



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

ABSTRACT

With the aging of the population and the long-time wearing of earphones, hearing health has gradually emerged as a worldwide health issue. Early detection of hearing health conditions would greatly reduce potential risks with timely medical intervention. This study proposes an earphone-based ear condition monitoring system, named EarHealth, which is low-cost, non-invasive, and easily usable in daily life. It can detect three major hearing health conditions: ruptured eardrum, earwax buildup and blockage, and otitis media. By analyzing the recorded echoes evoked by a chirp sound stimulus, EarHealth recognizes the distinguishable characteristics from ear canal structure and eardrum mobility. EarHealth achieves an accuracy of 82.6% in 92 human subjects, including 27 normal subjects, 22 patients with ruptured eardrum, 25 patients with otitis media, and 18 patients with earwax blockage. EarHealth is the first earphone-based system capable of monitoring hearing health conditions by utilizing the ear canal geometry and eardrum mobility. It is anticipated that EarHealth would provide pervasive and proactive protection for hearing health.

CCS CONCEPTS

- Human-centered computing → Ubiquitous and mobile computing.

KEYWORDS

Ear Disease, Earphone, Acoustic Sensing, Eardrum Mobility, Ear Canal

*This is the corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MobiSys'22, June 27–July 1, 2022, Portland, Oregon

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-9185-6/22/06...\$15.00

<https://doi.org/10.1145/3498361.3538935>

ACM Reference Format:

Yincheng Jin, Yang Gao, Xiaotao Guo, Jun Wen, Zhengxiong Li, and Zhanpeng Jin. 2022. EarHealth: An Earphone-based Acoustic Otoscope for Detection of Multiple Ear Diseases in Daily Life. In *Proceedings of The 20th ACM International Conference on Mobile Systems, Applications, and Services (MobiSys'22)*. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3498361.3538935>

1 INTRODUCTION

In the recent decade, 430 million people have experienced different levels of hearing loss and this percentage rapidly rises in the elderly population [23, 36, 40]. The high risk of hearing loss is associated with aging, the long-time wearing of headphones, insufficiency of health care, and chronic exposure to loud noises. According to the World Health Organization (WHO), it is estimated that 1.1 billion teenagers and young adults are at risk of hearing loss due to the unsafe and long-time use of personal audio devices and exposure to damaging levels of sound at noisy entertainment venues (such as nightclubs, bars, noisy factory, and sporting events) [39]. The WHO study indicated that among teenagers and young adults aged 12–35 years, approximately 50% were exposed to unsafe levels of sound from personal audio devices and around 40% were exposed to potentially damaging levels of sound in entertainment or working environments. Especially, along with the rapid increase in usage of earphones, the risk of hearing problems is continuously skyrocketing because of long-time wearing earphones.

In this work, we will explore the detection of ear diseases which happen in the outer ear and the connection of the outer and middle ear - eardrum [16], including earwax buildup and blockage, otitis media, and ruptured eardrum. It has been reported in the literature that earwax buildup and blockage and otitis media with effusion (i.e., before developing to acute otitis media) sometimes don't show clear signs for many patients. However, if individuals are not aware of abnormal ear conditions and thus don't receive timely, proper medical treatments, those ear diseases would lead to more severe problems, such as permanent hearing loss, mastoiditis, facial nerve palsy, or meningitis [14, 38]. Given the significance and increasing prevalence of those hearing health problems that may result in

hearing loss and other severe medical implications, it is imperative to seek solutions that help users monitor and detect their ear conditions in a timely manner.

Although some accurate ear examination techniques, such as pneumatic otoscopy or tympanometry [14, 43], have been extensively investigated and used in clinical settings, those devices are usually expensive and sophisticated for patients, which restricts their applicability in daily-life scenarios. Moreover, it is reported that some patients have experienced severe uncomfortable pain during and after the procedure and more severe ear damages have been observed [1].

To address the limitations of those specialized clinical methods, many research efforts have explored alternative approaches to detect various ear health conditions. For example, Teele et al. [48] proposed an acoustic reflectometry method to examine the middle ear status, and many similar research works have been done based on the acoustic reflectometry using the EarCheck Pro device [5, 6, 8, 26] or the smartphone with a specific funnel (i.e., a hardpaper-made, cone-shape accessory to be used as a speculum to direct sound into the ear canal) [14]. Those prior work have successfully demonstrated the potential of new wearable or portable devices that can be easily used in daily health monitoring and detecting abnormal ear conditions. However, these approaches were only shown to detect otitis media or otitis externa by assessing the mobility of the tympanic membrane (also called “eardrum”). As for other common ear conditions such as ruptured eardrum and earwax buildup and blockage, the eardrum mobility is quite insignificant, making it hard to distinguish these disease conditions from the normal ear condition and restricting the functionalities of those existing approaches. In addition, for most ear diseases, it is more effective to provide a convenient way for people to proactively monitor their ear conditions and detect the abnormalities to get immediate medical intervals. Although those portable, handheld devices (e.g., EarCheck Pro) have already offered viable solutions of detecting ear health conditions in daily-life settings, they can't be used as pervasive, continuous ear condition monitoring devices that provide a long-term assessment of an individual's ear health.

Besides, different ear conditions result in different symptoms, which exhibit diverse characteristics and require different recognition schemes, even with the same sensing modality. Traditional acoustic-based in-ear sensing utilized the unique ear canal structure [20], in-ear body-conducted voice sound [19], ear canal deformations [52], or eardrum mobility [14] for a range of applications spanning from authentication to ear health monitoring, but they are not applicable for detecting three different ear disease conditions. Given the limitations of existing works, it is thus imperative to seek an alternative approach to address the following challenges: (1) continuous monitoring of ear conditions using ordinary wearable devices could remove the burden of extra operational efforts in daily life; (2) custom-designed privacy-sensitive voice content could enhance privacy and trustworthiness of the monitoring procedure; and (3) one single versatile device could be applied for detecting various common ear diseases.

Therefore, to detect the aforementioned three prevalent abnormal ear conditions and differentiate them from the normal state of the ear, we propose a novel pervasive, affordable, user-friendly, earphone-based ear condition monitoring system, named EarHealth.

It takes advantage of a multi-view deep learning model to detect three different ear conditions based on the uniqueness of the ear canal structure (i.e., ruptured eardrum, earwax buildup and blockage, otitis media) and the degree of the eardrum mobility (i.e., otitis media, normal ear condition). As shown in Fig. 1, EarHealth is an earphone-based acoustic sensing system, combined with the smartphone for data computation and recording, for daily hearing health monitoring. At the earphone end, EarHealth utilizes the built-in speaker and microphone to collect the raw acoustic data within the ear canal. At the smartphone end, EarHealth analyzes the acoustic features and recognizes four different ear conditions. Specifically, in this study, we make the following threefold contributions:

- (1) We propose and implement a novel earphone-based system for monitoring three different ear diseases in daily life. The earphone-style design provides an easy-to-use, user-friendly and cost-effective solution to the general population to allow people to detect potential hearing health problems in daily life.
- (2) To the best of our knowledge, EarHealth is the first of its kind to find the mapping between different symptoms and acoustic features including FFT and Channel Response, and then utilize a multi-view deep learning model for monitoring and detection of three prevalent ear diseases (besides the normal state), leveraging the ear canal geometrical structures and eardrum mobility.
- (3) We evaluate the effectiveness and efficiency of the proposed EarHealth, based on a dataset collected from clinical settings (92 patients). The results reveal that, the earphone-based EarHealth can accurately capture the differences presented in different ear disease conditions. Our discovery will pave the way for keeping hearing healthy.

