

# 概述

从人的认知角度出发，人认知这个世界并做出反应有三步：

1. 人的感知系统（视觉、听觉、触觉、嗅觉、味觉）接收到外界的信息。
2. 人通过心理、认知、情绪等对信息进行处理。
3. 人通过语言、动作等实现对信息的反馈。

因此，推广至机器，机器的运转也有类似的步骤：

1. 利用各种类型的传感器接收来自外界的信息。
2. 通过程序算法分析信息模式。
3. 通过具体设备输出对信息的反馈。

以下，将把重点分别放在人与机器两个角度。

人：1.了解人的感知系统；2.了解人的心里模型、认知、情绪。

机器：1.各种传感器硬件常见模块以及工作原理；2.实际应用场景以及应用的优点、局限性；  
3.实际的应用案例；4.多源传感融合和交互协同；5.发展和设计的趋势（微型化、精细化、普适化）。

机器通过传感器获取数据，通过接口将数据导入处理器，通过算法处理数据、得到需要的操作指令或反馈信息，最后通过输出设备将反馈结果输出。

批注 [刘1]: “接口”就是用于连接机器与人的桥梁。

# 视觉部分

## 视觉机制

视觉信息占人体感知信息的 80% 以上。随着 CV 算法的发展和 AI 的强大能力，视觉感知是最准确最快的感知方式。

批注 [刘2]: 即 computer vision，之后在详细展开。

人体的视觉机制是，利用晶状体将进入瞳孔的光线聚焦到视网膜上，由视网膜上的视觉细胞将光信号转化为神经冲动再传送给大脑。

人眼的视野：宽  $160^{\circ}$ ，高  $135^{\circ}$ ，宽高比 1:1.2。

视网膜上由两种感受器：

1. 视锥细胞 cone。位于中央凹，主要负责注视及白天的成像，能够区分色彩，对黄色敏感。
2. 视杆细胞 rod。谓语边缘区域，用于夜间成像，无法区分色彩，对绿色敏感。

人眼对色彩的感知源于视觉细胞对 400-780nm 间的可见光波的感知。

人眼对 550nm 左右的黄绿色最为敏感。

人眼存在盲点，视网膜上视神经进入眼球处有一个凹陷点，此处没有感光细胞，因此无感光能力。

## 有一些与视觉设计相关的心理学知识：

### 1. 黑暗中人们更擅长利用余光。

由于感知强光的视锥细胞在视网膜中心，感知弱光的视杆细胞在外围分布，因此弱光中，更依赖外围的视杆细胞所提供的余光。

### 2. 人的视觉成像是二维的

大脑视觉皮质汇总所有的信息。这一点也有很多老师讲过，视觉设计的最大困难就在于三维与二维的转换。

### 3. 中央视觉

视网膜的分辨能力是从中央凹向外边缘急剧下降的，因此视觉系统的某一时刻只能处理1-2°的信息。通过注视的转移来切换视觉任务。在设计中，中央视觉是人们关注细节的核心，最为重要。可以利用注视点和注视时间（注视时间大于某个阈值或眼跳幅度小于某个阈值）来推测用户的兴趣点。

### 4. 余光与周围视觉

周围视觉在人们进行环境整体感知时更为重要。

### 5. 幻象、视觉错觉与全息显示

视觉错觉实际上就是利用反射或是折射，使得观众看到的图像所在位置与成像区的实际位置不同。最典型的应用就是现在的各类伪全息投影（具体在后面讲）。

## 技术引申

由以上生理基础内容可以引申出两个产品设计相关的内容：眼动技术和全息显示。

## 眼动技术（作为自适应交互、NUI、意图推理、兴趣点搜索、认知负荷的基础）

眼动仪的原理是将近红外光导向眼睛的瞳孔，在瞳孔和角膜中引起可检测的反射，通过计算瞳孔中心位置与角膜反射位置的相对差异，可以推测注视位置。目前的眼动仪大多都是使用这种“暗瞳技术”的。

其优点在于，由于红外线对人眼不可见，因此再进行跟踪测试时，被测不会分心。

其缺点在于，由于测量标准较少，因此必须严格校准。如果希望获得较为精准的结果，对测量环境也有较高的要求。由于对数据的分析是基于图形信息的，因此依赖特征提取算法。

批注 [刘3]: 具体原理下面会讲。

眼动仪的常用指标：

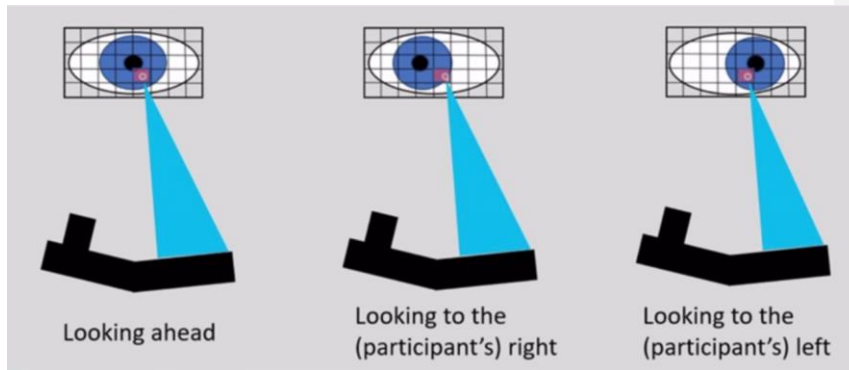
1. 凝视点 ( gaze point ) : 凝视点构成基本测量单位，即每个最小测量时间内眼球注视的位置。
2. 注视点 ( fixation ) : 如果凝视点持续一段时间，就认定某位置为注视点。
3. 感兴趣区域 ( AOI ) : 注视点集中区域。
4. 热图 ( heatmaps ) : 用于表示被试者在刺激材料上的注意力分布情况。
5. 固定顺序 ( fixation sequence ) : 用于确定注视发生的顺序和方向。
6. 首次注视时间 ( TTFF, time to first fixation ) : 被试在测试开始后多久才第一次注意到某个特定区域。

可以检测到的眼动包括：平滑追踪、扫视、注视、瞳孔直径。

1. 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。
2. 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。
3. 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。
4. 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 )。

#### 眼动仪的具体原理

眼动仪使用时会将红外光照射向人的眼睛，由于瞳孔内部可视作空腔结构，因此瞳孔并不会产生向外的反射；而由于角膜本身有反射能力，因此会对红外光产生向外的反射。理想状态下，由于人眼是球体，因此只要红外光位置不变，人眼球位置不变，相机所捕捉到的角膜反射位置也不会发生变化。此时就可以根据人眼瞳孔的位置与红外反射点的位置来确定具体的注视位置。



图中的红色点表示角膜反射位置，在眼球发生转动时基本不会改变位置；黑点代表瞳孔。

但由于人会发生移动，眼球位置会发生改变，因此通常情况下需要额外的算法来确定眼球的相对移动位置，进而做出补偿。

具体算法流程为：预处理（利用特征提取，确定眼球由于人体移动所造成的位置变化，在补偿眼球移动后获得相对固定的区域信息）-特征提取（根据图形灰度提取 PCCR）-计算（基于模型）-输出结果。

**批注 [刘4]:** PCCR, Pupil Center Corneal Reflection, 即瞳孔位置与角膜反射。角膜反射点通常也被称为普尔坎斑。

#### 眼动仪涉及的算法

**批注 [刘5]:** 具体我到图像识别算法里去说。

以 cv 算法为主，主要包括物体检测（Yolo、faster-RCNN）、物体识别、手势识别、姿势识别（openpose）、图像分割（unet）等。

#### 基于眼动仪的拓展应用

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。

现有的眼动仪类型：

Tobii 眼动仪( 京东 1800-3500 元 )、EyeSo 眼动仪、RED-M 眼动仪( 后面两个都没找到 )。

## 全息显示 ( 3d 投影 )

现在的很多全息显示技术都是使用所谓的视觉错觉技术的，实际上这些所谓的“全息”其实都是 3d 投影技术，与学术意义上的全息并不相同。

如佩珀尔幻象就是利用透明玻璃的反射来实现 3d 效果，最出名的应用应该就是春晚的节目和初音未来演唱会；

旋转 LED 显示技术，利用 LED 灯条的高速旋转来实现成像与隐藏显示设备的目；

光栅立体图，是通过将多个角度的图像以都切分成等距的细条，然后交替排列合并在一张图像中，再通过折射层保证不同角度只能看到其中某个角度的图像的细条（就是那种三维立体变幻卡）。

还有一些全息是通过向空气中喷射水或其他介质,在加上投影实现的;还有利用与海市蜃楼类似原理,通过空气密度不同来使光线“转弯”进而实现全息显示的。

真正实现不依赖投影的全息显示技术,日本的团队通过电离空气呈现 3d 画面的(利用激光电离空气,创造出大量漂浮的高温等离子体),但该技术存在明显的安全问题,且显示时间短暂。

此外,还有 looking glass 通过光场显示(与光栅立体图类似的思路),在一个透镜构成的显示器中实现了有区域性的 3d 投影(利用透镜将通常情况下显示器发出的无特定方向的光转变为有 45 个特定方向的光,配合以 45 个角度的图像,实现 3d 显示。具体技术细节我不是那么确定)。虽然这种方式并不是在空气中实现投影的,但我个人认为是有可能最快走向商业化的技术。其缺点在于透镜制作的复杂性及光场显示所需的强大的渲染能力。



## 图像识别算法 (cv 算法)

### 计算机视觉的主要技术：

1. 图像分类。根据图像内容，在已有的图像分类标签中找到并分配给未分类图像一个特定的标签。
2. 对象检测。对不同图像选定不同的目标对象，确定对象位置及边界。
3. 目标跟踪。在不同的图像集合中，对选定的目标图像进行跟踪检测，对目标对象进行信息的提取，并分析其在图像中的位置、大小等。结合检测到的结果，对目标对象进行观察分析，并预测其下一步行为诡计。
4. 语义分割。根据像素对图像的内容或对象进行划分，每个像素对应各自的标签与分类，从而实现图像不同内容或对象的类别的再分割。
5. 识别图像中不同实例元素，并对每个实例进行明显的分割，确定复杂背景下不同实例的边界。

### 计算机视觉应用算法包括以下若干个类别：

1. 图像识别：Alexnet、VGGNet、GoogleNet、ResNet、DenseNet。
2. 目标检测：Fast-RCNN、Faster-RCNN、YOLO、Retina-Net。
3. 图像分割：FCN、Mask-RCNN、U-Net。
4. 视频分割：SegNet。
5. 光流：FlowNet。
6. 图像生成：GAN、WGAN。
7. 目标追踪：Goturn、ECO。

**批注 [刘6]:** 图像识别这一块之后问问老师吧，我感觉自己找资料有点糊涂。  
这里主要是讲传感器的，这部分内容会不会放到别的地方去比较好？

**批注 [刘7]:** 来源：  
<https://kns.cnki.net/KXReader/Detail?TIMESTAMP=637387276245000313&DBCODE=CJFD&TABLEName=CJFDAUTO&FileName=WXXJ202011050&RESULT=1&SIGN=ojjs54YFXdBzBalN3u81aQdTcX4%3d>。

**批注 [刘8]:** 来源：  
<https://www.zhihu.com/question/342633770/answer/811211699>。

同一个需求，根据场景可能会有非常多的差异变化，从而会有非常多的实现路径。以人脸识别为例，人脸识别算法一共包含三个模块：

1. 人脸检测 ( Face Detection )：确定人脸在图像中的大小和位置，也就是在图像中预测 anchor。
2. 人脸对齐 ( Face Alignment )：找到人脸的若干个关键点 ( 也成基准点，如眼角、鼻尖、嘴角等 )，然后利用这些关键点通过相似变换 ( 旋转、缩放、平移 ) 将人脸尽可能变换到标准人脸。
3. 人脸特征表征 ( Feature Representation )：接收标准化人脸图像为输入，通过特征建模得到向量化的人脸特征，然后通过分类器判别得到是别的结果。关键在于如何得到不同人脸的有区分度的特征，如鼻子、嘴巴、眼睛等。

## 立体视觉 ( 获得深度信息的机制 )

### 深度信息的来源 :

根据津巴多先生的《心理学与生活》, 深度信息有三种来源 :

1. 双眼深度的信息来源是视网膜像差和视轴融合 ;
2. 单眼深度信息的来源是遮挡、相对大小、线条透视、质地梯度 ( 都需要参考对象 );
3. 运动视差。

### 深度相机 ( 按原理分类 ):

1. 飞行时间法 ( TOF , Time Of Flight )

通过连续发射经过调制的特定频率的光脉冲 ( 一般为不可见光 ) 到被测物体上 , 然后接收从物体反射回去的光脉冲 , 通过探测光脉冲的飞行 ( 往返 ) 时间来计算被测物体离相机的距离。

该方法优点在于测量距离较远 , 深度抗干扰能力较强 ( 在无人驾驶领域有较强优势 );

缺点在于只能在室内使用 ( 户外使用会受到日光的严重干扰 ); 由于对时间测量的精度要求高 , TOF 对硬件设备有较高要求 ; 在较短距离内 , TOF 的精度与其他精度相机相比还有较大差距。

目前消费级的 TOF 深度相机有微软的 Kinect 2 ( 京东 1500 元 ) , google 的 Google Project Tango 等 ( 不常见 ) 。另外 , 华为的手机面部识别就是通过 TOF 实现的。

2. 双目立体视觉法 ( 原理与人眼感受深度的原理类似 )

通过计算空间中同一个物体在两个相机成像的视差确定物体与相机的距离。

批注 [刘9]: 参考 :  
<https://zhuanlan.zhihu.com/p/32199990>。

该方法优点在于在户内外均可使用，且对相机硬件要求低，容易控制成本；

缺点在于由于该算法依赖特征匹配算法，因此受光照变化影响较大；在黑暗条件及缺乏纹理的场景下无法准确实现判断；计算复杂度较高，涉及极线约束、图像校正、图像匹配等复杂的算法；两相机之间的距离将显著影响测量范围。

较有名的产品有 STEROLABS 推出的 ZED 2K Stereo Camera( 淘宝 1700 元 )等。

### 3. 结构光法

该算法是为了解决双目匹配算法的复杂性和鲁棒性问题而提出的。结构光算法比依赖物体本身的颜色和纹理，采用主动投射已知的光图案（如离散光斑、条纹光、编码结构光等，通常为红外光）来实现快速匹配特征点的目的。结构光法似乎也仍然需要两个摄像头（好像也有单目摄像头就能解决的，且单目体积更小）。

根据知乎专栏，结构光法是在单目摄像头的基础上完成的。具体原理是在点阵投影仪发射的结构光，在被接受时投影点的相对顺序依然保持，但由于反射面的存在，导致相邻两点间的距离将发生改变（即结构光图形产生畸变）。通过对畸变的计算，可得出深度。

该方法的优点在于，由于结构光主动投射编码光，因此非常适合在光照不足或缺少纹理的场景使用；投射光结构通常经过精心设计，因此可以在一定范围内达到较高准确度；技术成熟，深度图像可以做到较高的分辨率。

缺点在于，投射的光斑容易受到强自然光的影响，导致无法被检测；测量距离较近，距离变远，测量精度也会变差；若反光面粗糙度低，投射的光斑会被反射影响。

较有名的产品就是 iPhoneX 的面部解锁。

**批注 [刘10]:** 散斑结构光对计算量要求较大，安全性高。  
有规律的编码结构光对计算量要求较小，安全性略差。

**批注 [刘11]:** 来源：  
<https://zhuanlan.zhihu.com/p/29971801>。

相机类型	TOF	RGB 双目	结构光
测距方式	主动投射	被动式，无主动投射	主动投射
工作原理	根据光的飞行时间直接测量	RGB 图像特征点匹配，配合三角计算间接测量	主动投射已知编码图案，提升特征匹配效果
测量精度	最高可达厘米级	近距离可达毫米级	近距离可达 0.1 毫米-1 毫米
测量范围	可测量较远距离，一般为 100 米以内（这种百米级别的其实就属于激光雷达了）	受摄像头间距离限制，只能测量较近距离，距离越远，测距越不准确。 在摄像头间距为 10 毫米的条件下，距离为 2 米以内	距离一般为 10 米以内
影响因素	不受光照变化和物体纹理影响。 受多重反射影响	手光照变化和物体纹理影响大，夜间无法使用	不收光照变化和物体纹理影响，受反光影响
是否支持户外	功率较小的情况下，受户外光影响较大	无影响	有影响，与编码图案设计有关
分辨率	低于 640*480	2K	1080P
帧率	较高，可达上百 fps	从低到高都有	一般为 30fps
软件复杂度	较低	较高	中等
功耗	很高，因为要实现全面照射	较低，因为只负责软件计算	中等，虽然也要进行投射，但只进行局部投射
材料成本	中到高	中	高

