2023 IEEE 43rd International Conference on Distributed Computing Systems (ICDCS) | 979-8-3503-3986-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICDCS57875.2023.00082

EarSonar: An Acoustic Signal-based Middle-Ear Effusion Detection using Earphones

Jingyang Hu*, Hongbo Jiang*, Daibo Liu*, Zhu Xiao*, Hangcheng Cao*, Yue Qi†, Schahram Dustdar‡, Jiangchuan Liu§.
*College of Computer Science and Electronic Engineering, Hunan University
†OPPO Research Institute
‡Computer Science heading the Research Division of Distributed Systems, Vienna University of Technology
§School of Computing Science, Simon Fraser University
Email: {fbhhjy, hongbojiang, dbliu, zhxiao, hangchengcao}@hnu.edu.cn, qiyue@oppo.com, dustdar@dsg.tuwien.ac.at, jcliu@cs.sfu.ca

***Abstract*—Middle ear effusion is a common symptom of otitis media, the reactive physical manifestation of otitis media (OM) in children's middle ear. However, diagnosing MEE for little children at home is troublesome due to their difficulty cooperating and the caregiver's lack of medical knowledge. To this end, we propose EarSonar, a novel acoustic-based MEE diagnostic system. The principle behind EarSonar is that the acoustic absorption effect exists in ear scenarios, and the volume of middle ear fluid can markedly affect the absorbed spectrum energy. By automatically eliminating the impact of potential interference factors and identifying the representative frequency range with the typical reaction of acoustic absorption, EarSonar captures fine-grained signal features on absorbed spectrum energy and models the intrinsic relationship between acoustic absorption and the volume of the filler fluid in the eardrum. On that basis, EarSonar extracts the features of the MEE signal segment and uses k-means clustering to classify middle ear effusion status. We conducted a test on 112 adolescents aged 4-6. We divided the degree of middle ear effusion into three grades. The final average detection accuracy rate exceeds 92%, which is 8% higher than the previous method. We have implemented a proof-of-concept prototype of EarSonar by building upon earphones embedded with a microphone and speaker. Experimental results demonstrate a feasible and effective way to turn earphones into potential home-use MEE screening tools.**

***Index Terms*—Middle ear effusion, Acoustic sensing, Device free, Healthcare**

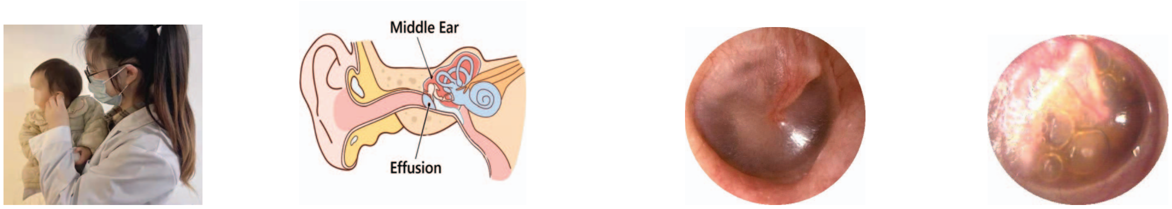
I. Introduction

Middle ear effusion (MEE) [1] occurs when fluid builds up in the space behind the eardrum. It is the reactive physical manifestation of inflammation in the middle ear, namely otitis media (OM). Although many cases can be cured at home, persistent infections can lead to severe complications such as impaired hearing, tearing of the eardrum, and meningitis [2]. OM is a middle ear effusion without signs of acute infection. As a result, it is difficult for patients to perceive the symptom of MEE at the early stages. However, recurrent infection harms infant development because it is associated with speech delay, sleep disruption, poor school performance, balance issues, and cause hearing loss [3]. According to recent studies [4], more than 62% children have had MEE within one year, and under three years is up to 83%.

Current methods of examination are usually pneumatic otoscope [5] and tympanogram [6], both of which are expensive and require medical knowledge. Moreover, it requires an ear specialist to perform. They are not suitable for home diagnostic use [7]. 2016, American Academy of Otolaryngology [8] Calls for researchers to focus on new methods for timely and accurate detection of MEE and new family strategies to help parents and caregivers monitor fluid effusions. Recently, significant progress has been made in acoustic-based ear canal state detection. EarHealth [9] proposes and implements a new earphone-based system for monitoring three different ear diseases (ruptured eardrum, earwax buildup, and blockage) in daily life. However, EarHealth cannot be fine-grained the classification of childhood otitis media, a difficult-to-detect disease. Chan et al. [10] used smart headphones to detect MEE, but they did not perform fine-grained segmentation and analysis on the signal, so the detection accuracy did not exceed 85%. Considering the diagnosis of MEE in infants is troublesome because they are difficult to cooperate with and caregivers lack medical knowledge. Therefore, developing an automated tool that is readily available and capable of accurately diagnosing MEE for home screening purposes is critical and urgent.

In this direction, we explore the possibility of turning a COTS earphone into a full-featured, calibration-free, and readily available MEE diagnostic device for daily healthcare. The idea stems from the fact that due to the inherent acoustic impedance [11], materials can take in sound energy when sound waves are encountered, namely acoustic absorption effect [12], as opposed to reflecting the energy. Acoustic waves penetrate and are reflected by the medium, and the impedance of the medium itself affects the degree to which the acoustic wave penetrates from itself or is reflected. (see Sec. II-A). Specifically, it will be reflected when an acoustic beam passes through the air and encounters an eardrum with effusion. The effusion absorbs part of the signal energy, as shown in Fig. 2(a). Although the intrinsic impedance of the effusion in the middle ear can not be directly measured by using the speaker and microphone on earphones, fortunately, the inherent impedance of the effusion will affect the energy distribution in the frequency domain of the effusion's reflected signals. The proportion of the energy absorbed by the acoustic wave in the medium is determined by the combination of the

2575-8411/23/$31.00 ©2023 IEEE
DOI 10.1109/ICDCS57875.2023.00082



(a) Usage scenario (b) Structural anatomy (c) Clear eardrum (d) Effusion eardrum

Fig. 1: EarSonar utilizes the COTS earphones with an inbuilt microphone and speaker to achieve automatic MEE diagnosis. (b) Shows the structural anatomy of ear with MEE, (c) The normal clear eardrum without effusion accumulation, (d)Fluid builds up in the space behind the eardrum.

acoustic impedances of the two different media.

In this paper, we present EarSonar, a readily available and automatic diagnostic system as illustrated in Fig. 1, which utilizes the extra inbuilt microphone and speaker of COTS earphone to capture acoustic echos that retain the traits of middle ear fluid status, therefore achieves full-featured MEE diagnosis for home screening purpose.