2 RELATED WORK

2.1 Ubiquitous Acoustic Sensing

Internet of Things (IoTs) have become increasingly popular and necessary components of our life nowadays and have stimulated many creative technologies in diverse fields. For example, a wide variety of novel sensing mechanisms have been explored and employed in almost every aspect of people's lives [12]. Among those emerging sensing techniques, acoustic sensing, given its ubiquitous nature and sensitive responses to environmental, structural, or disturbance influences, has gained extensive attention in recent years [28]. The applications of acoustic sensing span from structural health monitoring, gesture/behavior recognition, security, to medical diagnosis. For instance, Mirtchouk *et al.* [34] and Bedri *et al.* [7] utilized a microphone to record the sound and then detect the food types and eating episodes by analyzing the throat sound patterns, which help people monitor and take care of their own health conditions. Jin *et al.* [24] proposed a system to recognize different gesture patterns when performing American Sign Language (ASL). Shumailov *et al.* [45] presented a security threat, which was capable of capturing and recovering the typing behaviors on the smartphone through acoustic sensing. Another emerging and interesting application of acoustic sensing is the non-line-of-sight (NLOS) imaging which aims to recover 3D shape and reflectance information of objects hidden from sight [31]. Using inaudible sound, it's convenient for

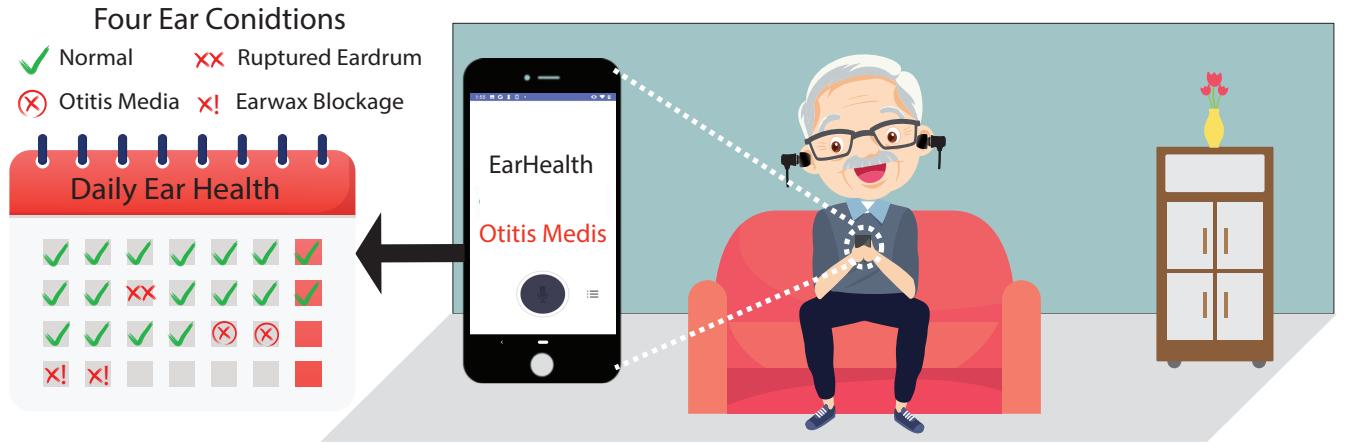


Figure 1: The EarHealth system for daily-life monitoring of hearing health.

people to enable monitoring and estimating the risk of Parkinson’s disease in daily life [54]. Based on the acoustic reflectometry, it has provided an at-home method to detect the middle ear fluid [6, 14, 17, 48–50]. Given the success of prior studies on ubiquitous acoustic sensing, it is anticipated that the increasingly prevalent earphones could provide a much broader functionality, in addition to being only a conventional music listening tool.

2.2 In-Ear Sensing

Given the widely usage of earphones, earables have gained tremendous attention for human sensing, such as blood flow [29], blood pressure [9], hear rate [11], dietary monitoring based on swallow [4], navigation [2], facial gestures detection [3], and authentication [18, 20, 52]. LeBoeuf *et al.* proposed a novel non-invasive in-ear sensing system to measure energy expenditure (TEE) and maximum oxygen consumption ($VO_{2\text{max}}$) from ear canal [29]. Bui *et al.* proposed a system to detect the blood pressure from the deep in-ear canal [9]. Butkow *et al.* explored the in-ear heart rate based on inner microphone [11]. Amft *et al.* utilized a microphone to record the chewing sound and then recognize different food categories [4]. Ahuja *et al.* investigated and proposed an inertial navigation system, and evaluated the performance based on one earable and two earables [2]. Amesaka *et al.* proposed a new system that utilized a speaker and microphone to sense the ear canal deformation and recognize different facial gestures [3]. Gao *et al.* and Wang *et al.* proposed two authentication methods based on the unique ear canal structure [20, 52]. Aligned with existing research in investigating next-generation earable sensing and computing platform, we propose a user-friendly, daily-use device based on commodity earphones to continuously and conveniently monitor individuals’ ear health conditions.

2.3 Hearing Condition Monitoring

Existing hearing condition monitoring and disease detection methods can be generally divided into two major categories: clinical detection methods and at-home detection methods. In clinical settings, physicians usually rely on otoscopy or tympanometry to

detect potential ear disease conditions, and then could analyze the results in remote locations [32, 35, 37]. However, those clinical devices are usually expensive and sophisticated to use, which makes those devices prohibitive for personal users at home. Moreover, it has been reported that some patients may have experienced uncomfortable pain during and after the procedure [1]. On the other side, convenient detection of hearing health conditions can provide significant benefits for patients to detect any adverse condition at an earlier stage for timely intervention and treatment. For instance, Teele *et al.* first presented how to utilize the acoustic reflectometer to detect the middle ear fluid [48]. Many similar studies later also used the acoustic reflectometry to detect the middle ear fluid [8, 17, 26, 49]. All of these works were based on a constant threshold, making it hard to get a good performance when being applied to different middle ear structures and can’t address any other middle ear problems. More recent work from Chan *et al.* [14] proposed to use some specific models of smartphones, where the speaker and microphone are located on the same side of the smartphone, to detect the middle ear fluid. However, this work can only recognize the middle ear fluid disease which is rarely happened in adult ears, some other common ear diseases (i.e., ruptured eardrum, earwax blockage) are more challenging to be recognized. Also, the changing shape of the funnel required in this system may dramatically influence the recognition accuracy of the same input samples. To this end, in this study we aim to propose an affordable, accessible, and pervasive device capable of detecting and recognizing three common ear diseases critical to hearing health.

3 PRELIMINARIES

3.1 Background of Ear Diseases

In this section, we provide some background information about ear diseases and their clinical symptoms related to ear structural characteristics. The human ear consists of three basic parts — the outer, middle, and inner ear. The ear canal is an S-shaped tube which is typically 2.5 cm in length and 0.7 cm in diameter and extends from the lateral porus acusticus externus to the eardrum. The eardrum is the junction of the outer ear and the middle ear.

When a sound comes in and your ear perceives it, the outer ear will receive the air vibration and relay it to the eardrum via the ear canal. Then the eardrum and three tiny bones behind the eardrum will amplify the vibration and relay the sound information to the cochlea. As mentioned before, we will explore the ear diseases that happen in the ear canal, including the three most prevalent medical conditions – earwax buildup and blockage, otitis media, and ruptured (perforated) eardrum [16], which are shown in Fig. 2.

- **Earwax buildup and blockage.** Earwax provides important protection to the ear, but too much earwax buildup could cause severe problems. Earwax buildup is a common reason for temporary hearing loss or tinnitus, because hardened and impacted earwax can form a plug to block the ear canal [53]. More importantly, many studies have confirmed that the use of earphones or headphones can increase earwax buildup because they tend to trap sweat and moisture within the ear canal, while the ear itself tends to generate more earwax for self-cleaning and lubrication [27].
- **Otitis media.** It represents a group of clinical inflammatory diseases of the middle ear, including acute otitis media (AOM) and otitis media with effusion (OME) [41, 43]. In particular, AOM (also called ‘middle ear infection’) is one of the most frequent diagnoses for children seeking acute medical care and the most common cause of children hearing loss [15].
- **Ruptured (perforated) eardrum.** Ruptured eardrum means a hole or tear in the eardrum, a thin membrane separating the ear canal from the middle ear. Because the eardrum is a critical part of the ear that can vibrate in response to sound waves, the ruptured eardrum may cause temporary or even permanent hearing loss if the perforated eardrum doesn’t heal and continuously allow bacteria to invade the ear [33].

3.2 Ear Disease Detection Based on Acoustic Sensing

In this study, we propose to monitor and detect ear diseases leveraging the in-ear acoustic sensing technique, by utilizing the built-in, inward-facing speaker and microphone on earphones. After the acoustic signals are emitted by the speaker, these signals propagate along the ear canal and are reflected/absorbed by the ear canal wall and eardrum [20, 52]. We treat the ear canal structure as a three-dimension space, and assume p be an acoustic scalar function $p = p(x, y, z; t)$. And we model the sound wave propagation within the ear canal as Equation 1.

$$\frac{\partial^2 p}{\partial t^2} = c^2 \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} \quad (1)$$

where c equals 343 m/s in a normal environment. On the other hand, some energy will be absorbed during the travelling and others will be scattered or reflected which can be described in Equation 2

$$\frac{P_r}{P_i} = |r_p| e^{j\phi\pi}. \quad (2)$$

where $|r_p|$ is the sound reflection coefficient with respect to the flexibility of the surface, ϕ represents the phase difference between P_i and P_r . We will investigate the different patterns based on the eardrum mobility and ear canal geometry.

