的解包通过相关算法压缩后经http协议上传云端--云端识别解析下发URL及CMD--WIFI端SDK解析URL返回MP3数据及cmd下发—本地端解包按照命令进行设备控制

拾⾳（模拟音频转换成PCM）--本地识别（若本地识别到了，以本地识别为主，不响应云端云端识别结果） --ASR--命令⽂本--控制命令匹配--控制设备

2）AI交互：拾音-- uart协议打包上传到wifi（比如Dueros）--wifi端相应的解包通过相关算法压缩后经http协议上传云端--云端识别解析下发URL—WIFI端SDK解析发挥MP3数据下发—本地解码mp3进行播放

5、拾⾳/播放：歌尔模组；⻨克⻛、喇叭是挂在歌尔模组上的吗？

R：是的。

1、 你们模组是只能发送和接收PCM数据，不做编解码？

R:语音模组以特定UART协议压缩音频包上传PCM数据，播放只支持MP3解码。

2、空调示例部分：

你们的模组会做本地ASR识别，如果识别结果⼤于阈值，则直接本地转⽂本，匹配控制命令？

R:是的。如果本地识别结果大于阈值，就直接以本地识别的结果为主，云端过来的指令直接不响应。

WiFi模组也会做⽂本与指令的匹配，对应打包⾳频包？这个你们的意思是wifi模组做控制命令匹配？

R:我们模组以特定UART协议格式对数据进行发送接收的，模组接收的数据包必须是它能解析的特定协议格式。Wifi模组发送给我们模组的数据包必须是与我们模组解析的协议格式对应。

3、播放⼩苹果示例部分：

需要的AI能⼒是哪些？除了语⾳识别，语义理解，是还需要云端的⾳频内容点播、百科、问答等能⼒？

R：内容点播、百科、问答等能力暂时不需要。

4、⼯作流程：

1）家居控制命令流程：拾⾳--⾳频编码--⾳频上传--ASR--命令⽂本--控制命令匹配--指令下发--控制设备

2) AI交互流程（点播⾳频、百科等）：拾⾳--⾳频编码--⾳频上传--ASR+NLU--AI交互匹配--⾳频下⾏--⾳频解码-播放

R：判断机制：

如果本地识别到了就以本地识别为主，忽略云端；

如果本地未识别到，云端识别到了，则就以云端识别为主；

如果本地未识别到，云端也未识别到，则云端需要下发对应识别不到的语音播报；

上述三种机制都需要相对应的语音播报。

工作流程：

1)家居控制命令流程：

--uart协议打包上传到wifi（比如Dueros）--wifi端相应的解包通过相关算法压缩后经http协议上传云端--云端识别解析下发URL及CMD--WIFI端SDK解析URL返回MP3数据及cmd下发—本地端解包按照命令进行设备控制

拾⾳（模拟音频转换成PCM）--本地识别（若本地识别到了，以本地识别为主，不响应云端云端识别结果） --ASR--命令⽂本--控制命令匹配--控制设备

2）AI交互：拾音-- uart协议打包上传到wifi（比如Dueros）--wifi端相应的解包通过相关算法压缩后经http协议上传云端--云端识别解析下发URL—WIFI端SDK解析发挥MP3数据下发—本地解码mp3进行播放

5、拾⾳/播放：歌尔模组；⻨克⻛、喇叭是挂在歌尔模组上的吗？

R：是的。

1、 你们模组是只能发送和接收PCM数据，不做编解码？

R:语音模组以特定UART协议压缩音频包上传PCM数据，播放只支持MP3解码。

2、空调示例部分：

你们的模组会做本地ASR识别，如果识别结果⼤于阈值，则直接本地转⽂本，匹配控制命令？

R:是的。如果本地识别结果大于阈值，就直接以本地识别的结果为主，云端过来的指令直接不响应。

WiFi模组也会做⽂本与指令的匹配，对应打包⾳频包？这个你们的意思是wifi模组做控制命令匹配？

R:我们模组以特定UART协议格式对数据进行发送接收的，模组接收的数据包必须是它能解析的特定协议格式。Wifi模组发送给我们模组的数据包必须是与我们模组解析的协议格式对应。

3、播放⼩苹果示例部分：

需要的AI能⼒是哪些？除了语⾳识别，语义理解，是还需要云端的⾳频内容点播、百科、问答等能⼒？

R：内容点播、百科、问答等能力暂时不需要。

4、⼯作流程：

1）家居控制命令流程：拾⾳--⾳频编码--⾳频上传--ASR--命令⽂本--控制命令匹配--指令下发--控制设备

2) AI交互流程（点播⾳频、百科等）：拾⾳--⾳频编码--⾳频上传--ASR+NLU--AI交互匹配--⾳频下⾏--⾳频解码-播放

R：判断机制：

如果本地识别到了就以本地识别为主，忽略云端；

如果本地未识别到，云端识别到了，则就以云端识别为主；

如果本地未识别到，云端也未识别到，则云端需要下发对应识别不到的语音播报；

上述三种机制都需要相对应的语音播报。

工作流程：

1)家居控制命令流程：

--uart协议打包上传到wifi（比如Dueros）--wifi端相应的解包通过相关算法压缩后经http协议上传云端--云端识别解析下发URL及CMD--WIFI端SDK解析URL返回MP3数据及cmd下发—本地端解包按照命令进行设备控制

拾⾳（模拟音频转换成PCM）--本地识别（若本地识别到了，以本地识别为主，不响应云端云端识别结果） --ASR--命令⽂本--控制命令匹配--控制设备

2）AI交互：拾音-- uart协议打包上传到wifi（比如Dueros）--wifi端相应的解包通过相关算法压缩后经http协议上传云端--云端识别解析下发URL—WIFI端SDK解析发挥MP3数据下发—本地解码mp3进行播放

5、拾⾳/播放：歌尔模组；⻨克⻛、喇叭是挂在歌尔模组上的吗？

R：是的。

控制设备：歌尔模组

R:语音模组将转换成的pcm音频数据按照uart协议打包发送到wifi,wifi端解包并通过相关算法（比如百度dueros语音算法）压缩通过http协议上传云端；wifi端sdk(比如百度dueros)解析下发的url返回MP3数据，并将uart协议打包发送到本地，本地语音模组解包并解码播放MP3。

⾳频编解码：庆科套⽚

⾳频上传、下⾏：庆科套⽚

ASR：歌尔模组+AI云

控制命令匹配：歌尔模组识别则歌尔匹配；AI识别服务端匹配

AI交互匹配：AI云------这块能⼒我们套⽚是整合了的

这是根据你们⽂档理的流程，你们看有⽆误解；

如果没有，以下是两种控制命令的匹配⽅式（匹配不放庆科套⽚做，通⽤），你们看下倾向于

哪种

1）歌尔可以做控制命令匹配，那么AI识别也建议歌尔做控制命令匹配；AI云做ASR、我们套

⽚直接传控制命令⽂本（若要加提示⾳可以本地加T 卡存）

2）歌尔不做控制命令匹配，命令匹配放到服务端处理，庆科套⽚定好和服务端的解析规则;匹

配都放到服务端处理，服务端可以做全所有家电控制命令（提示⾳可以AI返回）；服务端协议

和歌尔确定好，模组做透传

R:此控制命令匹配是指在线的吗？