However, several technical challenges must be addressed to realize such a system. First, in the human ear, besides the signals reflected from the eardrum, acoustic signals are reflected by other parts of the ear canal, such as the walls of the ear canal. It is challenging to separate multipath echoes from the ear canal and the eardrum. To overcome the multipath reflection, we design frequency-modulated continuous wave (FMCW) based chirp signals to sense the in-ear effusion status. Since the FMCW signal has high resolution in multipath reflections with different time-of-arrivals, EarSonar can effectively distill target signals reflected from the eardrum.

The second design challenge lies in the subtle energy absorption effect. To accurately identify MEE signs, EarSonar needs to reliably identify a representative frequency range with typical reaction and capture fine-grained signal features to characterize the acoustic absorption. To achieve this goal, we first conduct qualitative analysis (see Sec. II-B) to study the impact of effusion on acoustic absorption over different frequency ranges and observe that the amplitude of the FMCW chirp signals in 18 kHz (inaudible to human ears) is with apparent fading. Furthermore, the quantitative analysis reveals the impact of in-ear effusion volume on acoustic absorption variation. By analyzing the fine-grained signal features on absorbed spectrum energy, EarSonar can infer the in-ear effusion status.

Finally, to effectively identify the symptom of MEE, we have to model the intrinsic relationship between acoustic absorption and the volume of filled fluid in the eardrum. However, an effective diagnostic model still lacks that can accurately identify the symptom of MEE and apply it to various patients and diverse application scenarios. Meanwhile, different people's ear canals and other modes of wearing earphones will cause varying test results. To solve these problems, We extracted the MEE signal's statistical and MFCCs features. We used K-means clustering to classify and detect middle ear effusion.

We have implemented the prototype of EarSonar by building upon COTS earphones embedded with a microphone and speaker. We conducted experiments and confirmed clinical diagnoses in a pediatric hospital for over six months. Experiment results show that EarSonar can achieve median values for Precision, Recall, and F1score rates are 92.8%, 92.1%, and 92.3%, respectively. The results show that earphone has the full potential to become a tool for the initial screening of MEE in families.

Our main contributions are summarized as follows.

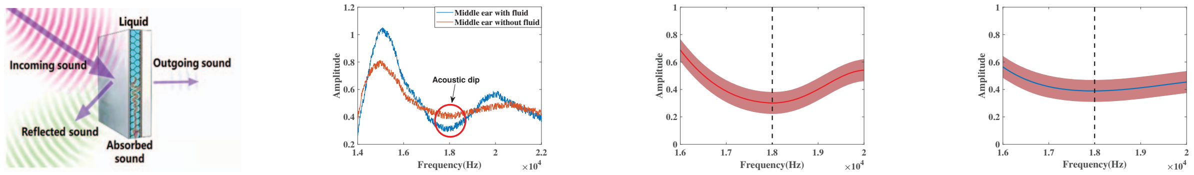
- We experimentally verified that the acoustic absorption effect exists in-ear scenarios. By identifying the representative frequency range with the typical reaction of acoustic absorption, EarSonar captures fine-grained signal features on spectrum energy to infer the in-ear effusion status.
- We model the intrinsic relationship between acoustic absorption and the volume of filled fluid in the eardrum. By designing a k-means Classifier, EarSonar can effectively identify MEE's symptoms and severity and apply diverse application scenarios to various patients.
- We design EarSonar to retrofit earphones into a wearable MEE diagnostic system by using the inbuilt microphone and speaker to capture acoustic echos that retain the traits of middle ear fluid status, therefore achieving full-featured MEE diagnosis for home healthcare.

II. Motivation

In this section, we present the principle of the acoustic absorption effect and conduct empirical studies to validate the phenomenon in the ear scenario and crucial findings.

A. Principle of Acoustic Absorption

Due to the inherent acoustic impedance [11], materials can absorb sound energy when encountering sound waves, the sound absorption effect [12], rather than reflecting energy. When sound waves are transmitted between different media, different degrees of reflection and refraction will occur due to the different impedance of the media. As sound waves travel from the ear canal to the eardrum with fluid, the fluid affects the reflection and refraction of the acoustic signal.



(a) The principle of acoustic absorption effect.

(b) The original waveform obtained by the acoustic signal when it enters the middle ear

(c) Middle ear with fluid

(d) Middle ear without fluid

Fig. 2: Feasibility analysis of EarSonar

Theoretical model. We now introduce how the impedance affects the acoustic absorption effect. The acoustic impedance of in-ear effusion describes the ratio of acoustic signal pressure P to the velocity U of molecule movement [13] can be expressed as: $Z_0 = \rho_0 c_0$, which is caused by the sound pressure of the medium, as well as the product of the density of liquid ρ_0 and the acoustic signal speed c_0.

Specifically, suppose the waves are directed vertically from the air to the effusion in the eardrum. In that case, the incident acoustic pressure is denoted as $P_i = P_0 cos(wt - kx + \varphi)$, where P_r is the reflected acoustic, P_0 is acoustic pressure amplitude, w is the angular frequency, k is wavenumber associated with wavelengths, and φ is the initial phase. The relationship between P_i and P_r can be expressed as

$$R = \frac{P_r}{P_i} = \frac{z_{fluid} - z_{air}}{z_{fluid} - z_{air}}, \tag{1}$$

where z_{air} denotes the acoustic intrinsic impedance of air, and the z_{fluid} is the that of fluid. R is the liquid reflectance, which is determined by the degree of effusion. Among them, under ideal conditions, the relationship between impedance Z and effusion thickness d is given by [14]

$$Z = \sqrt{\frac{\mu}{\xi}} tanh(\frac{2\pi d\sqrt{\xi\mu}}{\lambda}), \tag{2}$$

Where μ, ξ and λ can be constants, representing permeability, dielectric constant and signal bandwidth, respectively. In this sense, the impedance Z can be regarded as the tangent function of the thickness d. Under ideal conditions, as the thickness d increases, the impedance Z increases accordingly. Meanwhile, the reflected signal is

$$P_r = R * P_0 cos(wt - kx + \varphi). \tag{3}$$

Theoretically, by measuring and analyzing the reflected signals, we can describe the energy of reflected signals. With aspects to the ear canal, the sound is transmitted through the microphone and reflected at different objects (i.e., ear wall, eardrum, and foreign body) and finally received by the earphone's inbuilt microphone, assuming that the sinusoidal signal we transmit is

$$R(t) = \sum_{r \in M} P_r cos(2\pi f t + \varphi_i), \tag{4}$$

where f and φ represent frequency and phase, respectively. M denotes the set of all paths of acoustic signals. We use A_i to represent the amplitude of each path. The amplitude spectrum of a specific echo is

$$A(f) = \frac{FFT(R(t))}{N} = \sum_{r \in M} P_r = \sum_{j \in F} P_j + \sum_{k \in C} P_k, \tag{5}$$

Where $\sum_{j \in F} P_j$ denote the set of paths reflected by the eardrum, $\sum_{k \in C} P_k$ denote the set of paths reflected by the ear canal and foreign body in the ear. It can be seen from the above formula that the amplitude change of the received signal will be affected by the liquid volume and the propagation path.

