

EarHealth: An Earphone-based Acoustic Otoscope for Detection of Multiple Ear Diseases in Daily Life

Yincheng Jin
University at Buffalo, SUNY
Computer Science and Engineering
Buffalo, NY, USA
yincheng@buffalo.edu

Yang Gao
Northwestern University
Computer Science
Chicago, IL, USA
yang.gao@northwestern.edu

Xiaotao Guo
The First Affiliated Hospital of USTC
Otolaryngology-Head and Neck
Surgery
Hefei, Anhui, CHINA
gxt2012@mail.ustc.edu.cn

Jun Wen
Harvard Medical School
Biomedical Informatics
Boston, MA, USA
jun_wen@hms.harvard.edu

Zhengxiong Li
University of Colorado Denver
Computer Science and Engineering
Denver, CO, USA
zhengxiong.li@ucdenver.edu

Zhanpeng Jin*
University at Buffalo, SUNY
Computer Science and Engineering
Buffalo, NY, USA
zjin@buffalo.edu

1. 来源: mobisys 22年
2. 链接: [URL](#)

Abstract

论文提出了一种基于耳机的耳部状况检测系统: EarHealth。它可以检测三种主要的听力健康状况: 鼓膜破裂、耳垢堆积和堵塞以及中耳炎。

通过分析由chirp声刺激引起的记录回声, EarHealth 识别出耳道结构和鼓膜活动性的可区分特征, EarHealth 在 92 名人类受试者中达到 82.6% 的准确率。

EarHealth是第一个基于耳机的系统, 能够通过利用耳道几何形状和鼓膜移动性来监测听力健康状况。

EarHealth 是同类产品中第一个找到不同症状和声学特征(包括 FFT 和通道响应)之间的映射, 然后利用多视图深度学习模型来监测和检测三种流行的耳部疾病(除了正常状态), 利用耳道几何结构和鼓膜活动性。

Preliminaries

Ear Disease Detection Based on Acoustic Sensing

作者将耳道结构视为一个三维空间, 并假设 p 是声标量函数 $p = p(x, y, z; t)$, 将耳道内的声波传播建模为公式 1:

$$\frac{\partial^2 p}{\partial t^2} = c^2 \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2}$$

在正常环境中 c 等于 343 m/s (声音在空气中的传播速度)。另一方面, 一些能量会在传播过程中被吸收, 而另一些能量会被散射或反射, 这可以在方程式 2 中描述:

$$\frac{P_r}{P_i} = |r_p| e^{j\phi\pi}$$

$|r_p|$ 是相对于表面柔韧性的声音反射系数, ϕ 表示 P_i 和 P_r 之间的相位差。

这里的公式，作者没有更进一步的说明，我也没怎么理解

作者将根据鼓膜流动性和耳道几何形状研究不同的模式

1. Echoes Resulting From Varying Eardrum Mobility

较硬的材料会阻止声波穿透表面并导致更强的反射。因此，我们希望首先根据鼓膜的机械刚度（即鼓膜传播/反射声波的移动性）来区分不同的耳朵状况。正常耳膜具有更高的移动性，可以在多个声音频率下产生良好的共振，从而在频域中产生宽频谱。相比之下，其他条件下的鼓膜活动度较低，从而产生较窄的声学倾角。声学下降发生在耳道的共振频率处，其中chirp的四分之一波长等于耳道的长度。因此，声波倾角的形状主要取决于鼓膜的活性。所有三个目标异常耳朵状况都对鼓膜活动性和振动能力产生重大影响。具体来说，中耳炎会导致炎症和鼓膜后积液；耳垢堆积和堵塞会在耳道通路中形成坚固的障碍；鼓膜破裂通常会导致负压、脓性积液和中耳腔。中耳的压力、积液和结构变化是影响鼓膜活动性和声音传播的主要机制。

