



LD3320

声控智能产品语音界面设计指南

ICRoute 用声音去沟通
VUI (Voice User Interface)

Web : www.icroute.com

Tel : 021-68546025

Mail: info@icroute.com

目录

一. 简介	3
二. 语音识别技术 ASR 的技术特点说明	3
1. 语音识别 ASR 技术的主要特征:	3
2. 语音识别技术是基于特征匹配的技术	3
3. 说话内容不在列表中所引起的误识别	4
4. 语音识别应用的适用原则	5
三. 语音识别产品设计的思路	5
用垃圾词语吸收错误, 防止其他话语误触发	5
用发音相近词语来增加识别的容错性。	5
需要精确识别的场合, 尽量使用触发识别模式, 或者口令触发模式 .	6
设计合理的确认流程	6
四. 语音识别产品界面设计案例	6
1. 声控电视遥控器—按键触发模式	6
2. 声控咖啡机——“按键”触发模式: 结合产品特性来合理触发	7
3. 声控智能家居控制——口令触发模式: 设置合理确认流程	10
1) 设置前置口令, 第二级说“操作+设备”	10
2) 把操作(打开/关闭)作为口令, 第二级说“设备”	12
3) 第一级循环识别“操作+设备”, 第二级说“确定”	13
4. 声控智能固定电话——按键触发模式	14
5. 声控智能医疗护理产品——口令触发模式与按键触发模式相结合 ..	15

一. 简介

随着开发者对于 ICRoute 公司提供的语音识别芯片的开发深入，更多的讨论集中在如何在产品中应用好语音识别芯片。

尤其在目前的声控智能家居系统中，如何使用 LD3320 语音识别芯片完成实用的语音用户界面（VUI， Voice User Interface），成为技术开发者讨论的中心问题。

本文从语音识别技术的概念出发，结合目前在实际产品中的具体设计理念，给出了几种在智能家居系统中实用的声控界面设计。（有关语音识别芯片 LD3320 的各种技术资料请至 www.icroute.com 阅读和下载）

二. 语音识别技术 ASR 的技术特点说明

语音识别技术 ASR（Auto Speech Recognition）是基于关键词语列表的识别技术。（http://www.icroute.com/web_cn/Tech ASR.html）

1. 语音识别 ASR 技术的主要特征:

- ASR 只能识别设定的关键词语列表中的词语
- ASR 不是听写系统
- ASR 不能把人的声音转换为确定的拼音串
- ASR 不能任意地识别人的说话内容
- ASR 不能识别关键词语列表中的某个关键词语的一部分内容，除非这一部分内容自身也是一个关键词语
- ASR 不可以识别与关键词语列表中列出的关键词语不相符的情况
 - 比如，前后加了“嗯”，“阿”之类的语气词
 - 比如，只说出了关键词语中的一部分而不是整个关键词语
 - 比如，没有列在关键词语列表中的词语

2. 语音识别技术是基于特征匹配的技术

LD3320 作为非特定人语音识别技术的芯片，也是遵循了语音识别 ASR 技术的原理：（http://www.icroute.com/web_cn/LD332X_principle.html）：

把通过 MIC 输入的声音进行频谱分析->提取语音特征->和关键词语列表中的关键词语进行对比匹配->找出得分最高的关键词语作为识别结果输出。

这个也是目前世界上所有作 ASR 技术所遵循的原理。

从这里可以看到，语音识别并不是一个精确的识别，不能象在文档中搜索字符串一样作到精确匹配。

因为语音识别是在识别列表的各个关键词语之间取“得分最高（也就是最有可能最相似的意思）”的关键词语，作为识别结果（或者是识别结果的最佳候选之一）。

因此，学术界定义“识别率”时，是描述的当用户说的内容是在识别列表规定的范围内，可以达到识别率。（说明识别率是 90% 或者以上，都是按此定义）

语音识别并不存在一个把声音识别成拼音串的过程，而是直接用特征和特征进行匹配，打分，和得到结果。

因此，语音识别的识别体验，除了和外在因素，比如说话人的说话清晰程度，麦克风录音质量，周围环境干扰等因素相关。

也和内在因素，也即语音用户界面 VUI 的设计相关，具体说就是识别列表自身的内容和逻辑的设计相关。

本文主要探讨一些语音用户界面设计的思路和方法，以及案例介绍。

3. 说话内容不在列表中所引起的误识别

当用户说的内容是不在识别列表内的时候，此时必然会引起误识别。

由于语音识别 ASR 技术是基于关键词语列表进行匹配打分的技术，所以对输入的任何声音，都会去提取特征，并和设定的关键词语列表中的词语进行比对打分。不在识别列表内的发音，有可能会和识别列表中的某一个词汇很接近，同时又和其他词汇很不接近，此时就会产生误识别。

