

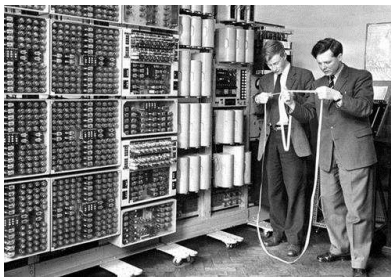
人机交互接口的发展

发展历程：NUI -> CUI -> GUI -> NUI

NUI：Null User Interface。

在计算机发展早期，并不存在用户交互界面，人们需要用打孔纸带与计算机进行交互。

早期的计算机并不能实现与人的交互，它只能在获得打孔纸带上的程序命令后完整地运行完程序并给出输出结果。



早期的计算机



早期计算机使用的打孔纸带

CUI：Command User Interface

随着计算机的发展，到 20 世纪 50 年代末期，人们可以通过命令行用简单的命令语句实现与计算机的交互。但由于当时的显示器价格过高且分辨率很低，CUI 交互并不在显示器上进行，而是在电传打印机（造型上有点类似老式打印机，但它提供了可供电脑实现输出的接口）上进行。用户通过键盘输入命令（同时打印到纸上），计算机再通过键盘将反馈打印在纸上。

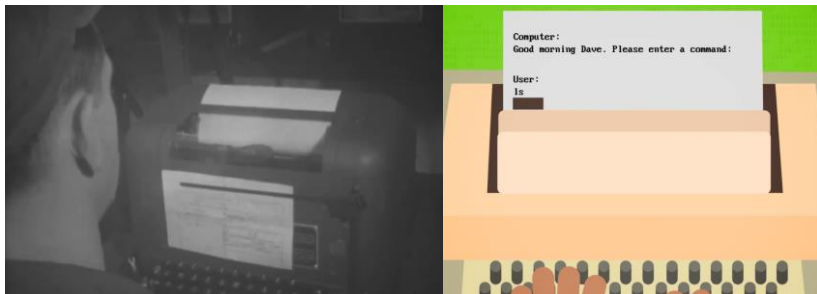
批注 [刘1]: 参考：

<https://www.bilibili.com/video/BV1EW411u7th?p=2>
2。（前后的相关章节都可以了解一下）

批注 [刘2]: 打孔纸带的具体原理：

<https://www.zhihu.com/question/22892716>。

批注 [刘3]: 此时的计算机被称为 teletype computer。



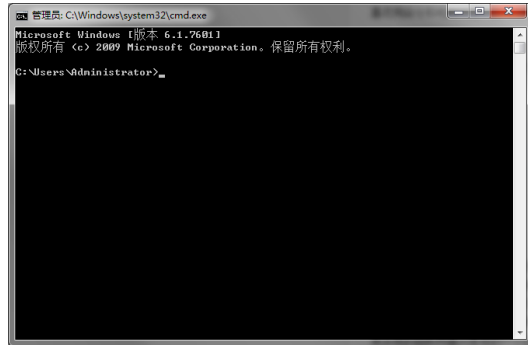
早期的 CUI 交互

20 世纪 70 年代，随着显示器技术的发展，计算机开始逐渐用显示器替代之前的纸作为计算机的输出终端。



早期带有显示器的输出终端

由于当时内存及显示器的限制，无法实现像素化的输出。显示器会确定一个字符集，并提前确定其对应的输出图形。显示器只能输出字符集中的字符。这种从内存读取字符代码、转化为光栅图形并显示到屏幕上的硬件被称为字符生成器（character generator）。在这样的基础上，出现了完全依托于字符集的操作界面。当前 windows 的命令提示符就是类似这样的交互界面。



现在的 Windows 的命令行界面

GUI: Graph User Interface

批注 [刘4]: 参考：
<https://zhuanlan.zhihu.com/p/51089574>。

早期的显示器在 20 世纪 60 年代就已经出现 (CRT 显示器), 但由于成像质量问题, 一般只用于显示需要实时更新的内容(因为将这些实时更新的内容打印到纸上会造成极大的浪费)。

第一台带有 GUI 的计算机是 1973 年的施乐奥托 (Xerox Alto), 但这一型号从未正式作为商品机出售过。直到苹果在 1984 年发布 Macintosh , 才出现第一台搭载图形交互界面的商品计算机。



1984 年的 Macintosh 与其操作界面

图形交互系统使得计算机能以图形显示操作界面, 并以鼠标的移动、点击作为基础操作模式, 与原先只能基于键盘的操作相比, 大大地增强了用户交互的直观程度, 实现了 “所见即所得” 的可视化交互理念。

如今我们看到的所有交互界面，不论是苹果的 IOS 系统，还是微软的 windows 系统，几乎都是基于 GUI 界面的。GUI 交互的到来大大降低了计算机的操作难度，是人机交互发展过程中具有革命性的变革。

NUI: Nature User Interface

自然人机交互界面。NUI 并没有脱离 GUI，只是在 GUI 的基础上拓展了交互的入口。GUI 指支持键盘、鼠标，而 NUI 希望增加手势、触摸、语言等各种不同的人机基本交流动作到人机交互之中。

VUI: Voice User Interface

语音交互界面。VUI 用于实现人类与电脑的语音交流。它包括通过语音识别获取理解用户的语言表达，并在获得用户的意图后执行特定命令或做出回应。

一些基本原理、概念与思想

交互发展与软硬件发展的辩证关系

交互的发展由软硬件的发展决定。交互的设计是基于当下的软硬件发展水平的，只有软硬件能够提供满足交互概念的功能，这种交互概念才有可能在产品中得以实现。

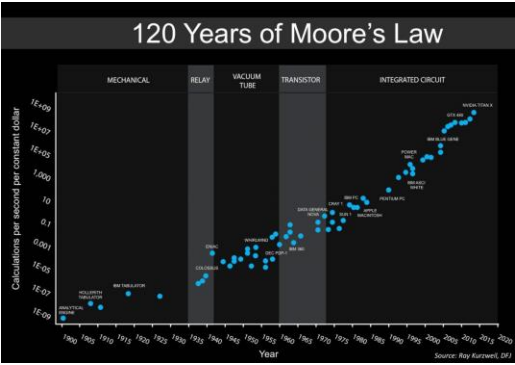
交互的发展会影响软硬件的发展方向。交互的理念会领先于软硬件的发展水平，向软硬件的发展提出要求，并引导软硬件的发展方向。

摩尔定律

该定律由英特尔的创始人戈登·摩尔提出，只是观测性的推论，不是严谨的定理。

具体内容为：“当价格不变时，相同大小的集成电路上可容纳的元器件的数目，每隔月18-24个月便会提高一倍，性能也将提高一倍”。

批注 [刘5]: 参考：
<https://zhuanlan.zhihu.com/p/21262505>。



莫尔效应发展示意图

摩尔定律的提出是在 1965 年，在 1965 年到 21 世纪初，半导体产业的发展都基本遵

循着这一推论。但在近些年，摩尔定律中提到的时间正慢慢被拉长。

摩尔本人对此现象的观点引用了霍金对这一问题的看法：“限制集成电路问题的极限是‘光的有限速度和材料的原子特性’”。

交互原语

这个概念来自《About Face 4：交互设计精髓》第 251 页。

交互原语指的是用于构成交互过程的最基本、最底层、无法再向下细分的动作或反馈。对于输入用户而言，这类动作包括点击、拖拽、滑动、旋转等；对于机器输出而言，这类反馈包括按钮触发、光标移动等。

批注 [刘6]: 参考：
<https://zhuanlan.zhihu.com/p/92506598>。

GUI 的标准化设备

目前的 GUI 标准输入设备就是键盘和鼠标，此外还有数位板、触控屏（在智能手机普及后，触控屏也慢慢变为主流了）等；标准输出设备主要是屏幕。

键盘

键盘按原理可以分为薄膜键盘、机械键盘、静电容键盘、静电容机械键盘、光轴键盘。

键盘的运行可以分为按键下压、触发、按键复原三个过程。就触发灵敏度来看，光轴键盘优于电容键盘，电容键盘优于机械键盘，机械键盘优于薄膜键盘。

除了以上这些传统键盘之外，现在还有激光投射键盘和隔空打字这样的新式键盘出现。

原理分类

薄膜键盘

薄膜键盘的每个按键下都有一个待连接的电路。当按键按下时，电路连通；当按键松开时，电路断开。

薄膜键盘的薄膜是指由硅胶膜、塑料膜、绝缘膜、塑料膜构成的四层膜。在两层塑料膜上印有电路，在按键松开时两层电路被带孔的绝缘膜隔开无法连通。当按键按下时，硅胶膜发生弯曲带动塑料薄膜弯曲，穿过绝缘膜的小孔与下方的塑料薄膜弯曲，电路连通实现触发。

按键的按压和复原是通过硅胶膜的形变和复原实现的。薄膜键盘可根据其键帽分为火山口、剪刀脚、公柱、蝶式等结构。

蝶式键盘是前几年苹果 macbook 使用的键盘，但被用户吐槽手感太差后在 2019 年的

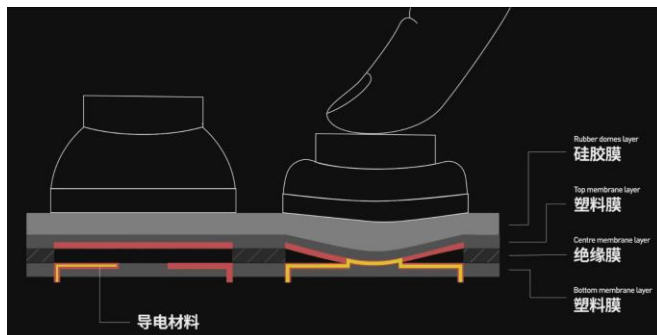
批注 [刘7]: 参考：
<https://www.bilibili.com/video/BV1Dt411A7PM>，

批注 [刘8]: 参考：
http://mouse.zol.com.cn/654/6546431_all.html，
<http://dongshuyan.com/keyboard.html>
。

macbook 中被替换。

相比机械键盘，薄膜键盘的厚度小、轻便、抗水性好、噪声小、价格便宜，在笔记本电脑中会较多使用；但敲击触感相对较差，且敲击次数较少，通常只有几百万次。

由于薄膜键盘的价格较低（几十元就能买到），它是现在主流的键盘。

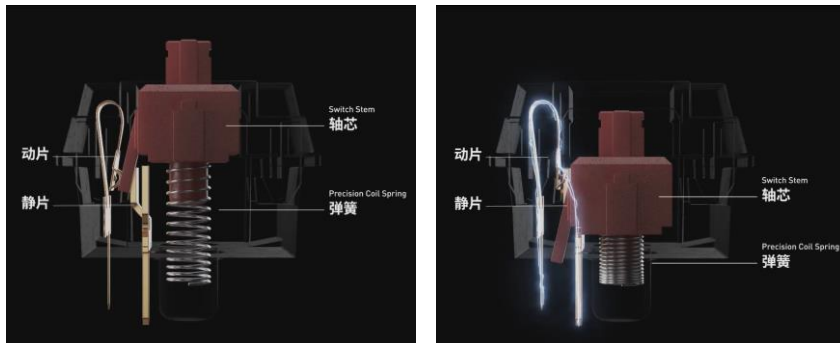


薄膜键盘松开与按下时的差别

机械键盘

与薄膜键盘相同，机械键盘的每个按键下同样有一个待连接的电路。当按键按下时，电路连通；当按键松开时，电路断开。

机械键盘的每个按键位置上都有两个固定的动片和静片。当动片与静片贴合时，能够构成回路；分开时，回路断开。当按键松开时，弹簧托起轴芯，使得轴芯卡住动片，动片无法与静片贴合。当按键被按下时，动片逐渐向静片靠近，当完全按下时动片与静片贴合，实现触发。



机械键盘松开与按下时的差别

机械键盘根据每个按键的轴芯不同会分为青轴、茶轴、红轴、黑轴等，不同的轴芯会带来不同的按压感觉。

机械键盘的按键按压与复原都是通过弹簧的形变实现的。

与薄膜键盘相比，机械键盘的重量更重、体积更大、敲击手感更好，敲击次数更多，平均可达 5000 万次以上；但对应的，机械键盘抗水性较差、噪声响亮、体积大、沉重，通常都作为单独的商品出售，一般不被集成进笔记本中。

机械键盘在游戏电竞的发展下，现在也慢慢晋升主流键盘的行列（一般 300 元也能买到）。

电容键盘及电容机械键盘

薄膜键盘和机械键盘本质都是通过触点连通形成回路实现按键判断的。与之不同的是电容键盘。电容键盘通过按键下压使得按键下方的电容极板的介电系数变化，进而导致电容变化从而实现按键触发。