2. Echoes Resulting From Varying Ear Canal Geometry

不同的耳道几何结构会在耳道内产生不同的回声模式。这种几何差异可能来自个体之间固有的独特性，也可能是由于多种因素导致耳道结构的变化（如正常状态、耳垢堆积和堵塞、中耳炎和鼓膜破裂）。在这些情况中，正常状态和中耳炎不会改变耳道空间和改变声音传播途径。耳垢堆积和堵塞显然是由于耳道内堆积过多的耳垢造成的，这会缩短和改变声波传播路径。鼓膜破裂是指鼓膜上有一个洞或撕裂，它将连接耳道和中耳腔，从而改变声波传输空间。需要区分三种不同的声音传播路径和模式。由于通道响应已被广泛用于感知耳道结构，在本研究中作者使用通道响应来识别不同的耳道几何结构。

Feasibility Study

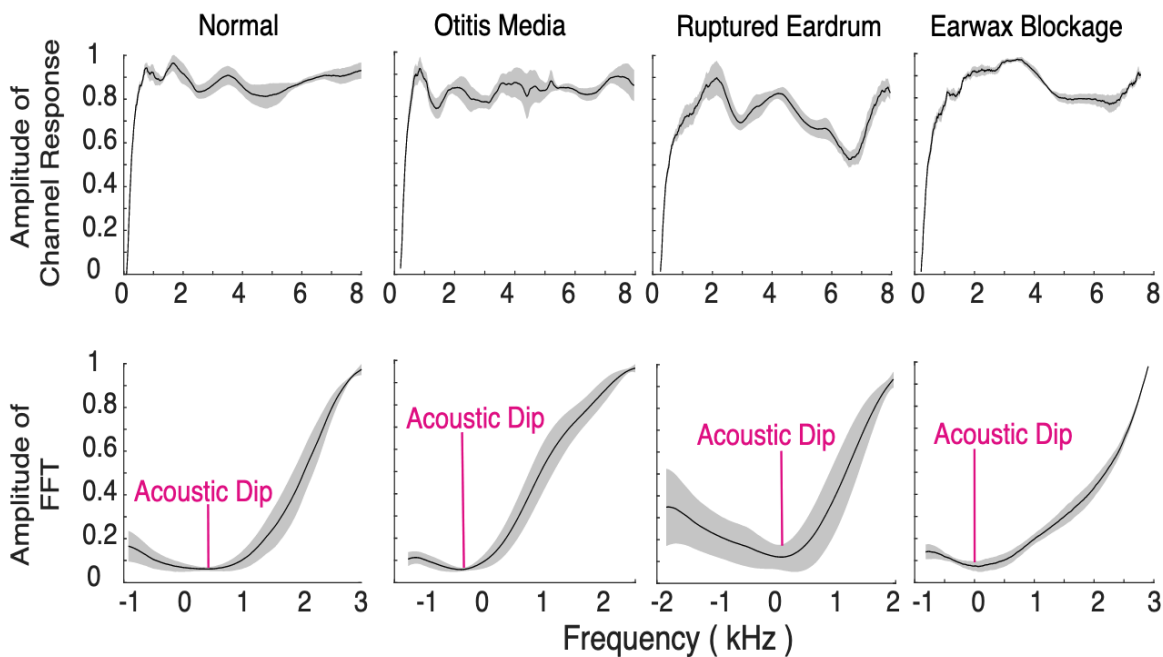


Figure 3: Average channel response and FFT amplitudes from five subjects for each condition as the black line, with standard errors as the gray shadow.

图3初步表明不同耳朵条件下耳内回声的各种可区分的特征模式，也就是说作者提出的想法具有可行性。

Implementation

EarHealth 旨在通过分析耳道内不断变化的回声模式来监测和检测耳朵状况，基于带有一对内置、面向内的扬声器和麦克风的耳机。扬声器将播放频率范围为 20 Hz 至 6000 Hz 的短啁啾，麦克风将记录相应的回声。然后提取一组精心挑选的特征来表示回声的物理和听力特征。这些特征将用于训练分类器，直到达到最小错误率。然后，探索一种物理数据增强方法来训练鲁棒模型并进行耳朵状况预测。系统图和方法流程如图 4 所示：

