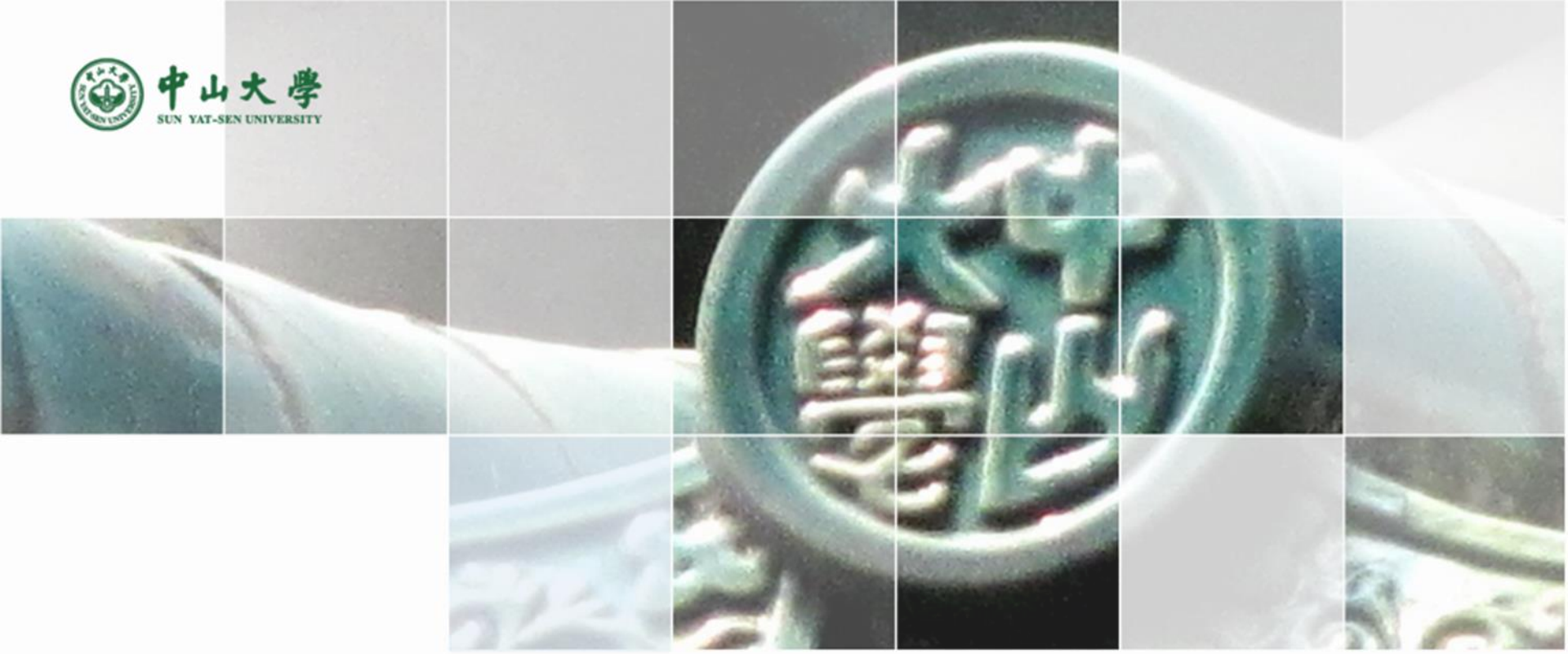


虚拟现实技术

Virtual Reality Technology

金枝

中山大学智能工程学院 2019秋季课程

本堂课内容

复习上堂课内容

虚拟现实的关键技术

1 三维虚拟声音技术；

2 人机交互技术；

复习题

填空题

1.立体视觉的基础：_____。

两只眼睛存在视差

复习题

填空题

1.立体视觉的基础：_____。

2.立体视觉的实现：_____和_____两步。

—对同一场景分别产生相应于左右眼的不同图像，让它们之间具有一定的视差。

—借助相关技术，使左右双眼只能看到与之相应的图像。

复习题

填空题

1.立体视觉的基础：_____。

2.立体显示的实现：_____和_____两步。

3.立体显示技术从时间特点上来分为：_____和_____。

— 同时显示（frame parallel）技术；

— 分时显示（frame sequential）技术；

复习题

填空题

1.立体视觉的基础：_____。

2.立体显示的实现：_____和_____两步。

3.立体显示技术从时间特点上来分为：_____和_____。

4. 三维建模技术分为：_____、_____和_____。

几何建模、物理建模和运动建模

本堂课内容

复习上堂课内容

虚拟现实的关键技术

1 三维虚拟声音技术；

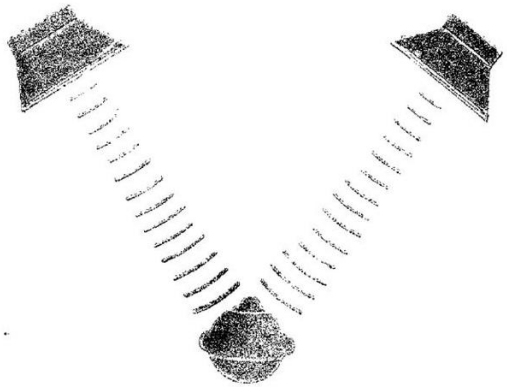
2 人机交互技术；

1 三维虚拟声音技术

- 在虚拟现实系统中，听觉信息是仅次于视觉信息的第二传感通道，听觉通道给人的听觉系统提供声音显示。
- 为了提供身临其境的逼真感觉，听觉通道应该满足一些要求，使人感觉置身于立体的声场之中，能识别声音的类型和强度，能判定声源的位置。
- 在虚拟现实系统中加入与视觉并行的三维虚拟声音，一方面可以在很大程度上增强用户在虚拟世界中的沉浸感和交互性，另一方面也可以减弱大脑对于视觉的依赖性，降低沉浸感对视觉信息的要求，使用户能从既有视觉感受又有听觉感受的环境中获得更多的信息。