Inspiration. Actually, it is not easy to directly measure the intrinsic impedance of filled effusion in the middle ear. However, by actively using a speaker to send a well-designed acoustic signal and exploiting a microphone to capture the reflected echoes, we can perceive the energy distribution in the frequency domain of the effusion's reflected signals, dependent on the inherent impedance of the volume of the effusion. The energy of the acoustic signal will be absorbed by the medium, and the amount of absorption is related to the impedance of the medium itself and the volume of the medium [15].

B. *Feasibility Analysis*

Based on the above principle, we conduct empirical studies to analyze the feasibility of using an earphone with an extra embedded microphone and speaker (The modified prototype is shown in Fig. 4) to identify the symptom of MEE.

To validate our previous analysis, we selected a patient with otitis media (female, four years old) at Children's Hospital and followed her continuously. To specify, we applied the designed EarSonar prototype to collect the ear canal data of the participants. We use a microphone to send out 14kHz-22kHz acoustic signals, perform FFT on the acoustic signals, and then perform data analysis. We compared the acoustic data of this participant when she was just diagnosed with otitis media and when she was fully recovered, as shown in Fig. 2(b). We found that in the entire frequency domain, the acoustic signals of different periods present different characteristics, and an apparent acoustic dip is produced near 18kHz. At the same time, we expanded the scope of the experiment and collected more participants (112 participants). We fitted the data and finally classified the symptoms of all patients when they developed MME symptoms and fully recovered, as shown in Fig. 2(c) and Fig. 2(d). In addition, we found that the signal followed different distributions when being with and without effusion.

Considering the user-friendliness and anti-interference performance, the frequency range of 16-20 kHz can be well used to transmit FMCW chirp signals for MEE sensing. When a chirp signal passes through the ear canal, the signal reflections from eardrum will be captured by inbuilt microphone. We finally chose the 16kHz to 20kHz signal as our test signal, which is relatively weak and beyond the range of human hearing.

III. System Overview

The schematic diagram of EarSonar is illustrated in Fig. 3. Fig. 5 shows an overview of the system design, including four main modules: *Acoustic signal collection*, *Signal preprocessing*, *Acoustic absorption analysis* and *MEE detection*.

Acoustic signal collection. As a terminal interface driven by EarSonar, the earphone with an inbuilt microphone and speaker is used to sense the effusion status inside the middle ear. EarSonar first uses a speaker to send an FMCW chirp signal over the 16-20 kHz frequency range, which the ear canal will reflect, eardrum, and the foreign objects in the ear. Then, the inbuilt microphone can collect these acoustic echoes.

Signal preprocessing. Given the received reflections, EarSonar first employs a bandpass filter to eliminate the impact of noises and adopts an event detection mechanism to extract each chirp and its corresponding echo signal. Then we use the correlation coefficient to separate echos reflected by different in-ear objects. Finally, it uses the parity decomposition method to identify the echoes reflected by the eardrum accurately.

Acoustic absorption analysis. By performing FFT on the eardrum-reflected echoes to distill the power spectral density, EarSonar extracts the fine-grained signal features on absorbed spectrum energy, which retains the trait of effusion status. To classify the ear effusion state, We extracted MFCCs features and statistical features of MEE signals. Then designed the feature vector.

MEE detection. Given the results of acoustic absorption analysis, EarSonar further models the intrinsic relationship between acoustic absorption and the volume of filled fluid in the eardrum. According to the selected features mentioned above, EarSonar uses k-means clustering to classify feature vectors.

IV. System Design

We present the technical details of EarSonarin this section. First, EarSonar processes the raw acoustic signal, detects the MEE signal, extracts features from the MEE signal segments, and cluster these feature vectors to determine user's middle ear fluid status. Next, We will detail how to properly use FMCW signals for eardrum echo detection and clustering techniques for MEE detection.

A. Acoustic Signal Collection

We design an FMCW chirp signal for client-side MEE status sensing. After that, the details of acoustic signal generation and processing are presented.

FMCW chirp signal design. The frequency of the FMCW signal (as shown in Fig. 6) varies linearly with time and has good autocorrelation properties. These characteristics make FMCW signals commonly used for multi-target detection. Comparing to other waveforms, FMCW signals are easy to demodulate due to their high spectral efficiency. The frequency f of the signal changes linearly with time as $f = f_0 + \frac{B}{T}t$, where f_0 is the initial frequency, B is the bandwidth, and T is the duration of the acoustic signal.

Signal generation. To convert earphone into MEE detection system. We need to design the acoustic signal so that, in principle, echoes from the eardrum can be detected without placing additional burden on the user. Consider that normal people can hear sounds in the frequency range of 20 Hz to 15 kHz. The sampling rate of current commercial smartphones is usually set at 48 kHz. According to the Nyquist sampling theorem, it is more appropriate to design the acoustic signal below 24 kHz. Therefore, the frequency range of the designed FMCW signal cannot exceed 24 kHz. Third, the frequency range of the design should be easily distinguishable from ambient noise. This is not easily disturbed by ambient noise and is easy to filter. Hence, EarSonar adopts FMCW chirps between 16 kHz to 20 kHz. To separate the echoes from the eardrum from the multipath signal from the ear canal, we intermittently send out the FMCW signal (Chirp). The duration of the chirp determines the distance resolution between the echoes. We need to ensure that the eardrum signal is not aliased too much in the multipath echo from the ear canal when the echo of the eardrum is detected.

To reduce aliasing of echoes from the ear canal and adequately receive echoes from the eardrum. We set the duration of chirp to 0.5 ms and the interval between adjacent chirps to be no less than 5 ms, so that we can capture all echoes in the range of 10 cm, and try to avoid the overlap between echoes . Note that the length of the human ear canal is usually 2 cm-3.5 cm [16] as shown in 7(a), we set the interval of each chirp to 5 ms. We set the initial frequency f_0 to 16 kHz, the bandwidth B to 4 kHz and the chirp duration T to 0.5 ms. When the signal is reflected by the eardrum, it produces an echo as shown in the figure 7(b).

B. Signal preprocessing

To capture the reflections from the eardrum, EarSonar has to perform the following processing on the received signal.