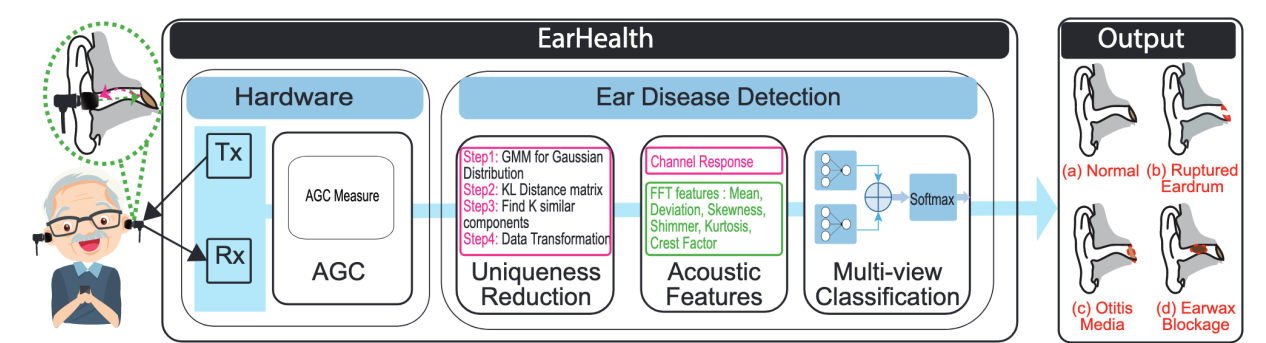


Figure 4: The methodological flow of EarHealth ear condition monitoring system. EarHealth first emits a sound and receives reflected echo signals, then extracts acoustic features, and finally utilize a well-trained classifier to determine the specific ear condition for one of the following : (a) normal (i.e., healthy), (b) ruptured eardrum, (c) otitis media, or (d) earwax buildup and blockage.

Automatic Gain Control (AGC)

每个音频设备都会产生不同的音量级别，这意味着回声会被设备本身放大或衰减。而且，大多数设备都是频率选择性的，这意味着在相同的音频刺激和音量设置的情况下，不同的设备可能具有不同的频率响应。在这项研究中，为了消除硬件造成的前端增益干扰，作者使用啁啾信号测量扬声器的频率选择性。根据测量的频率响应和音频的音量，作者通过补偿数字音频文件来估计扬声器的输出声音。

Ear Canal Uniqueness Reduction

为了克服由各种耳道结构引起的不同反射的巨大差异并促进可靠的耳部疾病检测，作者使用了一种数据转换技术，该技术有助于显着提高不同用户的准确性。其基本思想是为每种耳部疾病状况生成几个具有代表性的目标向量，然后将收集到的用户信号转换为具有目标向量特征的新信号。数据转换过程包括如下四个基本步骤：

1. 采用高斯混合模型（GMM）将输入信号表示为K个多元高斯函数之和；
2. 引入了基于 Kullback-Leibler (KL) 的距离矩阵，以在存储的模板中找到最相似的组件；
3. 然后，从高斯分布集中搜索距离矩阵找到与从收集到的数据中获取的K组件相比较最相似的K组件。在我们的例子中，距离最小的组件被认为是最相似的组件；
4. 最后，采用一个合适的数据转换函数将收集的数据 x 转换成目标向量 y'

参考语音变换的变换思想，作者采用的变换函数 $F(x)$ ，采用最小均方误差(MMSE)估计。采用期望最大化(EM)算法对GMM进行权值、均值和协方差矩阵的拟合。

Feature Extraction

声波包含大量代表耳朵声学特性的重要信息，包括耳道的长度、宽度和曲率，以及鼓膜的质量和硬度。作者采用Boruta 算法来确定 EarHealth 分类的所有相关特征。它依赖于计算高效的过程来迭代地丢弃不太相关的特征。在基于耳道结构和鼓膜活动性在不同条件下应用特征选择后，选择的主要特征如图5所示：