3.2.1 Echoes Resulting From Varying Eardrum Mobility. When sound waves reach the boundary between one medium and another, a portion of sound waves generate reflections. For hard materials, most sound energy, which travels through a medium and should vibrate an object or substance, will be reflected back [14]. That is to say, harder materials keep the sound waves from penetrating through the surface and result in a stronger reflection. Thus, we would like to distinguish different ear conditions first based on the mechanical stiffness of the eardrum (i.e., the mobility of the eardrum to propagate/reflect sound waves). As described in the paper [14], a normal eardrum has a higher degree of mobility to resonate well at multiple sound frequencies to create a broad spectrum in the frequency domain. In contrast, the eardrum with other conditions has a lower degree of mobility, thus creating a narrower acoustic dip. The acoustic dip occurs at the resonant frequency of the ear canal where the quarter-wavelength of the chirp is equal to the length of the canal [44]. Thus, the shape of the acoustic dip primarily depends on eardrum mobility. All three target abnormal ear conditions have significant impacts on the eardrum mobility and vibrational capacity [21]. Specifically, otitis media causes inflammation and a build-up of fluid behind the eardrum; earwax buildup and blockage forms solid obstacles in the pathway of the ear canal; and ruptured eardrum commonly leads to negative pressure, purulent effusion, and middle ear cavity. It has been reported that [22], pressure, effusion, and structural changes of the middle ear are the main mechanisms affecting eardrum mobility and sound transmission.

3.2.2 Echoes Resulting From Varying Ear Canal Geometry. It has been well proven from prior research that, different ear canal geometrical structures generate different echo patterns within the ear canal [13, 47]. Such geometrical differences may either come from the inherent uniqueness among individuals, or result from the changes of the ear canal structure by many factors. As mentioned before, we focus on four different ear conditions (i.e., normal status, earwax buildup and blockage, otitis media, and ruptured eardrum). Among those conditions, the normal state and otitis media will not change the ear canal space and alter the sound transmission pathways. Earwax buildup and blockage is obviously caused by excessive earwax accumulated in the ear canal, which will shorten and alter the sound waves propagation pathways. Ruptured eardrum means a hole or a tear in the eardrum, which will connect the ear canal and middle ear cavity and thus also alter the sound wave transmission space. As shown in the upper middle part of Fig. 5, there are three different sound propagation pathways and patterns to be distinguished. As the channel response has been widely used to sense the ear canal structure [20, 52], we would like to explore using channel response to recognize different ear canal geometrical structures in this study.

3.3 Feasibility Study

To examine the unique patterns of four different ear conditions based on the ear canal structure and eardrum mobility, we conducted a preliminary feasibility study to validate this hypothesis. EarHealth system incorporates both the inward-facing speaker and microphone. The speaker plays a short chirp of the frequency ranging from 20 Hz to 6000 Hz with a sample rate of 44.1 kHz, while the

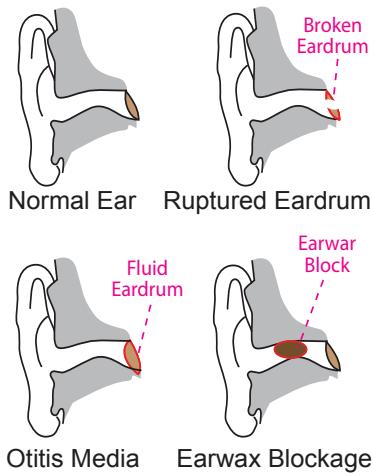


Figure 2: Illustration of four different ear-condition symptoms.

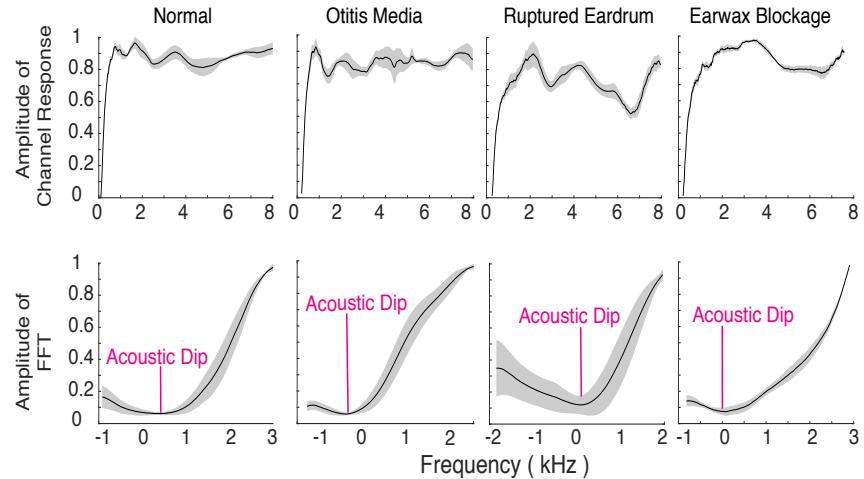


Figure 3: Average channel response and FFT amplitudes from five subjects for each condition as the black line, with standard errors as the gray shadow.

microphone records the raw, continuous echo data. After receiving the reflected sound waves, we calculated the FFT (Fast Fourier Transform) and channel response features that can provide a clue towards different ear conditions. In this preliminary experiment, we recruited five subjects for each ear disease condition and plotted the differences based on these two types of features. In the upper part of Fig. 3, the average channel response value of five subjects was plotted as the black line, and the standard error was shown as the gray shadow. It is observed that, the normal condition and otitis media are hard to distinguish based on the amplitude shape of the channel response features, while both ruptured eardrum and earwax blockage conditions have shown clearly distinguishable differences in the channel response amplitude.

In the lower part of Fig. 3, the average FFT absolute value of the five subjects (same subjects as the channel response analysis) was plotted as the black line, and the standard error was shown as the gray shadow. It is revealed from our preliminary study that, the FFT patterns of these four ear conditions in the acoustic dip area are similar to a valley shape. More specifically, the normal condition has a higher degree of eardrum mobility and thus, the sound energy would vibrate the eardrum very well, which results in a narrower and deeper acoustic dip. In contrast, the ruptured eardrum substantially (if not completely) degrades the eardrum mobility, which results in a broader acoustic dip accordingly. Given a large amount of accumulated earwax, the FFT pattern in the earwax blockage condition has only one peak which is easy to be distinguished from other ear conditions. In summary, the preliminary feasibility study described above have supported our research hypothesis and demonstrated the various, distinguishable feature patterns of the in-ear echoes under different ear conditions. In the following sections, we will discuss the research challenges when implementing EarHealth and provide the viable solutions that enable the deployment of EarHealth in real-world scenarios.

4 RESEARCH CONSIDERATIONS

4.1 Research Challenges

In our study, the (inward-facing) speaker emits chirp sounds and the (inward-facing) microphone records the sum of the reflected and incident chirps simultaneously. This sum reaches a nadir at a resonant frequency where a quarter of the chirp wavelength is equal to the distance from the microphone to the eardrum (i.e., the length of the ear canal). At this frequency, the reflected sound is maximally out of phase with the incident sound. Thus, although individual differences in ear canal lengths affect the location of the dip along with the frequency domain, the shape of the dip primarily depends on the eardrum mobility. However, the echoes inside the ear canal will be a mix of all of the propagation, absorption, reflection, and transmission of sound waves. Each ear canal is a unique, closed space consisting of many different reflection and absorption surfaces. On the other hand, several different conditions (i.e., ruptured eardrum, earwax buildup and blockage, and otitis media) will make the eardrum stiffer, which makes it more challenging to distinguish those conditions. Therefore, we propose the signal enhancement method and a multi-view classifier to distinguish the four different conditions.

4.2 Our Solution

To address these challenges, we adopt the following design and implement a multi-view deep learning model to detect multiple ear conditions based on the ear canal channel geometry and eardrum mobility. After obtaining the channel response and eardrum mobility features, each input is passed through the first part of our network respectively, and then aggregated at a fully connection layer. Moreover, this deep learning-based solution performs better than traditional approaches in the following three aspects: 1) the backpropagation allows to adjust weights for error correction; 2) multi-layer neurons can capture both the low-level and high-level features; 3) the activation functions (i.e., ReLU layer) provide the

ability to understand the non-linear relationships between the data. Finally, we utilize a data transformation mechanism to reduce the impacts of ear canal shape diversity and sensor position differences on the received signals.

