

蜂鸟 M 芯片方案 常见语音性能问题优化指导手册

V1-0

制订:	项目组成员	日期:	2021-01-18
审核:		日期:	
批准:		日期:	

*The information contained in this message is confidential and should not be disclosed to any third party whether or not you are the intended addressee indicated in the message.

*本文件所含内容为保密信息，未经授权请勿随意复制、编改和泄露给任何第三方。

重要声明

版权声明

版权所有 © 2020，云知声智能科技股份有限公司，保留所有权利。

商标声明

云知声智能科技股份有限公司的产品是云知声智能科技股份有限公司专有。在提及其他公司及其产品时将使用各自公司所拥有的商标，这种使用的目的仅限于引用。本文档可能涉及云知声智能科技股份有限公司的专利（或正在申请的专利）、商标、版权或其他知识产权，除非得到云知声智能科技股份有限公司的明确书面许可协议，本文档不授予使用这些专利（或正在申请的专利）、商标、版权或其他知识产权的任何许可协议。

不作保证声明

云知声智能科技股份有限公司不在此文档中的任何内容作任何明示或暗示的陈述或保证，而且不对特定目的的适销性及适用性或者任何间接、特殊或连带的损失承担任何责任。本手册内容若有变动，恕不另行通知。本手册例子中所用的公司、人名和数据若非特别声明，均属虚构。未得到云知声智能科技股份有限公司明确的书面许可，不得为任何目的、以任何形式或手段（电子的或机械的）复制或传播手册的任何部分。

保密声明

本文档（包括任何附件）包含的信息是保密信息。接收人了解其获得的本文档是保密的，除用于规定的目的外不得用于任何目的，也不得将本文档泄露给任何第三方。

本软件产品受最终用户许可协议（EULA）中所述条款和条件的约束，该协议位于产品文档和/或软件产品的联机文档中，使用本产品，表明您已阅读并接受了 EULA 的条款。

概述

本文档主要介绍了基于蜂鸟 M 芯片方案常见语音性能问题优化措施。

读者对象

本文档主要适用于以下工程师：

- 技术支持工程师
- 软件工程师

Unisound

修订记录

修订记录累积了每次文档更新的说明。

最新版本的文档包含以前所有文档版本的更新内容。

修订日期	版本	修订说明
2021-01-18	V1-0	初始版本

Unisound

目录

概述	III
读者对象	III
修订记录	IV
1. 离线语音识别基础知识科普	1
1.1. 基本概念	1
1.2. 语音误识别说明	1
2. 常见离线语音识别问题及优化思路	3
3. 蜂鸟 M 详细优化操作步骤	4
3.1. 阈值调整详解	4
3.2. 竞争词加入详解	9
3.3. 稳态噪声品类选用	12
3.4. 定制产品品类选用	13
3.5. 模型训练平台	14
附录：词条选择建议	15

