

研 究 报 告

R/TAF 003-2021



智能语音交互体验研究

Research on the experience of intelligent voice interaction

电信终端产业协会

2021 年 1 月 15 日

版 权 声 明

本研究报告版权属于电信终端产业协会，并受法律保护。转载、摘编或利用其它方式使用本研究报告文字或者观点的，应注明“来源：电信终端产业协会”。违反上述声明者，本协会将追究其相关法律责任。



目 录

| | |
|-------------------|-----|
| 前言 | III |
| 1 范围 | 1 |
| 2 规范性引用文件 | 1 |
| 3 缩略语 | 1 |
| 4 人机交互-语音交互 | 1 |
| 5 语音交互的形式 | 2 |
| 6 语音交互特点 | 2 |
| 6.1 快捷性 | 2 |
| 6.2 安全性 | 2 |
| 6.3 差异性 | 3 |
| 7 语音交互对话原则 | 3 |
| 7.1 简洁原则 | 3 |
| 7.2 准确原则 | 3 |
| 7.3 目标性原则 | 3 |
| 7.4 自然原则 | 4 |
| 7.5 友好原则 | 4 |
| 7.6 人格化原则 | 4 |

前 言

目前，智能语音识别技术已成为最为落地的人工智能方向之一，相关智能产品共同面对一个问题是：如何构建一种继承人类自然语音会话的用户体验模式。语音交互是一个全新的体验设计领域，目前各家机构、公司、设计组织还处于摸索设计范式的阶段。在计算机领域，从命令行界面进入到图形用户界面是一大突破，图形界面的学习成本较低。随后发展到目前的主流操作方式触控，使用手指在屏幕上滑动点按。语音交互界面、手势、动作、表情交互，甚至脑机接口，都属于自然用户界面。人类最早的交互方式就是语言和动作，在自然交互方式的趋势下，机器学习和人工智能激发着对话式界面和自然语言的进步，同时使对话式交互成为一种新的交互模式。

本研究报告涵盖了关于对话的核心机制，介绍了核心的设计原则，主要包括语音交互形式、语音交互的特点、语音交互的对话原则等。



承担机构：智能产品评测工作组（WG7）

研究单位：中国信息通信研究院、百度在线网络技术（北京）有限公司、 华为技术有限公司

项目负责人：董千洲

项目参与人：董千洲、李玮、刘毓炜、傅蓉蓉、李笑如

智能语音交互体验研究

1 范围

本研究报告主要研究语音交互的用户体验，主要包括语音交互现阶段的发展和场景，语音交互与其它交互区别的特殊性，语音交互对话的原则以及如何影响体验感等内容。

本文件适用于语音交互系统的语音用户界面设计。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 5271.29-2006 信息技术 词汇 第 29 部分 人工智能语音识别与合成

3 缩略语

下列缩略语适用于本文件。

| | | |
|-----|------------------------|--------|
| GUI | Graphic User Interface | 图形用户界面 |
| VUI | Voice User Interface | 语音用户界面 |
| CLI | Command Line Interface | 命令行界面 |
| NUI | Natural User Interface | 自然用户界面 |
| AR | Augmented Reality | 增强现实 |
| VR | Virtual Reality | 虚拟现实 |

4 人机交互-语音交互

人类与世界万物的交互过程大概是这样的：通过眼睛去观察人、事、物、环境等，再辅以耳朵听、鼻子闻、舌头尝、嘴巴交流、肢体触碰，而后大脑产生记忆和思考并做出动作、表情、语言和生理反馈，这个过程是循环和组合的。不同的环境和习惯会有不同，但基本是在这个范围。这个过程中，前半部分主要是接收信息，后半部分主要负责交流互动，而语言和动作是最主要的交流互动方式。人机交互角度看，用手操控是这个世界绝对的主流。