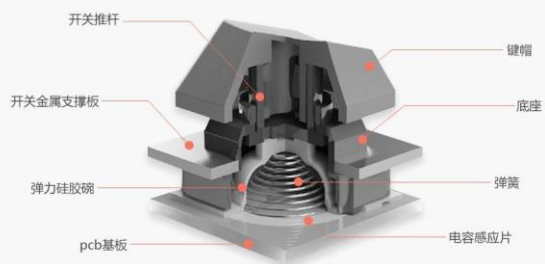
根据按键下压和复原的原理不同可以分为普通电容键盘和电容机械键盘，前者利用硅碗实现下压和复原，后者利用硅碗加弹簧实现下压和复原。

批注 [刘9]: 参考：

<http://dongshuyan.com/keyboard.html> ,
<https://www.bilibili.com/video/BV1Xt411B7xk?from=search&seid=16732766900777066086>。

静电电容开关

双色键帽是证明国产在努力的标志，双色键帽中最难做的是全封闭的键帽，如O键，更难的是pcb材料，而我们做的就是全封闭双色pcb键帽，希望用户能懂我们，也希望中国制造能成为品质的象征。



电容键盘结构图

电容键盘既有机械键盘的优秀触感，又能做到噪声小，且电容键盘使用寿命比机械键盘和薄膜键盘都长。其缺点就是价格太高。目前不是主流（一般价格在 800 元以上）。

光轴键盘

光轴键盘在键帽的轴芯下方放置一个光线发射器和接收器。当按键松开时，轴芯上的部件会遮挡发射器，使接收器无法接收到光信号；当按键按下时，轴芯下移会露出拱光线穿过的孔槽，使得接收器得以接收信号，实现触发。



光电键盘结构（此为按键未下压时的状态）

光轴键盘的下压与复原也是通过弹簧实现的。

光轴键盘的响应速度快、耐用且使用寿命长，但价格太高（没找到具体产品）。目前不是主流。

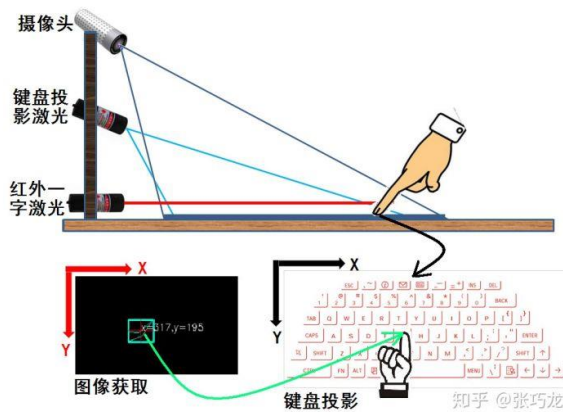
激光投射键盘

激光投射键盘是通过激光把键盘发射到某个表面，形成键盘的构造，在上方打字可以实现对应输入的键盘。

这种键盘一般为柱形，顶部配有键盘投射激光和红外摄像头，底部有一个一字红外发射器，用于发射平行于键盘投射面的红外光。在实际运行时，红外发射器会持续发射红外线，当有手指按键时，手指表面会反射红外线。由于顶部摄像头的位置固定，因此只要键盘投射的平面是平行于发射器发出的红外光，摄像头中键盘的位置就是相对固定的。此时，通过计算手指反射光线的位置在图像中的位置就可以确定手指敲击的是哪个按键。

批注 [刘10]: 参考：

https://www.bilibili.com/video/BV1hE411P7ES/?spm_id_from=333.788.videocard.4 ,
<https://zhuanlan.zhihu.com/p/42963682>。



激光投射键盘构造及原理图

这一键盘为手机屏幕较小、无法实现良好的全键盘输入提供了一种解决方案。但综合来看它还有诸多缺点：由于其原理较为简单，因此可靠性较差（一旦键盘投射面本身存在倾斜，或这个键盘投射装置被置于一定高度上时，投射出的键盘在摄像头中的位置就发生了变化，键盘就无法实现正确输入）；按键缺少一般键盘的反馈，使用体验不佳；续航能力不佳；这类键盘在淘宝的价格普遍在 100 元以上，售价太高。综上，很多时候这个激光键盘只能作为新鲜的玩具，难以起到真正的用处。

隔空打字

我找到两个隔空打字的案例。

一种是三星在 2020 年推出的通过前置摄像头拍摄用户双手，利用 CV 追踪手指运动并判断手势实现的隔空打字。这种隔空打字不需要额外购买配件，功能可以被直接嵌入到手机内，而且还可以配合隔空手势提高操作效率。

批注 [刘11]: 参考：

<https://zhuanlan.zhihu.com/p/31717144> ,
<https://www.bilibili.com/video/BV13W411V7KH> ,
<https://www.bilibili.com/video/av85397068/>.



三星的隔空打字

隔空打字的算法的一大难点在于如何确定用户的初始状态手指对应的按键位置。

从用户的角度看,由于在这种打字方式下,用户完全处于盲打状态,因此可能需要一段时间的训练才能完全适应。与激光投射键盘类似,这种打字方式缺少反馈也是一大问题。

另一种是 2017 年美国的公司 tap systems 做的,需要配搭额外的穿戴设备。tap 为每个字母和数字设置了对应的手势。虽然它是隔空打字设备,但仍然需要一个桌面,或其他用于敲击的表面,无法实现完全地悬空打字。此外,tap 的打字与一般的打字不同,它为每个字母和数字分配了特殊的手势,用户需要一定时间的学习才能实现熟练的打字。

tap 的每个手指的指套上都装有一个三轴加速度计,通过感知手指敲击到表面时的突然减速来判断运动的手指和具体的手势。



Tap system 的打字设备

这个设计看似在一般场景下不如三星的设备有效，学习难度很高，只能使用单手操作，但它们的主要目的是在 AR 和 VR 设备中实现打字输入，因此也具有不错的前景。（顺带提一句，这个键盘似乎还有鼠标功能）

键盘布局

字母键盘

目前的键盘普遍采用的都是 QWERTY 布局。这种键盘的打字速度并不是最高的，原因在于这种键盘的形成是在打字机时代，过快的打字速度会导致快速敲击相邻的字母时打字机的连杆发生碰撞和勾连，因此当时的人们对键盘的设计进行了调整。

但按理说这种问题在现在的键盘上已经不存在了，QWERTY 却仍然被保存了下来。原因在于这种键盘布局已经与大量的行业标准挂钩；且用户都习惯这种键盘的布置方式，不愿意耗费额外的时间成本去进行新型键盘的学习。

但根据上面的参考链接看，似乎关于键盘布局的内容还有可以扩展的空间。现在的 QWERTY 键盘的设计者是克里斯托弗·肖尔斯（Christopher Sholes），Sholes 用于改良打字机连杆碰撞问题的键盘被称为 QWE.TY 键盘。根据如今的统计学研究发现，QWE.TY 键盘导致的连杆碰撞概率显著降低，而左右手交替概率也显著上升。对于键盘来说，交替用手就意味着不会出现单手负担过重进而导致疲劳的问题，因此是有助于提高打字效率的。而 QWERTY 在 QWE.TY 的基础上进一步提高了字母相邻的概率，提高了打字效率。

由此看来，Sholes 对于键盘的改良设计始终是以高效为目的的。因此 QWERTY 键盘并不是一个为适应打印机时代而设计的低效妥协品，而是为了实现更高效的打字输入而设计出的产品。

以上所提到的都是一般的美式标准键盘，实际上在世界各地，由于语言文化的不同，不

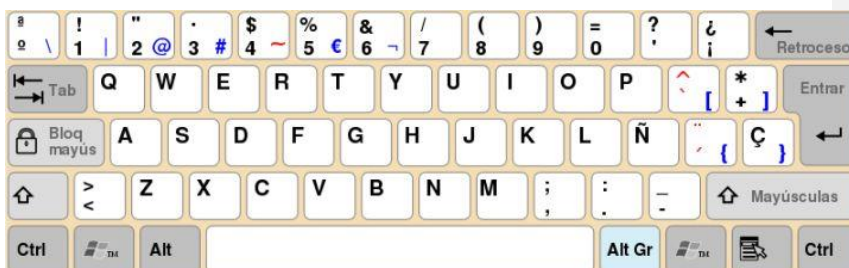
批注 [刘12]: 参考：

<https://www.zhihu.com/question/20195292>，（对 QWERTY 并非完全无效的设计的解释）
<https://www.zhihu.com/question/20121876/answer/129017959>。（对现有的不同地区的键盘布置的解释）

同地区的标准键盘本身也有一些细小的差异。



标准美式键盘



西班牙的键盘



加拿大法语区的键盘

计算器数字键盘与手机拨号数字键盘

仔细观察我们可以发现计算机的数字键盘是 789 在上、123 在下的，而手机拨号的数字键盘是 123 在上、789 在下的。

这里就没有字母键盘那么复杂的设计原因了，主要就是因为这两类布局分别对应了不同的工业标准。拨号数字键盘对应的是通信标准，而计算机数字键盘对应的是计算机标准。而这两个行业标准的不同其实也没有什么特殊的原因。计算机的数字键盘标准是由 1916 年的

批注 [刘13]: 参考：
<https://zhuanlan.zhihu.com/p/55900140>。

一个美国的 3*3 键盘布局的专利决定的；而电话的数字键盘标准是 AT&T 的贝尔实验室为了将旋转式电话调整为按键式电话，邀请实验用户进行了多轮实验比较最终得出的。

启发

HCI (Human Computer Interaction , 人机交互) 的注意点不仅仅是单一用户，而是整个社会群体。

人的习惯具有社会和历史的沿革性，一旦对某个产品形成了特定的使用习惯就会产生路径依赖，在之后很难再发生改变。与其强迫用户改变，不如改变设计。

鼠标

鼠标的是用于控制计算机游标的横纵位置的外接输入设备。鼠标的诞生是为了降低键盘操作的复杂性，简化计算机的使用。近些年来，鼠标的组成越来越复杂，中键、侧键等部件被不断地添加进来，鼠标的功能也越来越多样。鼠标最早在 1964 年发明，但直到 20 世纪 80 年代晚期才因为 PC 的普及而被大范围使用。

第一个鼠标是 1964 年由加州大学伯克利分校的道格拉斯·恩格尔巴特 (Douglas Engelbart) 博士发明的。Engelbart 发明鼠标时，世界上还没有出现图形操作系统。当时的主要交互方式都是 CUI。

Engelbart 的鼠标有两个滚轮，一个水平转动，一个垂直转动。通过跟踪它们的组合旋转来控制计算机屏幕上的光标移动。两个滚轮与一个按键一起被放在一个木盒子中，通过电线连接到电脑。由于两个滚轮的垂直摆放，这款鼠标只能实现横向或纵向的移动，无法进行斜向的移动。

批注 [刘14]: 参考：

<https://www.zhihu.com/question/320992975> ,
<https://www.bilibili.com/video/BV1fW41177Z1>。



Douglas Engelbart 发明的世界上第一个鼠标

根据鼠标的发展历史,大致可以把鼠标分为两类:机械鼠标和光电鼠标。此外,笔记本电脑上还有用于替代鼠标功能的触控板;与之类似的,IBM 还为笔记本电脑发明了指点杆 (track point 或 point stick,就是我们经常看到的 thinkpad 上的小红点)。

原理

机械鼠标

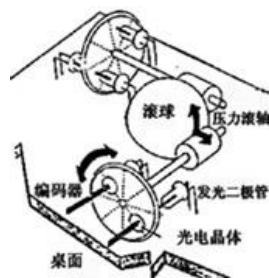
机械鼠标的核心是底部的滚球和与之相连的滚动轴。滚球可以实现任意方向的旋转,当滚球旋转时,横向或纵向的滚动轴也开始旋转。

最早的机械鼠标的滚动轴是有旋转极限的,换言之,鼠标只能在特定的区域范围内移动。鼠标通过确定转动轴的旋转程度确定具体的移动位置。

之后罗技公司在 1983 年设计了第一款光学机械式鼠标。这款鼠标的滚动轴不再有旋转极限,可以在任意范围内旋转。滚动轴上有两个带光栅缝的码盘,码盘两侧放有两个发光二极管和感光元件。滚球转动带动滚动轴旋转,滚动轴旋转带动码盘旋转,码盘上的光栅缝会使感光元件规律性的接收到“有光-无光”的信号。内部芯片通过分析感光元件的信号来判断鼠标的运动(具体原理我也有点没搞懂,不过我觉得应该不是重点,毕竟这个技术已经被淘汰了)。

批注 [刘15]: 参考:

<https://www.bilibili.com/video/BV1fW41177Z1>,
<https://www.bilibili.com/video/BV1PK411n7Qj?from=search&seid=3114692781373804338>,
<https://www.bilibili.com/video/BV1Zx411d7y4?from=search&seid=8378551119398206396>。