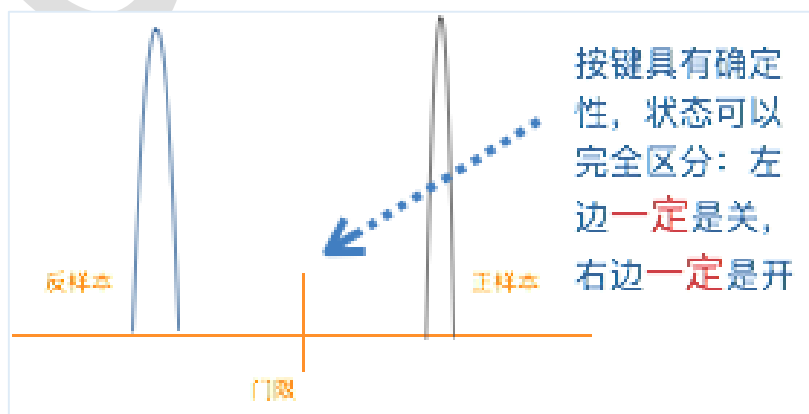
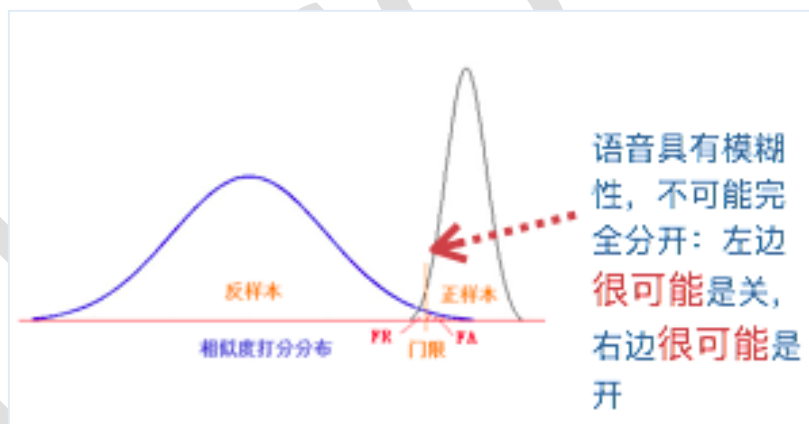
1. 离线语音识别基础知识科普

1.1. 基本概念

- (1) 离线语音识别的定义：从连续的录音中把用户说的目标关键词检测出来，并使程序作出反应
- (2) 对离线语音识别率的要求：唤醒率尽可能高，而误唤醒率尽可能低
- (3) 指标：唤醒率(100%-FR)：连续语流中唤醒词被检测出的百分比；误报率(FA)：每小时连续语流中，非唤醒词语音触发唤醒动作的次数；

1.2. 语音误识别说明

- (1) 为什么语音唤醒不能像电源开关那样准确？
 - 由于语音的模糊性（并不是非黑即白），正样本（真正的唤醒词）和反样本（杂音）得到的分数分布上如图所示是有交叠的
 - 好的算法和技术是最大限度地扩大两条曲线的距离，但两条曲线永远不可能完全分开而没有交叠

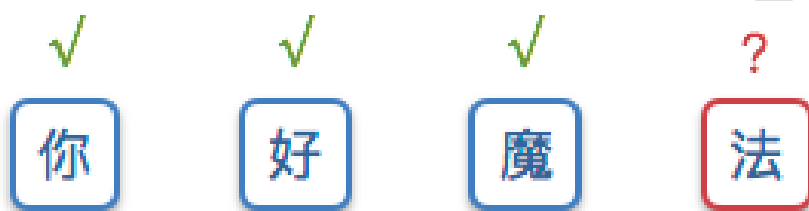


(2) 为什么误识别会永远存在？

- 识别率不会到 100% 的原因：识别词相似度分数小于门限；噪声、口音、混响等干扰导致失配
- 噪音或其他人声导致误识别的原因：非识别词相似度分数大于门限；发音相近词、频谱类似背景语音或噪声
- 人类也无法完全区分所谓『识别词』和『非识别词』，就好像人们永远无法确定所有案件的犯罪嫌疑人一样，所以才有了『无罪推定』和『有罪推定』一说。