举例说明，当关键词语列表中全部是单音节时，比如你设定了 50 个全部是“ao lao dao biao dang bang bai ” 这类的单音节

由于这些内容的相互差距非常小，所以此时进行识别，几乎是你任意发一个音，都有可能匹配出好几个相似的结果。

或者，设定了过百个二字词的列表，此时任意的两个音节的发音，也都有很大可能引起误识别和匹配。

对于此类错误引起的误识别，LD3320 芯片努力做到拒识。但是由于 ASR 技术整体的算法基础，是无法消除这种错误。所以此类错误是无法避免的，而且此类错误的错误方式是无法精确预测的。

开发者需要在设计产品的声控界面时，通过其他手段，诸如“垃圾词语”，“相似发音”，“二次确认”等手段，来消除这类错误。这些方法就在本文随后讨论的章节。

4. 语音识别应用的适用原则

综上：语音识别产品的默认的适用原则是，用户会知道自己需要说什么语音命令内容同时会遵循这个原则，不说其他的词汇，

那么这样会带来极高的识别率。

一旦用户不遵循语音识别产品预先设定的流程，而去任意说，那么会带来不可避免的误识别。

三. 语音识别产品设计的思路

在用 LD3320 芯片设计语音识别产品时，应该充分在产品的设计时，利用如下原则，来提升终端用户的使用体验，而不是一味地依赖于芯片自身的识别能力：

- **用垃圾词语吸收错误，防止其他话语误触发**

http://www.icroute.com/web_cn/Download.html#LD3320-Miji

语音识别芯片 LD3320 高阶秘籍.pdf

第二节 增添“垃圾关键词语”——吸收错误识别

- **用发音相近词语来增加识别的容错性。**

http://www.icroute.com/web_cn/Download.html#LD3320-Miji

语音识别芯片 LD3320 高阶秘籍.pdf

第四节 巧妙运用关键词语的 ID;

第七节 对于同一关键词 ID 设置多个习惯发音

- **需要精确识别的场合，尽量使用触发识别模式，或者口令触发模式**

http://www.icroute.com/web_cn/Download.html#LD3320-Mi_ji

语音识别芯片 LD3320 高阶秘籍.pdf

第一节 在识别精度要求高的场景中，使用“触发识别”模式；

第三节 口令触发模式

- **设计合理的确认流程**

在设计语音控制产品时，需要考虑到误识别/误触发可能带来的不良影响。所以需要考虑确认的流程。就是需要用户分两次或者两次以上说不同的语音命令，来完成对一个操作的确认。这样可以极大程度地消除误识别/误触发带来的误操作。

比如，在“口令触发”模式下，是先识别出口令，然后再在限定的时间内说出要操作的命令，这样分为两级的确认。

比如，也可以设计为在识别出操作命令后，需要用户在限定的时间内说出“确定”这个关键词语，来进行确认，同样是分为两级的确认。

比如，也可以设计更复杂的确认流程，先要识别出口令，触发后在限定时间内说出要操作的命令，再在随后说出“确定”来最终确定。这样的三级确认过程。但是由此可能会带来使用上的繁琐。

开发者应该结合具体的产品，来设定合理的确认流程。

四. 语音识别产品界面设计案例

1. 声控电视遥控器—按键触发模式

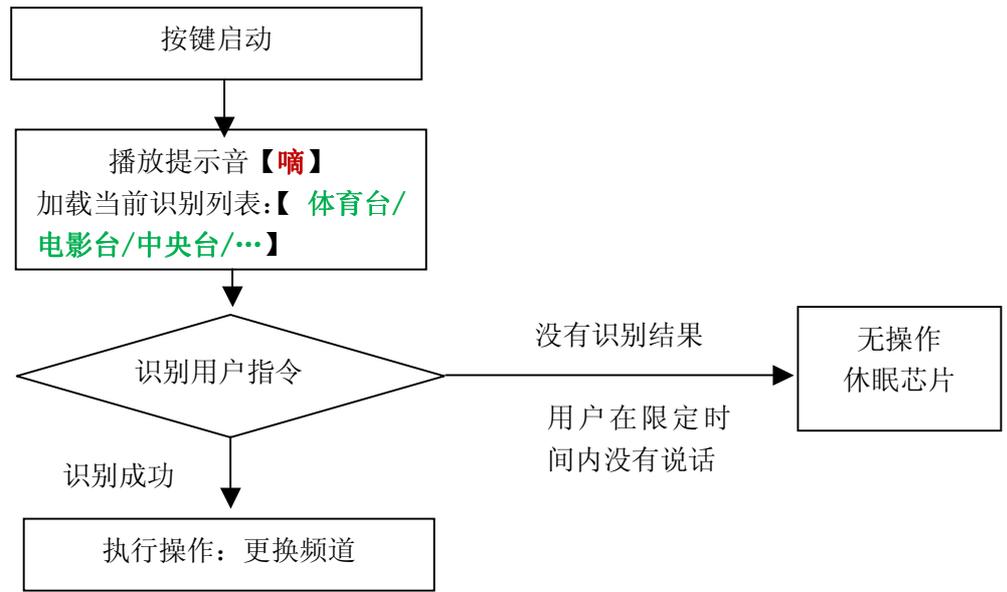
声控遥控器，最有用的地方是在更换频道时，可以直接说出频道名称，而不是去记忆频道名称和电视频道数字的联系。

