听懂声音—ADI公司的人工智能如何大幅延长设备的正常运行时间

作者：Sebastien Christian

简介

任何深谙设备维护必要性的人都知道，设备发出的声音和振动有多重要。通过声音和振动进行适当的设备健康监测，可以将维护成本降低一半，使用寿命延长一倍。实现实时声学数据和分析是另一种重要的基于状态的系统监测 (CbM) 方法。

我们可以学着了解设备发出的正常声音是什么样的。当声音出现变化时，我们可以确认出现异常。然后我们可以了解是什么问题，通过这样的方式把声音和特定的问题联系在一起。识别异常可能需要进行几分钟的训练，但将声音、振动和原因结合起来实施诊断可能需要一辈子的时间。经验丰富的技工人员和工程师可能具备这种知识，但他们属于稀缺资源。单单通过声音本身识别问题可能相当困难，即使使用录音、描述性框架或接受专家亲自培训也是如此。

因此，ADI 公司团队在过去 20 年里一直致力于理解人类是如何解读声音和振动的。我们的目标是建立一个系统，能够学习来自设备的声音和振动，破译它们的含义，以检测异常行为，并进行诊断。本文详细介绍了 OtoSense 的体系结构，它是一种设备健康监测系统，支持我们所说的计算机听觉，让计算机能够理解设备行为的主要指标：声音和振动。

该系统适用于任何设备，可以实时工作，无需网络连接。它已被应用于工业应用，支持实现一个可扩展的高效设备健康监测系统。

本文探讨了引导 开发 OtoSense 的原则，以及在设计 OtoSense 期间，人类听觉所发挥的作用。然后，本文讨论了声音或振动特性的是如何被设计出来的、如何从这些特性了解其代表的意义，以及在持续学习中如何不断改变和改进 OtoSense，用于执行愈加复杂的诊断，且结果更为精准。

指导原则

为了保证耐用、不可知且高效，OtoSense 设计理念秉持几个指导原则：

- **从人类神经学中获得灵感。**人类可以以一种非常节能的方式学习和理解他们听到的任何声音。
- **能够学习静态声音和瞬态声音。**这需要不断调整功能和持续实施监测。
- **在靠近传感器的终端进行识别。**应该无需通过网络连接远程服务器来做出决策。
- **与专家互动，向他们学习，前提是尽可能避免干扰他们的日常工作，**且过程要尽可能愉悦。

人类听觉系统和对 OtoSense 的解析

听觉是一种关乎生存的感觉。它是对遥远的、看不见的事件的整体感觉，在出生前就已成熟。

人类感知声音的过程可以用四个熟悉的步骤来描述：声音的模拟获取、数字转换、特征提取和解读。在每个步骤中，我们都会将人耳与 OtoSense 系统比较。

- **模拟获取和数字化。**中耳中的膜和杠杆捕捉声音，然后调整阻抗，将振动传输到充液腔道中，在那里，另一层膜会根据信号中存在的光谱成分选择性地移位。这反过来弯曲了弹性单元，这些单元发出数字信号，反映出弯曲程度和强度。然后，这些单独的信号通过按频率排列的平行神经传递到初级听觉皮层。
  - 在 OtoSense 中，这项工作由传感器、放大器和编解码器来完成。数字化过程使用固定的采样速率，可在 250 Hz 和 196 kHz 之间调节，波形在 16 位编码，然后存储到大小在 128 到 4096 之间的缓冲区。
- **特性提取**发生在初级皮层：频率域特性，如主频率、谐波和频谱形状，以及时间域特性，如脉冲、强度变化和在大约 3 秒时间窗内的主要频率成分。
  - OtoSense 使用一个时间窗，我们称之为“块”，它以固定的步长移动。这个块的大小和步长范围为 23 毫秒到 3 秒，具体由需要识别的事件和在终端提取特性的采样率决定。在下一节中，我们会就 OtoSense 提取的特性进行更详细地解释。