1) Noise removal: The FMCW chirp signals will get reflected by in-ear objects. To reduce the noise interference in the environment, we filter the received echo signal through a Butterworth bandpass filter. In addition, EarSonar passes the acoustic signal through a Hanning window [17] on each pulse to reshape the envelope of the signals and increase their peak-to-sidelobe ratio to obtain a higher signal-to-noise ratio (SNR).

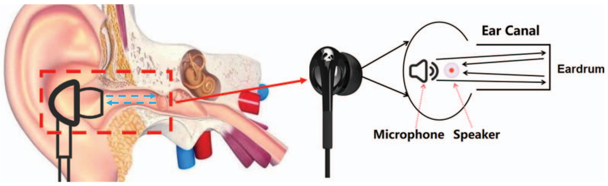


Fig. 3: schematic diagram of using earphone for MEE diagnosis.

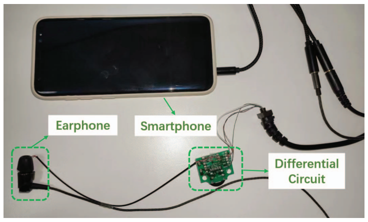


Fig. 4: The developed data recording prototype.

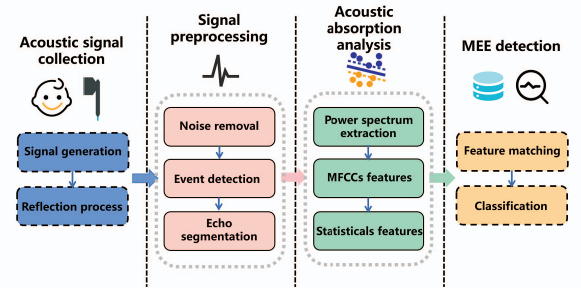
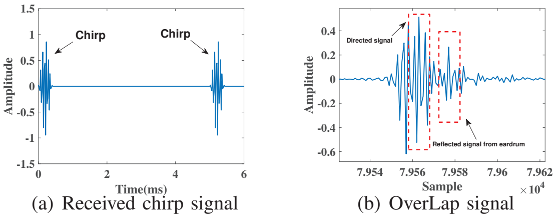


Fig. 5: The methodological flow of EarSonar.



(a) Received chirp signal (b) OverLap signal

Fig. 7: The captured chirp by micrphone.

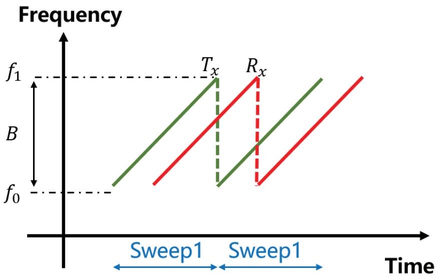
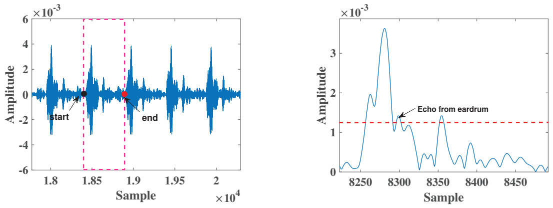


Fig. 6: The principle of FMCW chirp.



(a) Events detection. (b) Segmentation.

Fig. 8: Events detection and Segmentation.

2) Event detection: After noise removal, the next step is to segment and extract the signal of the echo band. Considering that the generation of echo events will have apparent energy compared with others, EarSonar uses an adaptive energy event detection scheme for event detection. Specifically, for a signal X(i), We use a sliding window scheme to calculate the average signal power for each window W, as follows

$$\begin{aligned}\mu(i) &= \frac{1}{W}A(i) + \left(1-\frac{1}{W}\right)\mu(i-1)\\ \sigma(i) &= \frac{1}{W}B(i) + \left(1-\frac{1}{W}\right)\sigma(i-1)\end{aligned} \tag{6}$$

where $\mu(i)$ and $\sigma(i)$ represent the mean and standard deviation, respectively. A(i) represent the cumulated power and and B(i) represent the overall standard deviation of signals within a sliding window. where

$$\begin{aligned}A(i) &= \frac{1}{W}\sum_{k=i}^{W+i}|X(k)|^2\\ B(i) &= \sqrt{\frac{1}{W}\sum_{k=i}^{W+i}\left(|X(k)|^2 - A(k)\right)^2}\end{aligned} \tag{7}$$

Then, the potential starting point of X(i) can be identified if $|X(i)|^2 > \mu(i) + \sigma(i)$, and the end point satisfy $|X(i)|^2 < \bar{\mu}$. $\bar{\mu}$ is the average signal power. Fig. 8(a) shows the example of event detection.

3) Echo segmentation: The goal of EarSonar is to extract the echo signal from the eardrum. The complete echo signal includes the direct signal (the speaker is directly transmitted to the microphone) and the multipath echo from the ear canal. We need to eliminate the influence of these multipath signals as much as possible. By the way, we need to create a uniform template and sequence length for the echo signal from the eardrum to ensure that the subsequent feature extraction can be performed usually. In general, we need to divide the obtained echo sequence, eliminate the overlapping interval as much as possible, and extract the signal from the eardrum.

To solve this problem, we propose a time series-based even/odd decomposition segmentation method [18] for echo segmentation. The core of the method is to concentrate the energy in the even and odd parts of the time series, and optimally place the center of symmetry in the even and odd parts. In this way, we can identify local symmetric intervals in the time series.

After event detection, we get each chirp signal and its echo x[n], $n \in [-T, T]$. This signal has limited support and time centering without loss of generality. By parity decomposition we represent x[n] as the sum of the even and odd parts, $x_e[n]$ and $x_o[n]$, respectively. In the discrete case, we can get $x[n] = x_e[n; n_0] + x_0[n; n_0]$

where n_0 is the point of symmetry, the $x_e[n; n_0]$ and $x_0[n; n_0]$ is given by

$$\begin{aligned} x_e[n; n_0] &= \frac{x[n]+x[2n_0-n]}{2} \\ x_0[n; n_0] &= \frac{x[n]-x[2n_0-n]}{2} \end{aligned} \tag{8}$$