三种深度相机原理及对比表

## 深度相机的应用场合

1. 人脸识别技术范畴：在获取人脸的具体深度信息后，就能有更多可能的操作。
  - (1) 根据深度信息更好地实现景虚化与人像光效（现在智能手机拍照应用中广泛使用）。
  - (2) 动画表情。
  - (3) 三维美颜，进一步增强美颜效果，并同步增强真实感。
  - (4) 人脸识别。
  - (5) 照片中对人眼视线的矫正。
2. 智能人机交互：通过深度信息获取较近距离内的人的四肢或手势信息，以实现进一步的交互优化。
  - (1) 人体骨骼的提取与跟踪。最常见的就是体感游戏。
  - (2) 近距离的手势识别与跟踪。在添加深度信息后，能获得比纯图像分析更加优秀的手势识别能力。
3. 三维重建：通过多角度扫描来实现三维信息的记录与重建。
  - (1) 三维空间测绘
  - (2) 三维物体重建
  - (3) 三维空间地图重建（其实就是基于三维空间测绘的）
4. 机器人与智能驾驶：通过对距离传感器的应用，极大地提升了机器人或汽车躲避障碍的能力。
5. AR 技术：高真实度的 AR 体验无疑需要结合深度信息实现高质量的渲染优化。  
( iPhone12 就通过深度摄像头实现了对 AR 使用的优化 )

## 相机传感器

### 相机的类型与应用场景

#### 1. RGB 相机：

正常的可以输出 RGB 照片的相机。广义上说，鱼眼相机、360 全景相机也都是 RGB 相机。基本原理是用透镜获取光线到传感器上，由传感器将光信号转化为电信号并进行保存的方法。

#### 2. 深度相机：

查看之前立体视觉中的部分。主要有三种原理，TOF、双目立体视觉法、结构光法。

#### 3. 鱼眼相机：

加装了鱼眼镜头的一般相机。鱼眼镜头是焦距为 16mm 或更短的、能实现接近 180°广角拍摄的镜头。由于其极大的视野范围，通常会造成图片的严重的畸变。

#### 4. 360°全景相机：

传统概念的 360°全景相机是以多镜头+后期合成（图像的变形、特征点匹配与拼接）的方式实现全景图像展示的。

但由于该技术的核心在于后期合成，因此很多只需要全景照片（不需要全景视频）的情况下，可以由一台相机进行多角度拍摄（通常需要 26 张图片），在统一由软件实现后期合成。

这个做法的进一步扩展就是手机中的全景相机功能，本质就是通过图像变形、特征点匹配、拼接实现的多图合一图。

#### 5. 红外相机：

通常现在的普通相机的传感器都是可以拍摄到红外线的，但为了成像质量考虑，会在镜

头上添加红外滤镜防止红外线对图像的干扰。

因此，在使用一般 CMOS 元件的条件下，只要不添加红外滤镜就可以实现对 780nm-1000nm 的红外线的捕捉。

但由于室温下的物体辐射的红外线波长通常在 8000nm-12000nm，因此一般 CMOS 是无法实现热成像的效果的，也无法实现红外夜视。

日常生活中对红外的应用可能主要体现在深度检测上。由于结构光法需要发射人眼不可见的光结构，通常选用的就是红外光，因此对应的检测设备就是红外摄像头。iphone 的深度检测使用的就是红外摄像头。

除此以外较多的就是在专业领域的，如夜视、利用红外热成像实现不接触式的测温（2020 年大量使用的体温测量图像等）、短波红外辅助雾霾环境的检测。

## 相机的几何标定

相机标定的目的是，建立相机几何模型（主要）并矫正透镜畸变。

相机的功能是将三维的场景转化为二维的照片，但这个转化过程是不可逆的，即我们无法简单地从二维照片反向获取原先的三维场景。相机标定的目的就是通过简单的数学模型来表示复杂的成像过程，并且求出成像的反过程。标定后的相机可以实现对三维场景的重建，从空间点的像素坐标映射到世界坐标。

由于相机使用透镜进光，而透镜的制造工艺会使成像产生多种形式的畸变，为了使计算机视觉分析更好地得到结果，需要矫正这种畸变。

### 几何标定的具体实现

具体实现需要通过三轮转换，将世界坐标系中的某个点转换到像素坐标系中，具体为，世界坐标系->相机坐标系->图像坐标系->像素坐标系。

这其中，涉及相机坐标系->图像坐标系->像素坐标系的转换的参数被称为内参，涉及

批注 [刘12]: 参考：  
<https://zhuanlan.zhihu.com/p/76183324>。

批注 [刘13]: 参考：  
<https://zhuanlan.zhihu.com/p/30813733>，  
<https://zhuanlan.zhihu.com/p/35223115>，  
[https://www.bilibili.com/video/BV18J411U7DS/?spm\\_id\\_from=333.788.b\\_7265636f5f6c697374.5](https://www.bilibili.com/video/BV18J411U7DS/?spm_id_from=333.788.b_7265636f5f6c697374.5)。



世界坐标系->相机坐标系的转换的参数被称为外参。

内参和外参可以通过对标定板的不同角度的照片分析得以确定。

畸变矫正问题通常只考虑由透镜与 CMOS 或 CCD 安装位置误差导致的切向畸变。畸变矫正可通过对标定板的不同角度的照片分析得以确定。

完成内外参确定和畸变矫正后,如果我们要实际确定图片像素对应的空间中坐标,必须要有两台相机,结合已求得的参数方可求解。

## 视频规格

### 视频大小

视频大小视频分辨率、压缩程度、色彩深度、帧率、时长等因素决定。

#### 1. 视频分辨率。

视频分辨率就是指视频的像素组合。当前的主流大小是 1080p,其中 1080 指纵向像素数;p 指逐行扫描 ( progressive scan ),与 i ( interlace scan ,隔行扫描,过去通常用于广播电视 )相对应。此外还有 720p、2k ( 2048\*1080 )、4k ( 4960\*2160 )、8k ( 7680\*4320 ) 等等。

#### 2. 压缩程度。

通常用视频码率来描述,单位通常为 Kbps ( kb per second ),具体就是指视频在压缩时,每秒的图像内容的大小上限。

#### 3. 色彩深度。

通常我们使用 RGB 色彩就代表色彩有 3 个通道,如果还需要透明,就还需要额外一个 alpha 通道。一般情况下会将每个通道划分为 256 级 (  $2^8$  ,正好可以用一个

Byte 表示，因此也成 8bit)，用于表示从明到暗。

如果一个像素点由 RGB 三通道构成，每个通道都是 8bit，那么这个像素点所需要的存储空间就是  $3 * 1 = 3\text{Byte} = 24\text{bit}$ 。

#### 4. 时长与帧率

帧率是指每秒中的画面数量，单位为 fps。25fps 就是指每秒会有 25 个不同的画面。

#### 5. 时长就是视频的持续时间。

### 视频的大小计算

如果不考虑压缩，那么视频的大小计算公式为，分辨率\*色彩深度\*帧率，结果以 Byte 为单位。以一般 1080P、fps25、时长 1 秒的视频为例， $1920 * 1080 * 3 * 25 = 148\text{MB}$ 。

如果考虑压缩，则公式为，码率 \* 时间 / 8，结果以 Byte 为单位。如 4000Kbps 的 1 秒钟的视频，大小约为 500KB。

以上计算都不包括视频中的音频内容。

**批注 [刘14]:** 这里其实还涉及视频编码压缩技术和海量数据处理能力，之后看老师会不会展开吧。

### CMOS 与 CCD

CMOS( Complementary Metal Oxide Semiconductor )，互补金属氧化物半导体，是一种集成电路的设计工艺，可以在硅脂晶圆模板上制出 NMOS ( n-type MOSFET ) 和 PMOS( p-type MOSFET )，由于 NMOS 和 PMOS 在物理特性上互补，因此被称为 CMOS。

CCD ( Charge Coupled Device )，电荷耦合器件，是一种探测元件的名称。

CMOS 和 CCD 本身都只是一种集成电路技术或元件的名称，如今对 CMOS 和 CCD 的

**批注 [刘15]:** 参考：  
<https://zhuanlan.zhihu.com/p/27463125>，  
<https://www.zhihu.com/market/pub/119640147/manuscript/1169593683740643328>，  
主要参考《现代传感器手册》，雅克布·弗雷登，p441。

**批注 [刘16]:** 最好把这个名字记一下。

较常见的应用是在相机的图像传感器上。CMOS 感光元件与 CCD 感光元件相对应，前者是现在的主流，后者在过去较为常用。两者的基础感光元件都是光敏二极管（在后面传感器中的光探测器中会讲到），每个光敏二极管就对应了图像的一个像素。

与 CCD 相比，CMOS 有更高的集成度、更低的功耗、更小的尺寸、更低的成本，且其成像质量并不显著落后于 CCD。这样的特性使得 CMOS 在要求传感器越来越微型化的今天，在越来越多的设备（尤其是手机，不过现在相机似乎也主要是用 CMOS 了）上被使用。

## 原理

### CCD

当光线照射到 CCD 传感器的光敏二极管单元时，每个像素都会有一个势阱用于收集光电效应产生的电子（电子量与光子量成正比）。

在每次曝光完成后，CCD 利用施加在芯片上的周期性时间电压将数据读出。CCD 的数据的读出节点很少，通常只有一个。假设该节点在顶端，所有的势阱中的电子都会同时向上移动，将顶端的电子挤出到读出行，然后读出节点从左向右读取每个节点中的电子量（有点类似逐行扫描），并通过电路将电信号转化为数字信号。

对上述过程反复循环，直至所有电子都被读出，即可在重构阵列上实现图像的储存。

实际过程中，为了减少读取时间，也会将多个节点组合在一起进行统一读取，这样可以加快读取速度，但会降低图片分辨率。

### CMOS

CMOS 与 CCD 不同，每个光敏二极管都有单独的内部集成放大器，可独立完成电信号的数字化。在光线照射后，每个光敏二极管中的光子产生的电子会直接通过放大器放大并转

**批注 [刘17]:** 这里光照射光敏二极管时产生电流的具体原理参考后面光探测器中光敏二极管的原理解释。

化为数字信号。但由于光敏二极管边存在这个附加电路 因此 CMOS 芯片的光敏度会降低，因为许多光子撞向了电路而非光敏二极管。

## 光学变焦与数码变焦

这里老师提到关于光学变焦和数码变焦，所以稍微展开一下。

光学变焦就是指通过镜头组中的透镜位置的变化实现成像大小的变化，其原理与望远镜相同。光学变焦不会导致图像质量的降低。

数码变焦就是简单的将图像拉大，通过将原先的单个像素点等比放大成多个像素点来实现“变焦”。数码变焦会导致图像质量的降低。

# 听觉部分

## 听觉机制

听觉的产生机制是物体振动产生的声波传到到人耳内,使人耳内的感受细胞兴奋并引起听神经的冲动,进而经估计听觉中枢分析后产生的感觉。

听觉在人类的感受带宽中排列第二,仅次于视觉。

人类能够听到的声波频率范围在 20hz-20khz,接受能力最高的范围在 1000hz-4000hz。

人耳的听阈指能引起人耳听觉反应的最小音量,正常为 25 分贝(也有说 0 分贝的,不过关于人耳能听见的最小声音这一问题会受到很多变量的限制,且 0-25 分贝基本属于宁静状态下的音量,因此不比深究)。

人耳的痛阈指挥让人耳感觉疼痛的最小音量,正常为 120 分贝(大货车鸣笛的声音,在这样的环境下呆超过一分钟即会导致暂时的耳聋)。

声波的主要指标包括,响度(一般也可理解为振幅)、频率(用于解码的重要的分辨信号)、波长。

人耳能够获取的声音线索包括单耳线索(可以判断距离,配合头部转动可以进行简单的方位判断)和双耳线索(可以较准确地定位,还可以加强声音信号获取)。

严格意义上说,双耳定位是依靠:双耳时间差、双耳相位差、双耳声级差和双耳音色差。

批注 [刘18]: 来源 :  
<https://zhuanlan.zhihu.com/p/59728295> ,  
<https://zhuanlan.zhihu.com/p/52854415>。

批注 [刘19]: 来源 :  
<https://www.zhihu.com/question/275780295/answer/403199450>。

批注 [刘20]: 来源 :  
<https://www.zhihu.com/question/340435642/answer/815855438>。

## 听觉感知硬件

用于感知听觉的主要就是麦克风。

使用场景不同的分类：

### 1. 普通麦克风：

通过空气振动，实现对机械波形式传输的声波的接收和感知。

### 2. 骨传导麦克风：

通过固体振动，实现对机械波形式传输的声波的接收和感知。骨传导麦克风可以听到摩擦的声音。如果在某硬件中放置骨传导麦克风，可以较好地实现对外部声音的隔绝和对硬件敲击声的检测。

**批注 [刘21]:** 骨传导这个词最早应该是用于做人体听觉测试的，由于测试通过颞骨传递振动，因此称为骨传导。

根据原理不同的分类：

### 1. 动圈式麦克风：

由放置在磁场中的膜片接收声波，声波引起膜片的振动，膜片振动带动与之相连的音圈切割磁感线，形成变化的电流，实现声信号到电信号的转变。

### 2. 电容式麦克风：

通过一个绝缘体将同一电源不同电极上的一个可动的隔膜片和一个固定的背板隔绝，形成一个电容。隔膜片接收声音后会产生振动，振动会改变电容，产生电压变化，实现从声信号到电信号的转变。

### 3. 电阻式麦克风：

基于压阻效应设计（材料在受机械式应力下所产生的电阻变化），半导体材料的压

**批注 [刘22]:** 除了下面三个以外，还有驻极体麦克风、激光麦克风等，这里不多赘述。

**批注 [刘23]:** 主要在以前的老式电话中使用。

阻效应远大于金属。具体为，在两块金属板间添加碳颗粒，给两块金属板通电，使之产生通过碳颗粒的电流，其中一块金属板，即金属膜片，会随传入的声波振动而向碳颗粒施加变化的压力，从而导致碳颗粒的整体电阻发生变化，进而使流过其中的电流发生变化，实现从声信号到电信号的转变。

现在的麦克风主要使用的是 MEMS 麦克风，即基于微型机电系统技术制造的麦克风，主要使用的原理是电容式麦克风的原理。由于将电容器直接集成到微硅晶片上，可以方便地实现与 CMOS 工艺和其他音频电路的结合。

## 麦克风阵列

现在通常会用 6-8 个的麦克风构成麦克风阵列，以实现更好的收音和定位效果。

### 1. 定位

2 个麦克风可以实现  $180^\circ$  内的声源定位，6 个麦克风可以实现  $360^\circ$  内的声源定位。

### 2. 增强收音

主要应用于远场交互场景，即用户与麦克风距离较远。

## 降噪

被动降噪：

用不同材质的阻塞物来放置噪音进入耳朵，通过噪音的声波在阻塞物中反复反射消耗实现降噪。

优点在于，对与高频率的声音比较有效，且价格也较为低廉。

批注 [刘24]: 参考：  
<https://zhuanlan.zhihu.com/p/95584370>。

缺点在于，降噪效果与阻塞物使用量正相关，在便携式耳机中，无法添加大量阻塞物，因此降噪效果并不优秀；用户在降噪效果不佳的情况下会调大音量，造成了间接地对用户听力的损伤。

主动降噪：

通过处理器分析外部传入的噪声，然后创建一个相位相反的声音，利用机械波相互叠加的原理将噪声进行中和消除，声学中称此为“相消干涉”。

优点在于，基于 MEMS 技术，即使在便携式耳机中也能添加处理器和微型扬声器，因此降噪效果不受设备大小制约；由于噪声被有效中和，用户不需要通过提高音量的方式消除噪声，有效保护了用户的听力。

缺点在于，由于设备设计复杂、质量要求高，因此成本较高；由于需要主动识别、分析、扬声，因此设备需要额外的耗能；主动降噪对于持续产生噪音的环境（如机舱内）有良好的效果，但对于无规则的响亮声音（如堵车中的喇叭声）则几乎没有效果。

环境音识别：

需要环境音识别的原因：

1. 由于主动降噪会很大程度上消除环境音，因此出于安全性考虑，必须有一个收集环境音的麦克风，确保用户处在较危险的环境时（过马路），耳机能快速地将危险的环境音通过扬声器传递给用户。
2. 主动降噪麦克风和耳机的主体扬声器距离较近，良好的环境音识别也有利于更好的降噪效果。
3. 有些设备有语音唤醒功能，如何在嘈杂的环境中实现准确的唤醒也是一个重要的问题



(鸡尾酒会问题)。

**批注 [刘25]:** 即使在嘈杂的鸡尾酒会中，人们也总能听到被人叫自己的名字。

## 环境音识别的实现：

**批注 [刘26]:** 参考：  
<https://www.zhihu.com/question/57672803/answer/154658778>，  
<https://zhuanlan.zhihu.com/p/22512377>，  
<https://www.bilibili.com/video/BV1t5411W7EU?from=search&seid=4228051116969293684>。