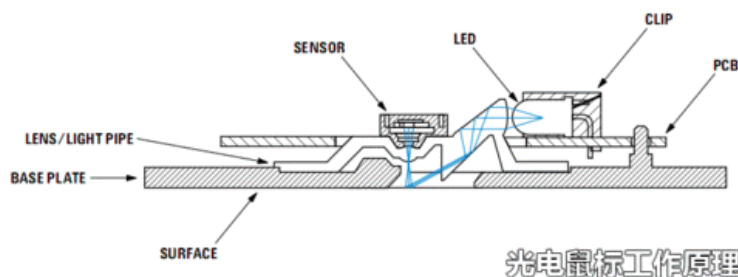


光学机械式鼠标内部结构

由于机械鼠标的滚球会直接与外部环境接触，因此很容易将灰尘带入内部结构中，导致鼠标灵敏度下降；而且由于为了保证滚球的良好旋转，滚球质量一般都比较比较大，这导致鼠标整体质量较大。因此在 90 年代后期光电鼠标商品化后，机械鼠标就被逐渐替代了。

光电鼠标

光电鼠标的本质是光学传感器通过不断拍摄鼠标底部的照片，图像分析芯片对高速拍摄得到的连贯图像的纹理进行分析，确定鼠标的移动。在此过程中，由于鼠标底部较为黑暗，需要用发光装置（LED 或激光）和棱镜将光线照射到拍摄位置提高亮度。



光电鼠标工作原理

光电鼠标工作原理

由于光电鼠标本质是通过照片纹理来判断鼠标移动的，在极为光洁或透明的表面上，光电鼠标难以区分移动前后的纹理区别，因此在那些表面上光电鼠标通常不能做出良好表现。

由于光电鼠标轻便、准确度高，因此在 21 世纪后，光电鼠标慢慢成为了主流。

触控板

触控板的原理与一般的电容屏幕原理相同。通过手指靠近改变触控板表面微小极板的电容来实现触控点的确认。当触控手指移动时，触控点也随即移动。芯片中的算法将触控点坐标与屏幕中的鼠标坐标相统一，就实现了触控板的功能。

现在的触控板通常还支持多点触控和手势技术，手势的实现本质也是通过算法对多个触控点之间的运动方式进行判断和处理。具体的实现方法和代码可以看上面的参考。

指点杆

指点杆的原理是，在指点杆下方的上下左右四个方向上都各有一个应变感应器（应该是压力陶瓷），在指点杆向特定方向倾斜时，对应方向上的感应器会感应到压力进而发生电效应。通过检测压电效应生成的电流可以确定指点杆移动的方向，进而可以转为鼠标的移动。

指点杆的商用基本都出现在 IBM 公司的笔记本上，其优势在于能在双手不离开键盘的基础上实现鼠标的操控。尽管如此，由于指点杆一直没有跳出 IBM 公司的产品，因此商品覆盖范围很小。不过在习惯它的用户群中，它一直拥有良好的口碑。

嘴控鼠标

嘴控鼠标一般是为残疾人设计的。使用时，用户需要咬住鼠标的控制杆，通过点头或摇头实现鼠标的上下左右移动，通过吸气或呼气实现点击或松开。

其基本原理应该并不复杂。在控制杆末端的四角添加压力传感器即可确定控制杆的移动；在控制杆内添加风扇，通过风扇的旋转方向即可确定呼吸和吸气。

目前最流行的嘴控鼠标是 quadstick（上面链接里的是类似这个产品的另一款产品 quadjoy）。

作为为残疾人设计的辅助装备，该鼠标显然并不适合一般用户使用。但以残疾人服务为目标进行评价，它的确实现了最大程度上的操作简化。而且，一般有意识的残障人士都至

批注 [刘16]: 关于手势控制可以参考：
<https://zhuanlan.zhihu.com/p/145135486>，

批注 [刘17]: 参考：
<https://www.youtube.com/watch?v=w3A7LDyizlc>，
<https://zh.wikipedia.org/zh-hans/%E6%8C%87%E7%82%B9%E6%9D%86>。

批注 [刘18]: 压力陶瓷会在收到压力的情况下产生电效应。

批注 [刘19]: 参考：
<https://www.bilibili.com/video/BV1wW411R7kx>。

少能保证嘴部还可以活动，因此嘴控鼠标也实现了对残障人群的大面积覆盖。

键鼠综合设备

在这里我想简单介绍一些用于辅助键鼠输入或适用于某些特定的、无法使用一般键鼠的人群的输入设备。这些设备为某些特殊人群或正常人群在某些特殊情境下的正常操作以及对正常操作的进一步优化提供了思路。

脑波输入设备

这部分内容我在《感知与传感器》中的 EEG 中也有提到。通过佩戴无侵入式设备，实现对脑波的检测，并通过用户行为产生的脑波的变化实现信息的判断与输入。通过配合其他设备，如眼动检测仪，脑波输入能够实现更自由地交互，让“所见即所得”的理念在实践中得到进一步的升华。

目前，脑波输入设备主要还是用于帮助残障人士优化交互设计体验，如：利用脑波控制假肢；利用脑波配合眼球检测装置实现打字等操作。但总体设备价格较高，暂时无法实现大面积商品化。

批注 [刘20]: 参考：

https://www.youtube.com/watch?v=FWo12rhz_TQ

,
<https://www.bilibili.com/video/av14995143>。

以陀螺仪为核心的佩戴设备

这类设备通常以陀螺仪为核心，通过佩戴器官（如头、手、胳膊、脚等）的倾斜程度实现不同的信息输入。

这类设备一般作为现有设备的输入补充，在双手都被键盘鼠标占据时，通过头或脚的倾

批注 [刘21]: 参考：

<https://www.bilibili.com/video/BV1ys411H7jd>。

斜实现额外输入。对于残疾人而言，也是增加交互便捷程度的设备。

脚部踏板设备

这类设备最常见的就是用于赛车游戏的油门踏板。这类设备的主要目的是在已有输入设备方案的基础上实现进一步的体验优化。

当然除此以外，也有为无手的残疾人士设计的踏板装置。通过在踏板的四周添加传感器，感应用户的踩踏方向，实现对鼠标的控制。

批注 [刘22]: 参考：

<https://www.youtube.com/watch?v=a1BMblGJp80>。

数位板

数位板也是一种鼠标的强化设备，通常被使用在需要手工做图的场合。

原理

数位板分两部分，一部分是画板，一部分是压感笔。

压感笔的主要结构有两个，一个是用于帮助定位的线圈，一个是用于判断压力的压力传感器。

画板内部有三成电路结构，一层激励线圈，一层 x 轴感应线圈，一层 y 轴感应线圈。在激励线圈中加载特定频率的电流，可以在数位板表面形成磁场。

当压感笔的电磁圈切割磁场时，会在线圈内产生电流（电磁感应原理）。这部分电流会再次产生磁场，此时画板上的 x、y 轴感应线圈通过感应这部分磁场确定压感笔所在位置（但是说实话我也不知道 x、y 轴的感应线圈是如何实现避免被自身发射磁场所干扰的）。

当笔尖触碰画板时，压感笔的压力传感器（应该是个陶瓷电阻）受压力影响产生电流，这部分电流被输入到之前切割磁感线的同一个线圈中，加强线圈产生的磁场信号。画板中的

批注 [刘23]: 参考：

<https://www.zhihu.com/question/20648055/answer/240842835>。

感应线圈根据磁场大小即可确定压力大小。

触控屏

早在 20 世纪 80 年代，其实就有类似触控屏的技术出现了，只不过当时需要额外的输入设备作为辅助。MIT 的伊万·萨瑟兰（Ivan Edward Sutherland）在 1963 年提出了基于光电笔（light pen）和 CRT 显示器的 SketchPad——一种用于计算机辅助设计（CAD）的交互式图形界面。

由于当时的显示器是 CRT 显示器，因此在较短的时间内屏幕上只会有一点存在。光电笔通过将它检测到的屏幕上的点出现的时间发送给计算器，再由计算机根据时间去反向推算出对应点所在的位置，就可以实现光电笔和屏幕坐标的统一。

现代的触控屏或触控板的使用原理与上述的光电笔已有很大不同。现在主流的触控屏都是电容式的，少数产品上还能看到电阻式、红外式、超声波式触控屏的使用。

原理

这些触控屏原理在感知与传感器的触觉传感器部分都详细讲过，这边再简单的重复一下。

电容式

屏幕的下方有两层 ITO（氧化铟锡）导电层，每层都由微小的菱形构成。一层 ITO 的菱形单位都同行相连，另一层 ITO 的菱形单位都同列相连。两层 ITO 层由绝缘介质隔开，通电后形成电容。

当手指靠近屏幕时，屏幕下的电容极板的电容量大小发生改变。检测电路检测到由电容量大小改变导致的电流变化，通过对两极板分别进行逐行和逐列的扫描，确定电容变化位置，

批注 [刘24]: 伊万·萨瑟兰的 SketchPad 参考：
<https://www.bilibili.com/video/av500059729/>。
原理可以参考：
<https://www.youtube.com/watch?v=Nu-Hoj4EjU>。

批注 [刘25]: 参考：
<https://www.bilibili.com/video/BV1DJ411T77Z?from=search&seid=12593938055198260023>。

最终确定触碰位置。

电容屏是现在触摸屏的主流，支持多点控制且价格便宜。其缺点是通常只能用导体作为触碰物，不能用绝缘体。且在手上沾水时，会导致大量触点被同时触发，产生误触。

电阻式

屏幕下方有两层 ITO（氧化铟锡）陶瓷层（应该与上面电容式中提到的 ITO 层的构造一致，即每层都由微小的菱形网格构成，一层行相连，一层列相连），通过千分之一英寸厚的隔离支点将两层分开。将上下两层 ITO 分别接入电路的两端，由于两层 ITO 被空气隔开，无法形成回路。当电阻屏受到按压时，压力使得上层 ITO 中的某个菱形网格被向下发生形变，与下层菱形网格发生接触，形成回路。通过形成回路的行列即可判断触点位置。

电阻屏现在不是主流，只在一些汽车上作为控制屏使用。电阻屏不支持多点控制，但价格便宜。

红外式

红外式分为以空气为介质的和以亚克力板（或玻璃板）为介质的。两种触屏都需要在方形触屏的两条相邻边上装上若干红外发射器，另外两条边上装上若干红外接收器。

以空气为介质的红外屏实际就是使用光线直线传播的原理。当手指靠近屏幕时，会遮挡特定位置的红外线的接收。通过确定未接收到信号的红外接收器，就可确定触碰的位置。

以亚克力板为介质的红外屏则略微复杂，利用了全反射的原理。当光线从光密介质（亚克力）射向光疏介质（空气）时，在入射角足够大的情况下，会发生全反射（即没有折射发生）。假设所有的红外发射器都已临界角想红外接收器发射信号。当手触碰到亚克力板时，触碰点的介质从空气变为皮肤。皮肤的折射率大于空气，这会使全反射发生所需要的临界角增大。也就是说现在的红外发射器发射的角度会导致光线发生部分的折射，产生逃逸。此时对应位置的红外接收器收到的信号就会减弱。依照这一原理，就可以确定触碰发生的位置。

批注 [刘26]: 参考：

<https://blog.csdn.net/qlexcel/article/details/82731055>。

感觉这个的原理其实更像《001_感知与传感器》中提到的开关式触觉传感器。

红外触控屏现在使用的也不多，但一般较大的屏幕会使用红外触控屏，因为价格相当便宜。主要缺点是容易发生误触，易受灰尘污垢的影响，且在室外环境下红外接收器会受到日光的干扰。

声波式

其原理与以空气为介质的红外触屏类似，只不过把红外发射器和红外接收器变位了声波发射器和声波接收器。

声波触屏也不是现在触屏的主流。其优点是，抗暴性强，适合公共场所；反应速度快、性能稳定；受灰尘、水滴等干扰小。其缺点是，体积较大；虽然有较强的抗灰尘能力，但由于通常是室外使用，还是需要较频繁的维护。

压力屏

压力屏是指在对屏幕施加不同大小的压力时，屏幕能感知出这一压力。压力屏大多是在屏幕下方或触控笔笔尖处安置压力传感器实现压力感知的。

屏下压力

屏下压力传感器的一个典型例子就是苹果的 3d touch。3d touch 就是在屏下添加一层薄膜状的电容式压力传感器实现的。通过压力产生的形变改变电容极板间的距离，进而造成电容的变化。通过电容变化导致的检测电路中的电流变化大小，就可以确定压力的大小。

但这种设计似乎因为造价太高、效用不足被苹果很快地抛弃了。

触控笔笔尖压力

还是以苹果的产品为例。apple pencil 能够实现笔尖压力感应，同时根据笔的倾斜制造不同的笔迹效果。这样的功能应该是在笔尖添加了压力传感器，这一压力传感器不仅能检测出笔身方向上的压力，还能感受到笔身倾斜时笔尖受到的侧向压力。apple pencil 将这些压