The choice of the symmetry point n_0 is not arbitrary. It needs to correspond to the position of the sample. Since the input acoustic signal sequence x[n] $n = 1, ..., L$ has finite duration, then $n_0 = \frac{k}{2}, k \in 2, ..., 2L$. When k is odd (or even), the folding point corresponds to the position of the half sample. Of course, the length of the odd and even parts may be different. For the parity energy E_e and E_o under the best symmetry point, the expression is

$$\begin{aligned} E_0 = \textstyle\sum_{n=-\infty}^{+\infty} \left|x_e\left[n; n_o\right]\right|^2 = \sum_{n=-\infty}^{+\infty} \left|\frac{x[n]+x[2n_0-n]}{2}\right|^2 \\ = \tfrac{1}{4} \textstyle\sum_{n=-\infty}^{+\infty} \left|x[n]\right|^2 + \left|x[2n_0 - n]\right|^2 + 2x[n]x[2n_0 - n] \\ = \tfrac{1}{2}E + \tfrac{1}{2} \textstyle\sum_{2n_0}^{1} x[n]x[2n_0 - n] \end{aligned} \tag{9}$$

Where [1, $2n_0$] is the non-empty support of the product in the summation. E_o is energy of the odd part and has the same expression, as long as the sign of the sum in the last line changes. Use the convolution definition of the energy sequence to generate

$$\begin{aligned} E_e &= \tfrac{1}{2}E + (x * x)[2n_0] \\ E_o &= \tfrac{1}{2}E - (x * x)[2n_0] \end{aligned} \tag{10}$$

After that we need to calculate the automatic convolution of x[n]. After obtaining the automatic convolution, the best candidate position can be expressed as $2n_0 = arg \max_m |(x * x)[m]|$.

In the previous process, we have identified the largest support segment that exhibits nearly perfect even or odd local symmetry for any sequence x[n]. To find fragments with strong symmetry and no overlap, these fragments cover most of the sequence. It requires three steps. First, we compute the autoconvolution of the sequence x[n], then locate all possible even or odd symmetry points. At this point, all local extrema become candidate points, and we store the positions of all candidate points in the list $\mathcal{C}$.

Second, we need to determine the symmetric range around the location of each candidate local extreme point. Here we need to set several parameters. First, each candidate location is marked as c_i, $c_i \in \mathcal{C}$. We define m_l as the minimum symmetry support and $p_t(0.5 < p_t < 1)$ as the even/odd energy ratio threshold. Then, we need to select a subsequence y[n] centered on c_i bits, and y[n] maintains a uniform length. We perform parity decomposition on y[n] and calculate the E_y, E_{ye} and E_{yo} of y[n], which respectively represent the energy of y[n] and the energy after parity decomposition. In order to divide y[n] into even or odd symmetrical segments, we need to check whether the energy ratio satisfies $E_{ye}/E_y > p_t$ or $E_{yo}/E_y > p_t$. If neither of these two conditions are verified, then c_i will be removed from the list of candidate positions.

Third, after we get the candidate sequence set, the potential best eardrum echo band follows the following principles (i) has a higher energy ratio (ii) keeps a distance of 2 cm-3.5 cm from the direct signal. We ensure that the eardrum echo maintains a high correlation with the direct signal for the first principle. A higher energy ratio can ensure that this signal segment is as complete as possible instead of overlapping with the multipath signal. Secondly, the second principle comes from keeping a distance between the earphone and the eardrum. This distance will be kept within a specific range for most people. We find the best distance from the direct signal within 2.5-5cm. Candidate point c_i, and create the best eardrum echo sequence m. The final confirmation process from the eardrum echo is shown in the Fig. 8(b).

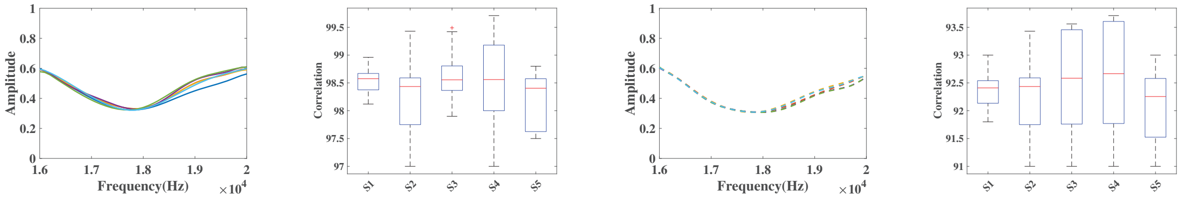
C. Acoustic Absorption Analysis

1) Echo power spectrum extraction: After signal segmentation, we get the peak point from the eardrum echo in the acoustic echo signal. For subsequent stable feature extraction, we need to specify a uniform window for FFT. EarSonar determines the peak from the eardrum. In the time domain, we take the peak sampling point of the eardrum as the centre and collect N sampling points on both sides of the fixed window. We obtain the power spectral density from the points collected by each chirp. Using this method, we can eliminate the impact of multipath signals on MEE detection as much as possible while ensuring that the collected signals are as uniform as possible.

We use samples of participant A collected in each of the six sessions($S_1 ... S_6$) in a quiet room accompanied by 20 - 30dB noise within the same day. Fig. 9(a) plots the power spectrum density (PSD) of the same person. It can be seen that the data we measured under different sessions maintained a high degree of consistency. Fig. 9(b) shows the correlation coefficients between S_1 to S_6 and other sessions. It can be seen that the echo of the eardrum and ear canal is relatively stable for the same person under normal circumstances.

On the other hand, We also need to prove whether the eardrum echoes of different people are similar under normal circumstances. Fig. 9(d) shows the correlation coefficients of another participant at six different sessions. It can be seen that the overall trend of participant B's PSD curve is similar to that of participant A, and the overall correlation coefficient between them is still higher than 90%. This provides the basis for our MEE detection.

The task of EarSonar is to detect whether the user has middle ear effusion. Usually, the middle ear effusion will last for 2-3 weeks. We tracked different participants to perform eardrum acoustic measurements at 8:00 am and 10:00 pm every day. Fig. 10(a) and Fig. 10(b) show the results from the time they enter the clinic until they recover, we can see that the participants signal patterns gradually return to normal levels (Fig. 9(a) and (b)), We divide the types of ear effusion into Purulent, Mucoid and Serous.



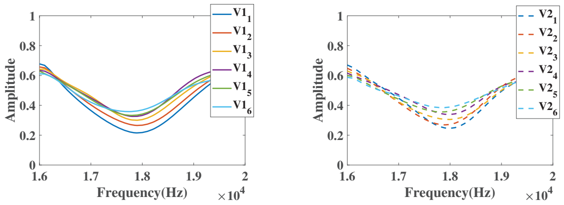
(a) The curves of participant A

(b) The correlation coefficients of participant A

(c) The curves of participant B

(d) The correlation coefficients between participant A and participant B

Fig. 9: The curves and correlation coefficients from different participants of middle ear without effusion

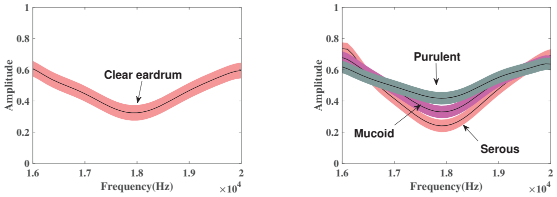


(a) MEE echo power spectrum of participant A.