5 SYSTEM IMPLEMENTATION

5.1 Overview

EarHealth aims to monitor and detect ear conditions by analyzing the changing echo patterns within the ear canal, based on an earphone with a pair of built-in, inward-facing speaker and microphone. The speaker will play a short chirp with the frequency ranging from 20 Hz to 6000 Hz, and the microphone will record the consequential echoes. Then we extract a set of carefully selected features to represent the physical and audiological characteristics of the echoes. The features will be used to train a classifier until the minimal error rate is achieved. Then, a physical data augmentation method is explored to train a robust model and make ear condition prediction. The system diagram and methodological flow is illustrated in Figure 4.

5.2 Automatic Gain Control (AGC)

Because every audio device will generate different volume levels, which means that the echo will be amplified or attenuated by the device itself. And also, most devices are frequency selective, which means that different devices might have different frequency responses given the same audio stimulus and volume setting [14, 30]. In this study, in order to eliminate the front-end gain interference caused by hardware, we measure the frequency selectivity of our speaker using a chirp signal. Based on the measured frequency response and the audio's volume, we estimate the output sound from the speaker by compensating the digital audio file.

5.3 Ear Canal Uniqueness Reduction

To overcome the vast diversity in different reflections caused by various ear canal structures and facilitate robust ear disease detection, we utilized a data transformation technique which is helpful to improve the accuracy across different users significantly [13]. The basic idea is to generate several representative target vectors for each type of ear diseases conditions, then transform the collected signal from users into a new signal with characteristics of the target vectors. The data transformation process consists of four basic steps.

- **Step 1:** A Gaussian Mixture Model (GMM) is adopted to represent the input signal as the sum of K multivariate Gaussian function.
- **Step 2:** A distance matrix based on Kullback-Leibler (KL) is introduced to find the most similar components in the stored templates.
- **Step 3:** Then, we search the distance matrix to find K components from the Gaussian distribution set that are most similar to the K components from the collected data. In our case, those with the minimum distance are considered as the most similar components.
- **Step 4:** Last, a suitable data transformation function to transform the collected data x into the target vector y' is adopted.

Table 1: Corresponding Features of FFT

Name	Calculation
mean	$a(f) = \frac{1}{n} \sum_{i=1}^n (f_i)$
deviation	$d(f) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (f_i - a(f))^2}$
skewness	$s(f) = \frac{1}{n} \sum_{i=1}^n \left(\frac{f_i - a(f)}{d(f)} \right)^3$
shimmer	$shim(x) = \frac{\frac{1}{n-1} \sum_{i=1}^n f_i^{i+1} - f_i^i }{\frac{1}{n} \sum_{i=1}^n n f_i^i}$
kurtosis	$k(f) = \frac{1}{n} \sum_{i=1}^n \left(\frac{f_i - a(f)}{d(f)} \right)^4$
crest factor	$c(f) = 20 \log \frac{\max_{i=1}^n f_i }{d(f)}$
absorption	$abp(f) = \frac{P_e}{P_p} = \frac{\sum_{i=1}^n f_i^2}{n}$
auto-corr	$AR_k(f) = \frac{\sum_{i=1}^{N-k} (f_i - \bar{f})(f_{i+k} - \bar{f})}{\sum_{i=1}^{N-k} (f_i - \bar{f})^2}$
mean-cross	$meancross(f) = \text{count}(f - \bar{f})$

Borrowed the transformation idea from speech transformation, we adopt the transformation function $F(x)$ assumed by the Minimum Mean Square Error (MMSE) estimation. The Expectation-Maximization (EM) algorithm is adopted to fit GMM with the weights, means, and covariance matrix.

5.4 Feature Extraction

The sound waves contain plenty of important information representing acoustic characteristics of the ears, including length, width, and curvature of the ear canal, as well as the mass and stiffness of the eardrum. The majority of the feature selection methods focus on finding a minimal-optimal subset based on the classification accuracy. However, a less desirable accuracy for a specific model is not sufficient to assert the irrelevance of the feature. Therefore, we employ the Boruta algorithm to determine all relevant features for the EarHealth classification. It relies on a computationally efficient process to iteratively discard the less relevant features. We utilize a two-step correction, i.e., the Benjamini Hochberg FDR for evaluating features against randomness and the Bonferroni correction for testing identical features repeatedly. After applying the feature selection on different conditions based on the ear canal structure and eardrum mobility, the chosen dominant features are shown in Fig.5. Specifically, in order to reduce the dimensions but with the same characteristics, we utilize the features listed in Table 1 to represent the shape of signals in different aspects [10].

5.4.1 Channel Response Feature. In order to explain how the ear canal deformation occurs, we leverage the channel response to estimate the shape of ear canal. Specifically, the channel response of the ear canal is the ratio of the reflected signal to the incident probe signals. It can be calculated through the Equation 3, where $c(s)$ represents the channel response, w represents the frequency, $r(t_s)$ represents the received signals by the microphone during a short period of time, and $i(t_s)$ represents the emitted signal by the speaker during the same period.

$$c(s)|_{s=jw} = \frac{r(t_s)}{i(t_s)} \quad (3)$$

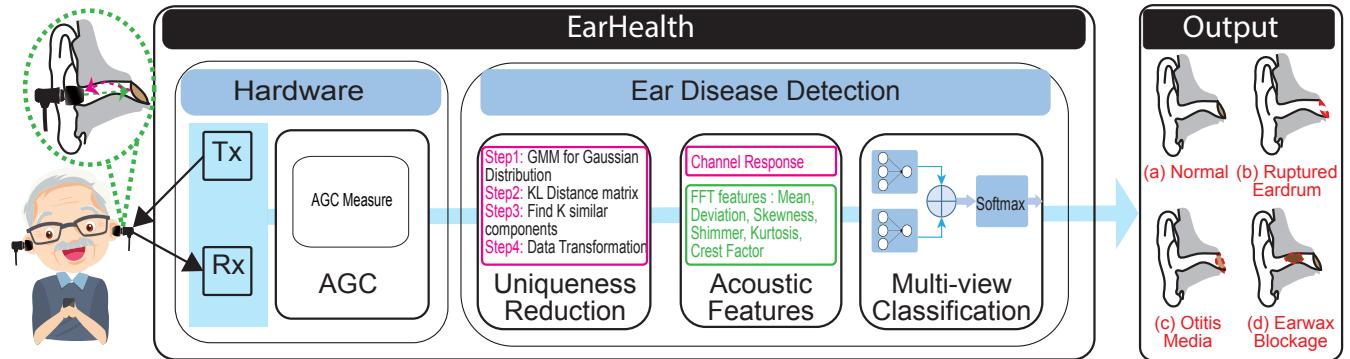


Figure 4: The methodological flow of EarHealth ear condition monitoring system. EarHealth first emits a sound and receives reflected echo signals, then extracts acoustic features, and finally utilize a well-trained classifier to determine the specific ear condition for one of the following : (a) normal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

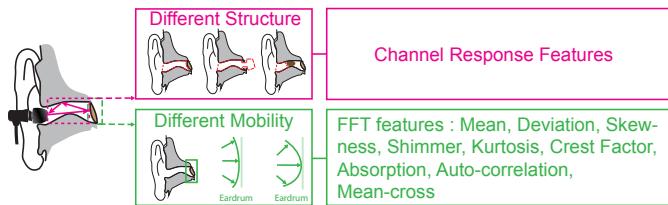


Figure 5: A taxonomy of feature descriptors corresponding to intrinsic ear conditions.

5.4.2 FFT Feature. In the paper [14], the author utilized the FFT shape near the acoustic dip point to distinguish the otitis media and normal ear which means different reflected capability. In this part, we introduce the features which are extracted from the FFT of reflected signals. As shown in Fig. 5, we extract several features to represent the eardrum mobility in the region of 1 kHz near acoustic dip which is the local minimum value. Several features will be introduced as follows.

Skewness is used to assess the symmetry between the left and right of the center point, *kurtosis* describes the probability distribution of the sound shape, *flatness* is used to indicate the degree of approximate Euclidean space in a specific dimension, *jitter* and *shimmer* are used to describe the vibration of the eardrum which would generate voice shaking, *crest factor* means the ratio of the peak values to the effective value, *auto correlation* is the similarity between observations as a function of the time lag between them, *mean cross* means the number of cross the mean of signals.

5.5 Multi-View Classifier

Given that we need to extract two dimensions of features which are not highly dependent on each other. Thus, it is hard to combine these two groups of features to feed into a single neural network. Traditionally, two popular solutions have been proposed to handle this problem [42]: ensemble classifiers (EC) which are designed for learning intra-sensor features, and feature concatenation (FC) which are designed for concatenating the separate sensor data

for learning corresponding features. Both of these two methods are widely used in lots of applications, such as acoustic sensing [51, 55] and Parkinson’s disease detection [54]. Since we want to extract the patterns of channel response and the shape of FFT in acoustic dips, we adopt the Ensemble Classifier (EC) to learn the intra characteristics.