我们每天使用的手机、电脑、相机、汽车，包括现在很火的 AR、VR 设备，基本上都要靠手去操控。（用手操控属于动作这个范畴）这跟人类的进化方式有很大关系，我们的祖先从学会制作工具开始，手就成为了人类接触万物的最主要工具。尤其是机械的出现，只有人类灵巧的双手才能精准的操控完成任务。

可是人的手有几个缺陷：不够长、不够多、需要配合眼睛，这给我们的生活带来了许多不便。因此，在人因工程的发展过程中，理念由人适应机器变为机器适应人类，以提高工作效率。从最原始的动作交互到命令行 CLI、降低学习成本的 GUI 以及电容屏手机触控，发展到了现在流行的 NUI，包括手势、表

智能语音交互体验研究

情、语音、脑机接口。其中语音交互因为其自身脱离双手的操作而逐渐成为主流的人机交互方式之一。

5 语音交互的形式

随着人工智能各种应用的不断渗透，智能语音交互形式可以简单的分为手机载体和其他硬件载体。

手机载体：包括以手机为载体的语音输入模块（语音输入模块和识别系统，输入法、搜索框等）、语音助手（如 Siri、Google Now、Bixby 等）、语音交互系统（如聊天、商务客服、医疗等）。

硬件类产品：包括车载语音助手（如车载语音控制系统，Command、Mylink、HiCar 等）、家庭语音助手（以语音交互技术为基础的家用语音交互产品，如音箱、电视、控制中心等）、可穿戴设备、机器人（具备语音交互功能的可穿戴设备、机器人等）。

6 语音交互特点

6.1 快捷性

语音交互之所以能够流行，其快捷性便是吸引用户的一大优势。以定闹钟为例，如果想通过传统方式定闹钟，流程是：亮屏-上划打开控制栏-点击图标-选择闹钟-定闹钟-结束。而如果通过语音助手，我只需要：嘿，相关语音助手唤醒词（启动语音助手）-帮我订一个明早 8 点的闹钟-结束。因此“语音交互”所需要满足的很重要一点就是操作快捷性，能动动嘴皮子就解决的事，往往会比动手来的轻松很多。而快捷性可以细分到三种特点：追求高效、偏向执行、设备优势。

6.1.1 追求高效

高效性主要是针对于场景，比如在办公场景：给 XXX 发送一封邮件，邮件内容是***；比如生活场景：我要去某地，请从我当前所在位置为我找一种时间最短的出行方式。诸如此类还有很多，用户追求的就是足够的快速，足够的方便。讲一句话执行操作的时间肯定比手动打字要高效、节省时间。

6.1.2 偏向执行

偏向执行就是结果导向，用户关注的是事情或者命令执行的结果，并不关心过程。比如：用户想要查询他买的股票是涨了还是跌了，对他来说也许关心的只是最后呈现的这么一个结果，那他只需要通过语音助手询问即可获知。因为本身通过“语音交互”执行命令时，用户就已经放弃了操作的过程，设备已经把所有的过程通过用户的一句话给省略了。

6.1.3 设备优势

即可以通过语音来实现远程控制设备，我们不需要去触摸设备，不需要有其他操作，只需说一声，设备就能运转起来。也许是简单的让放在桌上的手机设置一个闹钟，也许是让家中的电器开始运作。通过“语音交互”，我们确实能消除很多由于空间而带来的限制。

6.2 安全性

智能语音交互体验研究

安全性最直接的使用场景——开车。虽然明文规定开车的时候不许接打电话，但实际生活中仍有很多人还是会在驾驶途中接电话。即使有耳机，在有电话接进来的时候往往也需要我们再按一下相应的按键，才能接听。但在有“语音助手”的情况下，我们也许只需要说一声“接听”就可以了。包括我们临时有急事想要拨打电话给别人时，同样可以满足对应需求。

因此在很多时候，如果产品的语音交互功能完善，就可以为用户解决很多烦恼，同样也可以避免很多安全事故的发生，因为这个时候人的注意力不需要再集中在操作设备身上，只需要简单说几句话就可以解决一切，保证周围环境的安全。