同时，在使用电视遥控器时，遥控器是用电池工作，就不能让识别芯片一直处于工作状态。

因此在设计时，可以在遥控器上设置一个大大的按键，用户在使用时，按一

下这个按键，启动 LD3320 语音识别芯片，此时可以播放“嘀”的一声提示音，然后在限定的时间内，（比如 5 秒钟），接收用户的语音命令，并给出识别结果。比如用户说“体育台”，识别芯片把识别结果提供给遥控器的主控 MCU。随后遥控器的主控 MCU 就根据事先设定好的对应关系，发出对应频道的红外编码，实现换台。可以不加确认过程。直接换台。

终端用户可以根据自个人习惯来设定频道名称的内容需要终端产品提供输入界面，或者和 PC 机联机设置



说明:

1. 每次加载识别列表时，如果命令字不到LD3320 的最大容量，都可以添加垃圾词语吸收可能的噪音
2. 可以对同一识别词语设置多个不同的习惯发音，方便用户使用。

2. 声控咖啡机——“按键”触发模式：结合产品特性来合理触发

声控咖啡机，主要针对的是家庭和小型办公室场合使用的咖啡机。这类咖啡机的外观形体比较小，在工业设计上难以安排众多的按钮让用户直接选择；同时，由于使用者相对比较固定，使用环境比较安静，所以有利于发挥声控功能的效果。如果是象咖啡店使用的咖啡机一样对每个种类都设计一个按钮，就会出现小小的咖啡机上排满了按钮的情况。此时，说一句咖啡的名称，就可以让使用者快速选择自己需要的咖啡种类，就非常地实用。