(b) MEE echo power spectrum of participant B.

Fig. 10: The power spectrum of the reflected signal from the time of admission to the recovery of the two participants.

Finally, we perform FFT processing on the interpolated signal. After the processing of the above steps, we get the power spectrum interval of Middle ear without fluid and Middle ear with fluid, as shown in Fig. 11(a) and Fig. 11(b). the shaded area represents the range of the eardrum echo power spectrum in different states. We divide middle ear effusion into four states according to different middle ear effusion intervals, Clear, Purulent, Mucoid and Serous.



(a) Middle ear without fluid

(b) Middle ear with fluid

Fig. 11: The power spectrum of different state of MEE

2) Feature Extraction: In this section, for MEE detection and classification we need to obtain fine-grained features from MEE signal segments. We need to analyze the features of MEE signals and then extract the features.

MFCCs Features. Mel-frequency cepstrum coefficients (MFCCs) are proposed based on the auditory characteristics of the human ear and are widely used in speech recognition. It has a nonlinear correspondence with the signal in the frequency domain. Subtle changes and differences in acoustic signals can be represented by the MFCC. In order to obtain the MFCC of the MEE signal, we first need to perform fast Fourier processing on the segmented eardrum echo to convert the signal to the frequency domain. Then we need to split the frequency domain signal into multiple smaller frequency bins and then use a triangular filter on each frequency bin to calculate the short term power for each frequency bin. Finally, a discrete cosine transform (DCT) is used on these short-term power segments to obtain complex pairs. In EarSonar, the MEE signals from different periods of users have differences in the frequency domain, and the difference information can be reflected by the MFCC.

Statistic Features. We obtained the power spectral density profile of the echo from the eardrum under different effusion states by FFT calculation. We use the statistic features reflect the global characteristics of the MEE signal, We extract the following features from the Power spectrum sequence: 1)the mean and standard deviation, 2)the maximum and minimum value of x, 3)the skewness, 4)the kurtosis.

For MEE detection and classification, EarSonar constructs a 105-element feature vector for each MEE signal segment, which includes MFCC features and statistical features. In order to reduce the computational load of the model, we use the Laplacian score to measure the importance of features, and save the top 25 features with importance.

3) In-group k-means based clustering.: To accurately group different degrees of ear effusion, the k-means algorithm (a classic clustering method) is chosen due to its computational efficiency. The core of K-means clustering is to divide each data vector into the cluster represented by the nearest cluster center point. Given k values and k initial cluster center points, and assign all points to each cluster. After completing the above steps, recalculate the center point of each cluster according to all the vectors in each cluster, and then iterate the clustering process after each new point is added and continuously update the cluster center point. k-means clustering is proposed to partition n EarSonar samples $X = \{X_1, X_2, ...X_n\}$, where each object has attributes of m dimensions. The goal of the K-means algorithm is to cluster n objects into k clusters according to the similarity between objects, and each object belongs to one and only one cluster whose distance from the center of the cluster is the smallest.

For EarSonar, we have given four cluster centers according to the four different states of the effusion as $\{C_1, C_2, C_3, C_4\}$. Then we calculate the Euclidean distance from each object to each cluster center, as follows

$$\operatorname{dis}(X_i, C_j) = \sqrt{\sum_{t=1}^{m} (X_{it} - C_{jt})^2} \tag{11}$$

where X_i represent the i-th sample($1 \leq i \leq n$), C_j represent represents the j-th cluster center($1 \leq j \leq 4$), X_{it} Represents the t-th property of the i-th sample($1 \leq t \leq m$). C_{jt} represents the t-th property of the j-th cluster center. By comparing the distances of each object to each cluster center in turn, the objects are assigned to the cluster with the closest cluster center, and 4 clusters are obtained $\{S_1, S_2, S_3, S_4\}$. To obtain the best clustering results, we minimize the Euclidean distance of the combined features, the EarSonar samples in each cluster by satisfying

$$\min \sum_{i=1}^{K} \sum_{x \in C_i} \operatorname{dist}\left(c_i, x\right)^2 \tag{12}$$

After k-means clustering, the vast majority of EarSonar samples are correctly classified, while only a small fraction of corridor points are mixed.

4) Outlier remove.: K-means clustering can perform badly in the presence of outliers, and these individual data have a very large impact on the average value. Given this limitation of the K-Means algorithm, we need to pay special attention to these data noise and outliers when applying the algorithm. For data noise and outliers in the clustering, we use two strategies to remove outliers. First, outliers that are farther from the cluster center point than any other data point are directly removed. In order to prevent accidental deletion, data analysts need to monitor these outliers in multiple clustering loops, and then compare them with the results of multiple loops based on business logic, and then decide whether to delete these outliers. Second, the random sampling method can also better avoid the influence of data noise. Because of random sampling, the probability of data noise and outliers that are rare events can be randomly selected into the sample will be very small, so the randomly selected sample will be relatively clean. The clustering analysis of the random sample can not only avoid the misleading and interference of data noise, but also the clustering result can be applied to the remaining data set as a clustering model to complete the clustering of the entire data set.

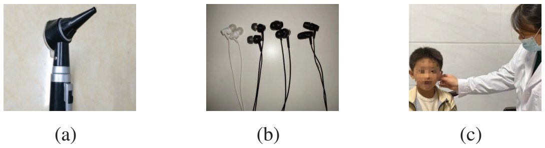
V. System Implementation

Hardware prototyping. Our experiment introduced an additional Micro microphone (low-cost) to maintain a parallel structure with the earphone speaker to facilitate the acquisition of echoes from the eardrum as shown in Fig. 3. We embedded it into various in-ear earphone as shown in Fig. 12(b), which signal-to-noise ratio is generally higher than 70dB, the sound input quality is high, and the sensitive fluctuation value is lower than -30dB. It can perfectly cover the acoustic signal with frequency response range of 20Hz-20KHz. Considering that most of our participants are children, we have customized silicone earplugs that are more comfortable and more in line with the structure of children's ear canals, which ensures good ambient noise isolation and a comfortable user experience. We connect EarSonar to the HUAWEI Mate 40. All acoustic data will be uploaded to an MSI Pro laptop with a 2.5 GHz Intel i7 CPU and 16GB memory as the server, and we use Python to process the algorithm we use in EarSonar. The server will finally feed back the detection result of MEE.