- **解析发生在联络皮层**，它融合了所有的感知和记忆，并赋予声音以含义（比如通过语言），在塑造感知期间起着核心作用。解析过程会组织我们对事件的描述，远远不止是对它们进行命名这么简单。为一个项目、一个声音或一个事件命名可以让我们赋予它更大、更多层的含义。对于专家来说，名字和含义能让他们更好地理解周围的环境。
  - 这就是为什么 OtoSense 与人的互动始于基于人类神经学的视觉、无监督的声音映射。OtoSense 利用图形表示所有听到的声音或振动，它们按相似性排列，但不尝试创建固定分类。这让专家们能够组织屏幕上显示的组，并为它们命名，而无需尝试人为创建有界线的类别。他们可以根据自身的知识、感知和对 OtoSense 最终输出的期望构建语义地图。对于同样的音景，汽车机械师、航空工程师，或者冷锻压力机专家，甚至是研究相同领域，但来自不同公司的人员，都可以按不同的方式进行划分、组织和标记。OtoSense 则与塑造语言意义一样，使用相同的自下而上的方法来给定意义。

从声音和振动到特性

经过一段时间（如之前所示，时间窗或块），我们会给某个特征分配一个单独的编号，用于描述该时间内声音或振动的给定属性/质量。OtoSense 平台选择特性的原则如下：

- 对于频率域和时域，**特征都应该尽可能完整地描述环境**，提供尽可能多的细节。它们必须描述静止的嗡嗡声，以及咔哒声、哗啦声、吱吱声和任何瞬间变化的声音。
- **特征应尽可能按正交方式构成一个集合**。如果一个特征被定义为“块上的平均振幅”，那么就不应该有另一个特征与之高度相关，例如“块上的总光谱能量”。当然，正交性可能永远无法实现，但不应将任何一种表述为其他特征的组合，每种特征都必须包含单一信息。
- **特性应该最小化计算量**。我们的大脑只知道加法、比较和重置为 0。大多数 OtoSense 特性都被设计成增量，这样每个新示例都可以通过简单的操作修改特性，而不需要在完整的缓冲区，或者更为糟糕的，在块上重新进行计算。最小化计算量还意味着可以忽略标准物理单元。例如，尝试用值（以 dBA 为单位）表示强度是没有意义的。如果需要输出 dBA 值，则可以在输出时完成（如果必要）。

在 OtoSense 平台的 2 到 1024 个特性中，有一部分描述了时域。它们要么是直接从波形中提取，要么是从块上任何其他特性的演化中提取。在这些特性中，有些包括平均振幅和最大振幅、由波形线性长度得到的复杂度、振幅变化、脉冲的存在与否和其特性、第一个和最后一个缓冲区之间相似性的稳定性、卷积的超小型自相关或主要频谱峰值的变化。

在频域上使用的特性提取自 FFT。FFT 在每个缓冲区上计算，产生从 128 到 2048 个单独频率的输出。然后，该过程创建一个具有所需维数的向量，该向量比 FFT 小得多，但仍能细致地描述环境。OtoSense 最初使用一种不可知的方法在对数频谱上创建大小相同的数据桶。然后，根据环境和要识别的事件，这些数据桶将重点放在信息密度高的频谱区域，要么是从能够熵最大化的无监督视角，要么是从使用标记事件作为指导的半监督视角来判断。这模拟了我们的内耳细胞结构，在语言信息密度最大的地方，语音细节更密集。

结构：支持终端和本地数据

OtoSense 在终端位置实施异常检测和事件识别，无需使用任何远程设备。这种结构确保系统不会受到网络故障的影响，且无需将所有原始数据块发送出去进行分析。运行 OtoSense 的终端设备是一种自包含系统，可以实时描述所鉴听设备的行为。