1 三维虚拟声音技术

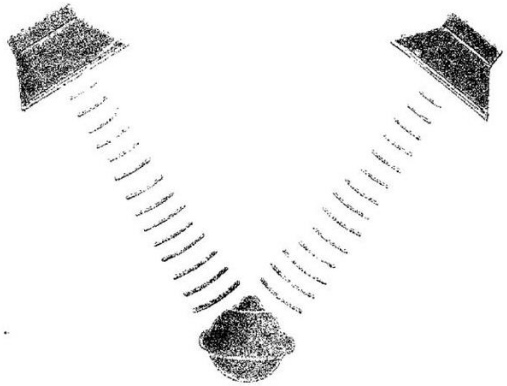
1.1 三维虚拟声音技术概念



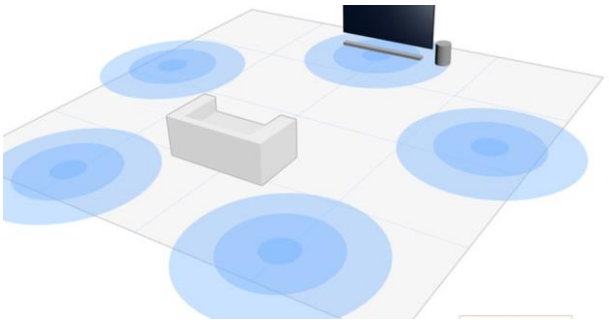
立体声

- 我们日常听到的立体声音(左右声道)虽然有左右声道之分，但就整体效果而言，我们能感觉到立体声音来自听者面前的某个平面；

1.1 三维虚拟声音技术概念



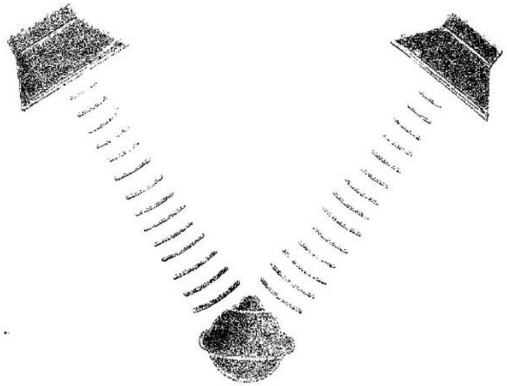
立体声



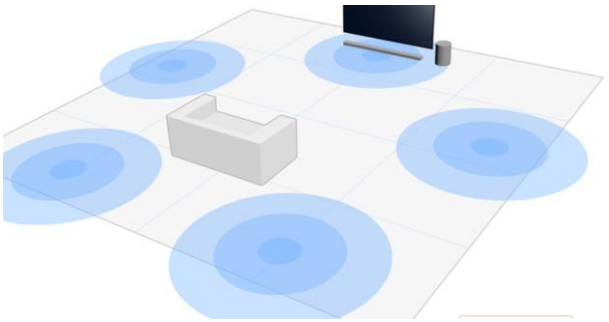
立体环绕声

- 立体环绕声保留原信号的声源方向感，并伴随产生围绕感和扩展感的音响效果。
- 在聆听环绕立体声时，聆听者能够区分出来自前后左右的声音，即环绕立体声可使空间声源由线扩展到整个水平面。