处于安全目的的环境音识别通常会识别车辆的行驶声或喇叭声，不过这项技术目前尚在中发展。

基于嘈杂环境的语音唤醒的环境音识别主要的处理目的是分离环境音。基础的语音识别的流程是，首先将通过回音消除等技术实现基础降噪的语音信号输入给模型（如神经网络），将信号转录成音素序列，然后把得到的音素序列转换成文字，用语言模型负责理解意思。

但在环境嘈杂的情况下，语音输入的数据会变得复杂。较好的解决方法是用麦克风阵列收集多个信号源的信号，结合声源定位和波束形成（beamforming），实现对特定位置声音的定向提取；此外就是进一步优化语音识别算法，使其对噪声有更好的过滤能力。

**批注 [刘27]:** 这个技术在 5G 中也有应用，简单的说就是将原先的广范围的信号发送和获取转变为某个特定范围的信号发送和获取，以此提高效率。此外，在声音部分的另一个应用是麦克风的收音加强。

## 基于 beamforming 的进一步拓展

### Beamforming 的原理

**批注 [刘28]:** 参考：  
<https://www.youtube.com/watch?v=A1n5Hhwtz78>

以接收为例，要实现 beamforming 必须要有多个接收器。当位置固定的信号源向接收器发射信号时，由于不同接收器的位置不同，它们接收到的信号在相位上会有所不同。为了保证信号的统一，会对不同接收器信号进行不同程度的偏移处理。这样的偏移处理将所有接收器接收到的信号同步，保证最终的信号是原信号的保留或加强。而这样的偏移处理，同时也保证了从其他角度来的信号将无法实现最优的偏移，产生由于高低位置的波形叠加导致的抵消，最终导致信号的减弱。因此，实现了信号的定向接收。

发射的波束成形也是同理，由于不同发射器在向特定位置进行信号发射时需要进行偏移，从而导致了只有指定位置会收到强信号。

基于 Beamforming 的声学应用：

批注 [刘29]: 参考：  
<https://zhuanlan.zhihu.com/p/241268888>。

#### 1. 在嘈杂环境下对话筒的定向收音：

由于我们之前提到，beamforming 技术能够实现定向的信号接收，在佩戴耳机时，嘴巴相对耳机的位置基本保持不变，因此可以利用 beamforming 技术定向接收来自嘴巴的声音，“屏蔽”周围的噪声。

#### 2. 结合 AR 技术的定向选择收音：

通过 AR 眼睛上的摄像头，我们可以通过特征识别的方式分析出当前正在发出声音的人或对象；通过用户的眼动和头动信息，我们可以判断出当前用户兴趣对象的位置；最后通过 beamforming 技术定向地对特定对象或区域进行收音并通过麦克风播放。这样一来，即使在即为嘈杂的环境下，用户同样能够清楚地听清目标对象所发出的声音。

## 相关产品

TWS ( True Wireless Stereo , 以苹果的 airpods pro 为例 )

批注 [刘30]: 这里有以下几个可能的考点，  
1.无线耳机使用了哪些传感器，作用分别是什么。  
2.VUI 交互的实现。  
3.耳姿交互。（这块其实就是考察传感器，只不过这里更侧重用户体验的设计）

抗噪：

被动降噪通过入耳设计加硅胶耳塞体现。

主动降噪方面，苹果采用了两个麦克风，外向式麦克风会检测外部声波，内向式麦克风会检测耳内的多余声波，然后分别用波形相反的抗造声波进行抵消。内耳式麦克风还会根据自己接受到的声波，对主动降噪进行进一步优化调整。降噪采样频率为 200Hz。

批注 [刘31]: 此外，内向式麦克风是在扬声器的内侧的，因此会对音频输出进行一定的优化调整。

环境音：

苹果设置了对主动降噪的开启和关闭。通过长按耳机柄上的压力感应器，可以实现降噪与通透模式的切换。此外，苹果的通透模式应该就是用麦克风对外部环境音进行主动接受，进行适当调整后在通过扬声器传出，并非简单的关闭降噪。

批注 [刘32]: 后面会说到这种感应器。

双波束成形麦克风：

使用两个波束成形麦克风，保证对人声的定向收音。

芯片：

采用苹果自行设计的 H1 芯片，有 10 个音频处理核心，保证处理的低延迟。

压力传感器：

这部分到后面再进行补充。

通气系统：

苹果通过气压平衡的系统使得主动降噪耳机带来的“耳压感”变得没有那么强烈。

## 智能音箱

智能音箱,即具有语音交互及信息处理能力的音箱。在未来物联网进一步发展的条件下,智能音箱可能与手机一样,成为有效的用户信息交换口,除了简单的语言交互外,还能在用户的命令下实现对家中任意智能家居的操控。

我个人的一些观点

批注 [刘33]: 2021.3.17 : 本观点略有落后, 请酌情参考。

我个人对国内大部分定价在 100-300 元范围内的智能音箱的定义是, 具有音箱功能的内植语音助手软件的设备。对于这类产品我的观点如下：

1. 目前这类智能音箱所提供的功能就只有简单的语音交互和信息检索。相比正式的产

品，它更像一个功能未开发齐全的工具。

2. 从功能方面，目前智能音箱的所有功能都可以被手机实现。与手机相比，音箱体积大、不易携带、缺少可视化交互(这意味着在存在用户选择时,信息交互效率极低)，唯一的优点是音量较大。
3. 由于缺少可视化交互，智能音箱的可播放内容大多是开发公司提前内置的，可播放内容及其有限。在前段时间的版权大战之后，智能音箱的可播放内容更是屈指可数。可以看出，依现在的趋势发展，智能音箱在未来只会成被 QQ 音乐、虾米、喜马拉雅等线上声音内容平台开发，并发展为其旗下的线下拓展硬件。
4. 结合 2 和 3 与其购买所谓的智能音箱，不如购买自身带有储存能力的便携式音箱。
5. 与 3 同理，在未来的物联网时代，只要智能家居拒绝向与自身品牌不同的智能音箱提供自身的操作接口，智能音箱就完全没有操作它们的可能。因此同理，智能音箱只会被智能家居企业开发，并发展为其旗下的硬件操作接口。
6. 这里我想在就语音操控做简单的展开。语音交互本身只是其他交互手段的一种补充，在非障碍条件下，它本身并不会成为信息交互的主流。原因在于，语言的产生是以人与人的交流为目的的，而人与设备之间只能称为交互。如果滑动一下手指能完成一项任务，说一句话也能完成一项任务，我一定还是会选择前者（对我而言，不论是划手指还是说话，都是我自己去做这件事，对设备说话并没有让我产生别人帮我完成某个任务的感觉）。语音输入之所以存在发展的必要，是因为很多迫不得已的情况下我们必须用语音（双手拿着东西或是在开车的过程中）。因此，如果智能音箱不是以某些障碍环境为前提进行而把语音交互输入作为主要信息交互方式的话，我认为这个设计方向一定是错误的。
7. 与其把智能音箱做成语音助手+大号扬声器，不如真正地实现音箱的智能化。如，

在多音箱条件下，通过对不同空间位置的音箱进行不同的发声任务分配，提供更优化的听觉体验。

#### 智能音箱发展的一些问题：

批注 [刘34]: 其实主要的问题都是语音识别上的问题。

##### 1. 必须使用唤醒词才能开启交互。

目前的智能音箱都必须要用唤醒词才能开启交互(这点倒不难理解,毕竟和人说话也要先说“你好”引起别人注意),且通常情况下,唤醒词是不允许更换的。这主要是因为语音模型是要提前训练的,更换唤醒词很可能导致新的唤醒词在原先的模型中并未经过训练。

##### 2. 对话过程中仍然需要用唤醒词进行激活。

在与人类交互时,除非我们明确结束了对话,否则一般会默认一切的对话都是与对方进行的,我们不需要频繁地说“你好”去开启对话。但和智能音箱或语音助手交流时,我们每次说完一句话,对话就会默认结束,下一次对话还需要再用唤醒词去唤醒它。问题主要的原因还是当前的语音交互流程被分为唤醒、响应、输入、理解、反馈,暂时无法实现连续的对话。听说腾讯和谷歌在这个问题上有突破,但还没有商用。

##### 3. 对话情景不连贯。

人与人的语言交流是连贯的,但人与机器的交流似乎连贯性很差。我们在前一句话中刚刚说过的内容,后一句话中就会被机器忘记。

##### 4. 无法区分打断输入的场景。

当我们在进行语音输入时忽然出现了需要打断输入的场景时,语音系统无法及时作出反应。如,当用户在向语音助手提要求时,忽然有人和用户进行了对话,用户作出了回应,那部分回应也会被当作有效输入。较好的解决思路有二,一是引入实时语义分析,判断是否对话内容与先前的内容相关(显然这一点已经遇到了3中的困难);二是通过其他输入设备来

判断当前用户的对话对象是否是硬件。

5. 以上的 1、2、3、4 其实本质上都是一点，无法实现全双工连续对话。

6. 监听安全问题

由于语音助手需要判断用户在任何时刻是否使用了唤醒词，因此相关智能设备实际上在任何时刻都是保持这录音设备的开放的，这就为安全问题带来了极大的担忧。

批注 [刘35]: 参考：  
<https://zhuanlan.zhihu.com/p/161563225>、

## 语音交互相关技术

声音处理技术：

1. 回声消除（AEC，Acoustic Echo Cancellation）：

回声的产生分声学回声（Acoustic Echo）和线路回声（Line Echo）。这里主要考虑由于扬声器发出的声音再次反馈到麦克风所导致的声学回声，而回声消除主要就是希望把这部分由扬声器传入麦克风中的信号消除。具体的做法是，通过记录扬声器之前发出的声波，将其振幅反转并进行适当调节后与麦克风中的声波叠加，实现大部分的回声消除。接下来，在经过 NLP（Non Linear Process）进行剩余回声的消除。

批注 [刘36]: 参考：  
<https://www.youtube.com/watch?v=BEBXj9A8jC8>。

2. 波束成形（beamforming）

在之前的“听觉感知硬件-降噪”部分已经讲过了。主要功能就是将原本辐射向各个方向的信号集中向一个方向发射，或将从任意方向收集信号的收集方式调整为从特定方向收集信号。

3. 降噪（NS，Noise Suppression）

在“听觉感知硬件-降噪”部分也讲过了。主动降噪的主要原理就是发出与传入噪声波形相反的声音，使之与噪声叠加，实现主动降噪。被动降噪的原理就是通过隔音材料避免外

部噪音进入耳朵。

## 语言处理算法：

### 1. 唤醒词（关键词）检索（KWS，Key Word Spotting）

唤醒词是用于唤醒语音交互设备，使之从休眠状态转为工作状态并准备开始进行语音交互的特定语音片段。

语音唤醒的整体流程都是，接收语音输入->根据特定算法判断确定是否需要唤醒->执行对应操作。语音唤醒能力主要依赖语音识别的算法。

### 2. 自然语言处理（NLP，Natural Language Process）

#### (1) 基本定义

自然语言处理就是用计算机来处理、理解以及运用人类语言，它的核心在于理解。

我们通常把人工智能分为运算智能、感知智能、认知智能和创造智能。自然语言处理属于其中的认知智能。比尔·盖茨曾说：“语言理解是人工智能皇冠上的明珠”。

#### (2) 发展与简单原理

自然语言处理主要有两大流派，一个是基于语言规则方法的符号派（也有叫理性主义的），一个是基于概率方法的随机派（也有叫经验主义的）。

符号派认为，人类语言是由语言规则来产生和描述的，因此只要能够以适当的形式将人类语言规则表示出来，就能理解人类语言。简单的说就是，通过理解句子的每个成分，再通过组合的方式，就能实现对整个句子的理解。然而现实世界中，语言的使用往往伴随大量的噪声和不规范，而符号派方法的鲁棒性较差，很难应对真实的语言情景；同时语言的复杂程度远远超过了科学家的一般构想，很难构造出有效且普适的语言模型。

批注 [刘37]: 参考：

<https://zhuanlan.zhihu.com/p/163047958>，其中讲述唤醒算法的部分其实应该是下面的 STT 的语音识别的算法。

批注 [刘38]: 参考：

<https://www.zhihu.com/question/28225747>，  
<https://www.jianshu.com/p/284a153c1cd6>。

批注 [刘39]: 参考：

[https://www.bilibili.com/video/BV17W411J75C/?spm\\_id\\_from=333.788.videocard.0](https://www.bilibili.com/video/BV17W411J75C/?spm_id_from=333.788.videocard.0)。  
（上面 NLP 标题中的第一个链接也提到了语音识别的发展）

随机派认为，这样从语言数据中获取语言统计知识，建立语言的统计模型，机器就能实现“语言理解”。可以看出，随机派的思路与当下采用机器学习实现识别的思路十分相似。然而在随机派被提出的年代，统计方法仍然较为原始，且尚未有足够的语言数据能用于统计学习，因此在当时随机派并非主流。

在二十世纪五十年代，从事自然语言处理的学者基本以基于语言规则的理性主义为主。但从二十世纪八十年代开始，理性主义方法不断遭受质疑，而基于统计的经验主义（即随机派）慢慢成为主流。尤其在计算机硬件大幅提升以及机器学习的方法兴起后，现在的方法几乎都是基于经验主义的。

### (3) 应用方向

目前自然语言处理主要的应用分两个方向，语言沟通和从大规模文本数据中提取信息。语言沟通还可被进一步分为智能对话和机器翻译等；从大规模文本数据中提取信息可进一步分为机器阅读理解、信息抽取、信息分析、文本分类等等领域。

## 3. 语音转文字（STT，Speech To Text）

STT 与 ASR（Automatic Speech Recognition）是相同的。

### (1) 基本定义

将用户输入的语音转化为文本文件以供后续使用的方式，也会被称为语音识别问题。

### (2) 发展阶段：

基于模板匹配。

通过特征提取将输入语音转化为特征序列，通过比较输入语音与词库中词语的特征模板的特征是否匹配来确定文字。

基于隐马尔科夫模型。

为每个基础音节训练一个模型，通过比较语音输入中的每个音节在各个模型的输出概率

批注 [刘40]: 参考：

《人工智能导论》，李德毅主编，p196-p200，  
[https://mp.weixin.qq.com/s?\\_\\_biz=MzI1NDY4NzUxNg%3D%3D&mid=2247483768&idx=1&sn=33777e5032567698f2b72930516704b5&scene=45#wechat\\_redirect](https://mp.weixin.qq.com/s?__biz=MzI1NDY4NzUxNg%3D%3D&mid=2247483768&idx=1&sn=33777e5032567698f2b72930516704b5&scene=45#wechat_redirect)，  
<https://www.zhihu.com/question/20398418/answer/18080841>，  
<https://www.bilibili.com/video/BV1FA411n7tt?from=search&seid=4244499031011170279>。



值，确定语音输入的具体音节，进而拼合成文字。

基于神经网络。

所有涉及神经网络的方法都属于此类。基于模板匹配的算法中，可以用神经网络实现特征提取；基于隐马尔科夫模型中，也可以使用神经网络模型进行概率输出。

### (3) 具体构成

批注 [刘41]: 我觉得只要记个大概的过程就好。

现在语音识别系统主要分为四个部分：特征提取、声学模型、语言模型和解码搜索。具体应用中，会先进行特征提取，再根据提取出的特征计算声学模型得分和语言模型得分，最后综合两个得分进行解码搜索

特征提取：

从原始语音信号中提取出与语音识别最相关的信息，滤除其他无关的干扰信息。具体做法为，将声音信号按照很短的时间切成一个个小段，称为帧。对于得到的帧，按照某种规则提取特征，将其变成一个多维向量，向量中的每一个维度可以看作描述了这帧信号中的一项特征。

声学模型：

处理编码得到的向量，将帧识别为状态（一种比音素更小的语音单位），将状态组合为音素（声音的最小结构单元，类似拼音中的声母和韵母等），将音素组合为单词。帧的状态的确定就是通过声学模型确定的，通过计算帧在不同状态的模型下的概率，其中最高的概率对应的状态即为所求。

语言模型：

调整声学模型得到的不合逻辑的字词，使识别结果变得通顺。

解码搜索：

搭建状态网络（先搭建单词网络，再由单词网络展开成为音素网络，由音素网络展开成

为状态网络)，通过在这条网络中搜索最佳路径，使得语音对应路径的概率最大，就是解码的过程。简单一点可以理解为，通过声学模型和语言模型来进行文字输出。

#### (4) 端对端识别

基于神经网络的发展，现在又衍生出了端对端的语音识别方式。简单的说就是输入端输入语音数据，输出端直接输出结果，中间的分析完全由神经网络完成。

#### 4. 文字转语音 (TTS, Text To Speech)

TTS 即将正常语言文本转化为语音的技术。文本转语音有两个步骤，文本处理和语音合成。

文本处理：

首先，需要保证文本中所有字词都转化文对应语言的单词（如数字 2020 需要被转化为二零二零或两千零二十）。接着把文本转化为音素序列，并标出每个音素的起止时间、频率变化等信息。这一步骤算是对语音合成文本的预处理。