(3) 为什么『挑战式地』尝试误识别是不科学的？

- 『挑战式地』尝试误唤醒的例子：唤醒词是：『你好魔方』，测试人员尝试『你好魔法』



相似度>75%，再加上语音本身的一些模糊性，很容易造成误唤醒
解决办法：通过设计不常出现的唤醒词来规避，但仍然无法解决用户故意挑战的情况。

(4) 关于误识别的结论：

- 对于语音识别而言，识别率永远不可能为 100%，而误识别率永远不可能为 0%
- 挑战式地尝试误识别是不科学的，对业界所有技术提供方这都是一个尚未解决的挑战

2. 常见离线语音识别问题及优化思路

常见问题	影响因素判断		优化思路	蜂鸟 M 详细操作步奏
识别率低	环境噪声	是否稳态：如电机、风噪等	选用稳态噪声品类	详见 3.3
		结构设计不合理：如阻塞、震动、干扰等	优化声学结构	可参照云知声《产品声学结构设计 checklist》
		超出单 mic 可以收敛的边界	选用多 mic 阵列方案	可选用蜂鸟等多 mic 方案
	灵敏度低	硬件设计不合理、mic 选型参数不合适等	优化硬件设计，选用合适灵敏度的 mic	可参照《蜂鸟 M 参考原理图》、《参考驻极体麦克风规格书》
		首次唤醒困难（长时间放置后）	调整深度睡眠唤醒阈值或延长进入深度睡眠时间	详见 3.1.1.2
		普遍识别率低或部分词条识别较差	调整阈值，提高灵敏度	详见 3.1
		调整阈值无法满足要求，确认是否云知声已有定制产品品类	选用线上定制产品品类	详见 3.4
		调整阈值无法满足要求，且无定制产品品类	可采用云知声模型训练平台进行采集录音，声学模型定制	详见 3.5
误识别高	误唤醒高	确认是否特定词条（非挑战测试且是高频词汇）造成误唤醒	加入竞争词	详见 3.2
		连续语流中出现误唤醒频次高	调整唤醒阈值或缩短进入深度睡眠时间	详见 3.1
	集外命令词误识别高	确认是否特定词条（非挑战测试且是高频词汇）造成误识别	加入竞争词	详见 3.2
		连续语流中出现命令词普遍误识别高或部分词条误识别高	调整阈值，降低灵敏度	详见 3.1
	集内命令词之间互相串扰	是否词条之间相似度太高	建议优化词条，增加区分度，可参照本文附录	详见 附录
		如非相似度高或词条固化	可烧录 debug 版本，根据实际测出的误识别阈值，针对性单独词条调整优化	详见 3.1.2.2

3. 蜂鸟 M 详细优化操作步骤

3.1. 阈值调整详解

云知声蜂鸟 M 芯片方案在生成时已经以通用场景为参照对阈值进行过优化，UniOS 开放平台也支持三档灵敏度配置，若需进行针对性优化，可在二次开发中修改配置文件实现。配置文件所在路径为 tools/scripts/aik_debug.json 和 tools/scripts/aik_release.json。

3.1.1. 唤醒

唤醒所涉及到的配置皆在 kws_lp.wakeup 对象中。

3.1.1.1. 全局调整

```

"wakeup": {
  "fa": "0.2",
  "fa_second": "0.4",
  "fa_sleep": "0.04",
  "time_to_sleep": 300,
  "time_second_wakeup": 5,
  "wuws": [
    {
      "wuw": "你好魔方",
      "fa_table": [
        2.89,
        2.08,
        1.14,
        0.94,
        0.25,
        -0.43,
        -0.89,
        -1.17,
        -1.4,
        -1.95
      ]
    }
  ]
}

```

默认阈值配置

阈值表

如上图所示，默认阈值配置“fa”的范围为{0.01, 0.02, 0.04, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.8}（后文中其他阈值配置的范围与此相同，不再重复列出），通过调整“fa”的值进行唤醒词的灵敏度配置（“fa”数值越高越容易唤醒，相对应地，误唤醒概率也会变大）。阈值表默认已经适用于通用场景，一般不建议进行修改。

3.1.1.2. 深度睡眠

```
"wakeup": {
  "fa": "0.2",
  "fa_second": "0.4",
  "fa_sleep": "0.04",
  "time_to_sleep": 300,
  "time_second_wakeup": 5,
  "wuws": [
    {
      "wuw": "你好魔方",
      "fa_table": [
        2.89,
        2.08,
        1.14,
        0.94,
        0.25,
        -0.43,
        -0.89,
        -1.17,
        -1.4,
        -1.95
      ]
    }
  ]
}
```

深度睡眠阈值配置
深度睡眠进入时间配置, 单位: 秒

本方案中深度睡眠是指设备一段时间“time_to_sleep”没有进行过唤醒或识别的操作后主动降低唤醒的灵敏度，以期获得更优的误唤醒指标。深度睡眠阈值配置“fa_sleep”的数值越小，误唤醒的概率越低。相对应地，从深度睡眠状态唤醒的难度也越高。