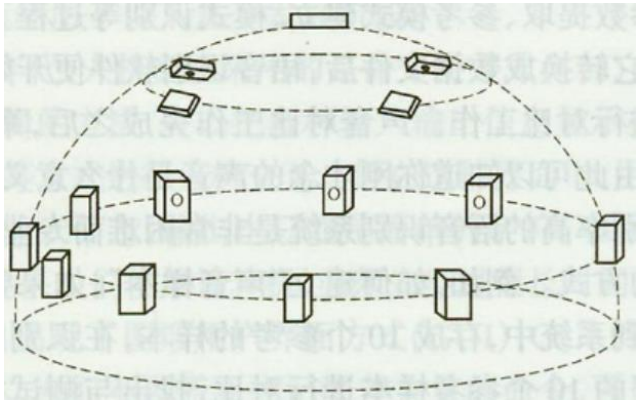
1.1 三维虚拟声音技术概念



立体声



立体环绕声



三维虚拟声音

- 三维虚拟声音是来自围绕听者双耳的一个球形中的任何地方，即声音出现在头的上方、后方或者前方。
- 在虚拟环境中，三位虚拟声音可以使用户准确判断声源的位置。

1.2 三维虚拟声音技术特征

在三维虚拟声音有两个主要特征：

- 全向三维定位特性
- 三维实时跟踪特性

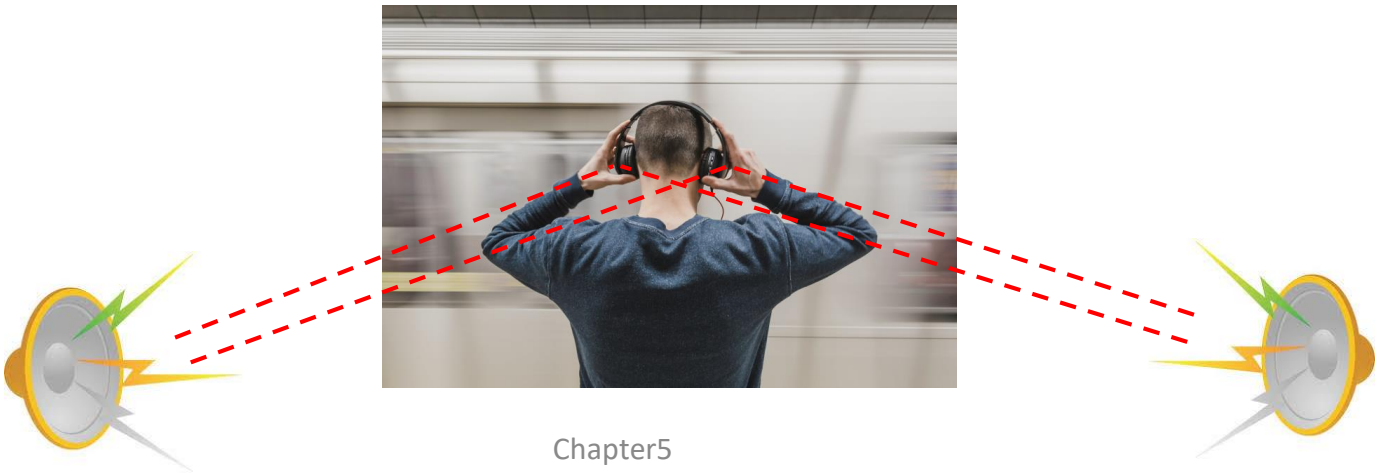
- 沉浸感与交互性 （与视觉系统共享）

1.2 全向三维定位特性

- 全向三维定位（3D steering），是指在三维虚拟空间中把实际声音信号定位到特定虚拟专用源的能力。
- 它能使用户准确地判断出声源的精确位置，从而符合人们在真实境界中的听觉方式。

1.2 三维实时跟踪特性

- 三维实时跟踪特性(3D Real-Time Localization)是指在三维虚拟空间中实时跟踪虚拟声源位置变化或景像变化的能力。
- 当用户头部转动时，这个虚拟的声源的位置也应随之变化。而当虚拟发声物体位置移动时，其声源位置也应有所改变。
- 因为只有声音效果与实时变化的视觉相一致，才可能产生视觉和听觉的叠加与同步效应。



1.3 语音应用技术

- 语音是人类最自然的交流方式，与虚拟世界进行语音交互是实现虚拟现实系统中一个高级目标。

1.3 语音应用技术

- 语音是人类最自然的交流方式，与虚拟世界进行语音交互是实现虚拟现实系统中一个高级目标。
- 语音应用技术主要是指基于语音进行处理的技术，在虚拟现实技术中的关键技术是语音识别技术和语音合成技术。

1.3 语音识别技术

- 语音识别技术(ASR, Automatic Speech Recognition)是指将人说话的语音信号转换为可被计算机程序所识别的文字信息，从而识别说话人的语音指令以及文字内容的技术。
- 一个完整的语音识别过程可大致分为以下三个部分。

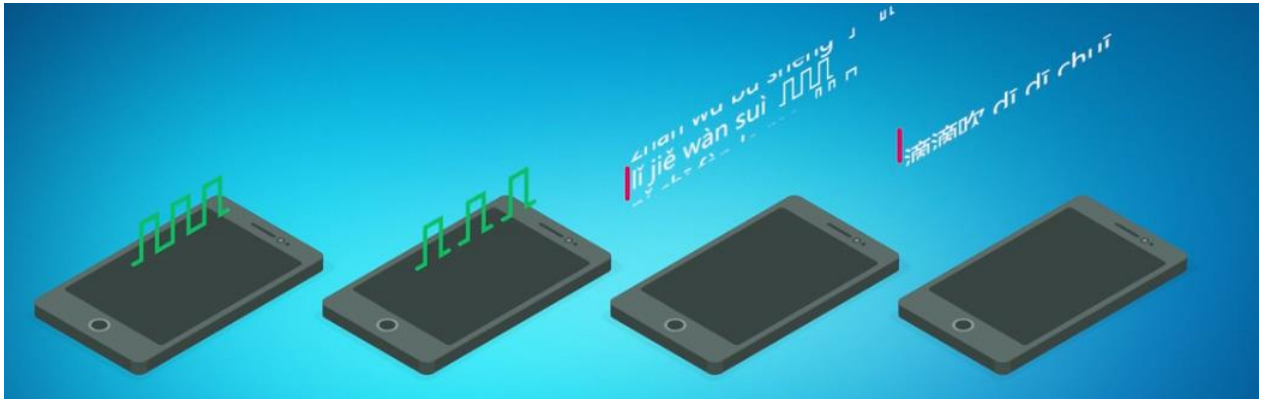
1）语音特征提取。

2）声学模型与模式匹配。

3）语言模型识别与语言处理。

1.3 语音识别技术

- 当通过一个话筒将声音输入到系统中，系统把它转换成数据文件后，语音识别软件便开始以输入的声音样本与事先储存好的声音样本进行对比工作。声音对比工作完成之后，系统就会输入一个它认为最“像”的声音样本序号，由此确定刚才的声音是什么意义，进而执行此命令。



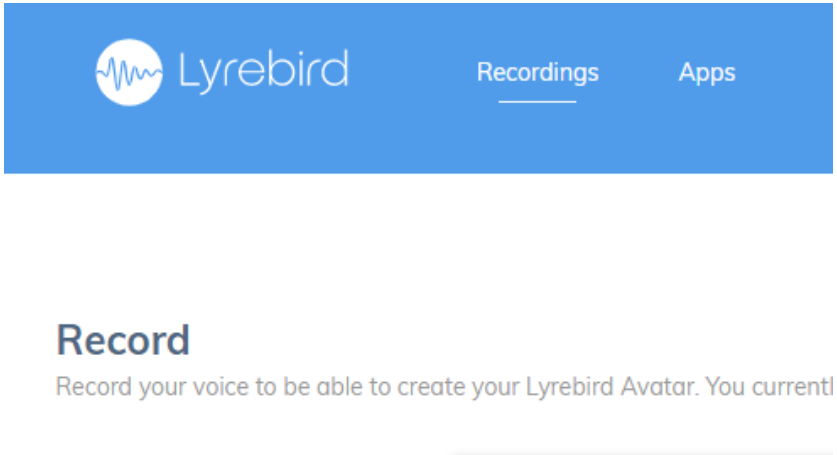
1.3 语音合成技术

- 语音合成技术(TTS，Text to Speech)是指用特定的方法生成语音的技术，当计算机合成语音时，如何能做到用户能理解其意图并感知其情感，一般对“语音”的要求是清晰、可听懂、自然、具有表现力。