语音合成：

语音合成这里又有两种思路，基于拼接的语音合成系统和基于参数的语音合成系统。

##### (1) 基于拼接的语音合成系统（拼接法）：

用大量事先录制的语音创建语音库，在需要合成时，通过文本中含有的音节、音素去库中进行搜索对应的音节、音素的发音，然后拼接成一段完整的语音。

其优点是，由于语音全部都是由真人录制的，因此语音质量较高。

缺点是，由于商用级别的语音库至少要 5 万句，因此成本较高。

##### (2) 基于参数的语音合成系统（参数法）：

通过统计模型来确定文本对应的声学特征（目前主流还是用神经网络实现），通过声码器（vocoder）把这些参数转化为波形来获得语音。

批注 [刘42]: 参考：

<https://www.zhihu.com/question/26815523> ,  
<http://www.woshipm.com/pmd/1381081.html>。

其优点是，对数据库要求较小，大约只要 5000 句即可。

其缺点是，与拼接法相比语音质量较差。

端到端的语音合成方法：

与语音识别一样，依赖深度学习可以实现从音素或字符直接合成语音的端到端合成。

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。

# 触觉部分

## 触觉机制

触觉也可被称为肤觉，是由分布在全身皮肤上的神经细胞感受器（机械感受器、温度感受器、伤害性感受器）接收来自外界的疼痛、温度、湿度、压力及震动的感觉。不同的感受器会对作用于皮肤的不同类型的刺激做出反应，如当有东西摩擦皮肤时，迈斯纳小体反应最为强烈；当一个小物体持续按压皮肤时，梅克尔触盘最为活跃。

皮肤对压力的敏感性在身体的不同部位的差异十分巨大，如指尖对刺激位置感觉的精确度是后背皮肤的 10 倍。

皮肤不同区域分辨两点刺激的最小距离叫两点阈。下面是皮肤不同区域的两点阈统计（单位为毫米）。

中指指尖	上唇	前颌	胸部	肩部	背部
2.5	5.5	15.0	36.0	41.0	44.0

## 相关应用

主要与触觉相关的传感器就是压力传感器，这块具体我到后面的传感器部分再展开。

触觉在设计上的应用主要是作为一种信息输入方式，如电容屏、电阻屏、3D touch 形变传感器。这些都统一放到后面讲。

# 嗅觉部分

## 嗅觉机制

嗅觉的产生是由于空气中的气味分子与鼻腔中嗅上毛膜上的受体蛋白发生作用 ,进而诱发神经冲动，刺激大脑额叶下方的嗅球而形成的。

## 相关硬件

### 电子鼻

电子鼻是通过化学传感器阵列和模式识别系统组成的 ,通过将样品的气味用真空泵吸入到传感器的气室中，在特定温度下与各个传感器进行作用，样品的“气味指纹” 就能被传感器感知并被识别为特定的模式，从而实现区分。

与其说电子鼻是鼻子，不如说它是气体检测器。

## 不足

嗅觉在设计领域的应用其实并不多，主要原因在与气味的难以复制。嗅觉与视觉不同，某个气味的产生依赖与许多不同种类的物质在数量上的不同配比 ,并不像视觉那样只需要能够显示 RGB 的像素点就能实现视觉信息的复制。嗅觉与味觉比较类似，都依赖复杂的工艺流程才能实现复制。

工业上，对嗅觉的应用一般也只在特定气体指标的检测等领域。

批注 [刘43]: 应用参考：  
[https://mp.weixin.qq.com/s/pg0c\\_hnhZh1BNMaHN9ZMYA](https://mp.weixin.qq.com/s/pg0c_hnhZh1BNMaHN9ZMYA)。

# 传感器

## 传感器的发展趋势

微型化：

传感器体积越来越小。

精细化：

传感器的精度（灵敏度）越来越高。

普适化：

传感器的使用场景越来越广，对环境的适应性和鲁棒性越来越强。

智能化：

由传统传感器向智能传感器发展，传感器本身除了获取信息外，还具有一部分的信息处理功能。

## 智能传感器

智能传感器的概念是，传感器不再只是信息接收和传递的入口，它还囊括了之后的分析、处理、储存等功能。我个人对此的理解是，这应该不叫智能传感器，而叫一体化设备。

## 传统传感器的基础知识

### 组成

敏感元件（输入）：

用于感受各种外界刺激信息，并将其转化为电信息的电子元件。

转换元件（输出）：

用于将敏感元件输出转换为适于运输和测量的元件。

基本电路：

用于配合实现信号输入和输出的电路。

### 传感器的物理输入与参数输出的关系

线性：

即输入与输出能以任何形式的函数曲线进行拟合。

非线性：

即输入与输出不能以函数曲线拟合。

### 静态特性与动态特性

静态特性：

在输入为不随时间变化的恒定信号时，传感器的输出与输入间的关系。可以简单地理解为一些描述传感器本身特征的属性。具体可以包括，量程、线性度、灵敏度、重复性、迟滞、重复性、分辨率、漂移等。

批注 [刘44]: 参考：

<https://www.zhihu.com/question/324331831/answer/684670915>。

这里简单对灵敏度、精度、分辨率做一下区分。灵敏度是指静态条件下，输出值的变化与输入量变化的比值；精度是指仪器在输出信号时的精确程度，简单的说就是小数点后的有效数字的多少；分辨率是指动态条件下，能引起传感器输出变化的最小输入变化值。

动态特性：

在输入为随时间变化的变化信号时，传感器的输出与输入间的关系。

## 传感器的选择标准

使用场景：

根据使用场景确定需要的传感器类型，能容纳传感器的大小，传感器是否需要与用户或外界环境相接触，具体价格，是否存在高温、潮湿、腐蚀性等特殊环境。

灵敏度：

根据预算和用途确定灵敏度。通常灵敏度越高的传感器价格越高。同时，灵敏度越高的传感器在现实环境中也有可能遇到信噪比较小的问题，主要是由于高灵敏度下传感器将环境中的噪声和有用信号一并接收造成的。因此，需要综合考虑灵敏度带来的利弊。

精度：

根据具体使用场合和预算价格确定，不需要无条件追求高精度。

稳定性：

传感器在不同条件下的鲁棒性。通常我们希望鲁棒性越高越好。



## 常见传感器（重点）

传感器中的具体原理大致了解即可，一般只要大概清楚有哪些分类就好。着重了解传感器的应用。传感器价格仅作参考。

批注 [刘45]: 推荐一本书：

现代传感器手册:原理、设计及应用，[美] 雅各布·弗雷登（Jacob Fraden）著。

## 温度传感器

用于将温度信号转化为可用输出信号的传感器。

### 分类

根据测量方式

根据测量方式可将温度传感器分为接触式和非接触式。

接触式：

检测部分与被测对象有良好的接触。传感器与被测物体通过传导或对流达到热平衡，从而使传感器本身的温度与被测的温度一致，实现测量。

优点是测量精度较高。

缺点是对于运动体、小目标或热容很小的对象会产生较大的测量误差。

非接触式：

检测部分与被测对象互不接触。一般的非接触式温度传感器都是使用黑体辐射原理，即任何温度高于绝对零度的物体都会不断向外发射红外辐射，通过获取红外辐射信号并进行放大、转化处理，获得温度值。

优点是可用于测量运动物体、小目标、热容小或温度容易迅速变化的对象的表面温度，可以测量温度场的温度分布。

批注 [刘46]: 参考：

<https://zhuanlan.zhihu.com/p/62344915>。

缺点是测量精度较低。

根据工作原理

根据电容方式可将温度传感器分为热电偶传感器、电阻式传感器、红外阵列传感器等。

这三种类型都在不同领域有所使用。

### 热电偶传感器

热电偶传感器利用的是塞贝克效应，即当给一个金属丝的两端（冷结端与热结端）施加不同的温度时，金属丝的两端会产生电动势，温差越高热电动势越大，闭合回路后金属丝中会有电流流过。具体测量温度时，将测量端与待测物体靠近，用待测物体发出的热辐射提升测量端温度，通过测量具体的电流（或电压）值，配合先前确定的温度与电流（电压）的相关表（或相关函数），确定具体的待测温度。

这里还有一点需要注意，由于金属丝两端的电势是与温差相关的，因此如果要确定待测物体的具体温度（热结端的温度），必须先确定冷结端的温度。实验条件下通常会令冷结端温度为零，但现实中这种方法一般无法实现。由于冷结端温度与环境温度一致，因此目前常用的做法是用一个额外的热敏电阻来测量环境温度。

具体应用时，会将金属丝缩小，然后多个串联，形成热电堆，用于提升对热辐射的吸收。此外还会在测量端前方用滤光片过滤掉待测量物体发出的电磁波以外的无效电磁波，以提高检测精度。

疫情期间常见的非接触式红外额温枪用的就是热电偶原理。那个红色的激光似乎只是用来帮助定位测量位置的。

热电偶传感器的优点在于测量温度范围广，耐用性强，可以实现非接触式测量。

缺点在于，信号调理复杂，精度较低，测量端易受腐蚀（测量端由两种金属构成）、抗噪性差。

批注 [刘47]: 参考：

<https://www.bilibili.com/video/BV1Qp4y1Y7aM?from=search&seid=13636110565596660460>，  
<https://zhuanlan.zhihu.com/p/48687237>，  
<https://www.bilibili.com/video/BV1q7411s7sR?from=search&seid=14128839594105582653>。

## 电阻式

使用热敏电阻作为输入元件，利用导体电阻率受温度影响的特性来实现测温的传感器。将测温电阻放到回路中，观察回路大小中的电流(或电压)变化，即可推知电阻的具体阻值，进而换算出温度值。

其优点在于，物理、化学性能稳定，工艺简单，价格较低。

其缺点在于：热电阻自身会发热，可能给测量带来一些误差；温度与阻值通常是非线性关系，对照关系无法依赖简单的函数，必须通过对照表；无法实现非接触式测量。

## 红外阵列传感器

同样是利用被测物体发出的热辐射来实现温度的测量的，不过此时的传感器不是金属丝构成的热结端，而是通过采集热辐射、利用光电效应形成电压的红外焦平面阵列。其主要特点在于，能够根据物体的红外辐射，形成具体的视觉图像，因此通常用于制作热成像仪。

从其原理看，红外热成像的技术似乎更像光探测器一些。

## 其他

除上述传感器以外，还有半导体 PN 结传感器（我怎么感觉这个好像就是 CMOS，也就是上面的红外阵列呢？）、光学温度传感器、声波温度传感器、压电温度传感器等，不过这些传感器不是主流，因此不过度展开。

## 价格

淘宝价格在几十到几百不等，最便宜的温度计也就 10 元。

## 湿度传感器

气体湿度为每百万立方英尺气体中所含的水汽磅数；液体和固体湿度表示为总重量(湿

批注 [刘48]: 参考：

<https://www.zhihu.com/question/381610829>。

其基础单元似乎是微测辐射热计，参考：

<https://zhuanlan.zhihu.com/p/28109727>。

重) 中的含水百分比。

湿度传感器的设计思路是,我们希望这个传感器对水分子有选择性,且其中的一些内部属性可以被水分子的浓度调节。

湿度传感器通常会与温度传感器放在一起,形成温湿度传感器。单独的湿度传感器只在部分领域有特殊用处,因此这里不会展开太多。常见的温湿度传感器由电阻式温度传感器和电阻式湿度传感器组成。

## 分类

根据原理分类,较为常见的湿度传感器可分为电容式、电阻式。其中较为常见的是电阻式。

电容式:

将某个湿敏材料制成的薄层(通常为高分子薄膜)放置到金属板点击构成的电容中充当绝缘介质,由于湿敏材料的介电常数会随湿度的变化而变化,电容的大小也会相应的发生变化(通常为线性关系)。通过检测电容大小的变化,就可以确定环境的湿度。

电阻式:

有些导体的电阻率会随湿度的变化而变化,通过将这种材料串联进电路中,检测电路的电流电压变化,计算材料电阻,来确定具体的湿度值。通常湿度越大,湿敏电阻的电阻值越小,但湿度与电阻值间并非线性。

其他:

除了以上的湿度传感器外,还有热导式湿度传感器、光学湿度计、振荡湿度计等,不展开。

价格

基础传感器检测模块在淘宝似乎 10 元以下就可以买到，成形产品价格几十到几百之间。

## 加速度传感器

加速度计用于测量物体直线的加速度。传感器在加速过程中，通过对质量块所受惯性力的测量，利用牛顿第二定律获得加速度。

任何加速度传感器的最基础的构造是一个两边牵引着尾部固定的弹簧的质量块，当质量块在弹簧方向上受力并发生运动时，一边的弹簧会被挤压，另一边的弹簧会被拉伸。通过确定弹簧的形变大小，结合胡克定律和牛顿第二定律，即可确定对应方向上的加速度。

由于一个加速度传感器（单轴加速度传感器）只能检测一组方向上的加速度变化，因此如果要检测空间中所有方向上的加速度变化，需要使用三个加速度传感器（三轴加速度传感器）。

加速度传感器可以实现 10kHz 的高频检测。

分类

对加速度传感器的分类主要体现在对于质量块的位移或质量块所受到的力的测量方法上。根据原理分类，加速度传感器可分为电容式、压阻式、压电式等。其中，较为常见的是电容式。

电容式：

传感器中的电容有两块固定极板和一块可移动的极板（可移动极板与特殊的弹簧相连，

批注 [刘49]: 参考：  
<https://zhuanlan.zhihu.com/p/169548572>。

批注 [刘50]: 参考：  
<https://www.bilibili.com/video/BV1EK4y1v75R?from=search&seid=16167164259378783294>，  
<https://www.bilibili.com/video/BV1NJ41117B8?from=search&seid=11993166544847936683>。

这里的可移动只是可在某一个方向上移动)构成。可移动极板在加速度作用下会发生位置上的偏移进而导致两块固定极板中的电容量的变化。通过对两块固定极板间的电容差的检测,即可确定物体在特定方向上的加速度大小。

#### 压阻式:

传感器中会在弹簧(或在现在较新的设计中,用单个横穿质量块的悬臂梁代替两边的弹簧)上添加一个应变片,用于检测弹簧(或悬臂梁)在收到加速度后所发生的形变所导致的电阻值变化(材料在外力作用下产生形变时,其电阻值也会相应发生变化,这种现象称作“应变效应”,也成“压阻效应”)。通过电阻变化值,反向推导具体的加速度大小。

简单的说,就是利用发生移动的质量块对电阻产生压力,压力作用导致电阻阻值发生变化,通过检测阻值变化可以反推压力大小,通过压力大小和质量块质量可以反推加速度大小。

#### 压电式:

直接用压电晶体替换原先的弹簧。在质量块收到加速度作用时,会对压电晶体施加力。压电晶体在受到力的作用下,内部会产生极化现象,与力同向的两个相对表面会出现正负相反的电荷(这种现象被称为“压电效应”)。通过电路进行电荷-电压转换,即可反向推知加速度大小。

#### 其他:

此外还有,热加速度传感器等,不做展开。

#### 应用

实际应用中,加速度传感器可用于检测物体任何直线方向上的速度变化,如物体的轻微移动、物体的振动、物体的碰撞等。

批注 [刘51]: 参考:

[https://www.sohu.com/a/107079934\\_454733](https://www.sohu.com/a/107079934_454733)。

批注 [刘52]: 参考:

<https://baike.baidu.com/item/%E5%8E%8B%E7%94%B5%E6%95%88%E5%BA%94/4515291?fr=aladdin>。

一般小型设备中：

kinect 的倾斜补偿，手机微信的摇一摇、记步数等功能。

汽车上：

汽车安全气囊、汽车 ABS 防抱死刹车系统，都是利用加速度的突然变化来做出“发生碰撞”或“需要紧急减速”的判断，进而启动气囊或 ABS 系统。

无人机上：

在失去 gps 导航时，无人机会切换到惯性导航。加速度计可以配合陀螺仪和磁力计实现惯性导航。

**批注 [刘53]:** 在惯性测量单元的应用里解释了什么叫惯性导航。

价格

淘宝价格在几十到几千不等。低价的基本在 100 元以内，中等价位一般在 500 元以下。

## 陀螺仪

陀螺仪早在 18 世纪就被发明了，当时的陀螺仪都属于机械陀螺仪，需要转子及用于固定转子的多层框架构成。在转子保持高速旋转时，只要旋转轴未受到垂直于自身的力（也可以说是旋转轴在转子旋转所在平面未受到任何力），转子就能使旋转轴始终保持某个特定朝向。也就是说，保持高速旋转的转子，配合多层外部固定装置，可以在陀螺仪所在空间发生角度变化时始终保持自身朝向稳定。这一特性被称为定轴性。

