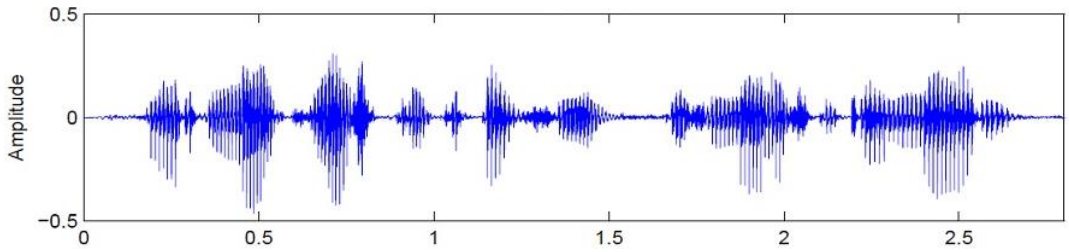
# 语音助手是怎么听懂人说话的？


当前很多手机安装了语音助手，如苹果的 Siri，华为的 HiAssistant。 这些软件和一个电子助手一样，可以和主人进行对话，帮主人做些简单的事情，如天气查询、电话呼叫等。

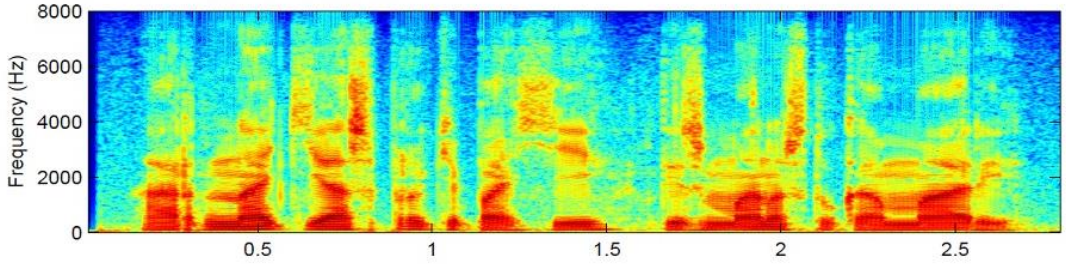
那么，语音助手是如何听懂主人命令的呢？这里就要谈到语音识别和自然语言理解两种技术，前者让机器“听到”命令，后者让机器“听懂”命令。

首先看语音识别。机器通过麦克风，把人的语音收集起来，得到如下图的声波，其中横轴是时间，纵轴是声波强度。




这一时域震动信号不好分析，科学家们将它变换到频域，发现不同发音内容在频谱上的表现形式是不同的。如下图所示，在一个较短的发音片段里，频谱特征保持不变，且不同时刻的频谱特征不同。一般来说，元音表现出较明显的粗横纹，而辅音则没有这些横纹结构。利用这些频谱上特性，即可识别出不同的音素，如 a, o, e 等。将这些音素组合起来，即可识别出词和句子。现代语音识别系统一般

基于复杂的统计模型，以处理发音上的各种变异以及发音之间的相关性。同时，还需要借助语言知识对识别结果进行约束，比如“我被鱼刺卡了”要比“我被鱼翅卡了”可能性更大，因此识别系统倾向于将前者作为正确输出。




语音识别得到一个句子，但并不太理解句子的内容，因此还不能说是“听懂”。自然语言理解技术利用大量语言学知识来解决这一问题。

以语音助手为例，语言理解主要包括两个方面：一是用户的意图，二是实现这一意图的关键信息。例如，用户说“请告诉我明天的天气”，这句话的意图是“天气查询”，而这一意图中包括的关键信息是“明天”（而不是“请”，也不是“告诉”）。

理解用户的意图一般基于意图分类：首先定义好一些具本的意图，然后设计一个模型来判断输入句子属于每一类意图的可能性，可能性最大的意图即是用户的意图。如果所有意图都不太可能，语音助手就会甜蜜地告诉你：“主人，我没听懂你的话哟”。

基于得到的意图，系统将试图定位该意图的关键信息。一种简单的方式是在句子中寻找具有相应功能的词。例如，如“天气查询”这一意图需要“时间”这一关键信息，而“明天”恰好就是代表时间的词，由此系统就知道了用户的目的是“明天”的天气。

本文的介绍只是语音识别和自然语言理解技术的一点点皮毛，实际中用到的技术非常复杂。事实上，让机器听懂人的语言从人工智能诞生那天起就是科学家们为之奋斗的目标。半个世纪过去了，这个目标终于慢慢成为现实。然而，直到今天，机器能听懂的也只是人类语言中非常有限的一部分。为了和机器愉快聊天这一伟大理想，无数科学家依然在日以继夜地工作着。