



## 智能音箱与声音传感器

(北京大学 朱星 编译自 Pip Knight. *Physics World*, 2020, (12): 25)

现代生活已经产生微妙变化。在节日期间，我们会与家中的智能音箱对话：“爱丽丝，请播放些圣诞歌曲”，“Google，请打开美妙的灯光”，“Siri，请问烤火鸡还需要多长时间出炉”，就如同与家里成员一样对话，而这种无形的指令几乎是瞬时得到执行。

包括亚马逊、谷歌、苹果的这类装置已经出现在英国五分之一的家庭中。2019年，全球已经销售了1.47亿套，2020年销量预计会增加10%。令人惊讶的是，智能音箱已经达到了很高的语音识别能力和精确性。这些都归因于高灵敏度的语音传感器和用来解释语音的复杂机器学习算法。

从正常讲话转换为文本需要两个过程：(1)一个语音传感器将接收的声波转换为电信号；(2)使用软件识别出语音中的词语。对于第二阶段，电信号首先由模拟信号转换为数字信号，然后用快速傅里叶变换找出不同频率信号随时间的振幅变化。用算法语言将单音(phones)与标

准的音素(phoneme)进行比较，由此构成完整讲话。在语音识别过程中，机器学习非常重要，可以用来提高精确度。计算程序会记住我们对所说话的修正，因此在解读我们个人的声音时变得更加精确。

### Audrey项目与电容式传感器

高灵敏度声学探测器源于19世纪末。最初的声音传感器是碳粉接触式麦克风，由美国的E. Berliner和T. Edison，英国的D. Hughes独立发明。这种麦克风将碳粉颗粒压缩在两片金属板之间，然后在两侧加上电压。传入的声波使得其中一个金属膜片产生振动。在压缩过程中，碳颗粒变形而增加了相互接触面积，使接触电阻下降引起电流增加。随着膜片运动使得声音可以通过电流变化记录下来。

然而，直到1952年才首次实现语音识别技术。美国的贝尔电话实验室设立了“自动数字识别机”(Audrey)项目，可以在普通电话中识别数字0—9，用于语音拨号，然而需要对用户声音进行训练，以及

许多其他电子仪器。

从Audrey设立以来，语音识别的计算方面已经有了长足的发展，语音传感器也接受了严格考验。出现了铝带式麦克风、动圈式麦克风、碳粒麦克风等，但先后淡出市场，而电容式传感器却一直是主流。1916年美国西部电子工程实验室的E. C. Wente发明了电容传感器，利用一个电容器平板之间的电压与间距有关的物理效应。在一个固定的背板和一个运动的薄膜两面加上电压，随着外部声波振动的薄膜引起电容两端电压的变化，由此可以计算出不同频率声波引起的振幅变化。

贝尔电话实验室的G. Sessler等人于1962年发明了驻极体电容麦克风(ECM)。驻极体材料(如聚四氟乙烯)具有本征的表面电荷，可以在电容器两端保持固定的电压，从而降低了输入功率。直径为3—10 mm的ECM占据了麦克风市场大约50年时间。然而，减小传感器尺寸会导致信噪比和稳定性下降，特别是

在温度变化的环境中。

当用于语音识别时，多数ECM传感器已经被微机电系统(MEMS)电容传感器所取代。智能音箱中的这种传感器直径约20—1000  $\mu\text{m}$ 。MEMS传感器与ECM的区别在于内部的模拟—数字转换电路。与ECM相比，MEMS器件对电子噪声不敏感，尺寸也更小，采用半导体工艺线加工，因而更易于批量制作。MEMS传感器的缺点是寿命不长，不适于恶劣的工作环境。沉积在膜片的颗粒、雨水和附着在膜片表面的空气层都会降低其灵敏度。