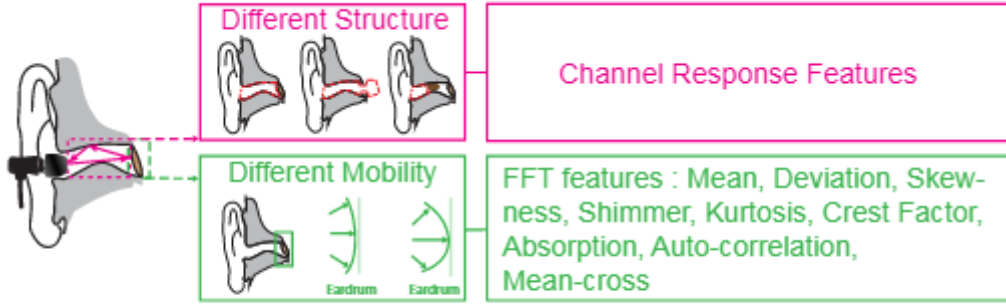


Figure 5: A taxonomy of feature descriptors corresponding to intrinsic ear conditions.

1. 通道响应特征：

为了解释耳道变形是如何发生的，作者利用通道响应来估计耳道的形状。具体来说，耳道的通道响应是反射信号与入射探测信号的比率：

$$c(s)|_{s=jw} = \frac{r(t_s)}{i(t_s)}$$

其中 $c(s)$ 表示通道响应， w 表示频率， $r(t_s)$ 表示麦克风在短时间内接收到的信号， $i(t_s)$ 表示同一时期扬声器发出的信号

2. FFT Feature

如图 5 所示，作者提取了几个特征来表示在接近声学倾角的 1 kHz 区域内的鼓膜移动性，这是局部最小值。Skewness 用于评估中心点左右对称性，kurtosis 描述声音形状的概率分布，flatness 用于表示在特定维度上近似欧几里得空间的程度，jitter 和 shimmer 用于描述声音形状的概率分布。鼓膜的振动会产生声音颤抖，波峰因数是指峰值与有效值的比值，自相关是观测值之间的相似性作为它们之间的时间滞后的函数，平均交叉是指交叉平均数的信号。

Multi-View Classifier

- 采用为学习传感器内特征而设计的集成分类器 (EC)。
- 通道响应特征通过三个卷积层，分别由 128 个滤波器、256 个滤波器和 128 个滤波器组成，每一层都包含 ReLU 激活函数。
- 另一方面，提取的 FFT 特征通过两个卷积层，ReLU 激活函数分别由 64 个滤波器和 128 个滤波器组成。
- 最后，我们结合这两个输出向量并将它们链接到一个 softmax 层以对四种不同的耳朵状况进行分类。