3.1.1.3. 二次唤醒

```
"wakeup": {
  "fa": "0.2",
  "fa_second": "0.4",
  "fa_sleep": "0.04",
  "time_to_sleep": 300,
  "time_second_wakeup": 5,
  "wuws": [
    {
      "wuw": "你好魔方",
      "fa_table": [
        2.89,
        2.08,
        1.14,
        0.94,
        0.25,
        -0.43,
        -0.89,
        -1.17,
        -1.4,
        -1.95
      ]
    }
  ]
}
```

二次唤醒的阈值配置

二次唤醒的间隔时间, 单位: 秒

本方案中二次唤醒是指设备在唤醒尝试因打分不够判定失败后，一段时间“time_second_wakeup”内再次收到唤醒尝试时主动提高唤醒的灵敏度，以期提高此次唤醒的成功率。二次唤醒阈值配置“fa_second”的数值越小，第二次唤醒的成功率越高。相对应地，误唤醒的概率也越高。

与全局阈值的配置相比，二次唤醒的阈值配置对误唤醒的影响较小。

3.1.2. 识别

识别所涉及到的配置皆在 `kws_lp.command` 对象中。

3.1.2.1. 全局调整

```

"command": {
  "command_tag": "ivm",
  "fa": "0.4",
  "fa_table": [
    7.98,
    6.72,
    5.44,
    5.04,
    3.76,
    2.4,
    1.62,
    1.03,
    0.59,
    -0.33
  ]
}
    
```

默认阈值配置

阈值表

如上图所示，通过调整识别默认阈值配置“fa”的值进行命令词的灵敏度配置（“fa”数值越高越容易识别，相对应地，误识别概率也会变大）。阈值表默认已经适用于通用场景，一般不建议进行修改。

3.1.2.2. 特定命令词调整

```
"command": {
  "command_tag": "ivm",
  "fa": "0.4",
  "command_custom": [
    {
      "word": "开机",
      "fa": "0.2"
    },
    {
      "word": "亮一点",
      "fa": "0.04"
    }
  ]
},
"fa_table": [
  7.98,
  6.72,
  5.44,
  5.04,
  3.76,
  2.4,
  1.62,
  1.03,
  0.59,
  -0.33
]
```