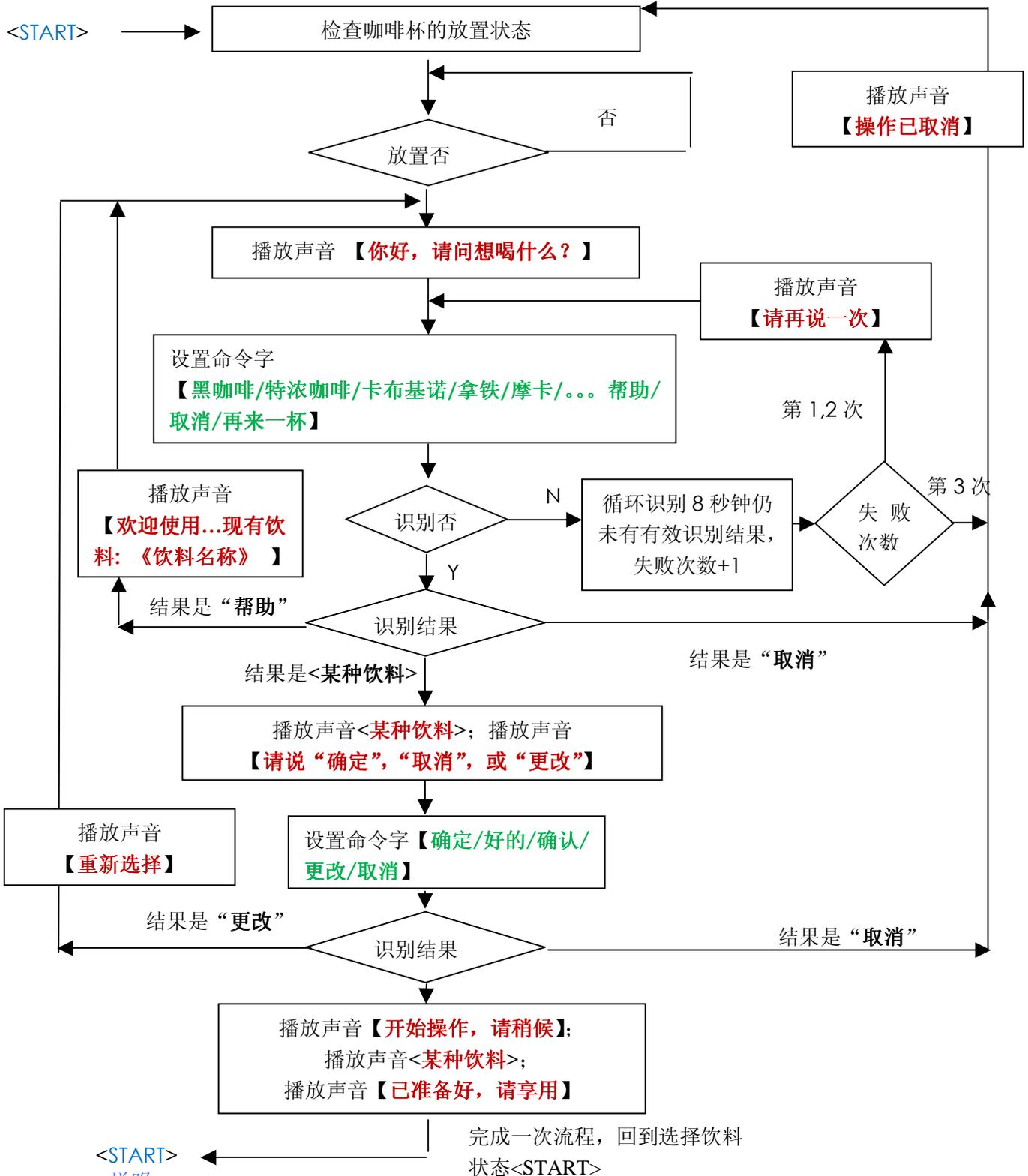
在声控咖啡机的语音声控界面设计时，充分考虑到咖啡机的实用特点：需要有咖啡杯来容纳咖啡。利用这一特性，就可以把“咖啡杯放置在托盘上”这一事件，作为语音识别开始的“按键触发信号”。

从而可以有效地避免语音的误触发。

在用户把咖啡杯放置到托盘上时，可以通过红外，重量等传感器检测到这一事件，并触发语音识别功能。播放提示音“请问需要什么咖啡”提示用户说出语音指令。在随后的 8 秒钟时间内，等待识别检测用户的语音指令。如果在 8 秒钟时间内用户没有说话，可以再提示“请重复”，如此重复等待 3 次后，就可以退出识别流程。

在识别出咖啡种类后，可以提示用户进行确认或者修改，播放提示音“***（重复识别出的咖啡名称），确定，取消还是更改”。在用户确定后，为用户生产一杯咖啡。

（*流程框图在下页*）



说明:

1. 每次加载识别列表时，如果命令字不到LD3320 的最大容量，都可以添加垃圾词语吸收可能的噪音
2. 可以对同一识别词语设置多个不同的习惯发音，方便用户使用。

3. 声控智能家居控制——口令触发模式：设置合理确认流程

声控智能家居控制，是在现有的智能家居控制系统中，增加了语音用户界面 VUI。

目前已有的智能家居控制系统，主要有两种方式：无线遥控和有线操作（电力载波或者中央布线）。都会有一个中央控制器或者中央遥控器来操作家里的各种家电。通过在中央控制器或者遥控器中集成引入 LD3320 芯片，可以很方便的为智能家居系统增加语音用户界面 VUI。

对于中央遥控器，如果是使用电池供电，那么就应该采用和<1. 声控电视遥控器>采用一样的按键触发模式。如果遥控器可以直接插在墙上的插座用交流电供电，那么就可以采用以下几种口令触发模式。

在设计这类产品的语音界面时，需要考虑到用户的需求，是希望在不接触控制器的状态下，通过语音来远程操作家里的各种家电。因此，口令触发模式就是最适合的模式。同时，一定需要设置确认的流程，务必要避免误触发/误识别的情况发生。以下以三个二级确认的流程来举例说明，开发者可以在此基础上进行修改，或者拓展成三级确认的流程。也可以在产品中支持多种流程，由终端用户根据自己使用的习惯来选择。