Overall Performance

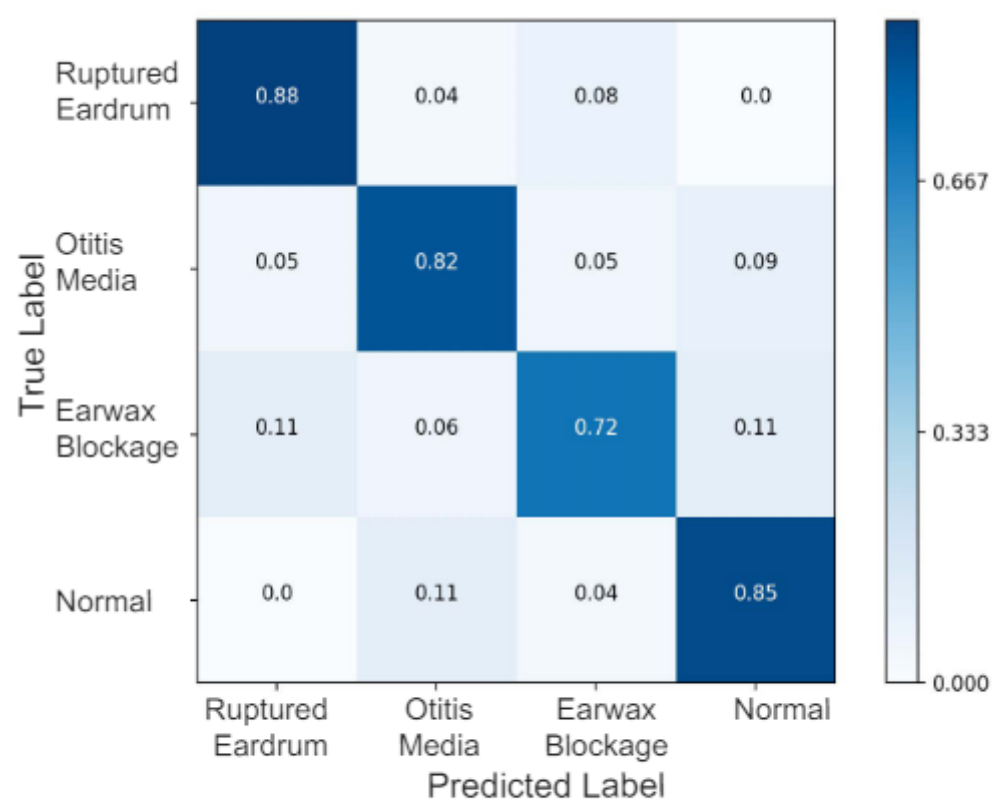


Figure 7: The normalized confusion matrix for detection of various ear health conditions.

EarHealth 对四类分类可以达到 82.6% 的准确率。识别异常耳朵状况（如鼓膜破裂、中耳炎、耳垢堵塞）的平均准确率为80.67%，识别正常耳朵状况（如正常）的准确率分别为85%。结果表明，EarHealth 系统将正常耳朵状况错误标记为异常耳朵状况（例如，鼓膜破裂、中耳炎和耳垢堵塞）的可能性较低。相比之下，EarHealth 将异常耳朵状况误检测为正常耳朵状况的概率略高。可能的原因是症状可能因人而异。中耳炎的鼓膜活动度不同，有些情况不能识别为中耳炎。耳垢堵塞情况的耳垢过多程度不同，可能无法完全堵塞耳道，因此可能导致错误分类。

Performance on Different Ears

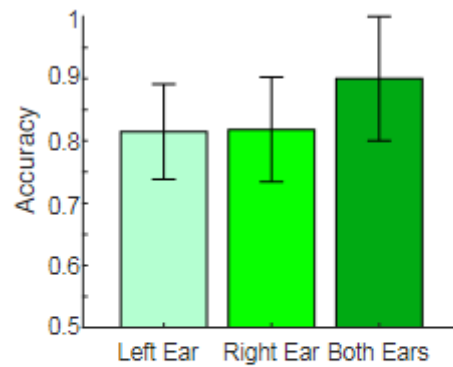


Figure 9: Recognition Accuracy of different ear sub-groups.

左耳和右耳的识别准确率几乎相同。但是双耳的识别准确率要高于单耳的识别精度。这可能是因为两只耳朵会产生更显著的反射信号模式，EarHealth System可以轻松识别这些模式。因此，EarHealth 系统可用于监测任一耳朵发生的耳部疾病。

Influence of Demographic Factors

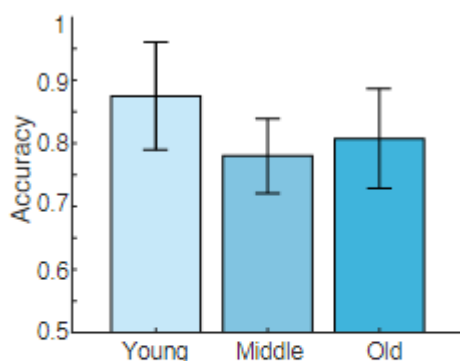


Figure 10: Recognition Accuracy of different ages sub-groups.

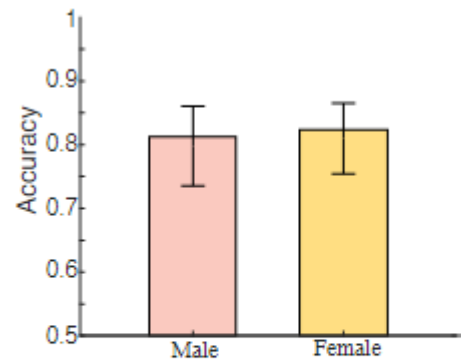


Figure 11: Recognition Accuracy of different gender sub-groups.