Participant recruitment. We recruited 112 participants (60 males and 52 females) from Children's Hospital to evaluate the performance of EarSonar. We distribute the ages of these participants between 4-6 years old, and our experimental scene is shown in Fig. 12(c). We pooled all data for cross-validation. We followed participants from diagnosis to full recovery (hospital discharge). The IRB approved our experiment with 112 volunteers (we followed each participant for at least 20 days) over 12 months. To increase the realism of the experiment, we provided participants with attractive reward gifts before and after the experiment.

VI. Evaluation

We implemented a proof-of-concept prototype of EarSonar using a micromodified earphone and smartphone. We developed a client application and set up the server to run on the user's smartphone. The EarSonar prototype uses the earphone's speaker and an additional embedded microphone for MEE detection. To verify the validity and accuracy of EarSonar, we recruited 112 participants.



(a) (b) (c)

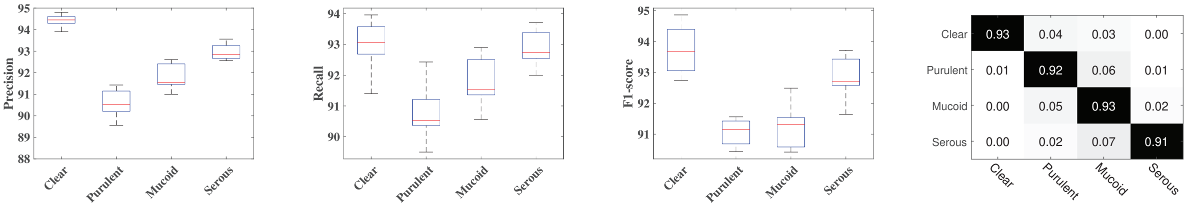
Fig. 12: (a)The pneumatic otoscope, (b)Differene kinds of earphones, (c)Experimental environment.

A. Experiment Setup and Metrics

Data collection. EarSonar labels MEE status as Clear, purulent, mucoid and serous. In the testing stage, we collected at least three MEE states for each participant. Acoustic signals for three week (20 days) of each participant, which completely covered the four states of the participant's middle ear effusion recovery, Acoustic data for 10s is collected every time at 8 am and 6 pm each day. We collected a total of 44800s (112 x 20 x 10 x 2) acoustic signal data. In order to verify the performance of EarSonar, we also set different experimental parameters, such as different room noises, different earphone wearing modes, etc.

Groundtruth data. Before each experiment, a trained nurse tests the patient, who is awake and remains or sitting. To verify the performance of EarSonar we will ask doctors to use the professional pneumatic otoscope(As shown in Fig. 12(a)) to detect the degree of eardrum effusion of participants as Groundtruth every time we collect data.

Evaluation metrics. In this paper, we use precision, recall, F-score, Confusion matrix as metrics to comprehensively measure the performance of EarSonar. For classification, we used the k-means clustering algorithm on the preprocessed acoustic data. We extracted the MFCC features and statistical features of the MEE signals. We train our algorithm on data collected in smartphones. To verify the effectiveness of the algorithm, we use leave-one-out cross-validation (LOOCV) for evaluation. Specifically, in each iteration of LOOCV, we use



(a) Precision (b) Recall (c) F1-score (d) Confusion matrix

Fig. 13: Estimation performance of EarSonar.

data from 111 of the 112 participants for training. Then output the prediction for the last participant. We repeat this process for all 112 participants to guarantee the accuracy of the model trained on all 112 participants.

B. EarSonar Performance

We first use leave-one-out cross-validation [10] to evaluate the overall performance of EarSonar. Since the recovery of middle ear effusion is a long cycle, we divided the types of effusion into serous, mucous, suppurative, and clear, and serous, mucinous, and suppurative represent the three stages of MEE. Clear represents a normal eardrum. We use Precision, Recall, F1-score to evaluate the overall performance of EarSonar, The results are shown in Fig. 13. The median values for Precision, Recall, F1score rate are 92.8%, 92.1%, 92.3% respectively. At the same time, we observe that for the four states of MEE, the Clear state has the highest detection accuracy, while the Purulent state The detection accuracy of the Mucoid state and the Purulent state is lower, because the Purulent state and the Mucoid state are prone to aliasing. Fig. 13 plots the confusion matrix of 4 states. The average accuracy rate is higher than 92%. The few that were misidentified were those states with similar characteristics. These results demonstrate the extent to which EarSonar can reliably detect MEE in real time.

C. Impact Quantification

1) Impact of angle.: Considering that we always wear earphones in a fixed way in the laboratory, in practice, when putting on and taking out the earphones, the relative position of the earphones and the ear canal will change, and even the behavior of wearing earphones will be irregular. We regard the plane of the human ear as a plane coordinate system to set the x-axis and y-axis, and we take the standard wearing headset posture as an angle of 0 degrees. And rotate the position of earphone. Six angles were tested, including 0 degrees, 10 degrees, 20 degrees 30 degrees, and 40 degrees. Table. I shows the echo detection accuracy under these five different angels. The precision results of the six cases were 92.8%, 91.3%, 90.2%, 88.5%, and 86.4%. EarSonar has the highest accuracy rate at 0 degrees. When participants place their earphones outside the effective area of 20-40 degrees, the multipath reflection in the ear canal will change significantly, resulting in a decrease in inaccuracy.

2) Background noise: Ubiquitous background noise can affect acoustic sensing systems. We examine our EarSonar in 4 different noise levels in a room with 40 dB, 55 dB, 65 dB, and 75 dB. To ensure reliability and control for experimental variables, we add additional background noise to the collected data to simulate the test environment under different scenarios of different sound pressure levels [19]- [20] with a smartphone one meter away from the participant.

TABLE I: The Acoustic Measurements Accuracy.

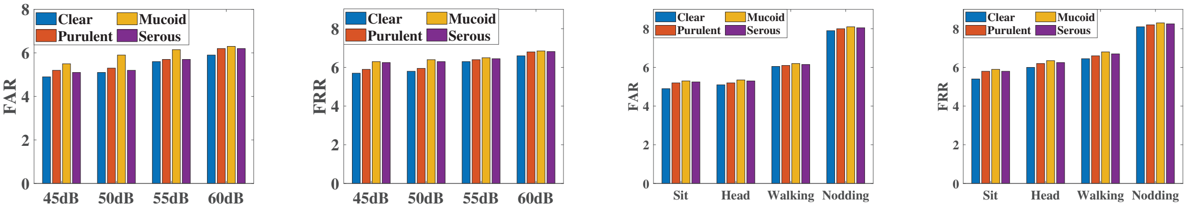
Angle	Axis0	Axis10	Axis20	Axis30	Axis40
Accuracy	92.8%	91.3%	90.2%	88.5%	86.4%

From Fig. 14(a), We observed a slight increase in FARs with increasing noise, but background noise sound pressure level (SPL) did not significantly affect FARs. However, as shown from Fig. 14(b), the FRRs will increase with the increase of noise. We recommend that users use EarSonar in a quiet room to prevent uncontrollable background noise from affecting the detection results.