On one hand, the channel response features are passed through three convolution layers consisting of 128 filters, 256 filters, and 128 filters, respectively, and every layer includes the ReLU activation function. On the other hand, the extracted FFT features are passed through two convolution layers with the ReLU activation function consisting of 64 filters and 128 filters, respectively. Lastly, we combine these two output vectors and link them to a softmax layer to classify four different ear conditions.

6 CLINICAL TRIALS ON SUBJECTS WITH EAR DISEASES

6.1 Participants

In order to validate the effectiveness and efficiency of our approach, we recruited 92 human subjects in the experiments (aged from 5 to 82 -years-old, 43 males and 49 females). All human subjects are patients who visited the Department of Otolaryngology at the First Affiliated Hospital of University of Science and Technology of China and voluntarily participated in this study. According to their ear conditions, the patients were labeled and classified as four different groups: normal ear condition (27 subjects), ruptured eardrum (22 patients), otitis media (25 patients), and earwax buildup and blockage (18 patients).

6.2 EarHealth Prototype

6.2.1 Hardware. We implemented EarHealth in a real earphone prototype to detect different ear conditions. Briefly, we embedded an MS-TFB microphone into an in-ear earphone (Bose SoundSport) which has a full, balanced sound and feature StayHear tips that could conform to ear shape and stay in place all the time (Fig 6). The microphone has an extremely high sensitivity (-32 dB), low noise for

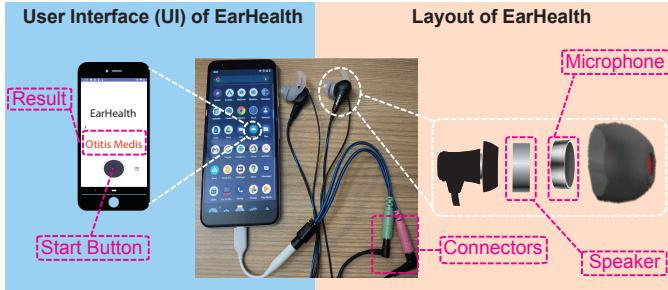


Figure 6: The User Interface (UI) of the EarHealth mobile app and the structural layout of the EarHealth device.

the quietest recordings (75 dB signal-to-noise ratio), and high maximum sound pressure level (115 dB with normal plug-in-power). The silicone earbud tips were used to mitigate the environmental noises and improve the users' comfort level. We used a Mobile Phone Audio Splitter Cable with one 3.5mm Male (TRRS) and two 3.5mm Female (TRS) connectors to connect the speaker and microphone to a smartphone (Google Pixel 3A).

6.2.2 Software and Auditory Stimuli. In order to monitor the hearing health conditions, we deployed an app on an Android smartphone to collect and analyze all data. A well-trained model (developed in Pytorch, version 1.3.0 and trained on the NVIDIA GeForce RTX 2080Ti GPU) was deployed. Sound stimuli signals and their evoked echoes were generated and recorded by our own developed EarHealth software. Considering the sound stimulus should cause valid eardrum mobility and vibration, thus we chose a relatively low-frequency sound which matches the hearing threshold of humans in this study. According to previous studies [14, 48, 50], our stimuli signal was designed as a train of chirp sound (1 s chirp repeats 20 times without interval, 20 seconds for the whole train) with the frequency ranging from 20 Hz to 6000 Hz, which falls into the most general and sensitive range of human hearing.

6.3 Data Collection

In this experiment, two types of data were collected for assessing ear diseases. All procedures were approved by the Internal Review Board (IRB) of The First Affiliated Hospital of University of Science and Technology of China.

EarHealth Assessment: EarHealth is an earphone-based real-time system that can provide timely detection of ear health conditions. We asked the participants to wear the EarHealth earphones, then played the gentle stimuli chirps and recorded chirp-evoked echoes using the developed EarHealth app. Specifically, we collected data using one side of the earphones, and the other side was left open in a large, empty space, to minimize the influence of ambient sound reflections.

Clinical Assessment: We obtained the ground truth of each patient which was evaluated and labeled through the clinical diagnosis of otolaryngologists. After the test of the audiometry, acoustic impedance, and otoscope, the physicians diagnose and determine one of the following ear health conditions: earwax buildup and blockage, eardrum ruptured (i.e., tympanic membrane perforation), and otitis media.

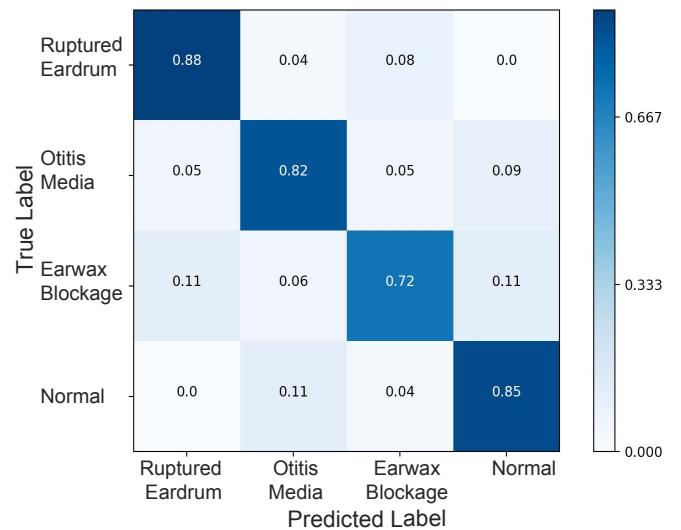


Figure 7: The normalized confusion matrix for detection of various ear health conditions.

7 PERFORMANCE EVALUATION

7.1 Overall Performance

We first evaluate our collected dataset on four different hearing health conditions. Traditionally, the leave-one-out cross-validation (LOOCV) strategy, which utilizes the data of (N-1) subjects as the training dataset and the data of the remaining one subject as the testing dataset, has been widely used for assessing the classification performance of machine learning-based models for medical problems or symptoms[25]. For example, researchers used the mobile acoustic sensing technique to detect ear fluid and achieved an accuracy of 89.8% (for a group of human subjects aged from 2 to 17 years) using the LOOCV evaluation [14].

Fig. 7 plots the normalized confusion matrix, and it is shown that EarHealth can achieve an accuracy of 82.6% for the four-categories classification. The average accuracy of recognizing an abnormal ear condition (e.g., ruptured eardrum, otitis media, and earwax blockage) is 80.67%, and the accuracy of recognizing normal ear condition (e.g., normal) is 85%, respectively. The result shows that the EarHealth system has a lower probability to mis-label the normal ear condition as an abnormal ear condition (e.g., ruptured eardrum, otitis media and earwax blockage). In contrast, EarHealth has a slightly higher probability to mis-detect an abnormal ear condition as the normal ear condition. The possible reason is that symptoms may be different from person to person. The mobility of the eardrum for otitis media is different, and some situations can not be recognized as otitis media. The excessive level of earwax for earwax blockage condition is different and may not fully block the ear canal, thus, it may lead to the misclassification.

7.2 Performance of Ear Canal Uniqueness Reduction

The proposed data transformation technique provides an efficient mechanism for EarHealth to reduce the impacts of ear canal shape

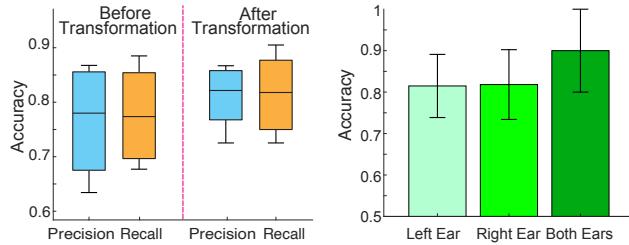


Figure 8: Results of data transform for ear canal uniqueness reduction.

Figure 9: Recognition Accuracy of different ear subgroups.

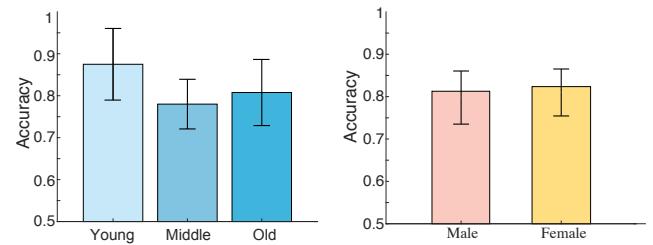


Figure 10: Recognition Accuracy of different ages subgroups.