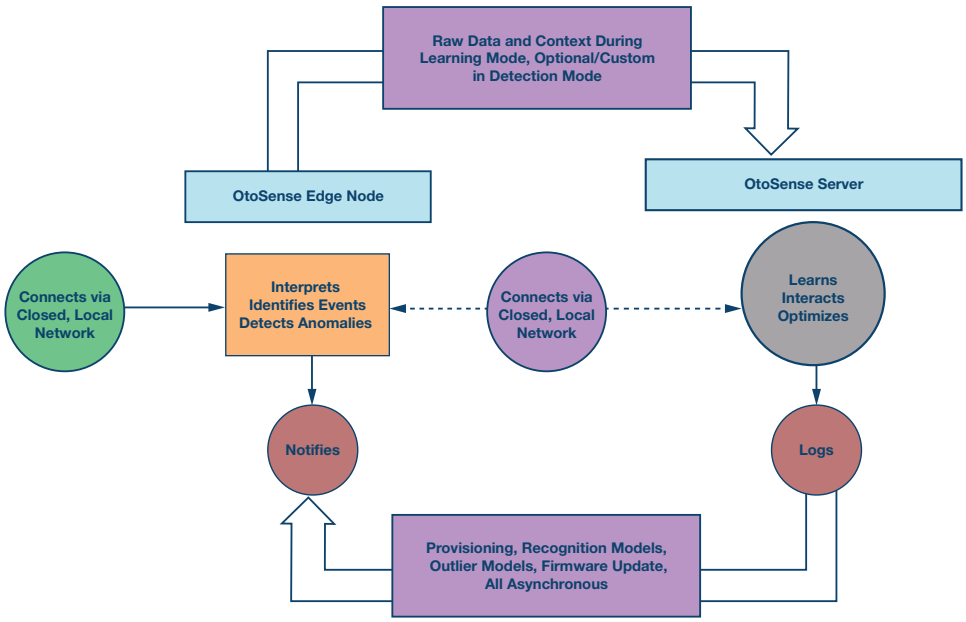

图1. OtoSense 系统。

运行 AI 和 HMI 的 OtoSense 服务器一般托管在本地。云架构可以将多个有意义的数据流聚合成为 OtoSense 设备的输出。对于一个专门处理大量数据并在一个站点上与数百台设备交互的 AI 来说，使用云托管的意义不大。

从特性到异常检测

正常/异常评估无需与专家进行太多交互。专家只需要帮忙确定表示设备声音和振动正常的基线。然后，在推送给设备之前，先将这个基线在 Otosense 服务器上转换为异常模型。

然后，我们使用两种不同的策略来评估传入的声音或振动是否正常：

- 第一种策略是我们所说的“常态性”，即检查任何进入特性空间的新声音的周围环境、它与基线点和集群的距离，以及这些集群的大小。距离越大，集群越小，新的声音就越不寻常，异常值也就越高。当这个异常值高于专家定义的阈值时，相应的块将被标记为不寻常，并发送到服务器供专家查看。
- 第二种策略非常简单：任何特性值高于或低于特性定义的基线的最大值或最小值的传入块都被标记为“极端”，并发送到服务器。

异常和极端策略的组合很好地涵盖了异常的声音或振动，这些策略在检测日渐磨损和残酷的意外事件方面也表现出色。

从特征到事件识别

特征属于物理领域，含义属于人类认知。要将特征与含义联系起来，需要 OtoSense AI 和人类专家之间展开互动。我们花了大量时间研究客户的反馈，开发出人机界面 (HMI)，让工程师能够高效地与 OtoSense 交互，设计出事件识别模型。这个 HMI 允许探索数据、标记数据、创建异常模型和声音识别模型，并测试这些模型。

OtoSense Sound Platter（也称为 splatter）允许通过完整概述数据集来探索和标记声音。Splatter 在完整的数据集中选择最有趣和最具代表性的声音，并将它们显示为一个混合了标记和未标记声音的 2D 相似性地图。