由于陀螺仪的定轴性，陀螺仪可以作为优秀的定向装置，也可以用于检测陀螺仪所在空间的角动量变化。

此外，陀螺仪还有进动性。这一特性语言描述比较麻烦，建议直接看上面的注释中的视频。进动性理解比较困难，实际使用也不是很多。

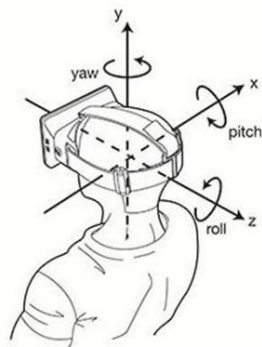
**批注 [刘54]:** 经典陀螺仪的基本原理涉及角动量及力矩等概念，相当不直观，建议参考下面这个视频中的 20 课和 24 课：

<https://www.bilibili.com/video/BV1u4411y7zq?p=24>。

此外，具体的陀螺仪的相关视频可以参考下面这个：  
<https://www.bilibili.com/video/BV1qJ411272r>。

现在手机或其他电子设备中虽然也有陀螺仪，但使用的已经不是传统的机械陀螺仪了，一般都采用 MEMS（微电机）陀螺仪，通过测量旋转系统中的科里奥利力来判断角速度，由此测量方向变化。

陀螺仪会分单轴陀螺仪和三轴陀螺仪，单轴即只能检测一个方向上的旋转，三轴即可以检测任意方向上的旋转。实际上，他们的差别就在于传感器数量的多少。单个陀螺仪传感器就只能检测一个方向上的旋转，所以就被称为单轴；而三个传感器就可以实现空间内所有方向上的角度检测，所以就被称为三轴。具体对单轴或三轴的选择根据实际的应用场景和预算规模确定。



三轴陀螺仪的三个轴分别称为 pitch（俯仰角）、yaw（航偏角）、roll（滚动角）。

## MEMS 陀螺仪

MEMS 陀螺仪依托于对科里奥利力的测量。科里奥利力是对旋转体系中进行直线运动的质点由于惯性相对于旋转体系产生的直线运动的偏移的一种描述。它的方向通常与旋转体系中的指点的运动方向相垂直。

科里奥利力的计算公式为  $F = 2 * m * \omega * v$ 。其中  $\omega * v$  是向量叉乘。该公式用于描述在某个平面中具有速度为  $v$  的点，在平面受到转速为  $\omega$  的、旋转轴与平面垂直的转动时

**批注 [刘55]:** 这里关于科里奥利力的公式推导我没有找到很好的材料。（内容了解即可）

有一些基本的 MEMS 陀螺仪设计的内容可以参考：  
[https://www.bilibili.com/video/BV1VJ411r7c7/?spm\\_id\\_from=333.788.videocard.6](https://www.bilibili.com/video/BV1VJ411r7c7/?spm_id_from=333.788.videocard.6) ,  
[https://www.bilibili.com/video/BV1Cb411Y73a/?spm\\_id\\_from=333.788.videocard.0](https://www.bilibili.com/video/BV1Cb411Y73a/?spm_id_from=333.788.videocard.0) ,  
<https://www.bilibili.com/video/BV1YK411A7HU>。



“受到”的科里奥利力的大小（严格的说，科里奥利力是惯性力，并不真实存在）。两边除以  $m$  后即可得到加速度的表达式， $a = 2 * \omega * v$ 。

传感器的设计及测量原理与电容式加速度传感器相似。传感器的电容有一个固定的极板，还有一个可在两个方向上自由移动的极板。在传感器开始运行后，会让可移动极板在某个方向上开始进行速度确定的周期运动，使移动极板获得某个确定的初速度。此时，如果发生了旋转轴垂直于传感器所在平面的旋转，可移动极板就会在“科里奥利力的作用”下产生垂直于之前周期运动方向（但不垂直于传感器所在平面）的加速度。通过电容的变化可以计算加速度  $a$  的大小，由于速度  $v$  是确定的，因此可以通过科里奥利力的公式反推得到旋转角速度  $\omega$ 。

不过，由于单个传感器只能实现对传感器所在平面的角速度大小的检测，因此如果需要实现对空间中三个维度的角速度的全面监测，就需要在各个方向上都安置一个陀螺仪传感器。

#### 激光陀螺仪

激光陀螺仪是基于萨格奈克（Sagnac）效应制作的。具体的，在一个由两个半圆构成的圆环形的光学环路中，在这两个半圆的一段向半圆内发射激光，则激光会在半圆的另一端交汇，且两束激光走过的距离是相同的。此时，如果让圆环发生旋转轴垂直于圆环平面的转动，再在某一位置发射激光，此时激光交汇的位置仍然是半圆的另一端，但两束激光走过的位置将不再相同，即光程大小将发生变化。此时，两束激光就会出现干涉。通过采集光波干涉条纹和对应的时刻，就能计算出圆环转动的角速度。

激光陀螺仪拥有非常高的可靠性和精度，且可以在及其恶劣的环境下工作。但其缺点是体积较大，难以在日常穿戴设备中应用。

批注 [刘56]: 参考：

<https://www.bilibili.com/video/BV1PK4y1E749?from=search&seid=12075678393827673851>。

## 应用

陀螺仪的应用一般都是利用了它的定轴性,利用陀螺仪检测设备的旋转并根据其旋转信息执行后续的操作。一般除了航天航空、航海、制导等领域会使用激光陀螺仪外,日常的中小型设备都使用 MEMS 陀螺仪。

### 导航定向

航天航空、航海、导弹制导领域中可以用陀螺仪实现稳定的指向导航,稳定性好、抗干扰性强。

另外在手机中其实也会使用陀螺仪帮助导航。在地图应用中,用于表示用户自身所在的位置的箭头是会随用户的转向而旋转的,这部分功能就是由陀螺仪实现的。

### 倾斜检测

手机或其他电子穿戴设备中会使用陀螺仪检测设备的旋转和倾斜,并将旋转参数作为一些应用的信息输入。最直接的引用就是手机从背面翻到正面会自动亮起屏幕,此外现在有许多游戏也将手机的倾斜和旋转作为操控手段辅助玩家操作。

### 相机防抖

#### (1) 电子防抖

通过在拍照时检测机身抖动来调节相机的快门速度、ISO、光圈大小等参数(如在检测到抖动时加快速度),从而实现更好的成像结果。

#### (2) 光学防抖

根据抖动对磁力悬浮镜片镜头进行位移补偿,复原光线因抖动而产生的偏移。

#### (3) 传感器防抖

根据抖动对可自由浮动的光学传感器(CCD 或 CMOS)进行位移补偿来实现防抖的目的。

批注 [刘57]: 参考:

<https://zhuanlan.zhihu.com/p/22942615>,

<https://zhuanlan.zhihu.com/p/267253413>。

相机防抖和云台稳定这块都是结合陀螺仪与加速度传感器共同作用的。

## 云台稳定

通过陀螺仪确定云台控制臂的旋转,通过多自由度的摇臂上的电机施加反向的动力,实现对云台的稳定。

## 惯性导航

在失去 gps 导航时,导航设备通常会切换到惯性导航。陀螺仪可以配合磁力计和加速度传感器实现惯性导航。

**批注 [刘58]:** 在后面“惯性测量单元”的应用里解释了什么叫惯性导航。惯性测量单元其实就是加速度传感器加陀螺仪。

## 价格

淘宝价格在几块到几百不等。低价的低于 10 元,中等价位大致集中在 400-700 这个区间段。

## 磁力传感器 (通常又称磁力计)

磁力传感器可以用于测量传感器周围的磁场强度大小。目前对磁力传感器的应用通常都是测量地球磁场,并将信息用于电子罗盘进行方位确认。

## 原理

目前大部分的磁力传感器都是基于霍尔效应的,因此通常也会被称为霍尔传感器。霍尔效应是指,如果将一个扁平的导电带置于垂直于导电带的磁场中,在其左右两端添加触点连接电压表,在其上下两端添加触点连接电源。在磁场的作用下,通过导电片的电子会向左右中的某一段聚集(根据磁场方向确定),随后在电压表两侧形成电压。通过测量电压大小,就能确定磁场大小。

**批注 [刘59]:** 参考 :  
<https://www.bilibili.com/video/BV1Fh411o76H?from=search&seid=12607881807158629394> ,  
[https://kreader.cnki.net/Kreader/CatalogViewPage.aspx?dbCode=CMFD&filename=1016785558.nh&tablename=CMFD201701&compose=&first=1&uid=WEEvREcwSIJHSldSdmVqMVc3ejUxUnlINWpMNURWUDRhM3JmamtkY2c2TT0=\\$9A4hF\\_YAuvQ5obgVAqNKPCYcEjKensW4ggI8Fm4gTkoUKaID8j8qFw!!](https://kreader.cnki.net/Kreader/CatalogViewPage.aspx?dbCode=CMFD&filename=1016785558.nh&tablename=CMFD201701&compose=&first=1&uid=WEEvREcwSIJHSldSdmVqMVc3ejUxUnlINWpMNURWUDRhM3JmamtkY2c2TT0=$9A4hF_YAuvQ5obgVAqNKPCYcEjKensW4ggI8Fm4gTkoUKaID8j8qFw!!)。

由于单个磁力传感器只能测量与其中导电带垂直的方向上的磁场大小,为了保证测量三维空间中各个方向上的磁场,通常会用三个磁力传感器构成三轴传感器。此外,还需要配合陀螺仪,实现对设备自身的旋转的补偿。

但由于地球磁场相对于设备内部元件产生的磁场较小,因此传感器对地磁的测量很容易受到设备本身磁场的干扰。所以要想使传感器获得良好的测量结果,必须进行校准。在实际使用手机的电子罗盘时我们也会发现,通常我们刚刚打开应用时的指向是存在偏差的,需要经过一段时间的使用后,方向的指向才会逐渐趋于正确。这就是因为电子罗盘在使用过程中在不断地进行自身校准。

## 应用

### 提供方向信息

手机中的电子罗盘就是通过磁力计感应地磁确定南北方向的,简单地说就是充当指南针的作用。

### 惯性导航

在失去 gps 导航时,导航设备通常会切换到惯性导航。磁力计可以配合陀螺仪和加速度传感器实现惯性导航。

### 车位检测

在车位下方地块内嵌入无线地磁传感器,当车辆靠近时,由于车辆本身的金属成分极多,会影响传感器上方的地磁场,因此可以确定某个车位上是否存在车辆。通过将相关信息上传到相关服务器,就能实现对车位的实时监测,确定车位上的车辆的停放时间;同时也能确定一个停车场内有多少车位空缺。

**批注 [刘60]:** 在惯性测量单元的应用里解释了什么叫惯性导航。

**批注 [刘61]:** 参考：  
<https://www.zhihu.com/question/274279808>，  
<https://zhuanlan.zhihu.com/p/73009743>。（后面这个简单了解一下就好）  
除了地磁传感器外，超声波传感器或一般的摄像头与分析算法的组合也能实现相关的功能。

## 价格

淘宝价格从几元到百元出头不等。有些商家会把磁力计与加速度计及陀螺仪一起组成姿态传感器一起出售，大概价格在 150 元左右。上面说到的车位上使用的地磁传感器价格在 200 元到 700 元不等。

## 压力传感器

用于将压力信号转化为可输出的电信号的传感器。一般的压力传感器都是检测固体压力的，最常见的应用就是手机的触控屏；不过也有一些情况会检测气体压力，如通过气体传感器确定海拔高度。

### 原理

#### 电容式

将固定电极作为电容的一极，易受压力发生形变的薄膜作为电容的另一极。当压力作用于薄膜时，薄膜的形变使得电容的电容量发生变化，通过测量电路可输出与压力成一定关系的电信号，从而确定压力的大小。

#### 压阻式

根据压阻原理，作用于导体的力会改变导体的电阻大小（一般为压力越大，阻值越小）。压阻式压力传感器将导体（通常为硅膜片）接入到电路中，通过记录导体的电阻变化，来确定作用于导体的压力大小。

在气压计、电子秤等硬件上，可以用压阻式压力传感器确定具体的压力大小。

**批注 [刘62]:** 关于电容式触摸屏和电阻式触摸屏的简单介绍：

<https://zhuanlan.zhihu.com/p/97130485>。

**批注 [刘63]:** 参考：

<https://baike.baidu.com/item/%E7%94%B5%E5%AE%B9%E5%BC%8F%E5%8E%8B%E5%8A%9B%E4%B%A0%E6%84%9F%E5%99%A8/8836217?fr=aladdin>。

下面这个链接中是电容式压力传感器在气压计中的应用：

<https://zhuanlan.zhihu.com/p/164696949>。

**批注 [刘64]:** 参考：

<https://www.bilibili.com/video/BV1Q54y1D7vg?from=search&seid=11092445362021764720>。

## 其他

此外，还有可变磁阻式压力传感器、光电压力传感器等等，这里不做展开。

## 应用

压力传感器在不同的行业中都会有独特的应用场景，这里只对一般日常生活中可能接触到的应用场景进行列举。

### 苹果手机的 3d-touch

3d-touch 就是通过在屏下添加一层（薄膜状的）电容式压力传感器实现的。通过压力传感器确定用户的触碰操作的具体压力大小，进而做出不同的反馈。不过这个功能由于作用有限，且成本不低，后来被苹果用长按的交互方式替代了。

### 接触式压力测试

最典型的应用是电子秤。当人站在电子秤上后，通过测量传感器中的电容或电阻的变化，确定电子秤上人的体重。

### 非接触式压力测试（气压计、液压计）

最典型的应用是气压计。除去一些特别的工业场合和气象场合，一般情况下，用气压计测得气压大小，进而确定海拔高度。

## 其他

压力传感器还有其他诸多应用领域，如开关，按钮等。

## 价格

用于电容式、电阻式压力传感器的价格大致都在几十元；气压传感器的价格大致也在几十元左右。

批注 [刘65]: 参考：

<https://www.zhihu.com/question/35535651> ,  
<https://product.pconline.com.cn/itbk/sjtx/iphone6s/1711/10253109.html>。

网上对 3d-touch 的原理的分析内容似乎很少，上面这些虽然提到了，但讲得也不是很清楚。根本原理还是压力导致电容变化。

## 触觉传感器

触觉传感器通常被认为是一种特殊的压力传感器。其特征是厚度小，且触摸发生时，与传感器的接触面积很小。触觉传感器一般检测是否发生触摸，而不检测触摸的力的大小。

触觉传感器与压力传感器存在着很大程度上的相似之处，因此不必过分区分两者间的差异。

现实中，触摸传感器较为广泛的应用就是触摸屏。

### 原理

#### 电容式

一般的电容式触摸传感器的原理和电容式压力传感器的原理类似，都是利用接触导致的电容大小变化来判断是否发生触碰或按压的。

现在大多数触摸屏都使用电容式触摸传感器，但原理略有不同。触摸屏由两层透明的、印刷在聚酯纤维（PET）上的菱形网格组成，其中间的聚酯纤维是一层光学透明的绝缘层。而两层菱形网格由一种叫氧化铟锡（ITO）的透明材料打印而成，作为电容的两极。由于人是导体，人的手指在触碰到屏幕时会与上层的菱形网格形成微弱的电流（不过到底是电子从网格流向人体还是人体流向网格我实在没找到具准确的答案，我个人倾向于从网格流向人体），进而导致上层菱形网格在触摸处的电容发生变化。通过扫描确定电容的变化的位置，即可确定触摸点的坐标。

戴上手套就无法使用电容式触摸屏，其原因就在于，手套通常是由不导电的材料制成的，因此无法与屏幕中的 ITO 层产生电荷交换、形成电流，手机也就无法确定具体的触碰位置。

另一方面，电容式触摸屏支持多点触控。

批注 [刘66]: 关于电容式触摸屏的较为深入的介绍：  
<https://www.bilibili.com/video/BV1DJ411T77Z?from=search&seid=12593938055198260023>。

批注 [刘67]: 具体原理我去知乎提了个问：  
<https://www.zhihu.com/question/427859615>。

当手上沾上水时，屏幕触碰操作就会失灵。其原因在于，平时我们与触摸屏的接触只有指尖的很小的一个面积，而沾上水后这个面积被大大扩大，从而导致处理器认定触摸范围内有大量点被激活，因此导致操作失灵。

批注 [刘68]: 参考：  
<https://www.zhihu.com/question/22497699>。

## 压阻式

压阻式触觉传感器有两种具体方式——利用压阻效应和利用弹性体。两者的基本思路都是导体电阻会随作用于导体的压力大小变化而变化。压阻式触觉传感器其实更像压力传感器一些。