特定命令词阈值配置

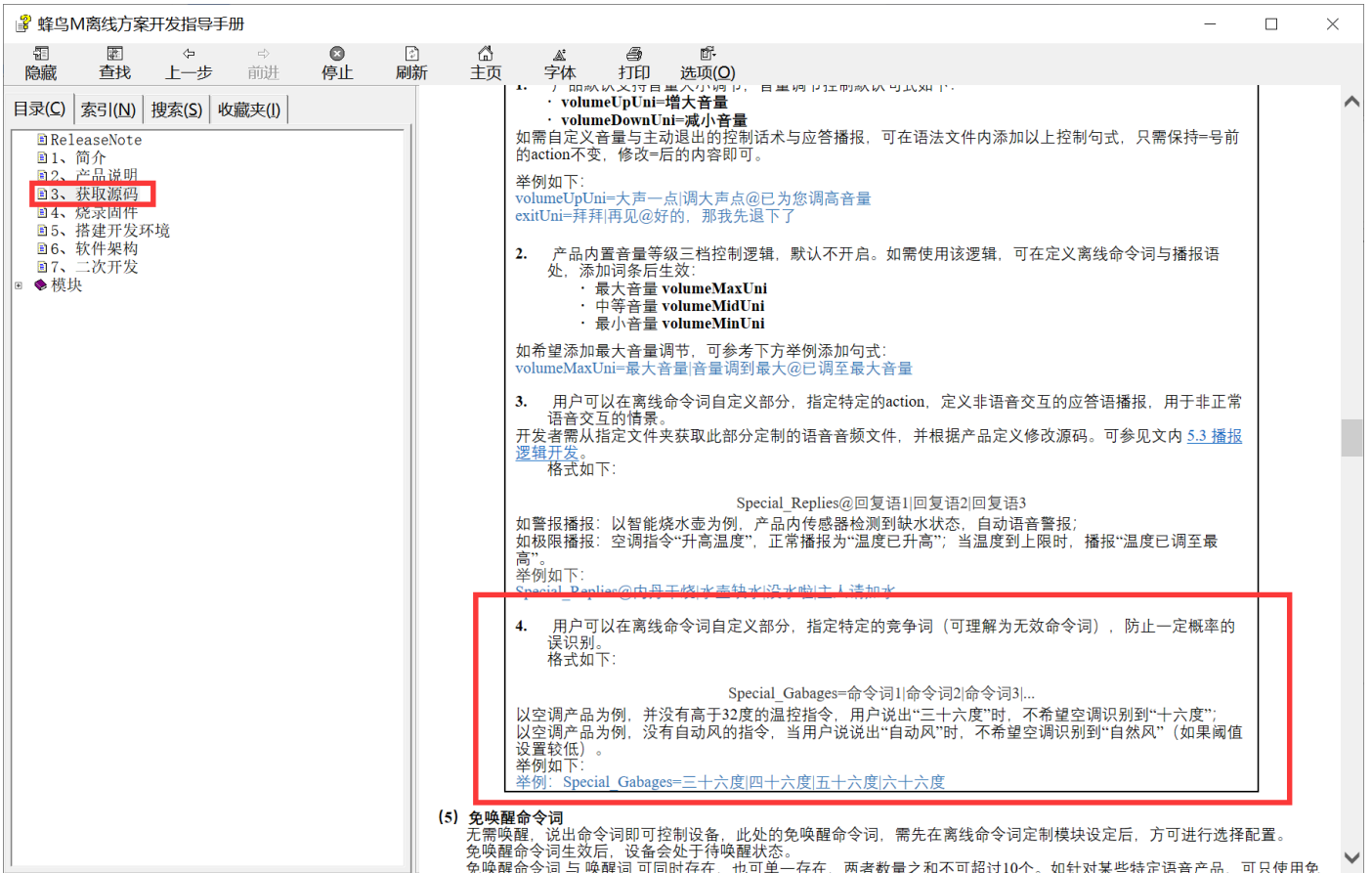
可配置多条

如上图所示，在"command"对象下新增"command_custom"对象，并填入需要特殊处理的命令词和对应阈值配置即可针对单个或多个命令词进行定制。

3.2. 竞争词加入详解

3.2.1. 平台添加

详细的操作步骤参照《蜂鸟 M 离线方案开发指导手册》里 3 章节『获取源码』中的第 2 段，如下图：



注：此方案会造成超时时间重新计算

3.2.2.手动修改

可在 include/uni_nlu_content.h 中注释 g_nlu_content_mapping 表单中的特定命令词以达到和竞争词配置相似的目的（即识别到此命令词后不做任何处理），此方法同时对唤醒和识别生效。

```
const uni_nlu_content_mapping_t g_nlu_content_mapping[] = {
    {2947460911U/*你好魔方*/, eCMD_wakeup_uni, NULL},
    {2497873774U/*退下*/, eCMD_exitUni, NULL},
    {2389288886U/*再见*/, eCMD_exitUni, NULL},
    {2438769531U/*开机 */, eCMD_ac_power_on, NULL},
    {3534892713U/* 打开风扇 */, eCMD_ac_power_on, NULL},
    {2593819740U/*关闭风扇 */, eCMD_ac_power_off, NULL},
    {2389495217U/* 关机 */, eCMD_ac_power_off, NULL},
    {3146628120U/*一档风 */, eCMD_ac_speed_1, NULL},
    {1429128034U/* 风速调到一档 */, eCMD_ac_speed_1, NULL},
    //{1460527232U/* 风速调到最小*/, eCMD_ac_speed_1, NULL}, 将“风速调到最小”作为竞争词
    {1024237780U/*一档风 */, eCMD_ac_speed_2, NULL},
    {1431332568U/* 风速调到二档 */, eCMD_ac_speed_2, NULL},
    {2544226657U/*三档风 */, eCMD_ac_speed_3, NULL},
    {1429396153U/* 风速调到三档 */, eCMD_ac_speed_3, NULL},
    {1460526884U/* 风速调到最大 */, eCMD_ac_speed_3, NULL},
    {2826737195U/*正常风 */, eCMD_ac_mode_nor, NULL},
    {2535485938U/* 打开正常风 */, eCMD_ac_mode_nor, NULL},
    {3631246374U/*自然风 */, eCMD_ac_mode_nat, NULL},