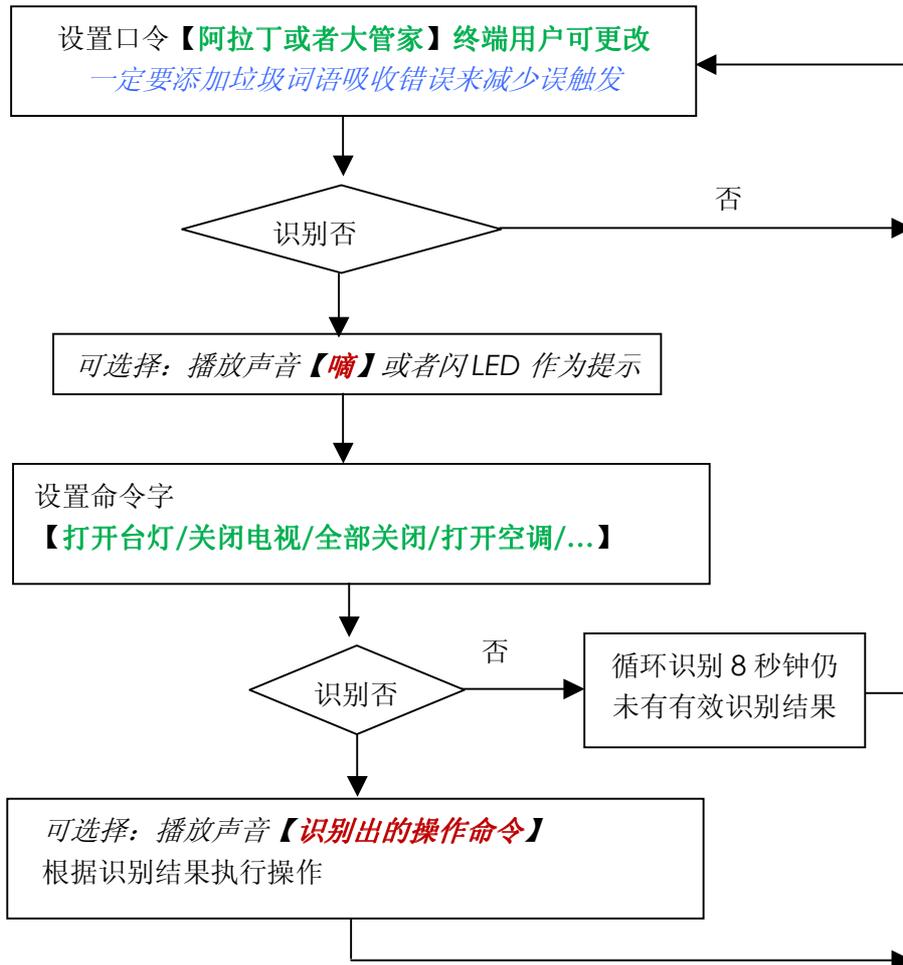
1) 设置前置口令，第二级说“操作+设备”

设置口令为“阿拉丁”或者“大管家”（口令可以选择发音响亮，琅琅上口的词语）。中央控制器处于一直上电的状态，所以识别芯片一直处于循环识别状态，只有在识别出是口令后，才进入下一阶段的识别流程。并且充分利用 LD3320 芯片的特性，在添加口令的同时另外增加垃圾词语直至达到 LD3320 芯片的最大容量，这样可以很大程度地避免其他声音带来的误识别。（在成功识别出口令后，可以选择播放很轻微的“嘀”的一声，或者闪一下小的 LED 灯，来作为提示）

在识别出口令后，在接下来的 8 秒钟时间内，等待接收用户的命令，比如“全部关闭”，“打开厨房灯”，“打开空调”等等。如果在限定的时间内，没有正确识别和接收到用户的命令，则取消本次识别流程。重新回到等待口令的状态。

（*流程框图在下页*）

设置前置口令，第二级说“操作+设备”