### 利用压阻效应

在导体在某个方向上受到应力时会产生压阻效应，即对应方向上导体电阻会随压力作用产生变化，通过检测电阻变化即可反推压力大小。

压阻式的触摸屏就是利用压阻效应的，但由于压阻式触摸屏只支持单点触控，并不支持多点触控。因此现在智能硬件领域使用压阻式触摸屏的产品并不多。

## 利用弹性体

批注 [刘69]: 主要参考《现代传感器手册》，雅克布·弗雷登，p258。

压阻式触觉传感器使用力敏电阻 ( FSR )，力敏电阻是导电弹性体或压敏浆料。力敏电阻的电阻随压力的变化既可以认为是压力造成弹性体与( 夹在压敏电阻两侧的 )导体板的接触面积的变化造成的，也可以认为是弹性体的厚度变化造成的。

利用弹性体制作的触觉传感器也是可以感知力的大小与电阻大小变化的，因此也可以认为它是一种压力传感器。

由于利用弹性体做成的触觉传感器面积大、具有柔性、成本低，因此被非常适于做机器人的电子压敏皮肤。

## 开关式

批注 [刘70]: 主要参考《现代传感器手册》，雅克布·弗雷登，p255。

这种触觉传感器比较简单。通过两个金属片和一个垫片组成，垫片上有圆形的孔，上方



金属片接地。当对传感器施加压力时，金属片会通过垫片中的圆孔的位置形成回路，由于上方金属片接地，此时电路中的信号为 0。没有压力时，由于电路中没有元件接地，因此信号恢复。

这种触觉传感器一般用于成本较低的消费级产品，如遥控器和玩具。而且现在使用的很少。

#### 压电式

在导体受到压力时，会产生压电效应，在导体的两个表面出现正负相反的电荷。它将影响原先作用于导体的电信号。通过电信号的变化可以就确定压力的大小。

一般的压电式触觉传感器都会使用压电薄膜设计。压电薄膜厚度小、灵敏度高、不产生能量消耗，因此在医学、护理领域有较多使用。

#### 光学接触式（红外）

从我收集到的资料来看，光学接触传感器同样也分为两种，一种是有亚克力板或玻璃板作为光线传递介质的，一种没有。

两种类型的光学接触传感器的基本构造都是在被触碰区域（假设为长方形）的相邻的两条边框上安装光线发射阵列（通常为红外发光二极管阵列，即发射红外线的 LED，Light Emitting Diode），再在其对面安装光电检测器阵列。

光学触觉传感器都支持多点操作。但光学触觉传感器也都有一个缺陷，由于太阳光中也有大量的红外光，因此在室外环境，光学接触式触觉感受器的灵敏度会收到很大影响。

#### 有亚克力板做介质

有亚克力板做介质的情况下，利用的是光线的全反射原理。全反射原理为，当光线从光密介质射向光疏介质时，如果入射角度较大，则可能不发生投射，只发生反射。刚好发生全反射时的最小角度被称为临界角，临界角大小与光线类型及两侧介质的折射率有关。

批注 [刘71]: 主要参考《现代传感器手册》，雅克布·弗雷登，p264。

在手未触碰亚克力板时，与亚克力板接触的介质是空气。令 LED 发射的红外光与亚克力板成一定角度。如果这个发射角度恰好为介质为空气时的临界角，则射出的红外光将在亚克力板中经过全反射，最后被另一侧的检测器检测，在此过程中光线强度不会发生改变。

而如果手指触碰到了亚克力板，此时与亚克力板接触的介质就不再是空气了。人体皮肤的折射率相对空气要更高，临界角更大。此时，原先的入射角小于现在的临界角，因此就不会再发生全反射，有部分光线会以折射的形式逃逸。在另一侧的检测器进行检测时，光线强度将变弱。

因此，在实际使用中，只要分别观察横向和纵向的哪一个光电检测器阵列中检测到的光线强度变弱，就可以确定具体的触碰位置。

无亚克力板做介质

无亚克力板做介质的光学接触传感器的原理更加简单。简单的说就是，当手从屏幕上方经过时，会遮挡一部分的红外光，只要分别观察横向和纵向的哪一个光电检测阵列中检测到的光线强度变弱，就可以确定具体的触碰位置。

由于无亚克力板做介质的红外光阵列必须再屏幕上方排布，因此很难实现无框屏幕；

应用

触摸屏

触觉传感器现在最广泛的应用就是触摸屏，而触摸屏中的主流是电容式触摸屏，原因在于：电容式触摸屏小巧、轻薄、检测精度高、受日光影响小。

另外，与电阻触摸屏相比，电容式触摸屏支持多点操作，而电阻触摸屏并不支持多点操作。

具体的原理可以参考上面的介绍。

**批注 [刘72]:** 我在人机交互接口中再细讲电阻触摸屏的原理。  
或者直接参考：  
<https://blog.csdn.net/qlexcel/article/details/82731055>。

## 指纹识别

指纹识别可分电容识别、光学识别和超声波识别。

### 电容识别

电容识别原理与触摸屏的原理类似,通过将电容极板做小,提升空间密度,当凹凸不平的指纹靠近识别装置时,指纹不同高度的凸起与电容极板的距离不同,因此各个极板间的电容也不同。通过识别电容大小,再进行额外的信号处理,就可以识别出指纹的纹路。

### 光学识别

光学识别的原理与光学触觉传感器类似,主要利用全反射原理。光学识别的指纹触碰处是一个棱镜,棱镜的一边放置光源,另一边放置光电探测器(或相机)。由于手指的触碰会使指尖凸起部分与玻璃接触,那些位置的全反射临界角将变大,因此部分光线会发生逃逸。此时在另一侧的光电探测器就能检测到与手指接触部分的反射光的强度下降,以此确定指纹。

现在流行的屏下指纹识别也是用这一原理实现的,只不过光源是 OLED 屏幕发出的光。

### 超声波识别

超声波识别指纹的原理与声呐类似。向手指表面发射超声波脉冲,由于指纹的凹凸不同,对超声波的反射也有所不同,通过传感器手机反射回来的信号,即可确定指纹的三维空间信息。

不过现在,使用这一指纹识别方式的三星手机似乎暴露出了重大技术漏洞。

### 婴儿床垫

通过在婴儿床垫下放置压电式触摸传感器,检测儿童睡眠时的呼吸频率,在儿童发生窒息时及时提出报警。

**批注 [刘73]:** 指纹识别除了电容识别外还有光学识别(屏下指纹识别原理)和超声波识别等方式,具体可参考:

<https://www.bilibili.com/video/BV1ZW411g7kJ?from=search&seid=6608823661819813988>,  
[https://www.zhihu.com/market/paid\\_magazine/1255913197112541184/section/1255913940376662016](https://www.zhihu.com/market/paid_magazine/1255913197112541184/section/1255913940376662016)。

**批注 [刘74]:** 主要参考《现代传感器手册》,雅克布·弗雷登, p264。

## 触觉手套

这是目前 MIT 在做的一个项目。在一个手套上继承 548 个触觉传感器（似乎是压阻式的），使机械手拥有与人的皮肤一样的高精度触觉感知能力。通过手套在操作物体时收集到的压力特征，结合机器学习，为高精度物体识别提供了新的可能性。

## 价格

淘宝触摸传感器价格在几元到几十元不等；触摸屏一般不会单独出售氧化铟锡层，都是直接出售整块屏幕的，价格大致在 100-300 元。

## 光探测器

光探测器能够将周围环境的光线强度改变转变为电信号的改变。光探测器很多都利用了光电效应。

## 原理

### 光敏二极管

正常情况下，由硅制成的二极管时无法导电的。通过向二极管两侧中添加 P 型杂质和 N 型杂质，可使得二极管的 P 侧有电子空穴，N 侧有电子。

若光敏二极管在电路中是反向偏置的（即二极管被接入电路中，正极与 N 侧相连，导致电子聚集在 N 侧，不向 P 侧移动，因此无法形成电流），当光线照射光敏二极管时，流经光敏二极管的电流会明显增加。

电流产生的原因是光子对二极管的碰撞会在 PN 结的两侧产生电子-空穴对（光电效应），电子涌向正极，空穴涌向负极，打破原先形成的内部势垒形成电流。（电流产生的原因不重

批注 [刘75]: 参考：

<https://zhuanlan.zhihu.com/p/67628663>。

批注 [刘76]: 有一些二极管的基础知识可以参考：

<https://www.bilibili.com/video/BV15x41147Yn?from=search&seid=7945846469318970661>，其中就讲述了什么是正向偏置、反向偏置、空穴、势垒等概念。更书面、系统一点描述可以看这个：<https://zhuanlan.zhihu.com/p/143791710>。

要，了解就好）

因此，可以通过检测二极管所在回路中的电流来确定光照的大小。

#### 光敏晶体管

光敏晶体管与光敏二极管原理类似，在光线照射晶体管时，流经晶体管的电流会增大，因此可以通过检测回路电流确定光照大小。光敏晶体管的具体原理也是在光子碰撞时产生电子和空穴，电子发生流动，进而形成电流。与光敏二极管相比，光敏晶体管的灵敏度更高。

#### 光敏电阻

光敏电阻的阻值会随表面光照发生变化（一般为光照越大阻值越小），因此通过检测光敏电阻的阻值就可以知道光照强度的大小。

#### 应用

##### 环境光传感器

手机听筒周围一般会放有环境光传感器，当环境光变强时，手机屏幕会变亮；环境光变弱时，手机屏幕会变暗。这一功能是屏幕亮度能够随周围环境的明暗调节，以达到降低人眼负担、减少设备功耗的目的。

#### 血氧传感器

不同状态的血液对不同波长的光线有着不同的吸收能力。氧化态的血液（动脉中的血液）会吸收较多的红外光，而还原态的血液（静脉中的血液）会吸收较多的红光。

一般医院中都会有脉搏测氧仪（夹在手指上的设备），通过在顶端释放红光和红外光，在底端接收，根据接收到的红光和红外光强度即可确定血氧含量。

而现在的智能穿戴设备一般是将光线发射器和检测器放在同侧。光线在经过皮肤时，有一部分发生穿透，一部分发生反射。检测器通过检测反射回来的光线强度确定血氧含量。

**批注 [刘77]:** 光敏电阻阻值变化的原理在《现代传感器手册》，雅克布·弗雷登，p439，这里不多展开。

**批注 [刘78]:** 参考：  
<https://zhuanlan.zhihu.com/p/259985457>，  
<https://support.apple.com/zh-cn/HT211027>。

以 apple watch 为例，现在的 apple watch 配备 3 个光线发生器：LED 红光、LED 绿光及红外光（LED 应该只适用于检测脉搏的）。除了血氧测量外，接下来要提到的脉搏测量也需要用到 LED 绿光。

#### 光学心率传感器

心率的测量方法为光体积描记法。血液能够反射红光吸收绿光，当血液流量大时，绿光的吸收量就大；当血液流量小时，绿光的吸收量就小。当心房收缩，血液被压出时，血液流量大；当心房舒张时，血液流量小。

通过向皮肤发射绿光并用传感器检测被血液反射回来的绿光，就可确定绿光的吸收量大小，进而计算出脉搏大小。

#### CCD 与 CMOS

CCD 与 CMOS 是两类主要的光学传感器，详细内容参考之前视觉部分。

#### 红外热成像

详见温度传感器中的描述。

#### 紫外线探测器

与 CCD、CMOS 及红外热成像类似，紫外探测器也是利用光电效应实现紫外线检测的，其基本单元也是光敏二极管。

与之不同的是，紫外线探测器需要实现对可见光的分离。为了实现对可见光的分离，紫外探测器需要使用禁能带（能隙）更长的材料。对于半导体或绝缘体的材料，电子必须通过禁能带才能到达导带，实现导电。禁能带越长，光电效应产生的电子就越难到达导带，实现导电。此时对于波长较长的可见光而言，光电效应产生的电子数量较少，很难在光敏二极管中产生电流，而对于波长较长的紫外线而言，光电效应产生的电子数量较多，可以形成电流。如此就实现了屏蔽可见光的目的。

批注 [刘79]: 参考：

<https://support.apple.com/zh-cn/HT204666>，

批注 [刘80]: 参考：

《现代传感器手册》，雅克布·弗雷登，p428，p444。

此外，在光敏二极管的透镜也需要使用对紫外光透明的材料来制作（通常为石英）。

紫外线探测器常被用于进行火灾检验。由于紫外线探测器的灵敏度高，它通常能对微小的火焰产生精确的警报。尤其在室外场景，通常烟雾浓度达到报警器报警阈值的时间远长于紫外线探测器探测到火焰发出的紫外线的时间。

## 价格

淘宝价格环境光传感器在几元到几十元不等；光学心率传感器价格在几十元到几百元不等；血氧传感器一般在一百元以内；紫外线探测器价格在几十到几千元不等。

## 距离传感器（雷达）

距离传感器与下面要讲的接近开关有些差别。距离传感器的功能是探测物体与传感器间的信号，且其探测距离远；而接近开关的功能不是探测距离，只是检测传感器周围是否有物体存在，且其探测距离一般很近（只有毫米量级）。

从原理上看，距离传感器一般都需要主动发射信号，通过被测对象对信号的反射来判断被测对象与传感器的距离。距离传感器其实就是雷达，雷达主要是根据发射信号的波长不同进行分类的。

雷达测距的基本原理都是 TOF，即通过计算发射与接收到的电磁波（或声波）之间的时间差，结合电磁波（或声波）的速度计算距离。雷达还可以利用多普勒效应，通过检测发射波与反射波的频率变化，确定障碍物的移动速度。

批注 [刘81]: 参考：  
<https://zhuanlan.zhihu.com/p/58945444>。

## 分类

### 激光雷达

激光雷达向目标发射激光，经目标反射后，激光会向各个方向散射，部分散射光回到传感器接收器。通过检测激光从被发射到被接收所经历的时间，可以确定传感器与目标的距离。

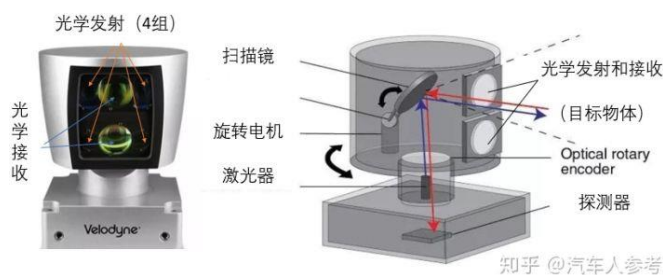
由于一个激光发射器只能探测物体的一个点到传感器的距离，这样的探测信息不足以让我们对物体产生明确的认识。因此实际使用中会通过调整激光雷达的结构，使得雷达能够获得更多的空间信息。简单的说就是，在一段时间内发射更多激光点到周围的环境中，这些探测点构成的数据又被称为点云。

激光雷达的具体结构分类如下。

### 机械式

雷达内部用旋转电机带着发射器高速旋转，发射器与旋转电机旋转轴同向放置，通过透镜反射使激光于旋转平面平行射出。在高速旋转下，特定时间内若干激光点就可在横向上连成一条线，反馈这条线上的距离信息。

进一步地，通过在纵向上布置多个激光发射器，我们就能将原先的线拓展成面，进而对雷达周围的物体信息有更完善的认识。



机械式激光雷达装置构造参考图

批注 [刘82]: 参考：

<https://zhuanlan.zhihu.com/p/33792450> ,  
<https://zhuanlan.zhihu.com/p/103579656>。

批注 [刘83]: 参考：

<http://m.elecfans.com/article/1198697.html> ,  
<https://zhuanlan.zhihu.com/p/139350599> ,  
<https://kns.cnki.net/KXReader/Detail?TIMESTAMP=637396768579375313&DBCODE=CJFD&TABLEName=CJFDLAST2019&FileName=GDGC201907004&RESULT=1&SIGN=7hUkfoW%2f7iazR5ABTqlgj86GZRq%3d>。



## 固态 MEMS

用 MEMS 扫描镜代替原先的机械式转动结构,保持激光发射器的固定,通过 MEMS 扫描镜的转动来将横向的扫描点“连成”线。这样的改进能避免发射器本身的旋转,降低整个系统出故障的几率,同时显著降低成本、减小体积。

## 光学相控阵 (OPA, Optical Phased Array)

由相位调制单元构成以为或二维的阵列称光学相控阵。当激光入射到光学相控阵上时,通过各个相位调制单元,调制单元会使激光的相位发生改变,进而造成激光的方向产生偏移。因此,通过控制不同相位调制单元的相位变化值,可以使激光依次射向不同的方向,实现范围扫描的目的。

## 泛光面阵 (Flash)

上面三种激光雷达的本质都是通过调整激光射出方向实现检测的扫描式雷达。Flash 则有所不同,它属于非扫描式激光雷达,在运行时,直接射出一大片覆盖探测区域的激光,然后通过接受阵列计算距离信息,完成对周边环境的绘制。