语音合成界的鼻祖

1.3 语音合成技术

- 语音合成技术的实现方法有三种：拼接技术、统计参数合成技术和基于深度学习的语音合成技术。



1.3 语音合成技术

- 拼接技术
  - 收集整理大量人们的语音，形成一个庞大的语音资料库；
  - 再根据人们输入的文字在资料库里搜索找到对应的语音资料进行整理拼接成一个完整的句子输出。

缺点：这种方法依托一个庞大的语音库耗时，且出来的效果逼真度差。

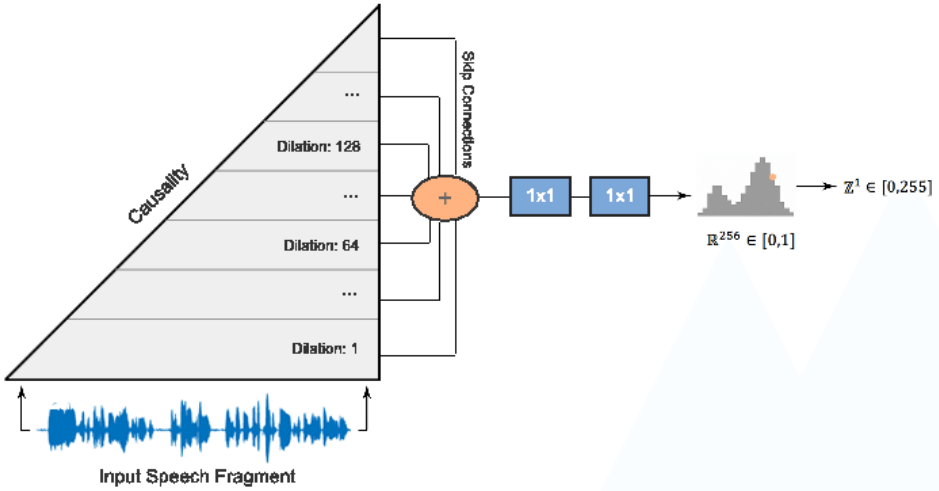
- 统计参数合成技术
  - 利用数学模型统计出来一个语言的特色；
  - 重建出声音的波形。

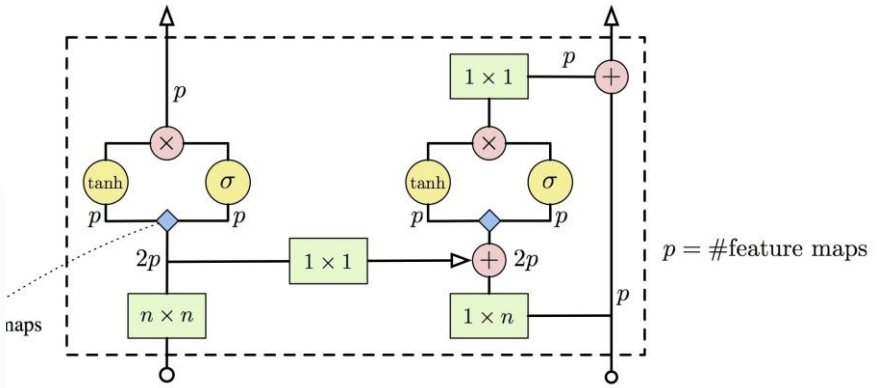
优缺点：好处是它不需要大量的资料库，但合成出的声音没有抑扬顿挫，并不真实。

1.3 语音合成技术

- 基于深度学习的语音合成技术

➢ WaveNet（google，2016）提出使用卷积神经网络来学习输入样本的特征，再理解一个语言的发音特征，然后对于输入的文字根据学习到的特征产生相应的语音。



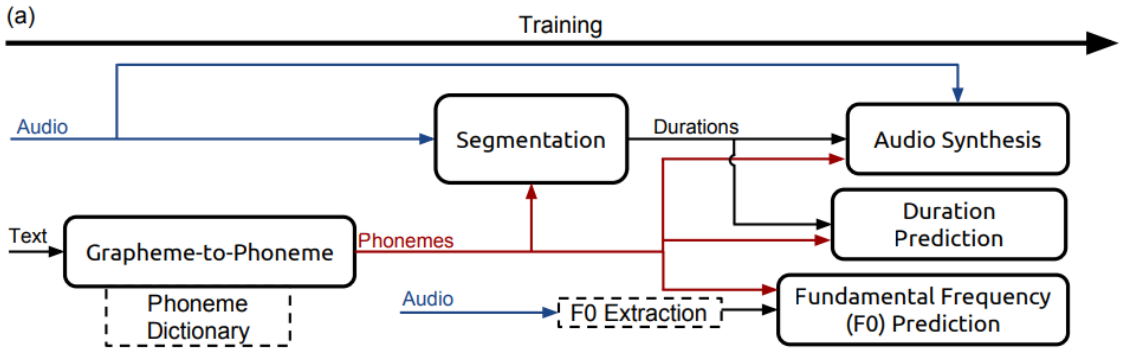


缺点：速度太慢；需要文本前端的支持，前端分析出错，将直接影响合成效果。

1.3 语音合成技术

- 基于深度学习的语音合成技术

➢ Deep Voice 1（百度，2017）百度deep voice的做法是仿照传统参数合成的各个步骤，将每一阶段用一个神经网络模型来代替。那整个模型就是一个大的神经网络。



缺点：百度提供了一套完整的TTS解决方案，用的人工特征少，而WaveNet, SampleRNN, Char2Wav这些方法需要依赖于一个现有TTS的部分功能模块为其提供特征。Deep Voice 1实时性更好。

本堂课内容

复习上堂课内容

虚拟现实的关键技术

1 三维虚拟声音技术；

2 人机交互技术；

2 人机交互技术

- 从计算机诞生至今，计算机的发展是极为迅速的，而人与计算机之间交互技术的发展是较为缓慢的，人机交互界面经历了以下几个发展阶段。

➢ 20世纪40年代到70年代，人机交互采用的是命令行方式(CLI)，这是人机交互界面第一代，人机交互使用了文本编辑的方法，可以把各种输入输出信息显示在屏幕上，并通过问答式对话、文本菜单或命令语言等方式进行人机交互。