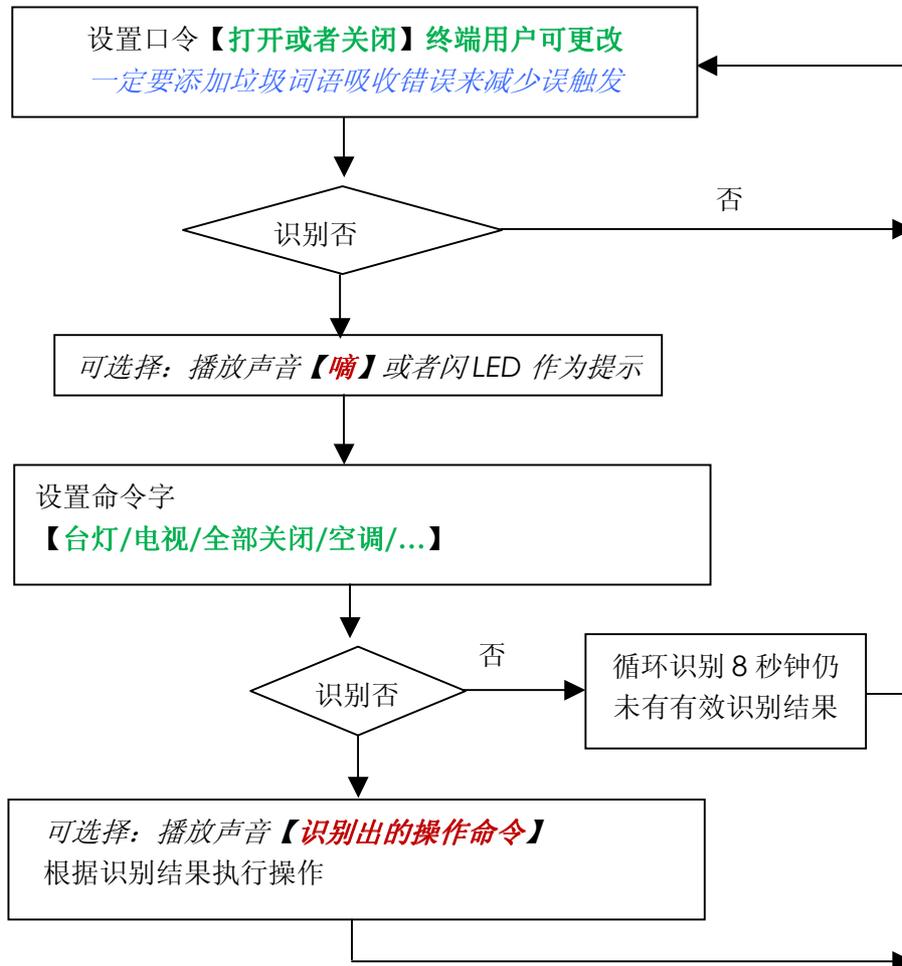


说明:

1. 每次加载识别列表时，如果命令字不到LD3320 的最大容量，都可以添加垃圾词语吸收可能的噪音
2. 可以对同一识别词语设置多个不同的习惯发音，方便用户使用。

2) 把操作（打开/关闭）作为口令，第二级说“设备”