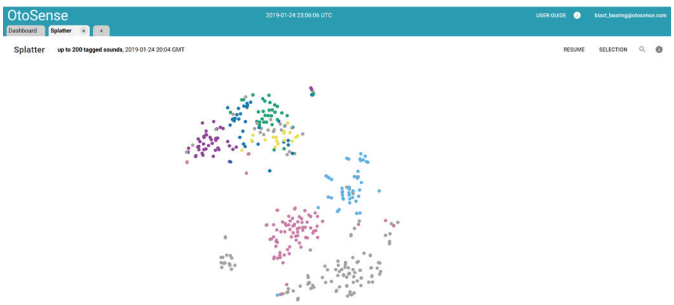

图2. OtoSense Sound Platter 中的 2D splatter 声音地图。

任何声音或振动，包括其环境，都可以通过许多不同的方式进行可视化——例如，使用 Sound Widget（也称为 Swidget）。

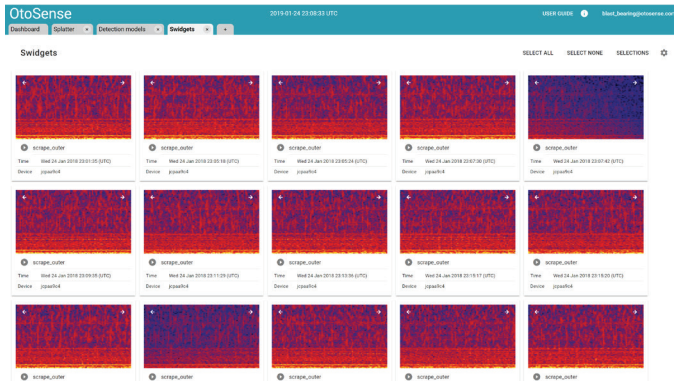

图3. OtoSense sound widget (swidget)。

在任何时候，都可以创建异常模型或事件识别模型。事件识别模型是一个圆形的混淆矩阵，它允许 OtoSense 用户探索混淆事件。

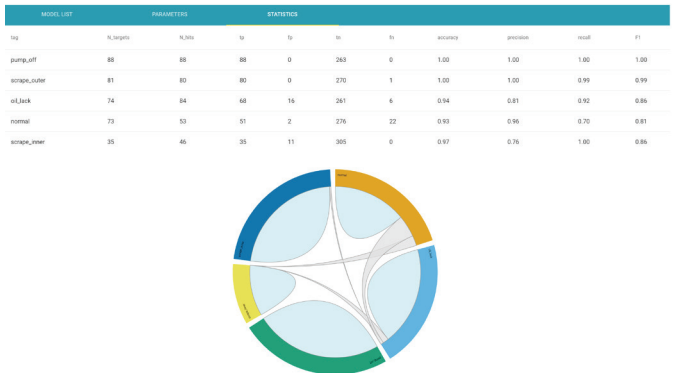

图4. 可以基于所需的事件创建事件识别模型。

异常可以通过一个显示所有异常和极端声音的界面进行考察和标记。

图5. 在 OtoSense 异常可视化界面中，声音分析随时间的变化。

持续学习过程—从异常检测到日益复杂的诊断

OtoSense 的设计初衷是向多位专家学习，并且随着时间推移，进行越来越复杂的诊断。常见过程是 OtoSense 和专家之间的循环：

- 异常模型和事件识别模型都是在终端运行。这些模型为潜在事件发生的概率以及它们的异常值创建输出。
- 超出定义阈值的异常声音或振动会触发异常通知。使用 OtoSense 的技术人员和工程师可以检查该声音和其前后声音信息。

- 然后，这些专家会对这个异常事件进行标记。
- 对包含这些新信息的新识别模型和异常模型进行计算，并推送给终端设备。

结论

ADI 公司提供的 OtoSense 技术旨在使声音和振动专业知识在任何设备上都持续可用，且无需连接网络来执行异常检测和事件识别。在航空航天、汽车和工业监测应用中，该技术被越来越多地用于设备健康监测，这表示，在曾经需要专业知识，以及涉及嵌入式应用的场景中，尤其是对于复杂设备而言，该技术都表现出了不错的性能。

参考资料

Sebastien Chistian，“文字如何创造世界。”TEDxCambridge，2014 年。

Sebastien Chistian [sebastien.christian@analog.com] 热衷于了解人类如何运用感知来创建内在可共享的世界模型，以及如何使用该模型来描述我们生活的世界。

Sebastian 获得了量子物理学硕士学位，随后获得神经科学硕士学位和语义学第三学位。他的教育结合了研究、开发和现场实验。作为语言和语言病理学家，他与精神病和聋哑儿童在一起度过了 10 年时间，这加深了他对基于感觉的意义创造和分享的理解，并重点关注听觉。Sebastien 说，他与同样年轻的病人一起工作了多年，这种经历让他将所有分散的知识整合成到一起，形成一个统一、连贯的画面。

同一时期，Sebastien 成为法国卫生部的专家并提出听觉损失政策，此外，他还在巴黎索邦大学医学院任教。2011 年，他创建了首个独立的私人研发实验室，致力于将受 AI 启发的创新技术带给存在感觉和认知障碍的人。

2013 年，Sebastien 完成了自己的机器听觉项目的完整原型，并因此获得了在马萨诸塞州剑桥市举办的 NETVA 科技竞赛的冠军。根据来自麻省理工学院 (MIT) 的同事和商界的积极反馈，他在 2014 年初创建了 OtoSense，并开发出首个专注于理解声音的 AI。这个机器听觉平台能够很好地适应复杂的环境，进行复杂的设备监测。

在获得了 2015 年 GSMA 全球移动大会上的年度最佳应用奖等多个奖项之后，OtoSense 将侧重点放在工业和交通垂直领域的设备监测上，并且其未来潜在的应用范围将会越来越广。

目前，Sebastien 就职于 ADI 公司，负责 OtoSense 内部产品开发。

Sebastien Christian