因此，这一时期的人机交互界面的自然性和效率较差。人们使用计算机，必须先经过很长时间的培训与学习 。

2 人机交互技术

- 到20世纪80年代初，出现了图形用户界面方式(GUI)，GUI的广泛流行将人机交互推向图形用户界面的新阶段。人们不再需要死记硬背大量的命令，可以通过窗口(Windows)、图标(Icon)、菜单(Menu)、指点装置(Point)直接对屏幕上的对象进行操作。

与命令行界面相比，图形用户界面采用视图、点(鼠标)，使得人机交互的自然性和效率都有较大的提高，极大地方便了非专业用户的使用。

2 人机交互技术

2 人机交互技术

- 到20世纪90年代初，多媒体界面成为流行的交互方式，它在界面信息的表现方式上进行了改进，使用了多种媒体。同时，界面输出也开始转为动态、二维图形／图像及其他多媒体信息的方式，从而有效地增加了计算机与用户沟通的渠道。

- 但无论是命令行界面，还是图形用户界面，都不具有以上所述的进行自然、直接、三维操作的交互能力。因为在实质上它们都属于一种静态的、单通道的人机界面，而用户只能使用精确的、二维的信息在一维和二维空间中完成人机交互。

2 人机交互技术

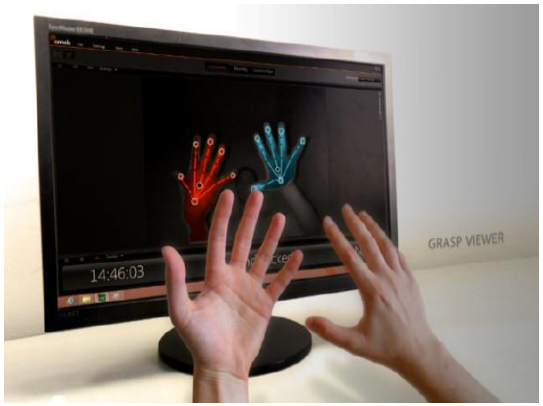
- 人机自然交互技术是指在计算机系统提供的虚拟环境中，人应该可以使用眼睛、耳朵、皮肤、手势和语音等各种感觉方式直接与之发生交互的技术。
- 在虚拟现实相关技术中，嗅觉和味觉技术的开发还处于探索阶段。
- 在虚拟现实领域中较为常用的交互技术主要有：手势识别、面部表情识别、眼动跟踪、语音识别等。

2.1 手势识别

- 手势识别系统的输入设备主要分为基于数据手套的识别和基于视觉(图像)的识别系统。

➢ 基于数据手套的手势识别系统，是利用数据手套和空间位置跟踪定位设备来捕捉手势的空间运动轨迹和时序信息。

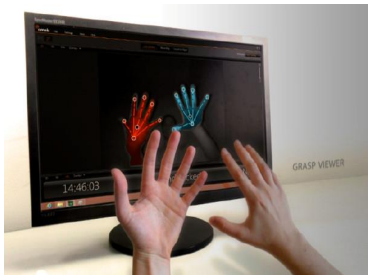
➢ 基于视觉的手势识别是通过摄像机连续拍摄手部的运动图像，然后采用图像处理技术提取出图像中的手部轮廓，进而分析出手势形态。