Figure 11: Recognition Accuracy of different gender subgroups.

diversity. Fig. 8 compares the leave-one-person-out validation results without and after data transformation. When we do not perform data transformation, the average recognition recall is 77.04%, and the average precision is 77.73%. The recall and precision of participants with the worst results are both a little bit lower. After data transformation, we observe a significant increase in recognition results, which reach 81.95% average recall and 82.18% average precision. It is shown that data transformation has high efficiency and is the key to accurate ear disease detection.

7.3 Performance on Different Ears

In this experiment, we asked the users to wear the EarHealth system randomly in either ear. Given that the received signals represent a mixed combination of both left and right ear echoes, it is hypothesized that the system could work for either ear side. To validate our hypothesis, we would like to investigate the detection accuracy of the left ear, right ear, and both ears. In our work, only five subjects suffered from ear disease in both ears (i.e., three with otitis media and two with ruptured eardrum), 34 subjects suffered from ear disease in the right ear and 26 subjects in the left ear. As shown in Fig. 9, it can be seen that the recognition accuracy of the left ear and the right ear is nearly the same. But the recognition accuracy from both ears is higher than the one relying on one single ear. This is probably because two ears will generate a more significant pattern of reflected signals which can be easily recognized by our EarHealth System. Thus, our EarHealth system can be used to monitor ear diseases which happen in either ear.

7.4 Influence of Demographic Factors

In this section, we study the influence of demographic factors on ear disease detection, including the impact of age and gender, respectively.

Impact of Age: It is well acknowledged that otitis media often happens at a lower age. On the other hand, due to unhealthy habits, improper way of cleaning the ear canal (such as using cotton swabs can push wax deeper in the ear canal), and long-time wearing of earphones, people can easily get infected by earwax buildup and blockage. Therefore, we are motivated to explore how age factors can influence the performance of our proposed ear diseases detection. We divide all the people with ear diseases into three groups (< 25, 26 - 59, and > 60), which the number of these groups is 21,

45, and 26, respectively. Results in Fig. 10 show that the ear conditions detection of EarHealth system is more accurate in young age, and the accuracy of the middle age group and old age group is not significant different. The possible reason is that the abnormal ear conditions will increase along with the increasing age. As the discussion in 7.1, the normal ear condition can be easily detected by our EarHealth system.

Impact of Gender: Gender, as well as aging, may be another factor that affects ear disease detection which is widely explored in Digital Health. In this section, we aim to explore whether different genders can influence ear disease detection. As shown in Fig. 11, the average accuracy of the male is 81.8% with a standard value of 6.3% higher than the one of the male, and the average accuracy of the female is 83.0% with a standard value of 5.5%. Our results show that, in most cases, there is the same possibility for the male and female to monitor different ear canal conditions. It is easy to say that the capability of detecting different hearing health conditions using EarHealth is irrelevant with the gender.

7.5 Comparison With Existing Works

We compared EarHealth System with other related works of ear disease detection methods, including the Pneumatic otoscopy & tympanometry, EarCheck Pro [8, 14, 26], and smartphone-based detection [14]. As shown in Table 2, despite the high accuracy, the pneumatic otoscopy & tympanometry are professional devices, which are normally used in clinical settings and not suitable for daily use scenarios. Commercially available off-the-shelf EarCheck Pro is specifically designed and used for detecting the otitis media based on traditional threshold-based rules. Smartphone-based solutions can also monitor otitis media with the machine learning method, but it needs a large amount of labeled data to monitor only one specific ear condition. Compared with those existing solutions, EarHealth can monitor hearing health conditions and detect three prevalent ear diseases with a daily-used earphone. More importantly, EarHealth is specifically designed to address the issue of insufficient labeled data and provide great generalization ability to accommodate the vast diversity of human ear geometries.

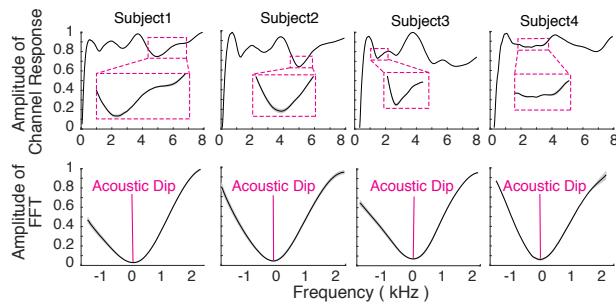
8 SYSTEM EVALUATION

8.1 System Stability and Reliability

As the EarHealth system is intended to be used in daily life, it is necessary to assess the stability and reliability of this system.

Table 2: Comparisons with related work

System	Abnormal Conditions	Cost	Portable	Ease of Use	Performance
Pneumatic otoscopy	N/A	High	No	No	High
EarCheck Pro [8, 26]	One	Medium	No	Yes	77.6%
Smartphone-based [14]	One	Low	Yes	No	89%
EarHealth (Ours)	Three	Low	Yes	Yes	82.6%

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

Hence, we conducted an experiment to show that our EarHealth system is reliable through analyzing the data stability from the same subject in different days. In this experiment, the data were collected from four subjects with normal ear conditions, in different days. Specifically, the subjects were instructed to record their data 20 times each day, in five consecutive days. It is anticipated that, the EarHealth system shall be able to obtain similar features over the observational period. We plotted the average channel response amplitude and FFT amplitude of these five days as the black line with a standard error shown as the gray shadow in Figure 12. It can be observed that the areas of the gray shadow for these two features are quite insignificant, even though we zoom in the gray shadow area, which means that these features obtained from different days possess a very high degree of similarity, with limited variations. It is indicated by the results that, our system can achieve stable and reliable data collection at different times and is applicable for daily-life usage.

8.2 System Overhead on Mobile Platform

Given the target application scenario where the EarHealth is expected to be connected with and executed on a smartphone, we deployed our neural network model onto different smartphone platforms and evaluated the computational efficiency. Specifically, we focused on three most significant factors that would affect the real-time usage of EarHealth, which consists of the execution time, memory usage, and CPU utilization. We set up the experiments to continuously execute our analysis module 100 times on the smartphone. Table 3 summarizes the system overhead on three different smartphone platforms.

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%

9 INSIGHTS ABOUT USER EXPERIENCE

It was reported that health monitoring systems with a high level of burden on their users can have a negative effect on adoption and user experience [46]. In this section, we intend to assess the usability and user burden of the proposed system in real-world settings by conducting a survey of two different groups of people about their view of technology. Specifically, we recruited two groups of participants: namely, 1) the *Common User* group that includes 20 volunteers who are self-identified without any ear disease, 2) the *Professional* group including 10 otologists and otolaryngologists. The participants were polled with three questions representing different usability metrics – Physical Pain, Difficulty of Use, and Effort [46]. Each question was rated on a 5-point scale with 1-point steps. A lower score indicates a superior usability with little pain and effort. For the Physical Pain, we asked a question that “EarHealth has made me feel physical pain.” For the Difficulty of Use, we asked a question that “EarHealth demands a lot of effort to use.” For the Time & Effort, we asked a question that “you spend too much time using EarHealth.”

Fig. 13(a) shows the user burden from physicians’ perspectives. This figure reports the average burden value in the physical pain is 1 with a standard value of 0.412, the average burden value in the difficulty of use is 0.7 with a standard value of 0.949, and the average burden value in the Time is 0.316 with a standard value of 0.900.

In contrast, Fig. 13(b) shows the user burden from common users’ perspective. This figure reports the average burden value in the physical pain is 1.375 with a standard value of 1.102, the average burden value in the difficulty of use is 1.050 with a standard value of 0.876, and the average burden value in the Time is 1.550 with a standard value of 0.959.

Regular users have a high level of awareness and concern about the physical pain, but physicians have less concern because this physical pain is much less than clinical ear disease detection. Meanwhile, regular users have a high level of difficulty of use and time, while physicians have less concern about this. I think the possible reason is that ordinary users would like to think it is better to play

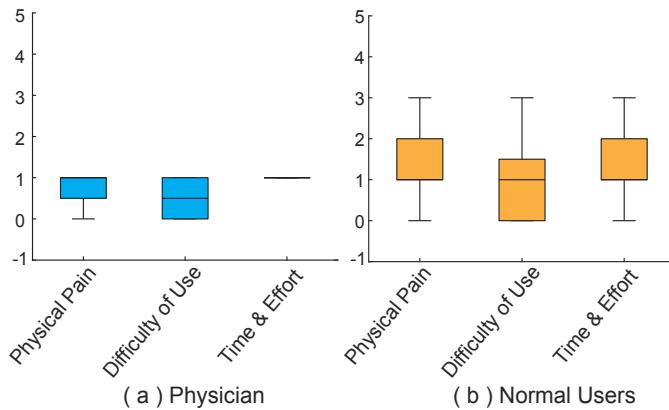


Figure 13: (a) The evaluation of EarHealth burdens from the physicians' perspectives. (b) The evaluation of EarHealth burdens from the ordinary users' perspectives. The Y-axis is from 0 to 5 to represent the level of burdens and efforts. A lower score suggests a better usability.

the one-second chirp to monitor ear conditions, but physicians think it is really easy to use compared with other clinical devices.