1. Impact of Age

图 10 的结果表明，EarHealth 系统的耳部状况检测在年轻时更准确，中年组和老年组的准确率没有显著差异。可能的原因是随着年龄的增长，异常的耳朵状况会增加。正如前面混淆矩阵的结果所示，EarHealth 系统更容易检测到正常的耳朵状况。

2. Impact of Gender

如图 11 所示，男性的平均准确率为 81.8%，标准值比男性高 6.3%，女性的平均准确率为 83.0%，标准值为 5.5%。论文的研究结果表明，在大多数情况下，男性和女性监测不同耳道状况的可能性相同。

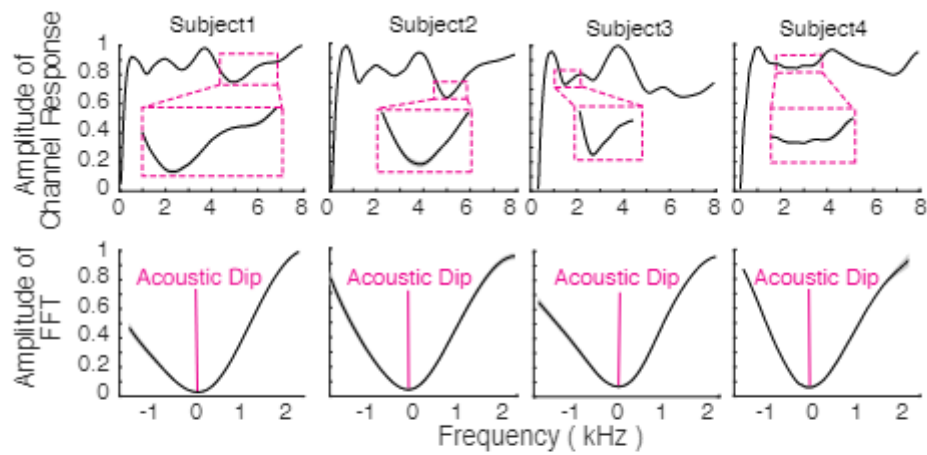


Figure 12: Average channel response and FFT amplitudes from five days (black line) with standard errors (gray shadow).

受试者被要求在连续五天内每天记录 20 次他们的数据。作者将这五天的平均通道响应幅度和 FFT 幅度绘制为黑线，标准误差显示为图 12 中的灰色阴影。可以观察到，这两个特征的灰色阴影区域非常微不足道，即使我们放大了灰色阴影区域，这意味着从不同日期获得的这些特征具有非常高的相似度，变化有限。结果表明，EarHealth系统可以在不同时间实现稳定可靠的数据采集，适用于日常生活使用。

2.System Overhead on Mobile Platform

在智能手机上连续执行我们的分析模块 100 次。表 3 总结了三种不同智能手机平台上的系统开销。

Table 3: Performance evaluation on mobile platforms

Smartphone platforms	Execution Time (ms)	Memory Usage (MB)	Average CPU Utilization (%)
Google Pixel 3A	343	29	43%
Google Pixel 3XL	220	36	30%
Google Pixel 4A	157	24	36%

Limitations

1. 精细化的耳朵结构建模

实验中没有考虑其他因素对耳内个体特异性回声的影响，例如耳毛、急性鼓膜炎症史和鼓室成形术。

2. 个人的耳朵结构

EarHealth 依赖于耳机原型接口隔离的密封入耳空间，但是文章中设计的原型耳接口不能完全贴合所有用户的耳朵，这导致了轻微的佩戴不稳定和信号泄漏。

3. 大规模的评估

在这项试点研究中，作者受到难以收集或创建大量具有不同听觉结构和特性的数据样本的限制。作者将继续通过添加更多主题数据来构建他们的数据集，包括其他常见的耳部疾病。

4. 长时间耳朵状况监测

用户不会一直带着耳机。

Conclusion

作者提出了 EarHealth 框架作为一种低成本、用户友好、耳机式、无处不在的耳部状况监测系统，可以轻松用于人们的日常生活中，无需任何额外的硬件，也无需额外的运营成本。给定一种特殊设计的声音刺激（具有广泛人类听觉频率的啁啾声），通过分析耳道内记录的回声并识别与不同耳朵状况相对应的可区分特征，可以有效地检测三种普遍的异常耳朵状况。

个人感觉这篇文章在一些细节上解释的不是很清楚，理解起来有难度