基于数据手套和基于视觉（图像）的两种手势识别技术

2.1 手势识别

VS

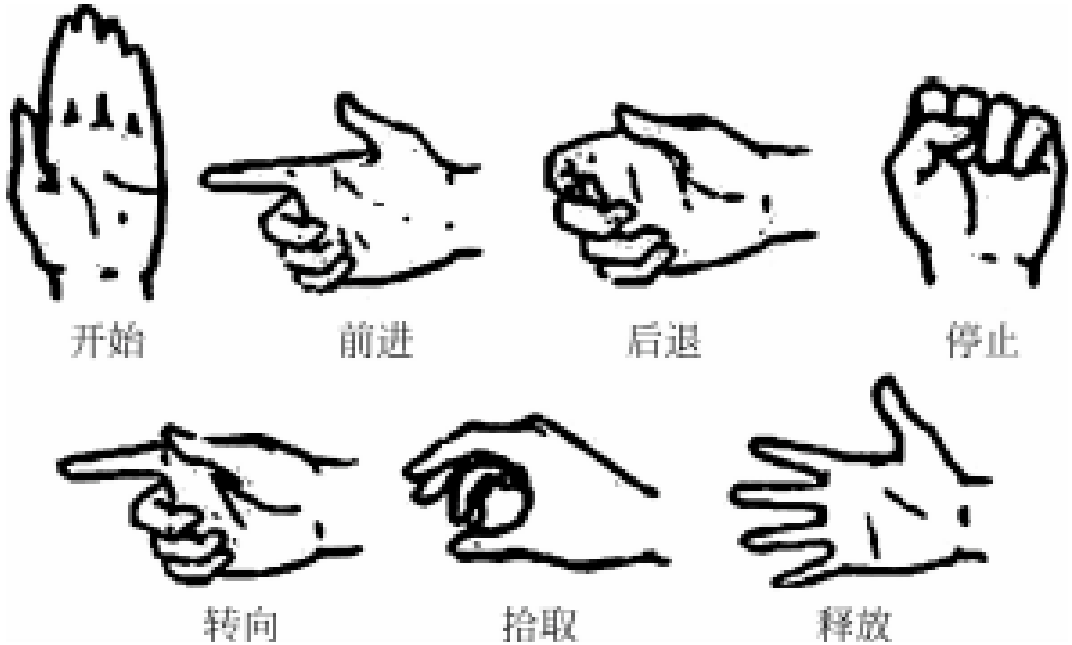


优点：系统的识别率高；
缺点：做手势的人要穿戴复杂的数据手套和位置跟踪器，相对限制了人手的自由运动，并且数据手套、位置跟踪器等输人设备价格比较昂贵。

优点：点是输入设备比较便宜，使用时不干扰用户；
缺点：识别率比较低，实时性较差，特别是很难用于大词汇量的手势识别。

2.1 手势识别

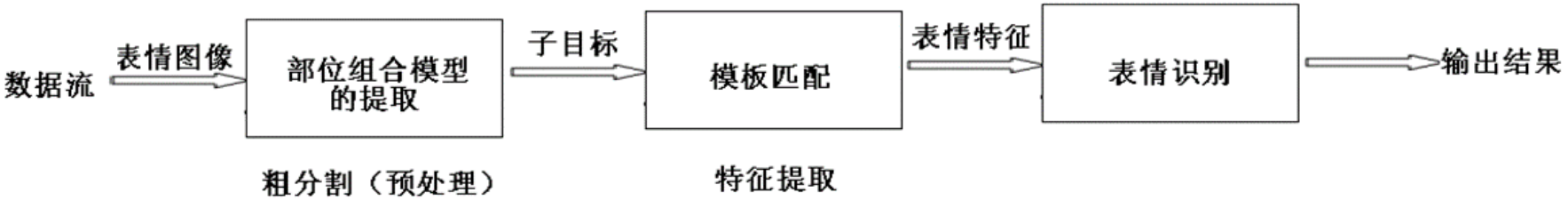
- 在手势规范的基础上，手势识别技术一般采用模板匹配方法将用户手势与模板库中的手势指令进行匹配，通过测量两者的相似度来识别手势指令。



2.2 面部表情识别

- 面部表情识别在人与人交流过程中传递信息时发挥重要的作用。
- 目前，计算机面部表情识别技术通常包括人脸图像的检测与定位、表情特征提取、模板匹配、表情识别等步骤。