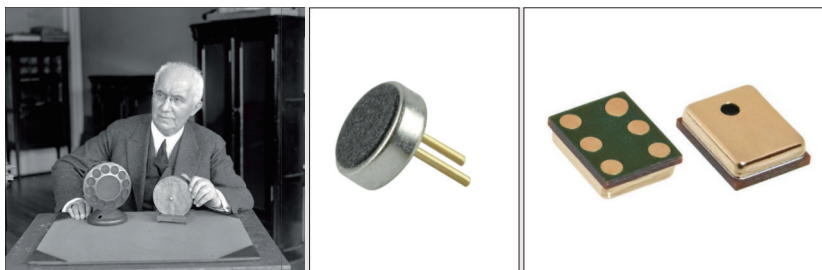
### 新的解决方案

尽管电容式传感器已经在工业界占据了数十年主导地位，但并非未来发展的首选。美国Vesper公司设计了压电语音传感器，成为新的解决方案。这家2014年建立的公司最初的设计是源于公司CEO Bobby Littrell博士的研究。

压电语音传感器采用压电材料制成的膜片，如锆钛酸铅压电材料，将机械能直接转换为电响应。当压电薄膜接收到声波，其内部离子间距离会增长，从而产生电偶极子，使得结构中的离子形成能量最低的分布。这种偶极子只能存在于非中心对称结构晶体单胞中。偶极子在晶体中的累积效果会产生电压，电压随着晶体内应变的变化而变化。

与电容式语音传感器相比，压电式传感器具有的优势是不会沾上污染物、空气或者水分，因此寿命更长。另外，这种器件是自供电的，节省了用于电池的空间。

然而，像这样的薄膜设备——以及电容式设计——往往很难制备，需要在高真空甚至超高真空环境。需要选择合适的衬底，按照单胞的某一晶体取向生长薄膜，以便



与时俱进的语音传感器。自从19世纪E. Berliner (左)，T. Edison和D. Hughes首次发明了碳粒麦克风以来，语音传感器经历了很大进展；(中)驻极体电容麦克风；(右)MEMS电容式麦克风

在机械应变条件下生长的偶极子均朝向同一方向。需要高温来提高原子的迁移性，使得原子在衬底的最低能量位置上形成理想点阵。然而，单晶的薄膜必须生长在有序的结构上，而柔性衬底是非晶结构，难于生长单晶薄膜。

### 向大自然学习

在语音识别领域中，韩国KAIST的团队发展了一种新的模仿人类听力的压电传感器。他们的压电传感器具有与人类耳蜗的基底膜类似的形状，因此，可以收集常规电容式传感器两倍的信息。这一优势源于，不仅可以收集含有所有频率的单一信号，从中提取频率与振幅信息，而且能在薄膜不同的位置获取多个信号。丰富的信息使得语音识别更加准确。这种设计的精确度和灵敏度占优势，可以获得远处的音频信号，并且能够分辨单个声音。

他们研究中的棘手问题是分析来自这些通道的信号，给出不同频率信号的相对振幅，这是由于振幅受到了通道共振行为的调制。该团队认为已经找到了适于这种探测器的通道数，但是必须在收集更多信息以提高精确度与适当大小的处理器之间取得平衡。

### 喉部传感器与新冠病毒

语音识别技术并不限于将传感

器放置在房屋各个角落，或者你的口袋里。用于探测喉部振动而不是探测声波的传感器，对于声音几乎无法传播的场合是非常重要的，如在嘈杂的工业环境，或者人们佩戴笨重的防毒面具时。2019年韩国浦项科技大学做出了突破性工作，研制出柔性并且可以贴在皮肤上的电容传感器。这种传感器通过探测喉部环状软骨上的皮肤振动来感知人的声音。由于喉部皮肤的加速度与声压存在线性关系，因此，可以通过测量电容值的变化感知喉部加速度，进而转换为声压。团队制备了厚度小于5  $\mu\text{m}$ 的环氧树脂薄膜，用以模拟自然界中柔性的聚合物材料。

这类喉部传感器或许还可以用于诊断类似COVID-19的疾病，其症状表现在声带的振动上。美国麻省理工学院的研究发现，COVID-19会通过影响喉部肌肉的复杂运动而改变发出的声音。他们将探测装置与手机小程序结合，可以用做早期COVID-19的筛选。

语音识别探测器的未来是面向应用的智能装置，如灵敏度高，能够识别作为密码或者指纹的个体声音信息。由于大流行，能够快速、批量检测呼吸道疾病，利用喉头监测器进行快速检测的技术很快会成为一种可行的诊断技术。