从 1) 的流程可以变化出另外一种流程：口令设置为“打开”和“关闭”，命令词语为“台灯”，“空调”等设备名称。

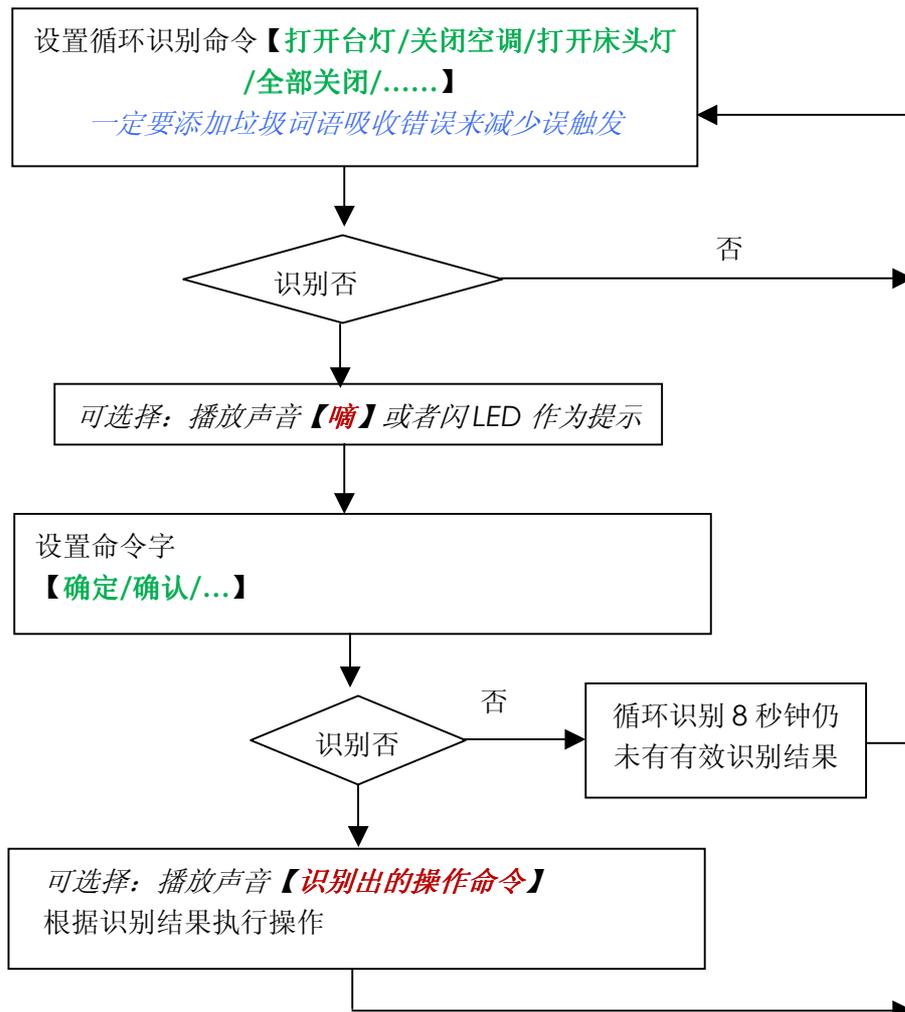


说明：

1. 每次加载识别列表时，如果命令字不到LD3320 的最大容量，都可以添加垃圾词语吸收可能的噪音
2. 可以对同一识别词语设置多个不同的习惯发音，方便用户使用。

3) 第一级循环识别“操作+设备”，第二级说“确定”

这一种是“后确认”的流程：在第一级循环识别时，识别内容是操作的名称，比如“打开床头灯”，“全部关闭”等等，识别出后，只有在接下来的5秒钟时间内说出“确定”这个关键口令，才进行正式操作，否则就取消本次识别流程，重新进入循环识别状态。



说明：

1. 每次加载识别列表时，如果命令字不到LD3320 的最大容量，都可以添加垃圾词语吸收可能的噪音
2. 可以对同一识别词语设置多个不同的习惯发音，方便用户使用。

4. 声控智能固定电话——按键触发模式

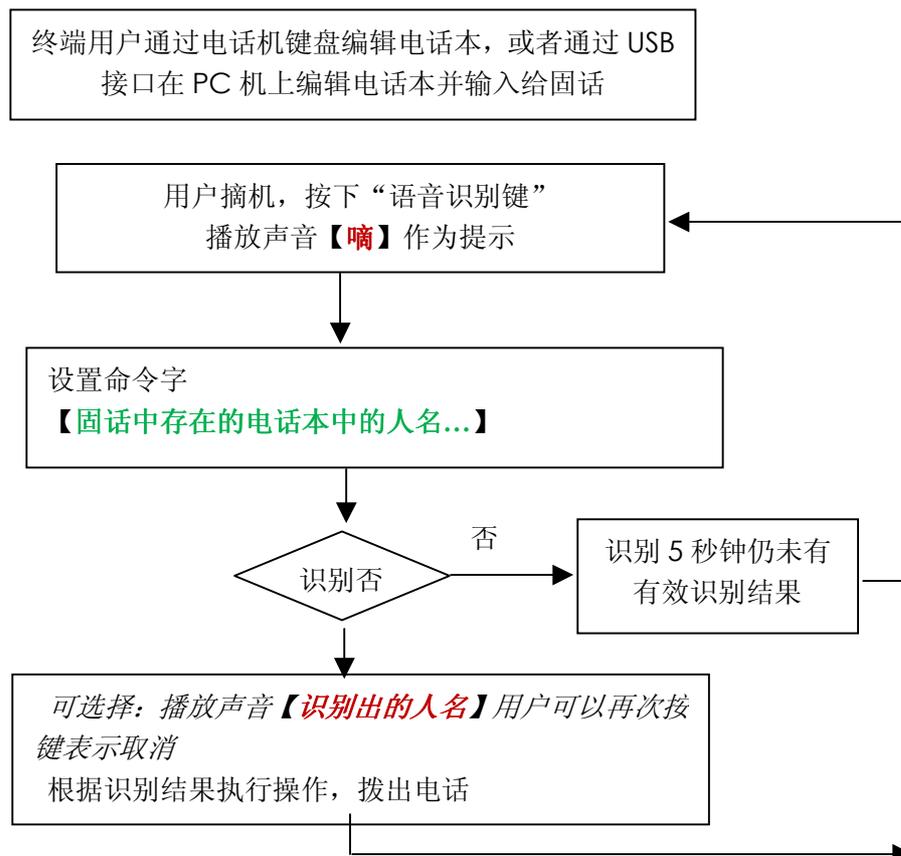
随着手机的大规模应用和迅猛发展，固定电话的发展已经停滞了很久。但是在家庭和办公场合，固定电话还是一个不可或缺的设备。但由于没有人再在固定电话上作开发，所以固定电话的人机界面很难用。