2.2 面部表情识别



- **人脸图像的检测与定位**

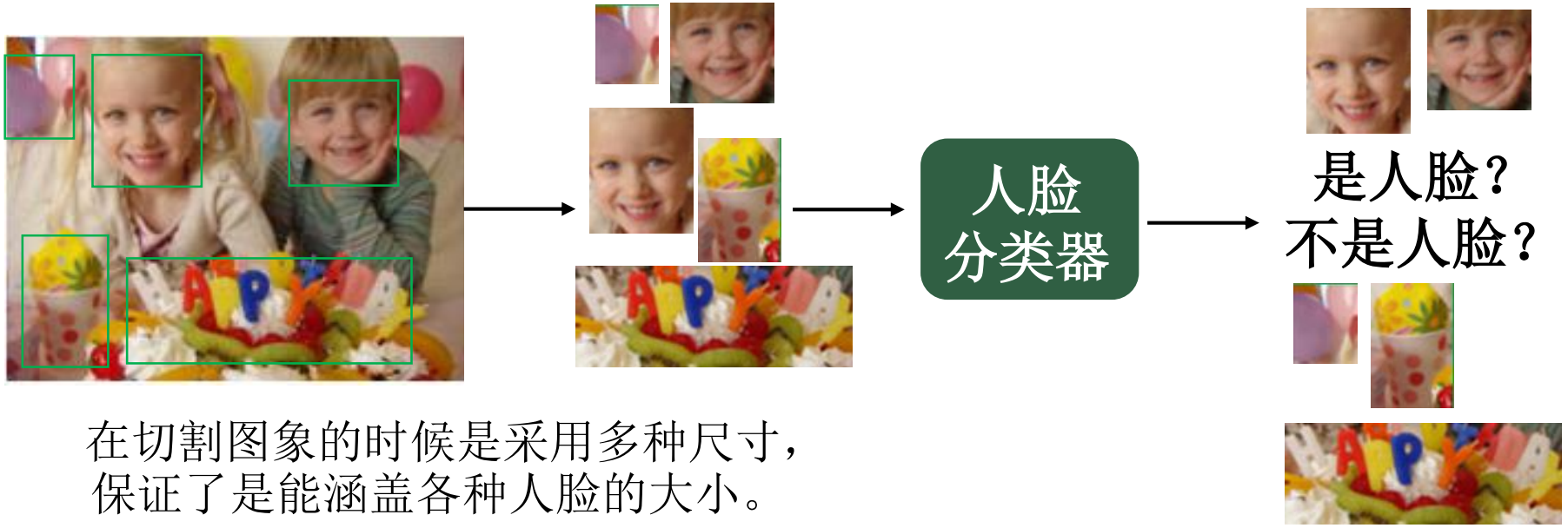
➢人脸图像的检测与定位就是在输入图像中找到人脸的确切位置。

➢人脸检测的基本思想是建立人脸模型，比较输人图像中所有可能的待检测区域与人脸模型的匹配程度，从而得到可能存在人脸的区域。

➢根据对人脸知识利用方式的不同，可以将人脸检测方法分为两大类：基于特征的人脸检测方法和基于图像的人脸检测方法。

2.2 面部表情识别

- 人脸图像的检测与定位



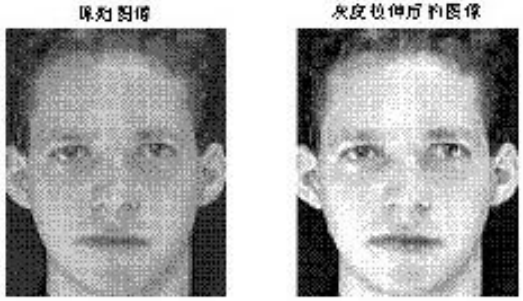
在切割图象的时候是采用多种尺寸，
保证了是能涵盖各种人脸的大小。

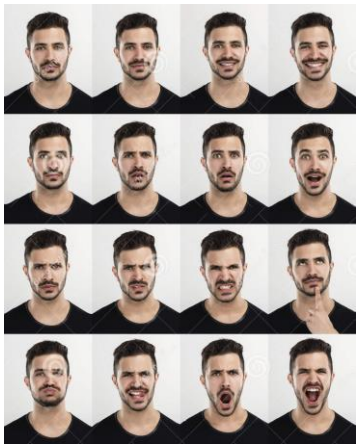
2.2 面部表情识别

- **表情图像预处理**

➢图像预处理常常采用信号处理的形式（如去噪、像素位置或者光照变量的标准化）， 还包括人脸及它的组成与分割、定位或者跟踪。

➢表情的表示对图像中头的平移、尺度变化和旋转是敏感的。为了消除这些不必要的变换的影响，人脸表情图像可以在分类前进行标准化的预处理。





2.2 面部表情识别

- **表情特征提取**

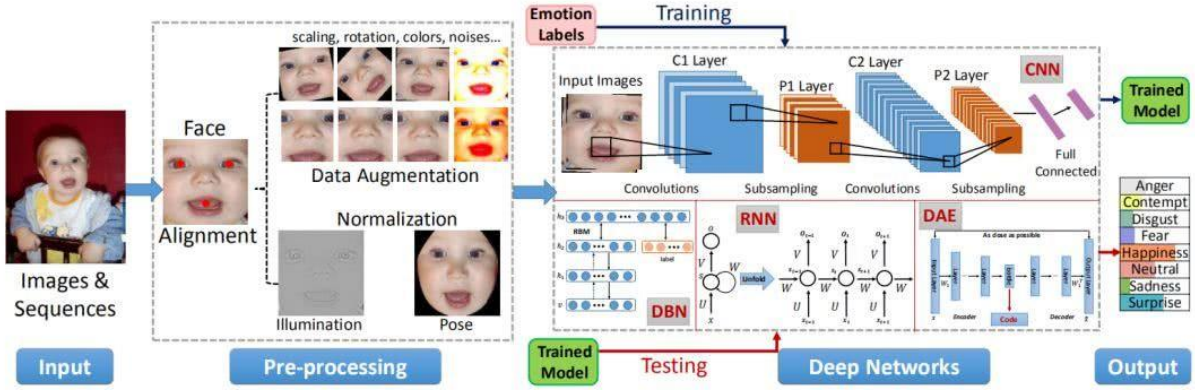
➢表情特征的提取根据图像性质的不同可分为：静态图像特征提取和序列图像特征提取。

➢特征选取的依据是：

尽可能多地携带人脸面部表情的特征，即信息量丰富；

尽可能容易提取；

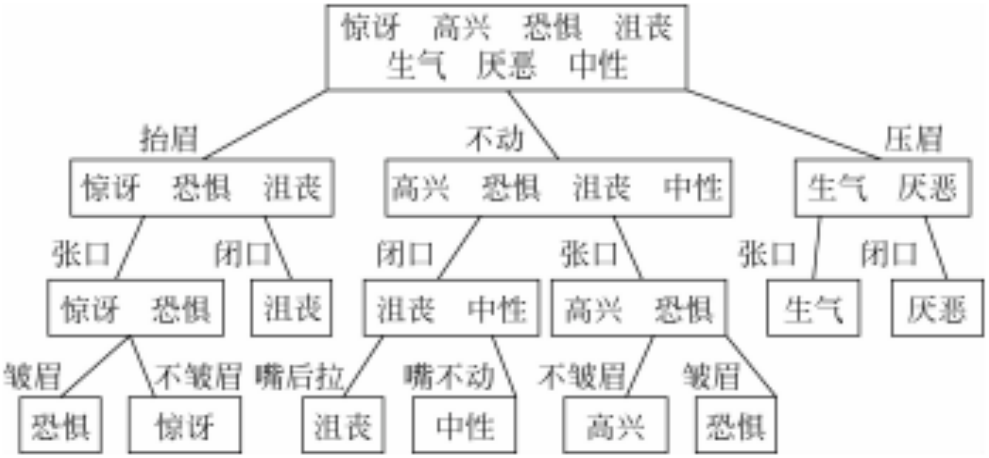
信息相对稳定，受光照变化等外界的影响小；



2.2 面部表情识别

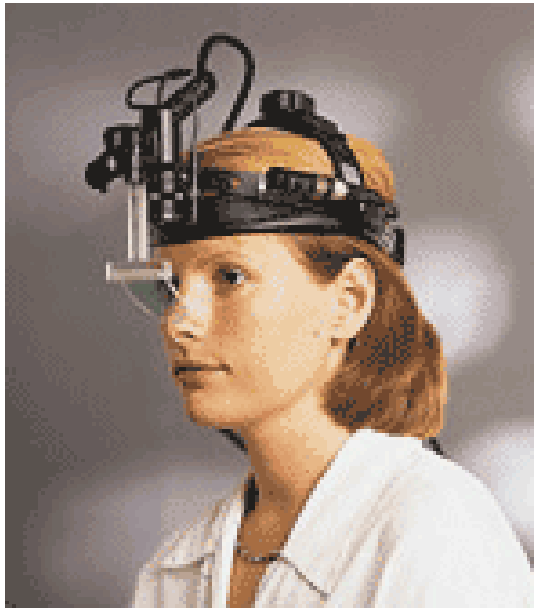
- 表情识别

➢基本方法是在样本集的基础上确定判别规则，对于新给定的对象根据已有的判别规则来分类，从而达到识别的目的。一个良好的分类器使分类造成的错误率最小，因此，分类器的设计也是表情识别的关键。