批注 [刘27]: 参考：

<https://www.zhihu.com/question/67483519>。

别理里面最高票的匿名用户的那个答案，看别人的回答。

力数据进行汇总 ,通过蓝牙传输到 ipad ,实现笔迹变化(侧向压力并不是由陀螺仪实现的)。

具体关于 apple pencil 所使用的传感器我找了很久也没找到靠谱的，所以只能给出一个大致的猜测。这里在简单展开一下关于 apple pencil 的定位原理和敲击笔侧换笔的可能原理：apple pencil 的定位原理应该还是利用了电容屏的定位原理(与手指触碰电容屏定位的原理一样)，与数位板的电磁笔定位的原理并不相同；敲击笔侧换笔的可能原理是在笔身内安置了加速度传感器，用于感知轻微的振动。

屏幕

屏幕是用于显示图像及色彩的设备。由于人类 80%的信息都是通过人眼获得的，因此屏幕是当下交互中最为重要的设备。

基本参数

占屏比

占屏比是指屏幕实际显示面积与屏幕的整体正面面积之间的比值。由于技术原因，屏幕通常无法实现完全无边框（即使实现左右无框，上下的边框还是存在），因此占屏比成为用户对屏幕使用体验（一般指手机）的一条重要影响因素。占屏比越大，屏幕边框越细；占屏比越小，屏幕边框越粗。

由于人类对大小的感知很大程度上受参照物影响，因此有时虽然屏幕的绝对大小很大，但较小的占屏比会使得用户感觉屏幕本身相对过小。以目前的趋势看，提高屏幕占屏比是当下所有手机厂商共同追求的目标。

屏幕大小

屏幕大小即指屏幕的面积（绝对大小）。不过一般对手机屏幕的大小不以面积做单位，而用对角线长度做单位，如 6.1 英寸就是指屏幕对角线的长度为 6.1 英寸（15.5 厘米）。

曾经史蒂夫·乔布斯曾提出，手机的黄金尺寸是 3.5 英寸（即对角线长度为 8.9 厘米），原因在于乔布斯认为这样大小的手机正好能允许用户在单手抓握的同时实现单手操作。不过目前的手机屏幕大小基本上都已经达到 5 英寸甚至 6 英寸。

DPI

DPI，即 dot per inch。DPI 通常用在打印行业，用于表示沿对角线的每英寸长度中所包含的点数。但实际上，DPI 的点是个很宽泛的概念，用到屏幕上，这个点可以指代像素发光单元；用到打印中，这个点可以指代喷墨打印的墨点；用到扫描中，这个点可以指代扫描仪的采样点。

PPI

PPI，即 pixel per inch。PPI 通常用于屏幕显示中，用于表示沿对角线的每英寸长度中所包含的像素（pixel）数。

DPI 与 PPI

在我们进行电子图像处理时，通常以 PPI 做单位。在电脑屏幕上我们通常考虑的是图像的分辨率，图像的每个最小单元就是一个像素点（pixel）。以扫描照片为例，假设一张照片横竖长为 1 英寸，扫描 PPI 为 100，则扫描得到的图片为 100pt*100pt。若扫描大小为 200，则扫描得到的图片为 200pt*200pt。

而在需要进行打印时，通常会使用 DPI 做单位。在打印时，打印机的 DPI 设置一般都是确定的。当我们用 300DPI 的打印机打印一张 300pt*300pt 的照片时，照片上点与点之间的距离是 1/300 英寸；当我们用 600DPI 的打印机打印一张 300pt*300pt 的照片时，照

批注 [刘28]: 参考：
<https://www.zhihu.com/question/23770739>。

片上点与点之间的距离是 1/600 英寸。

如果我们讨论手机屏幕，由于手机屏幕显示的像素就是手机屏幕的最小点（只对于一般的 LCD 屏幕而言；OLED 屏幕的像素点排列并不规则，因此并不满足这一说法），因此 DPI 和 PPI 其实是相同的概念。

手机屏幕的 PPI 计算公式为， $\frac{\sqrt{x^2+y^2}}{z}$ ，其中 x 代表横向分辨率，y 代表纵向分辨率，z 代表屏幕尺寸。 $\sqrt{x^2+y^2}$ 实际上就是用勾股定理求屏幕对角线的像素数。

以 iPhone 12 Pro Max 为例，其像素分辨率为 1284*2778，屏幕大小为 6.7 英寸，因此，按公式计算出的 PPI 为 $\frac{\sqrt{1284^2+2778^2}}{6.7} = 456.77$ ，其公布的 PPI 数值为 458 基本吻合。

另外，人眼视网膜识别的 DPI 上限为 300。因此，通常超过 300DPI 的显示屏幕都可以被称为 Retina（视网膜）显示屏（这一概念是在 iPhone 4 上市的时候被提出的）。

PPD

PPD，即 Pixel per Degree，它指 1° 视角角中所包含的像素数。这一单位一般在 VR 中用来作为评价分辨率好坏的标准。通常，我们认为 60PPD 是达到视网膜屏的标准。将 300PPI 的屏幕放在手机 40 厘米远处，换算为 PPD 大致就是 60。

刷新率

刷新率是指一秒内屏幕图像的刷新次数（正式的说法是电子束对屏幕上的图像重复扫描的次数）。刷新率越高，屏幕在表现连续动画时的自然程度就越高。

现在的屏幕一般都至少是 60Hz 刷新率的，部分有 90Hz 或者 120Hz（一加手机）。刷新率提高在带来更好的体验的同时也会造成更高的能耗。现在一般的手机会根据使用状态对屏幕刷新率进行动态变化，当手机在屏幕内容在变化时，提高刷新率，当手机屏幕内容静止时，降低刷新率。

通常情况下，手机的刷新率一般在 90Hz 就足以满足使用需求。电脑屏幕由于需要满足

电竞比赛的要求，现在普遍的刷新率是 144Hz，部分甚至可以达到 300Hz。

类型

E-Ink

E-Ink 即电子墨水屏，也可称作电子纸显示技术，由 MIT 的约瑟夫·雅各布森 (Joseph Jacobsen) 教授及其团队发明。最常见的商品应用就是 kindle。

国外有团队尝试将电子墨水植入皮肤下方，实现特殊的纹身显示。

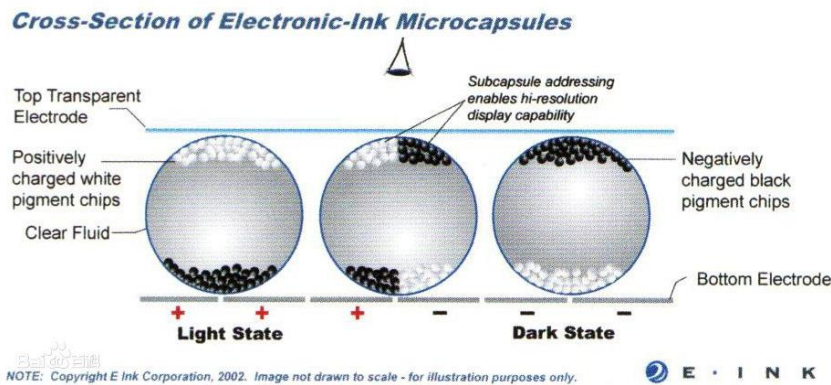
原理

一般的黑白墨水屏的原理如下：屏幕由上下两层基板组成，每层基板上各有一层电极（由许多小电极构成）。基板中间有大量微小的透明胶囊（正式的叫法是微胶囊，micro capsule）。胶囊中有黑色和白色粒子，胶囊内的其余空间用大量液体填满。胶囊中的黑色粒子带负，白色粒子带正电（我也有看到黑色带正电，白色带负电的，这不是重点）。当给上层基板的电极添加负电时，白色粒子就会被吸附到表面，黑色粒子就被吸附到下层，屏幕呈现白色；当给上层基板的电极添加正电时，黑色电子就会被吸附到表面，白色粒子就被吸附到下层，屏幕呈现黑色。通过调整基板上的电极的电荷分布，就可以调整胶囊中粒子的分布，从而实现图像的显示。

批注 [刘29]: 墨水屏厂家：
<https://cn.eink.com/technology.html>。

批注 [刘30]: 参考：
https://www.sohu.com/a/106881197_376384。

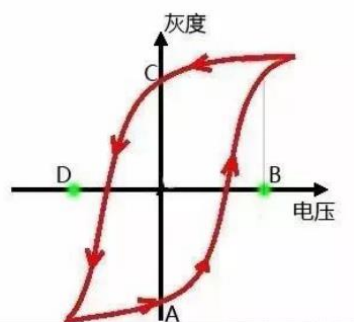
批注 [刘31]: 参考：
<https://www.bilibili.com/video/BV1d4411Y7BD?from=search&seid=15756583826739453666>，
<https://baike.baidu.com/item/%E5%A2%A8%E6%B0%B4%E5%B1%8F/8735573?fr=aladdin>。
磁滞效应参考：
<https://www.zhihu.com/question/20776757>。
彩色墨水屏：
<https://zhuanlan.zhihu.com/p/59457884>，
https://www.bilibili.com/video/BV1Gq4y1B7nZ/?spm_id_from=333.788.videocard.1。



墨水屏原理示意图

当胶囊中的粒子完成黑白分布后,即可切断基板电极上的供电。此时墨水屏中的黑白粒子并不会迅速地重新混合,由于电子墨水具有双稳态(磁滞效应),在一段时间内他们将继续保持之前的分布。

对磁滞效应的解释可以下图。当在电极上施加电压时,灰度变化走的时候 $0 \rightarrow B$ 的、下方的上升曲线。当电极上的电压切断时,灰度变化走的是 $B \rightarrow C \rightarrow 0$ 的上方的下降曲线,消失所需的时间会更长。(网上找到的答案清一色都是这么说的,但我其实挺不理解的。磁滞效应应该是针对磁中性物体的,但黑白色的粒子本身就是的带电荷的,并不算磁中性物体?所以应该是极板通电后产生了磁性?导致对粒子的吸附效果持续了一段时间?)



墨水屏灰度-电压变化图

彩色墨水屏的基本原理与双色墨水屏相同。此时胶囊中有三种颜色的粒子,分别为黄色、

洋红、青色，粒子大小各有差异。通过控制基板上电极的电压大小变化来调整不同颜色粒子的位置，以此实现不同的颜色显示。（具体的原理没找到特别靠谱的解释，上面参考里那个bilibili 的链接稍微提到了一点。）

优点

只需要更换显示内容时产生耗电，功耗低。在智能手表、电子书这样的产品中有广泛的应用。

缺点

刷新率低；色彩鲜艳程度不足；屏幕本身缺少发光能力，黑暗环境下的查看依靠辅助光照。在当前消费者对屏幕要求越来越高的条件下，墨水屏很难在手机这样的大众消费品中得到应用（虽然在 2020 年 CES 上确实出现了海信的彩色墨水屏手机）。

LCD

LCD (Liquid Crystal Display)，即液晶显示器。LCD 在电脑和手机等领域都有广泛的应用，常见的电脑显示器一般都是 LCD（近些年也有 OLED 的），前些年的手机屏幕基本都是 LCD（近些年的一些旗舰机则都是 OLED）。LCD 与后面会讲到的 OLED 目前在手机显示屏中都有广泛的应用，且目前位置任何一方都无法在功能或特性上完全取代另一方。因此，在阐述完两者的各自特性后我还会把两者放在一起做个对比。

原理

LCD 可以粗略地分为 7 层：背光层、垂直偏光片、正极电路、液晶层、负极电路、水平偏光片、彩色滤光片（红、绿、蓝）。

背光层又可以光源、导光板、匀光扩散膜、棱镜膜。背光板的主要目的就是将若干个发光点转化为完整柔和的面光源。背光板的光源只能发出白色的光。

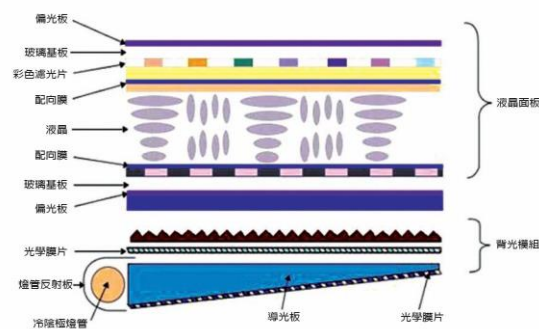
批注 [刘32]: 参考：

<https://www.bilibili.com/video/BV1Sb411W7Zf?from=search&seid=4737478505578821577>，
<https://www.bilibili.com/video/BV1Wz411B7Tf?from=search&seid=4737478505578821577>。

当光线从背光板射出后，会进入垂直偏光片、液晶层和水平偏光片。由于垂直偏光片只会保留垂直方向上的偏振光，因此，如果不存在液晶层直接进入水平偏光片，所有的光线都会被阻挡。此时屏幕会显示黑色。

但由于液晶层的存在，液晶分子的不同排布会使偏振光的方向发生改变（这里的具体原理我没有深入，感觉不用了解地这么详细）。而正极电路和负极电路又可以控制液晶分子的排布方式。因此如果垂直方向上的偏振光经过液晶层并产生了旋转，就可以顺利通过水平偏光片。最终的光强大小与旋转程度相关。