雷达内有多少激光发射器,我们就称这个雷达为几线激光雷达。现有的激光雷达有 4 线、8 线、16 线、32 线、64 线的,雷达价格与扫描距离范围、扫描精度、水平垂直角度范围等都随线数逐级递增。

激光雷达在测距、武器、通信等领域都有广泛应用。一般消费级的应用是无人驾驶中充当远距离的距离信息采集雷达。此外,在 iphone12 pro 中的后置激光雷达,用于获得周围环境的深度信息,辅助 AR 软件及其他图像处理功能。

优点:

测量精度高,抗干扰能力强,测量范围大(100-150 米)。

批注 [刘84]: 参考:

<https://kns.cnki.net/KXReader/Detail?TIMESTAMP=637396763514062813&DBCODE=CJFD&TABLEName=CJFD2012&FileName=GDGC201202020&RESULT=1&SIGN=PwAwLY2cKb4%2bhLGVMYM6fKq8Nok%3d>,  
<https://kreader.cnki.net/Kreader/CatalogViewPage.aspx?dbCode=CFD&filename=2010038185.nh&tablename=CFD0911&compose=&first=1&uid=>,  
<https://zhidao.baidu.com/question/577923825.html>

。

缺点：

在大雨、浓雾、浓烟的环境下，激光传播距离受极大影响，雷达准确度也因此大幅下降；造价太高，国内技术储备不足，量产能力差。

#### 超声波雷达

超声波是指频率高于 20000Hz 的声波。超声波雷达通过向外发射超声波，通过接收器接收反射回来的超声波，计算发射与接收的时间，确定目标对象距离。

常用的超声波探头的工作频率有 40kHz、48kHz、58kHz。一般来说，频率越高，灵敏度越高，但水平和垂直方向的探测角度越小。

现在，超声波已经被广泛应用在车辆的倒车雷达中。在自动驾驶领域，通常将超声波雷达放置在车体周围，用于检测车体附近的障碍物。但由于超声波的种种特性（下面会讲），通常只将其作为检测近距离物体的雷达。

优点：

由于超声波频率较高，因此其穿透性好，在各种天气情况下都能良好运行；近距离精度测量精度较高。

缺点：

由于超声波频率较高，波长较短，因此检测距离较短（一般在 10-500 厘米）；由于超声波传播速度较慢，在车辆高速行驶时，超声波测距的反馈速度无法跟上车辆行驶的速度变化；超声波雷达的波速还与温度有关，因此对于远距离对象，超声波的测量很可能因为环境温度的变化而产生较大的误差。

#### 毫米波雷达

毫米波是指频率在 10GHz-200GHz 间的电磁波，其波长在毫米量级。毫米波雷达的原理也是通过向外发射毫米波，用传感器接收反射回来的毫米波，通过发射与接收的时间差确

批注 [刘85]: 参考：

<https://baijiahao.baidu.com/s?id=1614186696271071348&wfr=spider&for=pc> ,  
<https://zhuanlan.zhihu.com/p/35177313>。

批注 [刘86]: 参考：

<https://zhuanlan.zhihu.com/p/34675392> ,

定具体的距离。

自动驾驶领域使用的毫米波雷达主要有 3 个频段，分别是 24GHz、77GHz、79GHz。

24GHz 的雷达用于检测近处的障碍物，通常被放在车辆的四角，帮助实现盲点检测、变道辅助等功能。

77GHz 和 79GHz 的雷达用于检测远处（200 米以内）的障碍物，通常被放在车辆的保险杠的内侧或外侧，帮助实现确认前后车辆距离及速度、紧急制动、自动跟车等功能。

目前无人驾驶领域主要使用的还是毫米波雷达。在手势识别领域也会使用毫米波雷达，如 google 的 soli。

优点：

测量精度高（但是还是比激光雷达差一点）；探测距离远；体积小；穿透力强，有良好的抗雨、雾、烟、尘的能力。

缺点：

毫米波雷达发射的是锥状的波束，因此覆盖范围较小；对金属敏感度高，很容易将路面出现的钉子、远距离的金属广告牌误认为是障碍物，产生误报警（且误报警的问题还是挺频繁的）。

应用


自动驾驶中的雷达

自动驾驶一般会在车辆上安装多个雷达，一般都会安装毫米波雷达和超声波雷达，部分车型会安装激光雷达。除了雷达外，自动驾驶还会使用大量的摄像头作为信息的补充。

批注 [刘87]: 参考：  
<https://zhuanlan.zhihu.com/p/21384539>。

批注 [刘88]: 参考：  
[https://www.sohu.com/a/256097077\\_465591](https://www.sohu.com/a/256097077_465591)。

批注 [刘89]: 参考：  
<https://zhuanlan.zhihu.com/p/35944238>。

品牌				
车型	 A8L	 Model S	 7 Series	 S class
激光雷达	1	-	-	-
前视摄像头	1	3	1	1
长测距雷达	1	1	?	?
中测距雷达	4	-	?	?
环视摄像头	4	5	4	4
超声波雷达	12	12	12	12
红外夜视摄像机	1	-	1	1

不同车型的雷达使用

自动驾驶中使用毫米波雷达的原因主要是其价格便宜且受环境影响小,通常用毫米波雷达做长距及中距的探测。毫米波雷达一般会在车辆的前后保险杠的内侧或外侧放置。

使用超声波雷达的原因主要也是价格便宜,一般的自动驾驶车辆都会在车身周围安排12个超声波雷达。而且用超声波探测近距离障碍的技术比较成熟。

自动驾驶中使用的激光雷达一般都是4线雷达,一般会放置在车头或车顶(线数越高越可能放在车顶)。限制激光雷达在自动驾驶中应用的主要因素是价格,16线的激光雷达耗价8000美元。此外,激光雷达在雨天、雾天、浓烟情况下的表现会出现大幅下滑。

iphone12 pro 中的激光雷达

2020年新发售的iphone12 pro中,在后置摄像头阵列中加入了一个激光雷达,用于探测用户使用手机时的环境深度信息。

照片或视频能与这些深度信息结合,获得更大的后期调整空间;深度信息也可以被直接用于空间测量或AR应用,以提供更优质的应用体验。

在低光条件下,利用深度信息可提高对焦精度,减少对焦时间。依托深度信息的对焦速度能比单纯依靠光学镜头对焦速度快6倍。

价格

激光雷达的价格在上面略有提及，可以看一下下面的这张图来了解。

			
特性	HDL-64	HDL-32	VLP-16
售价	8万美元左右	2万美元左右	7,999美元
激光束	64	32	16
范围	120m	100m	100m
精度	±2cm	±2cm	±3cm
数据类型	距离/密度	距离/校准发射率	距离/校准发射率
数据频率	1.3M像素/秒	700,000像素/秒	300,000像素/秒
垂直角度	26.8°	40°	30°
水平角度	360°	360°	360°
功率	60W	12W	8W
体积	203*284mm	86*145mm	104*72mm
重量	15kg	1kg	0.83kg

不同规格的激光雷达价格及性能比较

在百度采购上，毫米波雷达的价格在 40-50 元之间，在淘宝则是几百到几千元之间。

淘宝上，超声波雷达的价格在几十元左右。

接近开关

接近开关与雷达不同，通常它的探测距离在毫米量级。它的功能在于探测开关附近是否有物体存在，而非测量传感器与物体间的距离。

原理

按照接近开关的原理不同可分为电感式接近开关、电容式接近开关、霍尔式接近开关和光电接近开关。不同开关检测的物体对象也不同。

电感式

此类接近开关只能用于检测金属物体。

传感器中的线圈通过高频振荡产生高频磁场。当金属物体接近时，金属内部会产生涡流，

批注 [刘90]: 参考：

<https://kreader.cnki.net/Kreader/CatalogViewPage.aspx?dbCode=CJFD&filename=ZDBN200305016&tablename=CJFD2003&compose=&first=1&uid=>，  
<https://baike.baidu.com/item/%E6%8E%A5%E8%BF%91%E4%BC%A0%E6%84%9F%E5%99%A8/10492093?fr=aladdin>，  
<https://baike.baidu.com/item/%E7%94%B5%E5%AE%B9%E5%BC%8F%E6%8E%A5%E8%BF%91%E5%B3%80%E5%85%B3/18884011?fr=aladdin>。

较为简单的原理解释：

<https://www.bilibili.com/video/BV1p4411J7fZ?from=search&seid=5229925039754160937>。

应用讲解：

<https://www.bilibili.com/video/BV1pa4y1a7Yg?from=search&seid=4145203357115999968>。

涡流进而引发反向的感应磁场。振荡器受反向感应磁场的影响，逐渐减弱并停止振荡。通过振荡器振荡信号的有无可以判断是否有金属物体接近。

由于这一原理要求被测物体一定要有产生感应电流的能力，因此只能用于检测金属。检测距离与产生的感应电流强弱成正比。

#### 电容式

此类接近开关可以用于检测任何类型的物体（包括绝缘体）。

传感器的测量头与外壳是构成电容器的两个极板。当物体靠近时，外壳的介电常数会发生变化，通过检测电容的电荷变化造成的回路中的状态变化即可确定物体的接近。

以前我们曾说过，绝缘体是无法使电容式触摸屏起反应的。其实如果增大触摸屏灵敏度（通过调整介电常数）的话，是可以实现触摸屏受绝缘体影响的。只不过这样的设计通常会导导致日常使用中的大量误触，因此并不实际。而对应的思路则在电容式接触开关中被使用了。

但同样的，由于电容式开关可以检测任何物体，因此也很容易因为积灰等原因造成误触等问题，因此使用时需要频繁保养。

#### 霍尔式

此类开关只能检测此磁性物体。

霍尔式接近开关利用利用的是霍尔效应，通过在金属板两侧添加电流检测装置，当带磁性的物体靠近时，金属板两侧会因为霍尔效应产生电荷的极性分布，从而在金属板两侧的回路中形成电流。通过检测电流，即可确定物体的接近。

#### 光电式（红外接近开关）

此类接近开关可以检测反射率不高的物体（反射率较高会导致发射出的红外光被反射到特定的角度，无法正确进入接收器）。

当物体接近时，红外接近开关通过向外发射红外光并用接收器接收反射回来的红外光以

批注 [刘91]: 参考：  
<https://baike.baidu.com/item/%E9%9C%8D%E5%B0%94%E6%8E%A5%E8%BF%91%E5%BC%80%E5%85%B3/10619146?fr=aladdin>。

批注 [刘92]: 我原来把它放在距离传感器中的，后来感觉还是接近开关更符合一点。

确定物体存在。当物体离开时，由于没有物体能对红外光进行反射，因此接收器信号为 0。

红外接近开关的价格便宜，制作简单；但受日光影响严重，不宜在室外使用。

## 应用

### 智能手机中的红外距离传感器

智能手机通常通过在听筒旁安装红外接近开关来判断用户接电话时耳朵与手机屏幕的距离。当距离过小时，手机会自动黑屏，避免脸部对屏幕的误触。

### 工业生产流水线中的物体探测器

与智能手机相比，接近探测器最常见的使用还是在工业生产流水线中。生产中，利用接近探测器探测物品的靠近，并通过中控装置控制下一步的生产操作。

## 价格

淘宝价格在几元到几十元不等。

## ECG 传感器

ECG ( Electrocardiograph ) 即心电图。在苹果新一代 apple watch 上搭载的电极式心率传感器就是利用类似原理的传感器。

### 原理

心脏的每一次搏动都会伴随心脏电生理活动，即心脏的起搏点通过放电，使电流传导到每个心肌纤维，接收到电信号后，相应的心肌纤维完成一次收缩，心脏也就随之搏动一次。

批注 [刘93]: 参考：  
<https://zhuanlan.zhihu.com/p/44425434>。

心脏的电信号可以传导到体表皮肤，且不同体表部位所检测到的电信号表现不同。因此，在体表的特定部位放置电极进行检测，可以通过电极间的相互配合形成 12 段心电波形图（常称 12 导联心电图）。医院中使用的心电监测设备就是到处 12 导联心电图的，它能诊断各类型的心律失常、心肌病等心脏问题。

但一般的只能佩戴设备一般无法做到在身体上放置多个检测电极，因此基本都是所谓的“单导联心电图”。由于电极数量太少，因此单导联心电图无法给出非常详细准确的心脏运作信息，除了心率监测（这个用前面光探测器中的光体积描记法也能实现）外，只能对一些简单的心脏问题进行监测，不能代替正规的心脏检测设备。

## 应用

### 医院使用的正规心电图设备

医院的十二导联心电图也是以 ECG 传感器作为心脏起搏电流的获取传感器的。

### 智能穿戴设备的心脏监测设备

如苹果在 apple watch 上加装的电极式心率传感器就是 ECG 传感器，不过这种一般只是单导联心电图，功能有限，不能代替专业检测设备。不过国内该功能似乎因为尚未通过医疗监管部门审核所以暂时不能使用。

## 价格

淘宝上一般的带 ECG 功能的智能手表都在 1000-3000 元左右。



## EEG 传感器

EEG ( Electroencephalo-graph ), 即脑电图。脑电图用于检测人脑在各项运动时所发出的占主导作用的脑电波。

EEG 实际属于 BCI ( Brain Computer Interface , 脑机接口 ) 的一种。脑机接口可以分为非侵入式和侵入式, 而 EEG 属于其中的非侵入式。非侵入式就是指不需要通过手术将传感器植入大脑皮层, 只需在大脑外部即可获取需要信息的设备。

脑电图主要检测的对象是脑电波, 目前我们将脑电波按频率、振幅、波形模式分为  $\alpha$  波、 $\beta$  波、 $\theta$  波、 $\delta$  波、 $\gamma$  波等, 不同的波形会与人脑的不同活动相对应:  $\alpha$  波与放松冥想等活动相关;  $\beta$  波与专注意力或产生焦虑、恐惧感相关;  $\theta$  波与快速眼动睡眠相关, 同时它也与白日梦、记忆的形成和巩固相关;  $\delta$  波与非快速眼动睡眠相关;  $\gamma$  波与运动和感觉相关。

### 原理

脑电波来自于大脑神经元的电活动, 是神经细胞传递信号时会产生电位的变化。这样的电位变化形成的电流可以经过大脑皮层被贴在头皮表面的传感器检测到, 经过采样、放大、过滤后, 在脑电图上以波形的形式展现。由于人脑活动的脑电波在任何时刻都不是单一的, 任何类型的脑电波都会存在, 但强度不同。因此在获得脑电波后还要对混合的脑电波信号进行分类处理。

国际 10-20 系统是国际脑电图学会规定的标准电极放置法, 它会将若干电极按照特定位置用导电胶黏在人的头皮表面 ( 似乎也有其他的佩戴方式, 如制成电极帽等, 但检测单元分布不会变化 )。

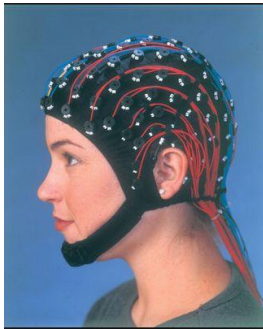
批注 [刘94]: 参考:

<https://www.bilibili.com/video/BV1iK411T7ZV> , ( 特别有用 )

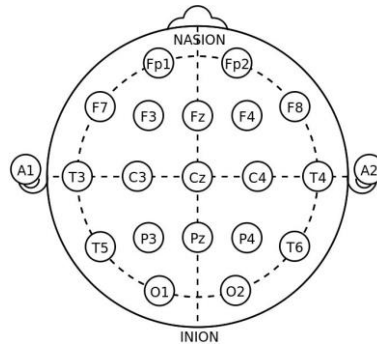
<https://www.bilibili.com/video/BV1Gz411q7yR> ,

<https://www.zhihu.com/question/22820093> ,

<https://zhuanlan.zhihu.com/p/29492075>。



使用国际 10-20 系统的检测电极的实际佩戴效果



国际 10-20 系统的检测电极分布。

由于这种检测系统要求的检测单元数量很多,且佩戴很不方便,因此也有许多其他的简化的佩戴设备,但相应的检测精度会差很多。



mindwave 的脑电波检测装置

## 应用

EEG 的主要使用还是在获取和解析脑电波上,仅依靠 EEG 并不能用外界手段反向影响人脑。反向影响人脑的装置是 TMS( Transcranial Magnetic Stimulation ),即颅脑磁刺激。

## 神经科学领域的研究

在神经科学领域,有大量的大脑与人类活动相关的研究需要 EEG 来帮助监测脑电波,最典型的试验就是检测不同睡眠状态下脑电波的变化。

## 麻醉检测

外科手术中，麻醉师会根据脑电波来判断患者是否处于昏迷状态，并根据脑电波的变化及时调整麻醉剂的使用。

### 利用脑电波作为输入信号

利用脑电波和眼动活动帮助渐冻症患者实现打字活动。清华大学脑机接口研究团队就在进行这一项目。

同理，也可以利用脑电波帮助瘫痪病人控制机械假肢。

再进一步，有实验尝试将两个玩家在游戏时的脑电波用 EEG 采集，然后通过 TMS 将信息传递给第三位玩家，结果发现第三位玩家在该情况下的游戏正确率比他单独游戏时的正确率要高。