10 DISCUSSIONS AND FUTURE WORK

Despite many attractive features and merits, EarHealth is still a proof-of-concept prototype in its preliminary stage. It is anticipated to go through more rigorous testing and evaluations based on a sufficiently large subject population, perform more optimizations on the aspects of user experience, reliability, computational and power efficiency, and capability of handling other ear conditions and diseases. Specifically, EarHealth suffers from the following major limitations:

- (1) **Refined Ear Structural Modeling.** In the current study, we considered the most dominant structural parameters that affect the acoustic characteristics of the ear canal. Other factors may also influence individual-specific echo behaviors within the ear, such as the ear hair, history of acute eardrum inflammation, and tympanoplasty surgery.
- (2) **Personalized Ear Interface.** The EarHealth depends on the sealed in-ear space isolated by the interface of the earphone prototype. We noticed some patient feedbacks suggesting that the ear interface of our prototype did not perfectly fit their ears, which caused a slight wearing instability and signal leakage. We believe this issue can be addressed by providing patients with multiple interface options in our future study.
- (3) **Large-Scale Evaluation.** In this pilot study, we were constrained by the difficulty of collecting or creating a large number of data samples with diverse auditory structures and properties. We will keep building our dataset by adding more subject data, including other common otic diseases.
- (4) **Long-time Ear Condition Monitoring.** The scope of this paper is to detect different ear diseases in real time. However, as one of the most popular daily-used wearable devices, earphones provide us with the opportunity for long-term

monitoring of ear health conditions, especially for the gradual progression of hearing loss. Thus, we plan to further expand our work to pervasive assessment of hearing loss. The user could wear our device daily to track their ear conditions and hearing levels, for assessment and treatment, without visiting physicians frequently.

11 CONCLUSIONS

In this study, we proposed the EarHealth framework as a low-cost, user-friendly, earphone-type, pervasive ear condition monitoring system that can be easily used in people's daily life, without any extra hardware and no additional operational cost. Given a specially designed sound stimulus (a chirp with a wide range of human hearing frequencies), three prevalent abnormal ear conditions can be effectively detected by analyzing the recorded echoes within the ear canal and recognizing the distinguishable characteristics corresponding to different ear conditions. In summary, based on the distinguishable features of different ear canal structures and eardrum mobility, EarHealth can achieve a reasonable recognition accuracy which provides a very promising solution for future personalized healthcare.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant No. CNS-1840790 and CNS-2050910. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The authors thank Dr. Wei Sun (Department of Communicative Disorders and Sciences, University at Buffalo) for feedback on the idea and facilitating access to the pilot experimental study. The authors also acknowledge the help from Manas Bedmutha (Indian Institute of Technology Gandhinagar) on early-stage prototype development.

REFERENCES

- [1] Penelope Abbott, Sara Rosenkranz, Wendy Hu, Hasantha Gunasekera, and Jennifer Reath. 2014. The effect and acceptability of tympanometry and pneumatic otoscopy in general practitioner diagnosis and management of childhood ear disease. *BMC Family Practice* 15, 181 (2014), 1–10.
- [2] Ashwin Ahuja, Andrea Ferlini, and Cecilia Mascolo. 2021. PilotEar: Enabling In-ear Inertial Navigation. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*. 139–145.
- [3] Takashi Amesaka, Hiroki Watanabe, and Masanori Sugimoto. 2019. Facial expression recognition using ear canal transfer function. In *Proceedings of the 23rd International Symposium on Wearable Computers*. 1–9.
- [4] Oliver Amft, Matthias Stäger, Paul Lukowicz, and Gerhard Tröster. 2005. Analysis of chewing sounds for dietary monitoring. In *Proceedings of the 2005 International Conference on Ubiquitous Computing*. Springer, 56–72.
- [5] Michael J Babb, Raymond L Hilsinger Jr, Harold W Korol, and Robert D Wilcox. 2004. Modern acoustic reflectometry: accuracy in diagnosing otitis media with effusion. *Ear, Nose & Throat Journal* 83, 9 (2004), 622–624.
- [6] Elizabeth D Barnett, Jerome O Klein, Kimberly A Hawkins, Howard J Cabral, Margaret Kenna, and Gerald Healy. 1998. Comparison of spectral gradient acoustic reflectometry and other diagnostic techniques for detection of middle ear effusion in children with middle ear disease. *The Pediatric Infectious Disease Journal* 17, 6 (1998), 556–559.
- [7] Abdelkareem Bedri, Richard Li, Malcolm Haynes, Raj Prateek Kosaraju, Ishaan Grover, Temiloluwa Prioleau, Min Yan Beh, Mayank Goel, Thad Starner, and Gregory Abowd. 2017. EarBit: using wearable sensors to detect eating episodes in unconstrained environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–20.