背光层发出的光线经过垂直偏光片、正极电路、液晶层、负极电路、水平偏光片后，呈现出明暗不同的白光。如果在最后通过水平偏光片之前再加上一个彩色滤光片，就能将的白光转变为明暗不同的特定颜色的彩色光。由于任何颜色都可以有红、绿、蓝三种颜色构成，因此将红、绿、蓝三种颜色的滤光片放在一起，配合与三个滤光片面积大小相同的偏光片和液晶层，就形成了一个 LCD 像素点。将大量这样的像素点结构进行排布，就能制成一张 LCD 屏幕。



LCD 基础结构图

优点

由于所有器件都是无机物，所以使用寿命长；由于全屏供光都有背光板完成，因此即使老化也是全屏统一老化，不会出现烧屏现象；制作工艺成熟，价格便宜；LCD 使用 DC 调

光，即通过改变电压调整屏幕亮度，不存在频闪伤眼问题。

缺点

屏幕厚度过大，挤占了手机等小型设备的其他元件的空间，且屏幕过厚导致屏下指纹识别很难实现。

由于光线的明暗是由偏振光实现的，难以做到 100% 的光线遮挡，因此无法做到纯黑色的输出，只是灰度较低的深灰色。这一特性导致 LCD 屏幕很难做到优秀的 HDR 显示。

由于供光是由完整的背光板实现的，因此只要屏幕启动就会持续耗电，无法像 OLED 那样通过只亮起屏幕中的部分像素来减少耗电（在智能手表中这一点尤其明显）；由于背光板不易发生弯折，因此 LCD 无法在手机这样的设备上实现曲面屏；在屏幕与边框的衔接处，背光层的背光很容易发生泄漏，在屏幕显示黑色画面时出现漏光现象。

液晶层的偏转决定了光线的明暗和一个像素点中具体颜色的显示。LCD 屏幕中的液晶层在低温条件下偏转速度慢，导致像素点转换速度慢，屏幕在显示动态图像时会出现明显的“拖影”。

虽然 LCD 屏幕不存在频闪伤眼的问题，但由于背光板中实际的发光部件是 LED 灯珠，这些 LED 灯珠会发出对人眼产生损害的高能蓝光。因此 LCD 屏幕对人眼也同样存在一定的伤害。

OLED

OLED (Organic Light Emitting Diode)，即有机发光二极管。它是近几年兴起的屏幕技术。在 OLED 技术的加持下，屏幕（不论电脑还是手机）能够实现更加优秀的 HDR 显示；手机能够实现更小的屏幕边框、柔性屏等等新形式。它的优点即为显著，但同样的，它的缺点也一样十分突出（后面会展开讲）。

批注 [刘33]: 参考：

<https://www.bilibili.com/video/BV1Wz411B7Tf?from=search&seid=4737478505578821577>。

关于 AMOLED 和 OLED 的区别可以看：

<https://www.zhihu.com/question/47064399>，
<https://www.zhihu.com/question/47064399>。

OLED 屏幕还可细分为 AMOLED(Active Matrix OLED)和 PMOLED(Passive Matrix OLED)。一般我们说的 OLED 都是指 PMOLED，即被动矩阵 OLED。而三星的一些手机会使用 AMOLED 即主动 OLED。两者的区别在于 AMOLED 采用的 TFT(Thin Film Transistor，薄膜晶体管)阵列来实现像素的显示控制(简单的说就是 TFT 就是控制电路)，而 PMOLED 采用阴极条和阳极条控制。相比之下，AMOLED 的屏幕响应速度更快，耗能更低。

原理

OLED 的基础发光元件是一个发光二极管，发光的亮度由通过的电流大小决定。每一个像素由三个分别为红、绿、蓝色的发光二极管构成。将这样的像素点拓展到整个屏幕，就是 OLED 屏幕。

优点

由于不像 LCD 那样有多层元件，OLED 的屏幕很薄，这为在手机中添加更多元件提供了可能。

由于发光元件是单个的二极管，因此 OLED 可以实现在屏幕上存在大量黑暗区域时，只亮起明亮区域二极管，大大减少了能耗；由于发光的强弱也直接由二极管控制，因此 OLED 屏幕的响应时间要比 LCD 短，不会产生明显的“拖影”；在需要显示对比度高的图片时，屏幕中的暗区显示可以通过直接降低二极管亮度来实现优秀的 HDR 显示效果。

由于发光元件不再是单独的一块发光板，因此 OLED 屏幕可以实现大幅度弯曲，进而使柔性屏、折叠屏、超窄边这样的设计得以实现。

缺点

OLED 的发光二极管是有机材料制成的，二极管频繁的电子迁移会加速有机材料的老化。因此 OLED 屏幕的寿命普遍较短。

另外，由于像素点都是独立发光的，且通常屏幕的特定区域显示的内容是不变的，这就

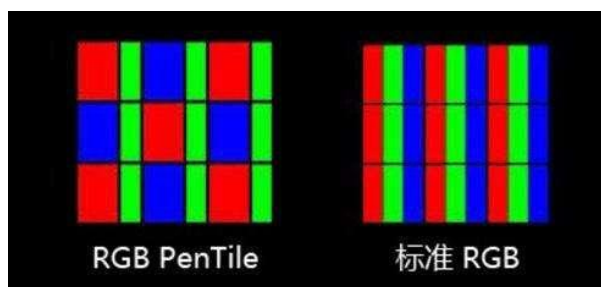
批注 [刘34]: 这边简单补充一下超窄边实现的原因。一般手机会把控制屏幕的电路放在手机屏幕的靠近手掌的那一侧，这就导致手机通常都有较大的“下巴”。而在 OLED 的条件下，屏幕可以通过折叠在底部实现一个 U 型的构造，将控制电路折到屏幕的正下方去。这样就能彻底消除“下巴”，实现超窄边。（来源与上面给出的参考视频）

导致不同区域的红、绿、蓝色组件老化程度不同。当之后需要显示其他颜色时，发光组件的老化程度不同，颜色显示也会出现偏差，即所谓的“烧屏”。

OLED 采用的调光方式是 PWM (Pulse Width Modulation，脉冲宽度调制)，通过占空比控制亮度。PWM 就是指以很快的速度开关屏幕，当频率很高时，人眼就会觉得屏幕是在一直发光（有点类似以前的 CRT 显示器）。当一个周期内开屏幕的时间较长时，整体显示就较亮；当一个周期内开屏幕的时间较短时，整体显示就较暗。由于 OLED 屏幕易老化，不能采用高频 PWM 调光，只能用低频 PWM 调光，因此这种低频的频闪很容易对人眼产生损害。

之前我说到 OLED 的发光基础元件是红色、蓝色、绿色的三个发光二极管构成一个像素的，但实际应用中并非如此。由于不同颜色的发光二极管老化程度不同及像素点种植成本问题，OLED 通常采用 Pentile 排列或 Delta 排列（也有人叫这个周冬雨排列或者小黄鸭排列）等非常规排列方式。在这样的排列方式下，实际像素与理论计算的像素值存在差异。这样的排列方式对数据显示也会产生一定的影响，在分辨率不高的情况下这种影响会产生明显的锯齿感。

批注 [刘35]: 参考：
<https://zhuanlan.zhihu.com/p/109841385>。



Pentile 排列与一般 RGB 排列

OLED 与 LCD

特点比较

	LCD	OLED
发光方式	依靠统一的发光背板供光，每个像素点只是偏振层、液晶层、滤光层的组合。较大的发光单元导致屏幕无法实现弯曲、折叠；且屏幕边缘可能出现漏光。	每个像素点自身就是独立的发光元件。较小的发光单元使得屏幕可以轻松地弯曲折叠；且不存在漏光的情况。
厚度	1.5-3.3mm，较厚	<0.5mm，较薄
像素点变化速度（拖影）	低温较严重	任何情况都保持良好
平均耗电	由于发光板需要持续亮光，耗电量较大。	由于发光板有时只许部分亮光，耗电量小。
使用寿命	元件都是无机材料，使用寿命长。	元件都是有机材料，且频繁地发生电子迁移，使用寿命短。
是否存在烧屏问题	不存在。只有一块发光板，因此全屏幕统一老化。	存在。不同发光单元老化速度不同，会导致烧屏。
调光方式	DC 调光，不存在频闪问题。	PWM 调光，存在频闪问题。
像素排布方式	一个像素点由一蓝、一红、一绿 3 个基础单元构成。	一个像素点可能由不同数量的蓝、红、绿基础单元构成。具体由使用的排列方式决定。像素点数量通常是 LCD 屏幕的 80%。

技术优化

经过之前的比较,可以发现 LCD 和 OLED 各有优缺点。于是有人提出,是否可以将 LCD 的发光单元缩小,与 OLED 一样,用三个极小的无机 LED 灯管组成一个像素,在铺展成一个屏幕?这种技术就是 micro-LED 技术。但目前这一技术尚未突破量产的技术壁垒,因此暂时只在实验室中存在;即使出现产品,价格也会偏高。之前听说 2020 年发布的 ipad 会采用 micro-LED 技术,结果并没有。

现在也有类似的过渡技术,即 mini-LED,即将原来 LCD 一整块的背光层进行切割,变为一个发光阵列(若干个发光点),根据屏幕不同区域的显示内容控制发光阵列中对应发光单元的亮度。但这一设计的缺点也很明显:由于并没有取出 LCD 中的液晶层,因此只能起到提高局部对比度的作用,并不能降低屏幕实际厚度,甚至可能增加;在显示小范围高亮、高对比度的图像时,周围会出现明显的光晕;发光阵列发热明显,甚至需要独立风扇进行主动散热。因此,这一技术目前基本只应用在显示器中,且只会充当一种过渡技术使用,并不会一直存在。

柔性屏与折叠屏

现有的柔性屏或折叠屏基本都是以 OLED 作为显示基础的(但似乎也有 LCD 柔性屏,CCTV10 还报道过)。两者虽然相似,但仍然存在差异。柔性屏(flexible)通常只要求能够实现弯曲;而折叠屏(foldable)对折叠次数有特殊要求,因此其基板材料、封装材料、触摸屏材料、盖板材料都与柔性屏有所差异。这里我不会过多地在两者的原理差异上进行展开。通常柔性屏支持曲面摆放;折叠屏则较多地使用在会发生刚性折叠的场合,如折叠手机中。

柔性屏和折叠屏通过在高分子材料制成的可折叠的屏幕框架中放置独立发光单元,实现了折叠与显示的统一。其实现依赖于独立的 OLED 发光单元、柔性基板与封装材料、柔性盖

批注 [刘36]: 参考:

<https://www.zhihu.com/question/313609236>,
<https://www.zhihu.com/question/313542802>,
<https://zhuanlan.zhihu.com/p/57585068>。

批注 [刘37]: 参考:

<https://www.bilibili.com/video/BV1c54y1d7WU?t=446>。

板。

此外，现在折叠屏使用的铰链也是技术的关键之一。它必须除了实现折叠、打开的功能外，还必须提供一定的阻尼，使得折叠手机在不同翻折角度下都能保持翻折状态。

折叠屏现在主要可以分为两类，这一种是外折叠，以华为的 Mate X 为代表；一种是内折叠，以三星的 Galaxy Fold 系列为代表。（我个人认为外折叠使用体验会更好一些，但由于脆弱的屏幕被直接暴露在外，很容易造成磨损擦伤等问题。）



三星 Galaxy Fold



华为 Mate X

缺点

这里先讨论折叠屏的缺点（基本都是以折叠手机评价对象的）。作为一个新兴的屏幕技术，折叠屏确实还显得十分不成熟。

首先是折叠中心的折痕问题。这一折痕在三星的 Galaxy Fold 中表现的非常明显，但在华为的 Mate X 中似乎没有那么明显。

批注 [刘38]: 优缺点都参考：

<https://www.bilibili.com/video/BV1iE411h7oc> ,
<https://www.bilibili.com/video/BV1WA411j7Kk> ,
<https://www.bilibili.com/video/BV1RJ41127ui?from=search&seid=5194501818089578745> ,
https://www.bilibili.com/video/BV1aJ411C7tb/?spm_id_from=333.788.videocard.0。

其次是软件的适配问题。折叠屏通常会有折叠、展开两种使用状态。这就要求应用对两种状态给出不同的交互界面设计。然而由于目前的折叠屏手机使用人群仍然有限,大部分应用无法支持双界面的软件设计。此外,展开后大屏幕的打字输入也成为了一个较为棘手的交互问题。

第三是屏幕及摄像头排布问题。这一问题针对三星的 Galaxy Fold。当折叠屏打开时,前置摄像头的位置会显得十分尴尬。它严重影响了用户对一整块完整屏幕的期待,同时影响了展开模式下用户的基本操作。但这一问题在 Mate X 中就得到了较好的解决。