    {3339995117U/* 打开自然风 */, eCMD_ac_mode_nat, NULL},
    {1454765812U/*睡眠风 */, eCMD_ac_mode_sleep, NULL},
    {1163514555U/* 打开睡眠风 */, eCMD_ac_mode_sleep, NULL},
    {2655813610U/*调大风量 */, eCMD_ac_speed_inc, NULL},
    {773838435U/* 风太小了 */, eCMD_ac_speed_inc, NULL},
    {4269386346U/* 风大点 */, eCMD_ac_speed_inc, NULL},
    {2269449286U/*调小风量 */, eCMD_ac_speed_dec, NULL},
    {763471167U/* 风太大了 */, eCMD_ac_speed_dec, NULL},
    {4279753614U/* 风小点 */, eCMD_ac_speed_dec, NULL},
};
```

注：此方案会造成超时时间重新计算

3.2.3.接口修改

通过云端构建竞争词后，在初始化的时候，调用 `user_asr_word_disable (const char *word)` 接口，禁用某个竞争命令词或唤醒词，注意参数 `word` 对应的中文编码需要 UTF-8。这种方式不会导致超时时间重新计算。

The screenshot shows a web browser displaying the 'Beebird M Offline Solution Development Guide Manual'. The left sidebar contains a table of contents with '语音控制接口' (Voice Control Interface) selected. The main content area is titled '语音控制接口' and lists several functions. The function `int user_asr_word_disable (const char *word)` is highlighted with a red box. Below the function name, it says '禁用识别词 更多...'.

3.3. 稳态噪声品类选用

登陆 UniOS 开放平台——>创建产品——>选择智能家居中文/英文版（纯离线）——>选择蜂鸟 M 芯片稳态降噪

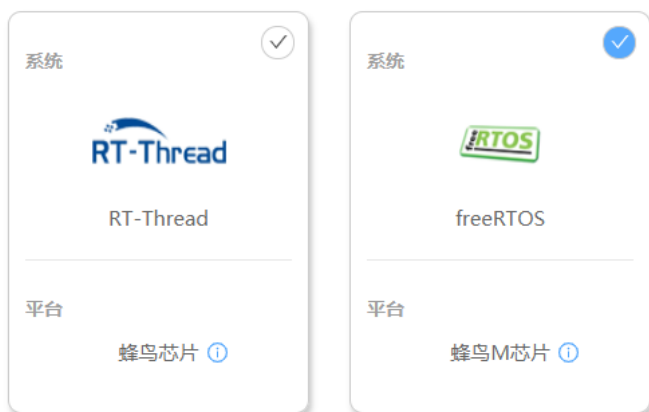
如下图：



3.4. 定制产品品类选用

登陆 UniOS 开放平台——>创建产品——>选择智能家居中文/英文版（纯离线）——>选择蜂鸟 M 芯片定制产品品类

如下图：



选择设备类型



3.5. 模型训练平台

详细的操作步骤参照《蜂鸟 M 离线方案开发指导手册》里 3 章节『获取源码』中的第 3 段，如下图：

目录(C) 索引(S) 搜索(O) 收藏夹(O)

键入关键字进行查找(O):