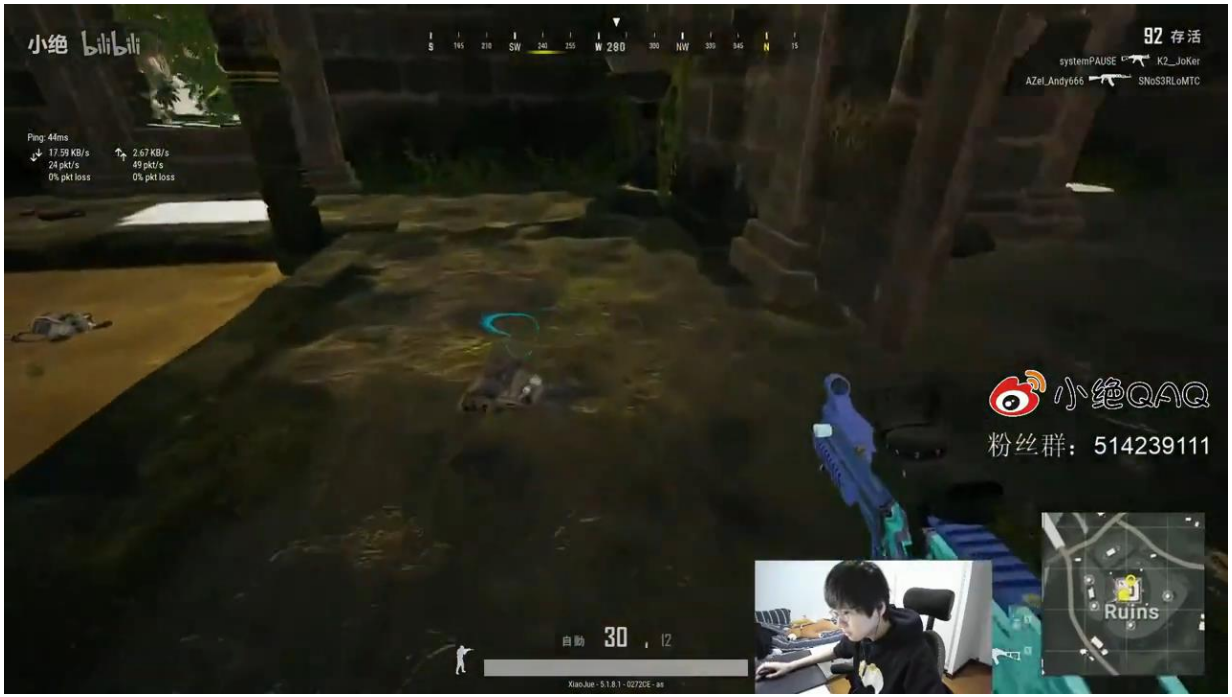


2.3 眼动跟踪技术

- 在虚拟世界中生成视觉的感知主要依赖于对人头部的跟踪，即当用户的头部发生运动时，生成虚拟环境中的场景将会随之改变，从而实现实时的视觉显示。
- 但在现实世界中，人可能在不转动头部的情况下，仅通过移动视线来观察一定范围内的环境或物体。在这一点上，单纯依靠头部跟踪是不全面的。为了模拟人眼的这个功能，引入了眼动跟踪技术。



2.3 眼动跟踪技术


小绝 bilibili
92 存活
systemPAUSE
K2_JoKer
AZel_Andy666
SNoS3RLoMTC
Ping: 44ms
17.59 KB/s
24 pkt/s
0% pkt loss
2.67 KB/s
49 pkt/s
0% pkt loss
小绝QAQ
粉丝群：514239111
Ruins
30


抖音
抖音号: zju1897

2.3 眼动跟踪技术


1 眼控仪内置红外光源、光学传感器、图像处理器以及视计算中心点
2 创建出对应的图像控射到人眼上
3 捕获用户头部、眼睛的图像信息
4 提取捕获图像的特征
5 精确计算注视点（Gaze Point）的位置
注视点
眼控仪

2.3 眼动跟踪技术

- 常见的视觉追踪方法有眼电图、虹膜-巩膜边缘、角膜反射、瞳孔-角膜反射、接触镜等。常见的几种视觉追踪方法的比较如表所示。

视觉追踪方法	技术特点
眼电图（EOG）	高带宽，精度低，对人干扰打
虹膜-巩膜边缘	高带宽，垂直精度低，对人干扰大
角膜反射	高带宽，误差大
瞳孔-角膜反射	低带宽，精度高，对人无干扰，误差小
接触镜	高带宽，精度最高，对人干扰大，不舒适

2.3 眼动跟踪技术

- 目前眼动跟踪技术主要存在以下问题：

➢数据提取问题：对大量采集的数据进行快速的存储和分析是一个困难的题目。

➢数据解释问题：目前，眼动跟踪数据的分析主要基于认知理论和模型的自上而下分析法和自下而上的观察法。由于眼动存在固有的抖动和眨动，导致从眼动数据中提取的准确信息较为困难。

2.3 眼动跟踪技术

- 目前眼动跟踪技术主要存在以下问题。

➢精度和自由度问题：以硬件为基础的眼动跟踪技术，其精度可以达到0.1°，但所应用的设备却限制了人的自由，使用起来很不方便。但以软件为基础的眼动跟踪技术对用户移动限制降低但精度也很低，只有2°。

➢米达斯接触（Midas Touch）问题：指的是由于用户实现运动的随意性而造成计算机对用户意图识别的困难。用户可能只想随便看看，计算机不必采取任何动作，这一点计算机还很难区分。

2.3 眼动跟踪技术

- 目前眼动跟踪技术主要存在以下问题。

➢算法问题：由于眼动跟踪技术还没有完全成熟，而且眼动本身的特点造成易数据中算，会存在许多信号干扰。另外，视觉模型要和听觉模型等其他模型配合起来才能发挥更大的作用，体术一个合理的整合模型和算法也是一个极大的挑战。

本章小结

- 本章介绍了VR系统中关键的三位虚拟声音技术和人机交互技术。
- 在三维虚拟声音有两个主要特征： 全向三维定位特性和三维实时跟踪特性。语音应用技术在虚拟现实技术中的关键技术是语音识别技术和语音合成技术。
- 人机交互技术在虚拟实领域中较为常用的技术主要有：手势识别、面部表情识别、眼动跟踪、语音识别等。

中山大學
SUN YAT-SEN UNIVERSITY

Thanks !