第四是屏幕过于脆弱的问题。由于屏幕需要折叠,因此使用的材质通常都是塑料材质而非玻璃材质,这就导致折叠屏幕很容易出现刮痕。

综上,目前来看,折叠屏在折起时,是一个加厚的、略微缩小的手机;在展开时,是一个带折痕的、略为缩小的、功能不完善的平板电脑。目前来看,折叠屏技术显然很不成熟,但似乎也不能说其独特的交互方式在未来一定一无是处。

优点

优点这部分的内容似乎都是针对 Mate X。简单地讲折叠手机的优点就是,折叠手机为手机与平板电脑建立了桥梁。

首先,折叠手机的交互方式充分利用了展开后“手机变平板”的思想,引入了平板电脑中的多线程同步工作的功能。这使得折叠手机真正地实现了功能上的飞跃。

其次,折叠手机的大屏幕对图像及视频展示带来了显著的效果提升,用户能够在不进行缩放或只进行少量缩放的情况下看清图像的细节,实现对细节与整体的双重把握。

综上,从使用体验上看,折叠手机最大的卖点无疑就是它接近平板的屏幕大小。因此,可以预见的是,在未来,折叠手机和平板的差异会逐步缩小,甚至最终融合。

透明显示屏

目前暂时没有用透明显示屏作为屏幕的手机，这类技术基本只用在电视领域。

透明显示屏包括基于 LCD 技术和基于 OLED 技术的两类。

透明 LCD

基于 LCD 技术的透明显示屏其实就只是把 LCD 中的偏振层、液晶层和滤光片构成的显示层单独取出，然后配备额外的后置光源实现显示，并没有特别不同的新技术。由于 LCD 透明屏还是需要后置光源，因此并不能实现理想中的那种独立的透明屏幕。

我看到的应用是将 LCD 透明屏作为电脑主机机箱的侧面板，然后在机箱内部加强光实现显示。此外索尼爱立信在 OLED 屏幕不成熟时开发过一款透明 LCD 手机 Xperia Pureness。

有一点需要注意，我们平时看到的那种在镜子上面显示东西的屏幕实际是使用了单面镜，通过将发光的 LCD 屏幕放在单面镜后面，即可实现那种悬浮在镜子上的显示。它并没有使用到任何透明显示的技术。

透明 OLED

基于 OLED 技术的透明显示屏的基本原理实际上和纱窗有点类似。简单地讲就是在 OLED 发光单元之间留出了透明区域，使得后面的光线能够穿透过来，实现透明。因此，透明 OLED 屏幕的主要技术难点在于，将 OLED 发光单元的驱动电路精确地隐藏到发光单元的后方。从目前来看，现有的透明屏幕制作厂家会为每个像素点配备占像素点大小 40%-50% 的透明区域。

目前从事开发透明 OLED 屏幕的厂家主要有 LG 和小米。其中小米已经推出了透明电视，49999 元，1080P，从显示器的性价比角度看显然略差，同时从另一个角度也看出，目前透明显示屏的商业化技术还欠佳。

批注 [刘39]: 透明 LCD :

<https://www.youtube.com/watch?v=Shp5JhOqdKI>。

透明 OLED :

<https://www.bilibili.com/video/BV1at411z7Zm> ,

<https://www.bilibili.com/video/BV1ca4y1s7qf> ,

<https://zhuanlan.zhihu.com/p/189444183> ,

https://weibo.com/ttarticle/p/show?id=2309404539221779284220#_0。

镜面上的屏幕制作 :

https://www.sohu.com/a/17700961_105527。

应用

透明显示屏显然在目前看来并不是一个用于日常工作或生活的设备,它更像是为了更优化的视觉体验而制作的奢侈品。单纯地以功能去对其进行评判似乎并不公平,透明显示屏为全新的视觉展示方式提供了一种可能。

目前对透明显示屏的具体应用大多集中在商业产品展示领域,通过将产品展示柜的前置玻璃更换成透明显示屏,是产品本身及产品的相关信息能够同步地呈现。

此外,北京地铁中也部分使用了 OLED 屏幕代替小型玻璃窗,在展示站台信息的同时透视窗外的实景。

在未来透明显示屏能够实现量产后,大量的透明显示屏组合将可能替代投影仪,在电子艺术中实现更优化的沉浸式体验;透明 OLED 也可能与汽车结合,通过在前挡风玻璃上添加透明屏幕,实现以前所设想的 AR 驾驶交互。

GUI 交互算法

根据触碰轨迹实现手势识别的算法

这类算法最常见的似乎就是用在浏览器的鼠标手势上。似乎有一款叫 magic touch 的游戏也用到了这个技术。

Rubine 算法

Rubine 算法是通过分析触控形成的轨迹的几何特征（起始角的正弦值、余弦值，手势总长度等）来判断输入手势与手势库中的手势的对应关系，并执行操作的。

\$1 算法

\$1 算法会先对输入的手势轨迹进行重新采样，将轨迹整体平滑化。然后将轨迹旋转、平移、缩放以实现标准化。最后将轨迹与现有轨迹特征集进行比对，确定轨迹对应的具体手势。

批注 [刘40]: 参考：

https://blog.csdn.net/weixin_42595525/article/details/89400859。

批注 [刘41]: 参考：

<https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CMFD&dbname=CMFD2007&filename=2007060011.nh&v=H%25mmd2BPhRzJIBLrz6RD7UimYC1zkqYMcZaWqm3q%25mmd2BjSHIOkLvEKNiMh4uxQ2NHykVWqZE>。

其中有的 2.3.2 节提到了 Rubine 算法的内容。

Rubine 本人的演示和介绍：

<https://www.youtube.com/watch?v=RdMUt0VHIP8>

。

批注 [刘42]: 参考：

<https://www.zhihu.com/question/29238666>。

体验\$1 算法：

<http://depts.washington.edu/accelab/proj/dollar/index.html>。（上面还有 JS 或 C#语言的源码）

原始论文：

<http://faculty.washington.edu/wobbrock/pubs/uist-07.01.pdf>。

语音交互（VUI）

一般的 VUI 实现需要 3 步：语音识别（STT/ASR）->自然语言处理（NLP）->语音合成（TTS）。其中很多内容我在之前的感知与传感器中曾经提到过（如果有冲突，以这个为准，因为这个是更后面写的）。

这里有一点要补充。NLP 又可以被分为 NLU 和 NLG（具体在 NLP 中会讲两者的意思）。NLU 和 NLG 与 STT 和 TTS（文字转语音，下面会讲）并没有特别明确的区分。NLP 本身也是可以直接以语音作为输入，然后在处理过程中实现信息的结构化，最终做出反馈。并非说 NLP 必须要求先将语音通过 STT 转化为文字，再将结果用 TTS 转为语音。我个人整理时，STT 和 TTS 中会主要着重如何对语音进行处理，NLP 中则是简单将如何理解（不深入，主要是因为我对 NLP 的了解也有限）。

目前的 VUI 只能作为交互的补充，还不能完全成为一种独当一面的交互方式。

预处理

这部分内容在“感知与传感器”中深入讲过，这里重复一下。

回声消除

这里的回声消除一般指声学回声消除。声学回声的产生是由于扬声器中的声音再次进入了麦克风。因此，其消除手法就是将之前输入的声音音频记录，然后进行波形反转，再叠加到输入的音频中，从而实现抵消的效果。

波束成形 (beamforming)

波束成形的实现需要麦克风阵列。麦克风阵列在接收来自特定位置的声音信号时,声音的音频在波形上会存在相位上的差别。为了保证最后输出的音频的统一,需要对每个麦克风进行一定的相位调整。这样的调整会导致来自之前位置的声音会被进一步地增强(相位调整对齐后的波峰与波峰相加、波谷与波谷相加),同样也会导致其他位置的声音被显著地减弱(相位调整后的波峰与波谷相加、波谷与波峰相加)。如此一来,就实现了信号的定向接收。

噪音消除

噪音消除的原理与回声消除类似。通过频率特征分离出传入声波的噪音,然后将这部分声波反转,叠加到之前的声波中,就实现了噪音的消除。

端点检测

有些语音识别技术会提到“端点检测”(也可以叫语音活动检测, Voice Activity Detection, VAD)。那其实是一个对语音输入声波的预处理,用于准确找出输入声波中的每次语音的开始点和结束点,去掉静音和噪音的部分。

批注 [刘43]: 参考:

<https://zhuanlan.zhihu.com/p/141194283>,
<https://zhuanlan.zhihu.com/p/180123232>。

语音识别 (STT/ASR)

STT (Speech To Text) 或 ASR (Automatic Speech Recognition), 即语音识别。语音识别可以分为四个过程: 特征提取、声学模型、语言模型、解码搜索。注意, 这里提到的是使用隐马尔可夫模型 (HMM, 在深度学习出现后基本就不提了) 实现的语音识别。

批注 [刘44]: 参考:

<https://easyai.tech/ai-definition/asr/>,
<https://www.bilibili.com/video/BV1qs411P7Ur?from=search&seid=4244499031011170279>,
<https://www.zhihu.com/question/20398418>。

实现过程

特征提取

特征提取是指将输入的语音波形按特定的时间长度分为若干帧,随后对每一帧信号进行傅里叶变换并将变换后的波形转化为一个若干维的向量。这个向量中的每一个维度就是一个特征。当对每一帧的信号都进行提取转换后,就会得到一个特征矩阵。

声学模型

当我们获得了特征提取得到的特征矩阵后,我们希望将若干帧进行组合,形成状态;然后将若干向量组合,形成音素;最后将音素组合,形成词汇。这里就出现了一个问题,如何对帧进行组合形成状态,如何对状态进行组合形成音素。如果组合的跨度过小,可能导致每个合成的音素都不是真正意义上的完整的音素;如果组合的跨度过大,可能导致部分音素被忽略。

解决的方法是使用隐马尔可夫模型,通过创建一个庞大的声学模型,将各种可能的单词展开变成音素,音素展开变成状态。在需要进行识别时,通过这个声学模型确定待处理语音的状态及音素的组合方式。最终将零散的帧组合成明确的若干个音素。

语言模型

在得到输入语音的音素组合后,既可以根据音素大致推断出可能的单词组合了(单词的音素组合都是确定的)。此时需要通过语言统计模型来确定这样的声音组合会对应什么样的文字,排除概率较小的组合。语言统计模型实际上就是判断某一个句子可能出现的概率高低,在解码过程中从语言层面上限制搜索路径。

解码搜索

其实解码搜索并不是一个完全单独的过程。之前在声学模型和语言模型中实现对音素确定、组合及具体词汇的确定的过程就是解码搜索。完成整个解码搜索的过程后,就实现了从

输入的语音波形到文字的转换。

唤醒词 (KWS , Key Word Spotting)

在我们日常使用语音交互设备时，通常都需要用唤醒词唤醒设备，才能开始交互。

语音交互设备其实是时时刻刻都处于录音状态的。由于我们日常生活中会发生很多对话，大部分都是与语音设备无关的，因此设备并不会启动。

但语音设备实际上时时刻刻在比对我们的语音输入与唤醒词是否相同。这种比对很可能就只是简单的波形比较（有些时候只需要“hey si”或“ey siri”就能够唤醒 siri，并不需要完整念出“hey siri”），并不需要复杂的识别过程。从这一点看，用户的语音中所包含的隐私信息并没有被实时识别并记录。

一般语音设备的唤醒词都是不允许自行设定或更换的，原因就在于一旦发生更换，唤醒词的波形就会改变，原先设计好的高鲁棒性的波形比对模型就会失效。

自然语言处理（NLP）

简介

我们知道，如果希望计算机能够处理特定的信息，我们必须保证输入的信息是结构化的，是符合计算机程序要求的。然而在我们身处的这个互联网时代，大约只有 20% 的信息是经过结构化处理的。也就是说，剩余的 80% 的信息都是无序的、以自然语言的形式存在的。因此，如果我们希望利用这 80% 的信息，将其从自然语言状态转化为结构化的信息，就是其中最关键的一步。这种将自然语言转化为结构化信息，再根据这些信息做出进一步反馈的过程就叫做自然语言处理（NLP，Natural Language Process）。

批注 [刘45]: 参考：<https://easyai.tech/ai-definition/nlp/#steps>，<https://kns.cnki.net/KXReader/Detail?TIMESTAMP=637406240057344063&DBCODE=CJFD&TABLEName=CJFDLAST2019&FileName=SDDZ201907054&RESULT=1&SIGN=UDAm1vPKASKRW%2bE%2bJxTG7iYPSyk%3d>，

NLP 有两个核心的任务：自然语言理解（NLU，Natural Language Understanding）和自然语言生成（NLG，Natural Language Generation）。自然语言理解着重于将无序化的语言进行结构化的处理；自然语言生成着重于根据结构化的数据执行反馈。当不做任何说明时，NLP 一般是指 NLU，即理解语言这一过程。