6.3 差异性

“语音交互产品”更可以解决不同设备之间的信息流转问题，这就是未来的智能家居概念，通过语音来控制所有的家具设备。因为不同的设备在输入方式的选择上可能会存在差异，比如：有些是按键，有些是触摸等，但如果所有家具都能利用“语音交互”来完成相应的控制，那一切就会随心所欲很多，而需求往往同样对应着合适的场景。

7 语音交互对话原则

话术作为语音交互过程中内容表达的载体，其重要程度在语音交互领域已有共识。平时有人说“会说话的人，一开口就赢了”，对于智能语音交互设备来说，也同样如此，说话的内容和表述方式直接影响着用户的感受和体验。为了做好 VUI 的用户体验，现阶段语音交互主要在对话方面着重下手。而以下几个原则也是现阶段提升用户体验的主要原则。

7.1 简洁原则

对于简洁原则，主要考虑的语音交互对话简洁，让用户能清晰明了的得到答案。比如在使用语音交互进行问答游戏时，一次提供的选项不宜超过三个，过多的选项会使得信息量过大，让用户无法一次性记住从而无法进行回答；比如在使用语音交互进行问题查询时不需要太多步骤，过多的步骤会引起用户的反感，降低用户体验。比如在使用语音交互播放有声资源时，用户只关心是否播放成功，不会在意播放资源的细节，因此在达到用户目的便不再发出声音。

7.2 准确原则

准确原则简单来说就是对于用户来说，语音表达准确无误，避免笼统模糊，用户能够第一时间明白内容。比如在语音交互问答时，如果技能说“说继续进入下一题”，用户很容易误解为自己要说“继续进入下一题”，但技能的原本意思是让用户说“继续”两个字，技能会“进入下一题”，如果把话术改为进入下一题，请说“继续”，就很明确了；比如在播放资源时，应该结合前后语音交互语句和语境，播放用户想要的资源；比如在问答时，给用户的示例答案要无歧义，宜直白明了。

7.3 目标性原则

在语音交互时，如果遇到网络系统等 bug、用户回答不准确等客观问题，强制让用户重新语音交互会在一定程度上降低用户体验。语音交互宜清楚传达技能接收情况，不确定用户意图时进行“意图确认”。

智能语音交互体验研究

比如智能音箱识别到语音答案但是因为系统问题无法做出语音回应时，不应显示未识别内容，对于技能状态的说明一定要清晰明确；比如在进行多轮对话时，需求未满足时，不轻易终结对话，告诉用户接下来怎么做。

7.4 自然原则

自然原则，其实就是让语音交互更口语、自然化，让用户感觉不到与机器进行对话。语音交互对话宜措辞口语化，避免使用专业的、晦涩的语言，比如下达命令语句改为日常陈述，非编程用语；在回答用户时，宜加语气助词让用户明白指令已被听到，使得句式自然，使用对话标记；比如在表达同样意思时可以提供多样化的表达，同一内容提供多样化表述。

7.5 友好原则

友好原则主要体现在出现错误的时候，不宜过多的把责任推脱到用户身上。比如在在语音对话过程中遇到问题时，设备或技能首先要主动致歉或认错，不是直接指责和归咎于用户，宜主动认错，不责怪用户；比如在语音对话过程中识别错误，设备可使用适当策略处理改正错误，或提供帮助避免重复错误，不要将技术错误“事件”视为用户行为错误。

7.6 人格化原则

人格化原则类似自然原则，更多的则是锦上添花的功能，让用户高兴，增加体验感。比如日常的问答对话可适时幽默，有时能化解无法满足用户需求的尴尬气氛或者为每个用户设计不同的语音交互体验，优先考虑个性化。



电信终端产业协会研究报告

智能语音交互体验研究

R/TAF 003—2021

*

版权所有 侵权必究

电信终端产业协会印发

地址：北京市西城区新街口外大街 28 号

电话：010-82052809

电子版发行网址：www.taf.org.cn