3、获取源码

1、简介
2、产品说明
3、获取源码
4、烧录固件
5、搭建开发环境
6、软件架构
7、二次开发

audio_end
audio_play
AUDIO_PLAY_MUSIC
AUDIO_PLAY_REPLY
AUDIO_PLAY_SKILL
audio_相关宏定义
checksum
cmd
event_custom_setting_t
event_goto_awakend_t
event_goto_sleeping_t
cmd_code
custom_setting
Demo相关宏定义
eTIMER_IDX
event_audio_end_t
is_auto
event_audio_play_t
file_name
event_custom_setting_t
cmd
reply_files
event_goto_awakend_t
cmd
reply_files
trigger
event_goto_sleeping_t
cmd
reply_files
trigger
EVENT_TRIGGER_ASR
EVENT_TRIGGER_AUTO
event_trigger_t

提供选择任一历史版本进行迭代开发的选项，默认为基于最新修改的版本，也可根据需求进行其它版本的选择。

操作系统: RT-Thread

AppKey: gwhxco5mppcx633kgczd2n3xyi4thilua2ua

AppSecret:

版本号: V 1.1.1

基于版本(O): V1.1.0
V1.1.0
版本描述: V1.0.1
V1.0.0

下一步

3. 模型训练平台介绍

蜂鸟芯片方案支持通过云知声模型训练平台对声学模型进行定制。

当系统默认的声学模型性能无法满足用户需求时（注：此处用户需求，指经过语音性能量化测试的明确指标），可以使用模型训练平台，采集定制，可有效提升语音识别准确率。

云知声 Unisound UniOS 开放平台 产品输入 技能开发 技能商城 文档中心 开放平台首页

模型训练平台

用零延迟的流式文件，快速定制声学模型，有效提升语音识别准确率。

平台使用旧版本教程

附录：词条选择建议

中文

1. 音节覆盖尽量多，差异大，长度建议 4-6 字，不在范围内尽可能少，但不建议 2 字以内的唤醒词及命令词。
 - a) 针对汉语以音节为发音单位，至少需要发音不重复的汉字；
 - b) 相邻音节要规避，例如：语音（yǔ yīn）。
2. 尽量选取不常用的词语、低频词；尽量规避高频词，比如你好 xx。
3. 尽量选取开口音，不要选取闭口音（所谓闭口音，就是发音的时候，嘴唇是闭合的）。
 - a) 开口音发音清晰，响度大，容易辨识，例如电饭煲是开口音；
 - b) 闭口音导致气流受阻，容易和噪声混淆，不容易辨识，比如 物语（wù yǔ）。
4. 不要选取迭词，比如空调空调（名为四个音节，有效音节只有两个，大大增加识别难度）。
5. 唤醒词发音组合，不能超过 6 种。

英文

1. 词条应该是英文单词或短语。说该词尽量不超过两秒钟。
2. 音节覆盖尽量多，差异大，一个词条元音在 3 个以上，4-7 个音节的单词效果最好。例如，“Hey, Computer”是一个很好的唤醒词，而只是“Hey”则是一个糟糕的唤醒词。
3. 词条应遵循常见的英语发音规则。
4. 遵循常见英语发音规则的独特或甚至虚构的单词可以减少误报。例如，“computerama”可能是一个很好的唤醒词。
5. 不要选择常用词。例如，“eat”和“go”是人们在日常对话中经常说的词。它们可能是设备的错误触发器。
6. 避免使用可能有其他发音的词条。用户必须知道“正确”的发音才能使他们的设备做出响应。例如，“509”可以发音为“five zero nine”、“five oh nine”或“five hundred and nine”。“R.E.I.”可以发音为“R E I”或“Ray”。“Live”可以发音为 [lɪv] 或 [li:v]。
7. 不要使用特殊字符、符号或数字。例如，“Go#”和“20 + cats”不会是好的词条。但是，“go sharp”或“twenty plus cats”可行。你仍然可以在品牌中使用这些符号，并通过营销和文档来强化正确的发音。