有一点值得注意的是，NLP 不止可以应用于语音输入的自然语言，也可以应用于一般的以文字形式输入的自然语言。

应用方向

NLP 的应用都依赖于对输入的无序化数据的结构化处理。NLP 有三个典型的应用：信息摘要与分析、机器翻译、自然语言人机交互。信息摘要与分析指对结构化的数据进行进一步的提炼与概括（如文章的自动摘要和情感分析）；机器翻译指将一种语言构成的句子或文章翻译成另一种语言，实际上就是根据获得的结构化数据，用另一种语言实现意思上的重构；自然语言人机交互指计算机基于提炼出的结构化信息执行具体的预置操作（智能语音助手），同时，自然语言人机交互还能实现联系上下文对话内容。

NLU

早先的实现 NLP 的方法主要是机器学习，而近年来，主要的实现方法都是深度学习。两者的相同点在于，两者都必须对输入的文字信息进行预处理。两者的不同点在于，机器学习需要由人给出特征，结果的有效性很大程度上依赖于人对数据表示和输入特征的设计；而深度学习则可以自动完成输入表示和特征提取工作。（这部分内容我自己了解地也不深入，所以选择性地看看就好。）

预处理

由于中文和英语的语言特性不同，两者在 NLP 的预处理中所需要进行的程序也有所不

批注 [刘46]: 一般别人还会把语音识别也加进去，但我感觉语音识别应该只通过 ASR 就能实现，并不需要 NLP。
不过我对应用的概括也只是通过简单的文献阅读得出的，不一定准确。

批注 [刘47]: 参考：<https://upload-https://www.cnblogs.com/mantch/p/11385113.html>。

批注 [刘48]: 参考：<https://www.youtube.com/watch?v=5ctbvkAMQO4>，。

同。英语中的预处理可以分为分词 (Tokenization)、词干提取 (Stemming)、词形还原 (Lemmatization)、词性标注 (Parts of Speech)、命名实体识别 (Name Entities Recognition)、分块 (Chunking), 共 6 步。中文的预处理可以分为分词、词性标注、命名实体识别、去除停用词, 共 4 步。

我会简单解释一下各种预处理所做的工作是什么。

分词 (Tokenization)

分词是将句子拆分成以字词为单位的结构, 方便后续的分析处理。比如将“北京欢迎您”拆分为“北京”、“欢迎”、“您”。

分词需要注意寻找合适的最小单位 (粒度)。如果粒度过小, 会使得原本应该合成在一起的字被拆分, 如“鼠标”被拆为“鼠”、“标”; 如果粒度过大, 会导致词本身的意思太复杂, 没有起到拆分的目的。

英文的分词相对中文要更为简单。因为英文的单词本身就是各自独立的, 单词与单词之间不需要过多担心组合问题; 而中文由字组成词, 字与字之间的组合问题就十分复杂。

词干提取 (Stemming) 与词形还原 (Lemmatization)

这部分是英文中才特有的预处理, 原因在于中文并没有复杂的词态的变化 (当然这也带来了词性标注上的困难。)

词干提取就是指通过去除前缀或后缀实现获得词根的过程。如将“going”转变为“go”。

词形还原是指基于词典, 将单词的复杂形态转化为基本形态的过程。如将“went”转变为“go”。

词干提取和词形还原本质都是为了将句子中形态不同的词转化为其最基本的形式, 便于之后的处理与分析。

词性标注 (Parts of Speech)

将句子中的不同单词标注为其对应的词性。

中文的词性标注是一个相对困难的问题。由于中文不存在词态变化，因此中文的词性标注无法直接从词的形态上来判断。而且中文中常用词兼并现象很严重，即常用词常常会不只有一种意思，或不止有一个词性。如，“把”字，在“把手正过来”这一句话中，既可以充当动词，以“手”为宾语；也可以和“手”字一起构成“把手”，充当名词短语。

命名实体识别 (Name Entities Recognition)

命名实体识别也可称为专名识别。一句话中可能会出现一些平时不常见的专用名词，如果将这一个单词分开理解，可能无法正确地理解它的意思。因此需要用命名实体识别专门将这些词汇识别出来。

分块 (Chunking)

将已经确定词性的、独立的词组再进行合并，组成单元更大的句子成分。如，将“一只”（量词）、“美丽的”（形容词）、“小鸟”（名词），组合形成“一只美丽的小鸟”（名词短语）。

去除停用词

去除对文本特征没有实际贡献的字词，如标点符号、语气词、部分人称等。

特征工程

特征工程又可分为特征选择和特征提取，目的是将经过预处理的文字转换为一种计算机能理解的形式（通常为向量）。特征选择有若干种方法，它将决定文本将以什么样的规则被转化为向量。对于后续的不论是机器学习还是深度学习而言，特征工程都将深刻地影响输出结果。

常见的特征选择方法有 DF、MI、IG、CHI、WLLR、WFO。常见的特征工程模型有词

批注 [刘49]: 参考：

https://blog.csdn.net/qq_27713281/article/details/72818848 ,

<https://mofanpy.com/tutorials/machine-learning/nlp/intro-search/>.

袋模型和词向量模型（上面的参考链接中有简单介绍）。

基于机器学习或深度学习

在实际应用中，两者实际上在实际使用时都只要输入特征矩阵就能直接输出结果（不过好像对深度学习而言，已经有直接端到端不需要特征提取的网络了）。因此两者的重点都在如何选择学习模型并对模型中的网络进行训练、提高准确率上。

适用 NLP 的机器学习模型有 KNN（最近临近节点算法）、SVM（支持向量机）等，深度学习模型有 CNN（卷积神经网络，使用 CNN 进行语言文字处理需要将语音或文本转化为类图片的格式）、RNN（循环神经网络）等。为了更好地实现上下文理解，可能还会用 LSTM（长短记忆网络，是为了解决 RNN 中过远距离的上下文无法正确反映正确相关关系的问题的）等。

神经网络最终的输出是由功能决定的。以基本的神经网络为例，如果我希望一个神经网络帮我检查大量的微博，然后归纳每一条所表达的感情，那么我的输入就是这些微博文字，输出就是预先确定的若干种感情（如开心、难过等）。在之前训练这个模型时，我会先准备好很多标注了最终应该属于什么感情的微博文字，让模型学习。当完成训练时，就能实现放入文字，得出对应感情。

上下文理解

如果需要 NLU 实现上下文理解（如在翻译或语义理解中），通常需要使用 RNN，即循环神经网络。循环神经网络在每次进行隐含层的传递时，除了考虑上层隐含层的参数外，还会考虑先前数据带来的影响，以此实现意思在上下文的传递。但由于 RNN 本身的设计问题，前后文的关系可能无法很好的传达，出现梯度消失或梯度爆炸的问题。为此，LSTM（Long Short Term Memory）被设计出来，用于解决这一问题（不展开）。

不过有一些时候在我们命令语音助手执行某些特定工作时，他们似乎也能逐步要求我们

批注 [刘50]: RNN 和 LSTM 都可以参考：
<https://www.bilibili.com/video/BV1Vx411j7xF?t=224>。

给出它们需要的信息，看似也实现了上下文的理解。然而那种前后情景的结合实际上是基于预先写好的程序的，并不是真正意义上的实现上下文理解。

全双工

语音识别中的全双工就是指用户与设备的语音交互是双向的、持续的（上下文有联系的）、可打断的。目前来看，全双工的实现还存在诸多困难，比如我们很难保证在一般情况下语音和我们的交互能根据我们上文提到过的内容给出不同回答；比如目前实现的语音交互在每次和设备对话前都需要主动触发唤醒词或点击特定按钮；比如设备无法区分在用户和它交互的同时与其他人进行了对话时，哪些内容是不该被记录的。因此，全双工只能说是未来语音发展的一个目标和方向。

NLG

NLG 实际上可以理解为将结构化的信息进行重新组织的过程。较为常见的 NLG 包括，利用结构化的信息生成自然语言文字（自动生成新闻）；自动根据用户输入判断回复（聊天机器人）等。

感性理解

我个人感觉写完了上面的内容之后，对 NLP 深入内容的迷惑使我并不能很好地理解 NLP 本身。这边我想借用《人工智能导论》（李德毅主编）p178 给出的一个对话理解的例子来帮助理解什么是 NLP。

我个人认为从感性角度理解，NLP 就是包括把数据结构化，再利用结构化数据做出反馈的过程。下面将给出《人工智能导论》中的例子。

假设我有一个智能助手，现在我对它说：“我想预定明天下午三点在王府井附近的全聚德烤鸭店”。

在听到这句话后，智能助手会将这一输入判定为一个命令（而非普通的闲聊），进而分析这句话的内容。分析内容后，它会判断出我要做的是餐厅预订，因此执行预先设置的餐厅预订的若干流程。

假设预设的餐厅预订流程需要用户提供时间、地点、具体餐厅名称、用餐人数这四个信息，人工助手会将我说话中的“明天下午三点”、“王府井”、“全聚德烤鸭店”都“填入”要求的栏目里。

然后，人工助手会发现我并没有给出用餐人数的信息，因此会触发询问的输出指令。此时就需要用生成自然语言的系统构建一个问句，他可能会问我：“请问有多少人前来就餐呢”。当我给出“4人”的回答后，它会重复之前的步骤，将信息填入之前的空缺栏目中，并继续执行后续的指令。

以上就是我简单概括的《人工智能导论》中给出的一个语言交互的案例。我个人认为，NLP 这个内容如果要深入了解的话会很难，但对于我们而言可能只需要一个感性的认识，知道它在语言处理的过程中能扮演什么样的角色就可以了。

文字转语音（TTS）

TTS (Text To Speech)，即文字转语音。文字转语音的重点在语音合成上，目前的语音合成有两种思路，一种是基于拼接的语音合成系统，一种是基于参数的语音合成系统。

基于拼接的语音合成系统

通过大量录制语音创建语音库，在需要合成时，通过文本中含有的音节、音素去库中搜索对应的音节、音素的发音，然后拼接成一段完整的语音。

其优点是，由于语音全部有真人录制，因此语音质量高。

批注 [刘51]: 参考：

<https://www.zhihu.com/question/26815523/answer/220693948>。

其缺点是，由于商用级别的语音库至少需要 5 万句，因此成本比较高。

基于参数的语音合成系统

确定文本对应的声学特征，通过声码器把这些特征参数转化为波形，从而获得语音。

其优点是，对数据库要求较小，大约只需 5000 句即可，成本较低。

其缺点是，语音质量相对较差。

端对端的语音合成

与语音识别一样，现在一来深度学习一样可以实现用音素或字符直接合成语音的结束。

2016 年，谷歌的基于深度学习的 WavetNet 语音生成模型，可以对经过预处理的文本进行直接的原始语音数据建模，避免了对语音进行参数化时的音质损失。同样是属于谷歌 Tacotron 可以在不对文本进行预处理的前提下，直接从字符合成语音。两者的最终合成效果都非常优秀。

体感交互

体感交互主要可以区分为手势交互和姿态交互。

手势交互

手势交互的重点在于实现从复杂的环境信息中将手势进行分离。在完成分离后，对手势进行处理（特征提取），之后可以用模板匹配、隐马尔可夫模型或神经网络（现在显然都是用神经网络的）实现手势的匹配。

手势交互可以根据其实现原理分为惯性传感器方案、可见光方案、深度信息方案、肌电

方案、毫米波雷达方案等。

惯性传感器方案

这一类的手势识别是别的大多是手掌或手臂移动形成的手势。用户需要佩戴内置惯性传感器的设备,在进行手势动作时,惯性传感器内部的加速度传感器会记录手掌或手臂的运动,从而判断出大致的运动轨迹。根据这个运动轨迹,与手势库中的手势进行比对,即可实现手势判断。

还有一类与此设备类似的设备,即之前提到的隔空打字键盘那样的设备。通过在手指上套上惯性传感器,根据手指做出手势时与刚性平面的碰撞来确定发生碰撞的手指,从而确定手势。

此外,类似 wii 这样的游戏手柄或一般的手机这种通过内置惯性传感器来检测转动或振动的设备,也可以被认为是基于惯性传感器方案的手势交互。

可见光方案与深度信息方案

可见光方案和深度信息方案(深度信息的获取方式去参考“感知与传感器”中的立体视觉部分)其实是类似的,两者的基本流程都是先将手从环境中提取出来,然后持续跟踪被分离的手,用机器学习或深度学习的方式来对手势进行判断。

两者的不通点在于如何实现对手的分离。

可见光方案可以从静态分离和动态分离两种方向去进行分离的实现。静态分离主要是利用人皮肤的颜色实现区分,其缺点是当有较多皮肤且相互重叠时,区分效果较差。动态分离可以利用背景差分法和光流法实现分离,它的计算量会相对较大一些,且无法识别静态手势。