## 价格

之前提到过的简化版的 mindwave 的脑电波检测装置的价格似乎是 99 刀。

## 生物阻抗传感器

生物阻抗传感器一般被用于使用生物阻抗分析法（BIA，BioImpedance Analysis）分析人体脂肪含量的设备（体脂秤）中，其本质是通过测量人体的电阻率，反推人体的脂肪含量。

### 原理

人体本身是导体，人体内的脂肪及肌肉的含水率不同，因此电阻率也不同。肌肉含水率

批注 [刘95]: 参考：  
<https://www.bilibili.com/video/av14995143>。

批注 [刘96]: 参考：  
<https://www.zhihu.com/question/31791800>，  
<https://www.zhihu.com/question/40241431/answer/90659278>。

更高，因此电阻更低；脂肪含水率更低，因此电阻更高。

生物阻抗传感器就是通过在人体表面放置多个接触电极，分别输入微弱的电流，通过测量电流和电压的数值，确定人体的总体电阻值。然后根据经验公式反推人体的脂肪与肌肉的比例。

生物阻抗分析法易受条件影响：接触电极数量、放置位置、与皮肤的贴合程度，测量室温度、湿度，被试的水合状态（含水量多少，由被试测量前是否饮水、运动出汗、排便等决定）等因素都会对最终的测量结果产生影响。此外，最终通过电阻值推算人体成分比例的经验公式有许多，公式的选择也会对结果产生影响。

现在常见的体脂秤设备通常只有两个电极，且被试直接站在两个电极上，测量前对水合状态不做控制，因此测量的误差极大。

据上面参考中做的试验来看，一般体重较大的人测量的准确度较高。

如果要获得精确的人体成分含量数据，还是要用双能量 X 射线吸收法（DXA），不过这种设备一般只在医院中才有，且会发出低剂量的 X 射线。

## 价格

淘宝体脂秤价格一般都在 100 元以下。

## 传感器的组合与应用

### 惯性测量单元 ( IMU , Inertial Measurement Unit )

IMU 实际上就是整合了加速度计和陀螺仪的测量单元。从理论力学的角度来看，物体的运动只能分解为直线运动和旋转运动。因此，通过三轴加速度计测量物体的直线运动信息，通过三轴陀螺仪测量物体的旋转运动信息，结合两者，我们就能得到物体在三维空间中的运动姿态。

通常的 IMU 是 6 轴传感器，即 3 轴加速度计加上 3 轴陀螺仪。也有 9 轴的 IMU，即在加速度计和陀螺仪的基础上，再添加一个 3 轴磁力计。

#### 原理

IMU 的原理是基于加速度计、陀螺仪和磁力计的。通过加速度计获得的加速度，通过积分可以确定物体运动过程中的速度和各个方向上的运动距离，从而确定现在的位置与运动开始时的位置；通过陀螺仪可以确定物体的角速度，从而确定物体的旋转状态；通过磁力计可以确定北极，结合角速度信息，可以确定当前运动的方向。

在上面那个参考链接中提到了航姿参考系统（也叫航姿模式，AHRS，Attitude and Heading Reference System），链接中将 6 轴 IMU 称作 IMU，将 9 轴 IMU 称作 AHRS。因此，虽然两者看上去名称不一致，但本质上并没有太大差别。

批注 [刘97]: 参考：

<https://zhuanlan.zhihu.com/p/28262130>。

下面这个是详细讲解如何用 IMU 进行动态姿态的计算的，简单看一看就好：

<https://zhuanlan.zhihu.com/p/54500745>。

应用

### 惯性导航（或称航姿模式）

惯性导航（INS，Inertial Navigation System）或航姿模式是指不依赖外部导航信息（一般就是指 gps 信息），仅使用设备自身运动信息实现的导航。通常，无人机或地图导航软件只有在失去 gps 信息时，才会被迫切换到惯性导航模式。

磁力计会配合加速度计、陀螺仪，共同提供用于实现导航的信息（如果是无人机的话还会添加额外的气压计用于确定高度）。

但由于设备误差等原因，在缺乏 gps 的信息校准、修正的情况下，纯惯性导航很容易导致误差的积累，从而导致由惯性信息计算出来的位置与实际位置的差别越来越大。因此，很多情况下，惯性导航只是暂时失去 gps 信息下的一种辅助功能。

惯性导航可以被理解为按照那种向东 30 步、向南 40 步的藏宝图寻宝，由于缺少准确的定位信息，细微的错误会不断积累，导致最终与目标位置相差很远。

常见的应用如手机上的地图 app 在进入隧道失去 gps 信息后，仍能持续进行导航，这就是依赖惯性导航实现的；此外，无人机又有时会因为信号屏蔽等原因失去 gps 信息，此时依靠惯性导航仍然能够对无人机进行操纵。

### 自动驾驶系统

自动驾驶通常使用的定位信息是 gps 定位。但由于 gps 的信号更新频率较低（通常为没 100ms 更新一次，即 10Hz），因此在道路状况复杂的情况下，很难实现对车辆的精确控制。

而 IMU 能提供重要的运动参数，且信号更新频率高（通常为 10ms 更新一次，即 100Hz），因此通常与 gps 一同构成自动驾驶的运动信号处理。

根据两者的信号更新频率可知，在 100ms 内，会用 IMU 进行 9 次位置信号与分析，

批注 [刘98]: 参考：

<https://www.zhihu.com/question/27847666>，  
<https://www.zhihu.com/question/26717929/answer/35992240>。

还有一个比较专业的内容，可以简单参考：

<https://kns.cnki.net/KXReader/Detail?TIMESTAMP=637394823218750313&DBCOD=CFD&TABLEName=CFDAUTO&FileName=CGQJ202011035&RESULT=1&SIGN=2VvEQfGQ6yiK5J7S1rzFGOTzDLg%3d>。

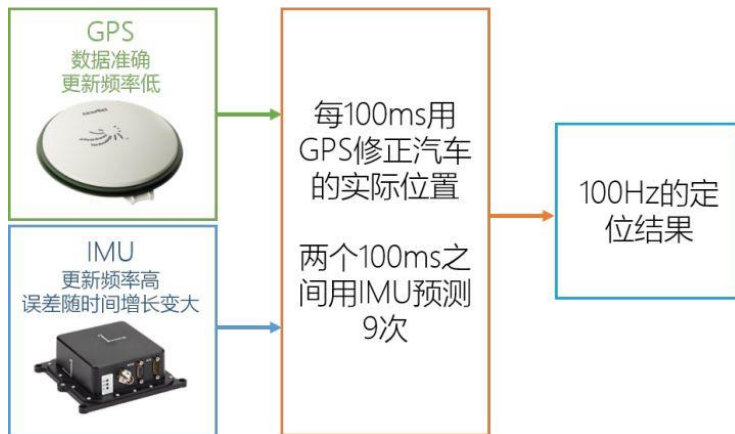
批注 [刘99]: 参考：

<https://www.zhihu.com/question/26717929/answer/35992240>。

批注 [刘100]: 参考：

<https://zhuanlan.zhihu.com/p/32693377>。

然后进行一次 gps 信号校准，确保车辆在行驶过程中的位置准确。



gps+IMU 融合原理图

手机、游戏手柄、VR 眼镜

现在也有不少手机、游戏手柄或 VR 眼镜中放置了 IMU 进行运动信息的采集，以确保获得手机的加速度和角速度信息，或通过磁力计获得磁场信息，从而将这些信息技术作为交互信息传入软件以实现操控、画面转动、防抖等功能。这部分在之前的加速度传感器、陀螺仪、磁力传感器中都提到过了，不多赘述。

价格

IMU 可以根据应用场合分为消费级、车规级、航天航空级等类别，一般消费级的 IMU 在 100 元以内就能买到。

## MEMS

批注 [刘101]: 参考：

<https://www.bilibili.com/video/BV1mJ41137r6?p=1>

MEMS ( Micro Electro-Mechanical System ) , 即微电子机械系统 ( 也称微机电 )。它是指由微加工技术实现的, 包括微结构、微传感器、微执行器、控制处理电路等组成部分, 能够实现测量、处理或执行等功能的, 尺度为微米至毫米的微小集成器件或系统。同时, 实现这种微型化的技术也可被称为 MEMS。

简单的说, MEMS 就是把传感器、处理器、执行器都做小, 利用集成电路 ( IC ) 和微加工技术 ( MEMS ) 整合到同一块芯片上的微型化系统。

MEMS 应用领域广泛、功能多样、精度高、适合大批量生产、体积小等优点使 MEMS 成为了现在智能设备的主要发展方向。

### 应用简述

在之前将传感器的部分时就有提到目前大量的传感器其实都是 MEMS 传感器。

MEMS 是实现智能的物质基础, 是当代工业不可缺少的基础设备, 在通信、工业控制、智能电网、消费电子、能源化工、航空航天、生物医学、军事国防、精准农业、环境监测、科学研究、汽车电子等诸多领域都有应用。一部智能手机需要十几个 MEMS 传感器, 一辆汽车需要上百个 MEMS 传感器, 一架飞机需要上千个 MEMS 传感器。

然而, 一项 MEMS 技术从研发到产品化再到市场化, 往往需要 20-30 年的时间, 技术难度相当高。再加上 MEMS 产品虽然种类繁多且利润率高, 但由于应用情景还是相对有限, 因此总销售额仍处于攀升阶段。这导致了从事 MEMS 设计、生产、销售的公司大都集中在欧、美、日, 国内的 MEMS 产业还在发展阶段。



## ASIC

ASIC ( Application Specific Integrated Circuit ), 即专用集成电路。它指应特定用户要求和特定电子系统的需要而设计、制造的集成电路。

最常见的就是前几年用于比特币挖矿的矿卡。矿卡只能用于比特币挖矿的 SHA256 算法, 无法执行除此以外的任何工作。

### 优点

硬件设计由具体任务决定, 因此处理效率高、能耗低。

### 缺点

硬件只能适应单一任务, 无法处理此外的任务, 泛化能力差; 此外, 对于特定用途需要进行专门的算法和硬件设计, 成本较高。

## 全球卫星导航系统

全球卫星导航系统 ( Global Navigation Satellite System ) 是利用多颗轨道卫星实现对地球上的任意位置的具体位置信息进行实时定位、导航的地理信息交互系统。

不同的国家或地区组织都有各自的全球卫星导航系统, 如美国的 GPS、中国的北斗、俄罗斯的 GLONASS、欧洲的 Galileo。不同国家的系统使用的卫星数量也有所不同, 其主要原因是不同国家的卫星所在的轨道高度不同。此外, 由于不同国家地区对卫星的发射方案、功能设计、备用卫星数量、精度要求不同, 卫星的数量也会有所差异。系统卫星的总数量决定了 GPS 定位设备在地球上任意位置能接触到的卫星个数, 进而影响了系统的覆盖范围和定位精度。

俄罗斯的 GLONASS 需要 24 颗卫星, 轨道高度为 19100km; 美国的 GPS 需要 24 颗

批注 [刘102]: 参考 :  
<https://www.zhihu.com/question/21092045>.  
( 关于卫星数量 )

卫星，轨道高度为 20200km；中国的北斗系统需要 35 颗卫星，轨道高度为 21500km；欧洲的 Galileo 需要 27 颗卫星，轨道高度为 23222km。

民用 GPS 的定位精度在 5-10 米，军用的精度在 1 米左右。

## 原理

以下解释都以 GPS 为例。

当一个 GPS 定位设备在地球上的某个未知位置想要确定自身的地理位置坐标时，他实际上希望得到经度、纬度、高度这三个参数。假设三个地理位置坐标参数为  $x_0$ 、 $y_0$ 、 $z_0$ ，此时 GPS 设备的时间坐标为  $t_0$ （确定）。该 GPS 设备将自身的设备时间通过无线电（速度为光速  $c$ ）发送给它能接触到的卫星，卫星通过比较设备的时间与卫星自带的原子钟的时间  $t_1$ ，就可以确定卫星与 GPS 设备之间的距离  $d = c * (t_1 - t_0)$ 。而由于卫星是持续在绕地轨道上运行的，它自身的位置坐标  $x_1$ 、 $y_1$ 、 $z_1$  是实时确定的，因此根据勾股定理可以得到卫星与 GPS 设备之间距离的第二个表达式  $d = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2 + (z_1 - z_0)^2}$ 。

通过联立之前两个等式，可得  $\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2 + (z_1 - z_0)^2} = c * (t_1 - t_0)$ 。此时可以发现等式有  $x_0$ 、 $y_0$ 、 $z_0$  三个未知数待求解，似乎只需要再向 2 个不同的卫星发射自己的时间信息就能实现等式求解。但由于 GPS 设备本身的时间并不精确、电磁波在大气层中的速度也存在波动、GPS 卫星位置本身存在一些误差，为了弥补以上的误差、实现精确定位，GPS 设备需要同时向 4 个设备发射信号，才能实现最终的定位。

换言之，要实现精确的 GPS 定位，地球上任意位置都必须保证至少有 4 颗卫星覆盖。这也是全球卫星定位系统卫星数量要求的主要原因。

以上的定位方法被称为单点定位法，为了提高精度，还会使用差分定位法。差分定位法，即在一个固定位置放置一个定位基站，在原有的定位流程上，再让基站与 4 颗卫星进行单点

批注 [刘103]: 参考：

<https://www.bilibili.com/video/BV1Ts411L73G?from=search&seid=3213555170536669783>，  
<https://www.zhihu.com/question/20903715>，（这个比较简单）  
<https://zhuanlan.zhihu.com/p/71385507>。（这个更深入一点）

具体的设备与卫星间时间差的测量需要使用测距码，参考：

[https://www.bilibili.com/video/BV1TV411z7aB/?spm\\_id\\_from=333.788.videocard.3](https://www.bilibili.com/video/BV1TV411z7aB/?spm_id_from=333.788.videocard.3)。（3:30）

定位，并与设备进行一次定位，以此大致确定定位的误差，以实现精度的提高。

## 应用

全球卫星导航系统在如今的应用十分广泛，除了军事、工程、农业领域的应用外，比较常见的应用就是手机或其他智能穿戴设备的全球定位。

现在设备获取全球定位系统的卫星信号是免费的，但由于获得的具体位置数据一般需要在地图应用中才能得到使用，所以在一定程度上还是依赖数据网络。

## 双星定位

关于北斗系统，我们有个说法叫双星定位。如果我们至于两颗卫星进行信息交互，那么根据我们到卫星的距离，两颗卫星能够各自划出以距离为半径的球体。两个球体相交的圆弧就是我们可能在的位置。此时，如果配合地面的电子高程地图，在预先就确定好不同经纬度的高度信息，就可以根据圆弧与地球不同高度表面的交点确定我们所在的具体位置。这就是所谓的双星定位。

批注 [刘104]: 参考：  
<https://zhidao.baidu.com/question/488705955.html>。

# 跨模态交互

跨模态交互就是指多种感官组合的交互方式。

人脑对于环境信息的处理本身就是多通道、跨模态的，人可以通过视觉、听觉、嗅觉、触觉、味觉感受周边的世界，因此我们希望交互体验也从单一模态向多模态、跨模态发展。（注意，多模态指强调存在诸多感知、反馈的通道；而跨模态则强调多种感知、反馈通道间的交叉，能够通过一个模态的信息的获取做出对另一个模态的信息的判断）

人的不同感官的信号带宽、接口面积和数据属性都不尽相同。实现跨模态交互需要做到多感官融合，融合的目的是感官对信息的传达更加自然、更加高效、更易于理解、对于体验者而言更具沉浸感。

某一感觉的信号带宽可以理解为人获得的信息有多少是由这一感觉带来的（如人眼提供了 80%的信息）。接口面积感受起某一感觉的器官所占据的人体面积（如视觉有两个眼球引起，听觉由两个耳朵引起，触觉由覆盖全身的皮膚引起）。数据属性就是指不同感官对人体带来的刺激是不同的。

想要实现感官同步，不同信号通道的延迟时间需要控制在 20ms-100ms 以内，否则体验者就会感受到明显的错位与不同步。

跨模态交互的主要难点在于，环境信号的采集时存在的噪音难以分离、动态的多模态信息不容易实现匹配等。

批注 [刘105]: 简单了解就好。

批注 [刘106]: 电影有图像加音乐就是多模态。

批注 [刘107]: 吃鸡这种游戏，可以通过声音判断别人的位置，就是跨模态。Airpods pro 的空间音效应该也能算，它实现了声音与视觉的统一。