- [8] Stan L Block, Ellen Mandel, Samuel McLinn, Michael E Pichichero, Shelly Bernstein, Sandra Kimball, and Joseph Kozikowski. 1998. Spectral gradient acoustic reflectometry for the detection of middle ear effusion by pediatricians and parents. *The Pediatric Infectious Disease Journal* 17, 6 (1998), 560–564.
- [9] Nam Bui, Nhat Pham, Jessica Jacqueline Barnitz, Zhanan Zou, Phuc Nguyen, Hoang Truong, Taeho Kim, Nicholas Farrow, Anh Nguyen, Jianliang Xiao, et al. 2019. eBP: A wearable system for frequent and comfortable blood pressure monitoring from user's ear. In *Proceedings of the 25th Annual International Conference on Mobile Computing and Networking (MobiCom)*. 1–17.
- [10] Jamie Bullock. 2007. LibXtract: a Lightweight Library for Audio Feature Extraction. In *Proceedings of the International Computer Music Conference (ICMC)*.
- [11] Kayla-Jade Butkow, Ting Dang, Andrea Ferlini, Dong Ma, and Cecilia Mascolo. 2021. Motion-resilient Heart Rate Monitoring with In-ear Microphones. *arXiv preprint arXiv:2108.09393* (2021).
- [12] Chao Cai, Rong Zheng, and Menglian Hu. 2019. A survey on acoustic sensing. *arXiv preprint arXiv:1901.03450* (2019).
- [13] Yetong Cao, Huijie Chen, Fan Li, and Yu Wang. 2021. CanalScan: Tongue-Jaw Movement Recognition via Ear Canal Deformation Sensing. In *Proceedings of the 2021 IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 1–10.
- [14] Justin Chan, Sharat Raju, Rajalakshmi Nandakumar, Randall Bly, and Shyamnath Gollakota. 2019. Detecting middle ear fluid using smartphones. *Science Translational Medicine* 11, 492 (2019), eaav1102.
- [15] D. K. Cherry, D. A. Woodwell, and E. A. Rechtsteiner. 2005. National Ambulatory Medical Care Survey: 2005 summary. *Advance Data* 387 (2005), 1–39.
- [16] Mayo Clinic. 2019. Hearing loss. <https://www.mayoclinic.org/diseases-conditions/hearing-loss/symptoms-causes/syc-20373072>.
- [17] Jerome T Combs, Hugh W Busey, and Kresimir Uzraincik. 1999. Device and process for generating and measuring the shape of an acoustic reflectance curve of an ear. US Patent 5,868,682.
- [18] Andrea Ferlini, Dong Ma, Robert Harle, and Cecilia Mascolo. 2021. EarGate: gait-based user identification with in-ear microphones. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*. 337–349.
- [19] Yang Gao, Yincheng Jin, Jagmohan Chauhan, Seokmin Choi, Jiyang Li, and Zhanpeng Jin. 2021. Voice In Ear: Spoofing-Resistant and Passphrase-Independent Body Sound Authentication. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1, Article 12 (2021), 25 pages.
- [20] Yang Gao, Wei Wang, Vir V. Phoha, Wei Sun, and Zhanpeng Jin. 2019. EarEcho: Using ear canal echo for wearable authentication. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–24.
- [21] Xiying Guan, Yongzheng Chen, and Rong Z. Gan. 2014. Factors Affecting Loss of Tympanic Membrane Mobility in Acute Otitis Media Model of Chinchilla. *Hearing Research* 309 (2014), 136–146.
- [22] Xiying Guan and Rong Z. Gan. 2013. Mechanisms of Tympanic Membrane and Incus Mobility Loss in Acute Otitis Media Model of Guinea Pig. *Journal of the Association for Research in Otolaryngology* 14 (2013), 295–307.
- [23] Preben Homoe, Kari Kværner, Janet R Casey, Roger AMJ Damoiseaux, Thijs MA van Dongen, Hasantha Gunasekera, Ramon G Jensen, Ellen Kvæstad, Peter S Morris, and Heather M Weinreich. 2017. Panel 1: epidemiology and diagnosis. *Otolaryngology—Head and Neck Surgery* 156, 4_suppl (2017), S1–S21.
- [24] Yincheng Jin, Yang Gao, Yanjun Zhu, Wei Wang, Jiyang Li, Seokmin Choi, Zhangyu Li, Jagmohan Chauhan, Anind K Dey, and Zhanpeng Jin. 2021. Soni-cASL: An Acoustic-based Sign Language Gesture Recognizer Using Earphones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–30.
- [25] Michael Kearns and Dana Ron. 1999. Algorithmic stability and sanity-check bounds for leave-one-out cross-validation. *Neural Computation* 11, 6 (1999), 1427–1453.
- [26] Sandra Kimball. 1998. Acoustic reflectometry: spectral gradient analysis for improved detection of middle ear effusion in children. *The Pediatric Infectious Disease Journal* 17, 6 (1998), 552–555.
- [27] Helene J Krouse, Anthony E Magit, Sarah O'connor, Seth R Schwarz, and Sandra A Walsh. 2017. Plain Language Summary: Earwax (Cerumen Impaction). *Otolaryngology—Head and Neck Surgery* 156, 1 (2017), 30–37.
- [28] Nicholas D Lane, Petko Georgiev, and Lorena Qendro. 2015. Deepear: robust smartphone audio sensing in unconstrained acoustic environments using deep learning. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 283–294.
- [29] Steven F LeBoeuf, Michael E Aumer, William E Kraus, Johanna L Johnson, and Brian Duscha. 2014. Earbud-based sensor for the assessment of energy expenditure, heart rate, and VO_{2max}. *Medicine and Science in Sports and Exercise* 46, 5 (2014), 1046.
- [30] Hyewon Lee, Tae Hyun Kim, Jun Won Choi, and Sunghyun Choi. 2015. Chirp signal-based aerial acoustic communication for smart devices. In *Proceedings of the 2015 IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 2407–2415.
- [31] David B Lindell, Gordon Wetzstein, and Vladlen Koltun. 2019. Acoustic non-line-of-sight imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 6780–6789.
- [32] Tal Marom, Oded Kraus, Nadeem Habashi, and Sharon Ovnat Tamir. 2019. Emerging Technologies for the Diagnosis of Otitis Media. *Otolaryngology—Head and Neck Surgery* 1 (2019), 10.
- [33] Ritvik P. Mehta, John J. Rosowski, Susan E. Voss, Ellen O'Neil, and Saumil N. Merchant. 2006. Determinants of Hearing Loss in Perforations of the Tympanic Membrane. *Otology & Neurotology* 27, 2 (2006), 126–143.
- [34] Mark Mirtchouk, Christopher Merck, and Samantha Kleinberg. 2016. Automated estimation of food type and amount consumed from body-worn audio and motion sensors. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 451–462.
- [35] Aaron C Moberly, Margaret Zhang, Lianbo Yu, Metin Gurcan, Caglar Senaras, Theodoros N Teknos, Charles A Elmaragh, Nazhat Taj-Schaal, and Barth F Essig. 2018. Digital otoscopy versus microscopy: How correct and confident are ear experts in their diagnoses? *J. Telemedicine and Telecare* 24, 7 (2018), 453–459.
- [36] Lorenzo Monasta, Luca Ronfani, Federico Marchetti, Marcella Montfico, Liza Vecchi Brumatti, Alessandro Bavcar, Domenico Grasso, Chiara Barbiero, and Giorgio Tamburini. 2012. Burden of disease caused by otitis media: systematic review and global estimates. *PloS One* 7, 4 (2012), e36226.
- [37] American Academy of Pediatrics Subcommittee on Management of Acute Otitis Media et al. 2004. Diagnosis and management of acute otitis media. *Pediatrics* 113, 5 (2004), 1451.
- [38] World Health Organization. 2004. *Chronic suppurative otitis media: burden of illness and management options*. Technical Report. Geneva, Switzerland.
- [39] World Health Organization. 2015. *Hearing loss due to recreational exposure to loud sounds: a review*. Technical Report. Geneva, Switzerland.
- [40] World Health Organization. 2022. *Hearing Loss*. who.int/health-topics/hearing-loss#tab=tab_2.
- [41] Ali Qureishi, Yan Lee, Katherine Belfield, John P Birchall, and Matija Daniel. 2014. Update on otitis media—prevention and treatment. *Infection and Drug Resistance* 7 (2014), 15.
- [42] Valentin Radu, Catherine Tong, Sourav Bhattacharya, Nicholas D Lane, Cecilia Mascolo, Mahesh K Marina, and Fahim Kawsar. 2018. Multimodal deep learning for activity and context recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–27.
- [43] Anne GM Schilder, Tasnee Chonmaitree, Allan W Cripps, Richard M Rosenfeld, Margaretta L Casselbrant, Mark P Haggard, and Roderick P Venekamp. 2016. Otitis media. *Nature Reviews Disease Primers* 2 (2016), 16063.
- [44] Edgar AG Shaw and Ryunen Teranishi. 1968. Sound pressure generated in an external-ear replica and real human ears by a nearby point source. *The Journal of the Acoustical Society of America* 44, 1 (1968), 240–249.
- [45] Ilia Shumailov, Laurent Simon, Jeff Yan, and Ross Anderson. 2019. Hearing your touch: A new acoustic side channel on smartphones. *arXiv preprint arXiv:1903.11137* (2019).
- [46] Hyewon Suh, Nina Shahriaree, Eric B Hekler, and Julie A Kientz. 2016. Developing and validating the user burden scale: A tool for assessing user burden in computing systems. In *Proceedings of the 2016 CHI Conference on human factors in computing systems*. 3988–3999.
- [47] Wei Sun, Franklin Mingzhe Li, Benjamin Steeper, Songlin Xu, Feng Tian, and Cheng Zhang. 2021. TeethTap: Recognizing Discrete Teeth Gestures Using Motion and Acoustic Sensing on an Earpiece. In *Proceedings of the 26th International Conference on Intelligent User Interfaces*. 161–169.
- [48] David W Teele and John Teele. 1984. Detection of middle ear effusion by acoustic reflectometry. *The Journal of Pediatrics* 104, 6 (1984), 832–838.
- [49] John H Teele. 1986. Ear pathology diagnosis apparatus and method. US Patent 4,601,295.
- [50] Heikki Teppo, Matti Revonta, Henriikka Lindén, and Arto Palmu. 2006. Detection of middle-ear fluid in children with spectral gradient acoustic reflectometry: A screening tool for nurses? *Scandinavian Journal of Primary Health Care* 24, 2 (2006), 88–92.
- [51] Yanwen Wang, Jiaxing Shen, and Yuanqing Zheng. 2020. Push the limit of acoustic gesture recognition. *IEEE Transactions on Mobile Computing* (2020).
- [52] Zi Wang, Sheng Tan, Linghan Zhang, Yili Ren, Zhi Wang, and Jie Yang. 2021. EarDynamic: An Ear Canal Deformation Based Continuous User Authentication Using In-Ear Wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–27.
- [53] Tony Wright. 2015. Ear Wax. *BMJ Clinical Evidence* 07, 0504 (2015), 1–24.
- [54] Hanbin Zhang, Chen Song, Aosen Wang, Chenhan Xu, Dongmei Li, and Wenya Xu. 2019. Pdvocal: Towards privacy-preserving parkinson's disease detection using non-speech body sounds. In *Proceedings of the The 25th Annual International Conference on Mobile Computing and Networking*. 1–16.
- [55] Qian Zhang, Dong Wang, Run Zhao, and Yinggang Yu. 2021. SoundLip: Enabling Word and Sentence-level Lip Interaction for Smart Devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–28.