深度信息方案直接通过判断特定距离之内的深度信息,就可以实现对手部信息的分离。深度信息方案在获取手部信息后,还是会将其转为图像的形式,用于后续的处理。

批注 [刘52]: 参考:

<https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2016&filename=JSJA2016S1027&v=sHio613OpzNBiDWBwQMhFT6WD0Lr1uKS6wjrh5YWS1Hfz6phY6PyT7KfdePo3C5>。

将手部信息提取并转化为图像后,会对获取的图像进行进一步处理,以保证被分离的有效区域中只保留手掌(不保留手臂)。

此后,传统的利用机器学习的识别会对手势进行建模。建模可分表观建模和三维建模,简单地讲就是寻找进行特征提取。有一些方法还可以实现对手势中手指的关键点识别(一般是 21 个关键点,5 个手指,每个 4 个点,再加一个手掌的点)。实现特征提取后通常会用一些机器学习的算法去进行识别。

在深度学习技术发展之后,可以直接利用经过训练的深度学习模型实现手势的识别。

肌电方案

由于人类对肌肉的控制是通过神经元传递微弱的电信号实现的,因此可以通过在手上放置电极进行电信号采集,并对电信号进行特征提取、识别,从而实现对手势的识别。

此类方案目前有一个较大的问题是需要额外的、较笨重采集设备才可以实现良好的信号采集。在实际应用到交互中时,很难有良好的交互体验。

毫米波雷达方案

设备会向用户发射毫米波,毫米波在用户手表面发生反射后回到设备的接收端(传感器阵列?)。接收端通过传感器阵列接收到的波的特征判断用户的手势。

这一类方案最典型的设备就是谷歌的 soli。Google 给 soli 做的演示是 soli 能够实现类似一般触摸屏那样精细的手指手势,但目前来看谷歌的 Pixel 4 似乎还是只能实现简单的手掌摇动等手势。这一技术要完全达到成熟还需要一段时间。

姿态交互

姿态交互主要是指人体关键点检测。此外,由于头部姿态信息有时也有特殊的作用,因

批注 [刘53]: 参考:

<https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2020&filename=CGQJ202006041&v=GTjTNIU%25mmd2BercHDzqHVZIMoBj5RIzts6XNIp%25mmd2BTV7LJU02P1z7fLTpnN3xRLnkY6gY。>

批注 [刘54]: 参考:

<https://zhuanlan.zhihu.com/p/21384539>、

此也有专门用于检测头部姿态的技术。

头部检测

头部检测主要检测头的 pitch (俯仰角) yaw (航偏角) roll (滚动角) , 既通过佩戴设备中的惯性传感器来实现 , 也可以通过计算机视觉提取和分析特征点的方式实现。

通过佩戴设备实现头部检测的主要应用范围就是 VR , 通过确定头的转向来确定输出的显示画面。

通过计算机视觉实现头部检测可以应用在对长途司机驾驶时的注意力检测 , 也可用于对教室内的学生上课是否有东张西望、交头接耳的行为的检测。

批注 [刘55]: 参考 :

<https://zhuanlan.zhihu.com/p/51208197>。

人体关键点检测

人体关键点检测主要是依托视觉识别技术 , 实现对若干个人体中的重要节点的位置的持续标记。宽泛的讲 , 之前提到的手势交互应该就属于人体关键点检测的一种。

由于图像种可能不止存在一个人 , 因此人体关键点检测技术可以分为自上而下的 (top-down) , 先对人进行区分 , 再逐一进行关键点检测 ; 也可分为自下而上的 (bottom-up) , 先将所有关键点确定 , 再根据特定原则进行关键点分组。

目前的人体关键点检测经典解决方案有 openpose 和 yolo (2020 年出了 yolo v5) 等。

人体关键点检测能够应用于诸多领域 , 如在家庭中对老人或幼儿的摔倒检测 ; 对肢体语言的理解、手语的翻译 ; 步态身份的识别 ; 应用于 CG 制作的动作捕捉 ; 应用于体育训练或乐器训练中的姿态调整等。

批注 [刘56]: 参考 :

<https://www.jianshu.com/p/3aa810b35a5d> ,
<https://www.bilibili.com/video/BV1Bf4y1971y?from=search&seid=525686909153585097>。

眼动交互

这部分内容也在“感知与传感器”中有讲过。这边简单的重复一下。

原理

眼动仪在使用时会把红外光照射向人的眼睛，由于瞳孔内部可以视为空腔结构，因此瞳孔不会向外反射；而角膜本身具有反射能力，因此角膜会对红外光进行反射。理想状态下，人眼是球体，因此如果人眼的相对位置不发生改变，则角膜上红外光的反射位置（PCCR，Pupil Center Corneal Reflection，即普尔钦斑）也应该不会发生改变。此时通过瞳孔的位置变化，就能确定具体的注视方向。

现实使用中，一般难以保持人眼完全不动。因此，还需要额外的算法确定眼球的相对移动位置，并对计算结果做出补偿。

检测指标

眼动仪能够实现对人的扫视、注视、瞳孔直径、平滑追踪的检测。

平滑追踪

人眼在追踪运动物体时，可以实现匀速而持续的运动（smooth pursuit eye movement）。而有一些患有精神病或情感障碍的病人，他们的平滑追踪机制就存在问题。病人的眼动无法追上物体，因此会产生停顿和扫视来修补误差。

注视

人眼停顿下来注视某一点。熟练阅读者的平均注视时间大约是 225-300ms。通常认为超过 100ms 就算是注视。

眼跳、扫视 (saccade)

幅度可以反映眼睛采样的空间稀疏度。持续时间在 30-120ms 间。

瞳孔直径

当视觉行为处于不同的认知状态时，瞳孔的直径也有所不同，可以用于衡量认知的负荷即被试的情绪状态（观察情感强烈的图片会让瞳孔直径比观察中性图片时更大）。

应用

1. 注视点渲染技术 (英伟达)，由于人眼在注视时，视野范围只有中心点是清晰的，因此可以通过眼动仪获得人眼注视信息的条件下，就可以选择性地分配对于画面的渲染解析度，以最大限度地提升对渲染能力的利用。
2. 在实验室内，可作为大量的基于视觉的试验的基础，如探究人的阅读与注视点的关系、探究人对图像的识别与注视点的关系、利用眼动时间来建立人的认知模型等等。
3. 利用眼动仪来确定消费者对特定空间布置、平面布置等的注意关键点及其顺序，以实现更好的定价策略（交互研究）。
4. 通过眼动仪了解司机驾驶时的关注焦点，为更好地道路交通建设提供数据。
5. 利用眼动数据辅助游戏操作。
6. 利用眼动信息提升智能交互效率。
7. 在航空航天领域，利用眼动提供的注意信息实现对大量仪表盘的优化设计。

脑机接口

批注 [刘57]: 参考 :
<https://www.bilibili.com/video/BV1SK411N7Fj>.

脑机接口可以根据接口位置分为侵入式 (invasive) 和非侵入式 (non-invasive)。非侵入式指接口在大脑之外, 通常有 EEG 脑电图和 TMS 颅脑磁刺激两种; 侵入式脑机接口通常需要通过手术将电极或芯片植入到大脑中。由于手术不可避免地会带来风险, 侵入式接口一般只在病情严重且没有其他保守替代方案的情况下才会被使用。

用于对大脑进行扫描的非侵入式的工具除了 EEG 外还有 CT、MRI、FMRI、PET 等。CT、MRI 通常用于检测大脑的结构, 而 FMRI 和 PET 通常用于检测大脑的功能活动。这些设备的大小一般和价格使得它们一般只有在医院或实验室中才会被使用(他们的原理简单看看就好)。

接口类型

EEG

在“感知与传感器”中提到过, 这边简单重复一下。

EEG 即脑电图, 用于检测人脑再进行各项活动时发出的占主导作用的脑电波。脑电波其实就是来自于大脑神经元的点活动, 是神经细胞传递信号时会发出的电位的变化。通过在大脑表面放置检测电极, 就可以将脑电波以波形的形式呈现。通过对不同类型的脑电波进行分析, 可以确定大脑目前的活动状态。

EEG 最主要的一种用途就是在外科手术中, 麻醉师会利用脑电波来判断患者是否处于昏迷状态, 并根据脑电波变化来调整麻醉剂的使用。

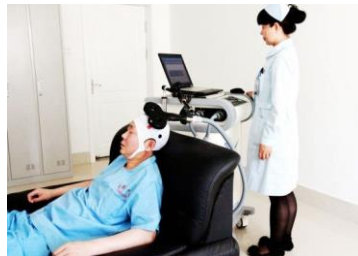
ERD 与 ERS

老师讲到了 ERD 和 ERS 这两个概念，所以我就补充一下。ERD (Event-Related Desynchronization)，即事件相关异步化，指单手准备和执行运动时，该手对侧的大脑感觉运动皮质的脑电信号中 α 频段和 β 频段出现振幅降低。ERS (Event Related Synchronization)，即事件相关同步化，指单手准备和执行运动时，该手同侧的大脑感觉运动皮质的脑电信号中 α 频段和 β 频段出现振幅增加。

TMS

TMS (Transcranial Magnetic Stimulation)，即颅脑磁刺激(也叫经颅磁刺激)。TMS 将磁信号通过颅骨传递到大脑，刺激脑神经。通过调整磁信号，可以实现对大脑皮层兴奋性的降低或增加。

TMS 在临床中被应用在治疗帕金森或重度抑郁症等疾病。



临床上使用经颅磁刺激设备治疗病人

EEG 与 TMS

曾经有一个试验，要求两个受试者在独立环境下同时玩某一个游戏。通过用 EEG 收集某一位被试 A 的脑电波，再将脑电波信号传递给另一位被试 B，并用 TMS 刺激被试 B 的相同脑区，使他按动按键。被试 B 在此条件下能够达到 83% 的正确率。

用类似的做法，通过在被试 A 的手部采集电信号，再在被试 B 的手部输出，可以让 A 在控制自己手腕弯曲的同时，使 B 的手腕也发生弯曲(本质是电信号的传递导致肌肉收缩)。

批注 [刘58]: 参考：

https://blog.csdn.net/qq_24163555/article/details/88783189。

批注 [刘59]: 参考：

<https://www.youtube.com/watch?v=rSQNi5sAwuc>

CT (Computerized Tomography)

CT，即电子计算机断层扫描。通过一个 X 射线发射器围绕患者旋转来获得患者大脑的逐层扫描。由于人体不同组织对 X 光的吸收能力不同，因此可以通过 X 光的衰减吸收率来得到组织的信息。CT 的优点是扫描速度快，可以在紧急情况下快速地找出脑组织异常。

CR 主要用于扫描人脑组织结构。

MRI (Magnetic Resonance Imaging)

MRI，即核磁共振成像。其原理与外磁场与原子核的自旋相关。通过在患者周围施加外部磁场使得患者体内原子中的质子朝向外磁场产生向量进动。在关闭外部磁场后，质子会恢复原本的状态，在这个恢复过程中会向外释放无线电信号。MRI 的优点是成像清晰度高，不存在辐射，但是一般时间较长，价格较高，且如果体内有金属（如做过心脏搭桥）的患者是不能使用 MRI 的。

MRI 主要用于扫描人脑组织结构。

fMRI (functional Magnetic Resonance Imaging)

fMRI，即功能性核磁共振成像。FMRI 通过测量血氧含量变化来检测大脑活动。结合氧分子的血红蛋白是反磁性的，释放了氧分子的血红蛋白是顺磁性的。结合 MRI 对原子核自旋的检测，fMRI 就能实现对血氧浓度的检测。通过将不同状态下的大脑的 fMRI 进行比较，就可以确定在特定活动下那些大脑区域活动比较活跃。

fMRI 主要用于扫描大脑活动功能。

PET (Positron Emission Computed Tomography)

PET，即正电子发射性计算机断层显像，利用放射性元素衰变来检测大脑活动。组织代谢活动会改变葡萄糖、蛋白质等物质的浓度，通过给某个特定活动相关物质打上放射性标记

后注入人体中，就可以通过对其扫描实现对代谢活动的检测。

PET 主要用于扫描大脑活动功能。

侵入式接口

目前侵入式接口的主要应用还是在医疗领域。在帕金森治疗领域，通过在大脑中植入电极，科学家可以通过抑制肌肉的过度放电，使得帕金森病人能够停止不自主的肌肉颤抖，实现平稳的动作。

应用

脑机技术现在还不是特别成熟的技术，一般的应用大多是医疗或实验领域的，如使用脑机接口实现假肢控制，或使用脑机接口治疗帕金森等疾病。

虽然 2020 年马斯克提出了向猪大脑植入脑机的技术，但谈论较为商业化的应用目前还为之过早。

情绪识别

情绪识别的方法目前可以分为以下几种：人脸表情识别、语音情绪识别、语义情绪识别、神经信号识别。

从技术角度看，这些识别无非都是依赖之前提到的语音识别、视觉识别或神经信号识别。只是在识别的具体结果上会输出不同的结果。