通过增加 LD3320 语音识别芯片，就可以非常实用地为固定电话增加声控的界面。简单来说，就是可以让拨打电话者说出电话本中的人名，就可以直接拨打出去。

非常适合家中的父母使用，也适合办公室环境中拨打常用的内线电话。

需要解决的问题是让终端用户如何输入电话本，可以通过增加拼音输入法，让用户通过电话机键盘输入；也可以通过增加 USB 接口，通过 PC 机修改。

在有了电话本后，每次用户提起电话机后，按动电话机上某个特点按键，则播放“嘀”声提示，用户在随后的 5 秒钟时间内，说出电话本中的某一个人名。在有识别结果后，可以回放一次该人名，如果识别有错误，用户就重新按一次按键，取消该次识别结果并重新说人名，如果用户没有按键取消，就直接拨出该人名对应的电话号码。



说明：

1. 每次加载识别列表时，如果命令字不到LD3320 的最大容量，都可以添加垃圾词语吸收可能的噪音
2. 可以对同一识别词语设置多个不同的习惯发音，方便用户使用。

5. 声控智能医疗护理产品——口令触发模式与按键触发模式相结合

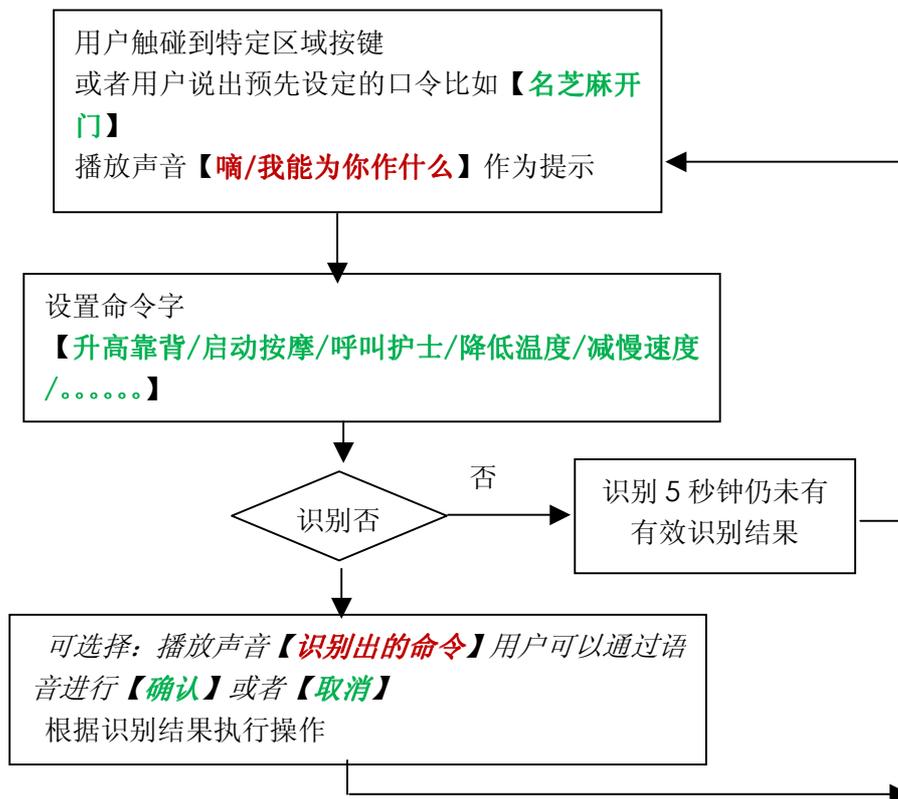
随着人们生活质量的提高，对于身体的健康问题也是日益重视，因此很多家庭都购买了医疗保健产品，比如按摩椅，跑步机，足浴盆等。但是每次都要从享受护理的状态变成去找按键操作机器，带来了很大的不方便。

在医院中，有很多病人需要医疗护理设备，但是这些病人往往很难象正常人一样去伸手按键进行操作，比如操作护理床升高降低等；也有一些残障人士，无法有效地去用手操作这些设备，使得这些设备的效果大打折扣。

在这样的设备中，引入语音用户界面 VUI，可以很好地解决这些问题，为用户提供更加方便的操作界面。

在设计这类产品时，可以考虑口令触发模式和按键触发模式相结合。在产品上设计一个超级大的触摸区，只要接触到这个区域，就会启动语音识别流程，引导用户通过语音进行操作。同时，也支持口令触发，用户也可以通过说出特定口令，来进行操作。

这样的结合模式，可以充分照顾各种不同人士在使用医疗护理产品时的需求。



说明：

1. 每次加载识别列表时，如果命令字不到LD3320 的最大容量，都可以添加垃圾词语吸收可能的噪音
2. 可以对同一识别词语设置多个不同的习惯发音，方便用户使用。

(更多技术资料请到 www.icroute.com 阅读和下载)