3) Body movement: Involuntary movements of the user while using the EarSonar may affect the detection results. In order to evaluate the robustness of the system under different slight movements of the user, we prescribe several body movements such as sitting, slight head movements, walking and slight nodding. Considering the young age of the participants. All actions are carried out with the assistance of a physician.

Fig. 14(c) and Fig. 14(d) show the detection performance of EarSonar under different body movements. We can see that the EarSonar maintains relatively good robustness under the two actions of sitting down and slight head movement. However, the two actions of walking and nodding affect performance. This shows that EarSonar is robust to slight motion noise. When the degree of motion increases, the relative position of the earphone and the ear canal may change, resulting in a decline in system performance. What needs to be pointed out is that we recommend that when using EarSonar for MEE testing, participants should maintain a sitting posture.

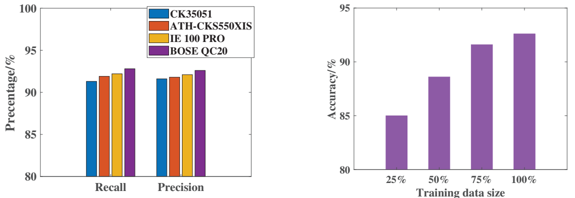
4) Impact of the device: Earphones with different configurations may affect the performance of EarSonar. Therefore, we use four other earphones simultaneously for data collection. Specifically, we use CK35051, ATH-CKS550XIS, IE 100 PRO, and BOSE QC20 for testing. The prices of these devices are different, and the hardware functions will differ. Fig. 15(a) shows a comparison of the recall and precision of EarSonar under these four other earphone usages. The results show that EarSonar can adapt to different earphones and run robustly



(a) False Acceptance Rate (b) False Rejection Rate (c) False Acceptance Rate (d) False Rejection Rate

Fig. 14: The impact of Background noisy and Bodymovement.

under commercial earphones.



(a) Impact of the different earphone (b) Impact of Training Size

Fig. 15: Impact of the different earphone and Training Size.

5) Impact of Training Size.: The proposed detection of MEE states is based on the fact that the middle ear effusion states from different states will absorb and reflect sound waves to different degrees. The model is actually trained to recognize these varying degrees of middle ear fluid. To investigate the user's burden on training data collection, we train the model with different numbers of data samples. As shown in Fig. 15(b). The average accuracy rate of EarSonar increases with the increase of the training data set, but when we use 50% of the data volume, the average accuracy rate of EarSonar reaches 91.6%, and then with the increase of the data volume, the accuracy rate improves greatly Small.

6) Power and latency Measurement.: EarSonar uses Earphone and smartphone to perform MEE detection. Currently, there is no additional sensing module configured on earphone for MEE detection. In fact, Most earphones today offload data directly to your phone or over the web. We evaluated the power consumption and latency of EarSonar on smartphone. We train machine learning models on laptops and implement MEE recognition (including bandpass filtering, feature extraction, and inference) on smartphones. When we perform MEE detection on mobile phones, the latency of different operation is shown in Table. II. In fact, the recognition (feature extraction and inference) time is usually very short, so the actual energy consumption will be much lower as shown in Table. III.

VII. Related Works

MEE Detection. Many methods are focusing on how to detect MEE more conveniently at present. Chan et al. [10] use a smartphone to detect middle ear fluid. They only used an extra piece of paper to convert the smartphone into a tool detecting MEE. Song et al. [21] use the water stream in the syringe as the medium to detect the influence of the middle ear effusion on the ultrasonic signal. Ozana et al. [22] use cheap optical instruments to assist medical staff in detecting MEE. P3Q-2 [23] highlights that the condition of the middle ear is determined by analyzing the ultrasonic echo collected from the tympanic membrane and middle ear cavity. J et al. [24] developed a handheld optical coherence tomography (OCT) system to monitor in vivo response of biofilms and MEEs in the OM-induced chinchilla model, the standard model for human OM. Different to existing works, EarSonar is the first earphone-based MEE diagnostic system, exploiting the inbuilt microphone and speaker of COTS earphones to capture acoustic echos that retain the traits of middle ear fluid status achieve full-featured MEE diagnosis.

Earphone Sensing. Nokia Bell Lab's has developed a smart earphone [25] prototype esense to expand the development of earphones in the field of human health monitoring. Min et al. [26] use embedded sensors in the headset to detect speaking activity and participant emotions separately. Hossain et al. [27] designed a multi-sensory device based on eSense and an activity recognition framework was proposed. It has a microphone, 6-axis inertial measurement unit and dual-mode Bluetooth. they use eSense accelerometer sensor data to detect behavioral activities related to the head and mouth. HeadFi [28] made a circuit change for cheap earphones. This makes unintelligent earphones smart, which greatly reduces the cost of potential smart earphones in the future. EarEcho [20] used unique acoustic signals in the ear canal to extend earphone to identity authentication devices. EarDynamic [29] leverages ear canal deformation that combines the unique static geometry and dynamic motions of the ear canal when the user is speaking for authentication. EarphoneTrack [30] proposed a new mode of acoustic motion tracking using earphones. Earphone-based tracking alleviates the limitations associated with traditional smartphone-based tracking.

VIII. Conclusion

In this paper, We recommend using earphones to detect middle ear effusion and extend this function to the health monitoring function of smart earphones. EarSonar uses earphone's internal speakers and requires an additional embedded microphone (low cost), no special sensor is required. Specifically, EarSonar exploit the inbuilt microphone and speaker of COTS earphones to capture acoustic echos that retain the traits of middle ear fluid status, therefore, achieve full-featured MEE diagnosis. We implement the prototype of EarSonar and conduct extensive experiments in clinical diagnosis. The results show that EarSonar can effectively diagnose MEE at

TABLE II: Latency of EarSonar for different operation.

Operation	Latency(ms)
Band-pass Filter	1.32
Feature Extract	35.89
Inference	1.2

an accuracy up to 92.8% and it is 8% higher than the previous method based on acoustic detection of MEE.

IX. Acknowledgments

The authors thank the anonymous reviewers for valuable and constructive comments. This study is partly supported by NSFC under Grant U20A20181 and 62272152, The National Key Research and Development Program of China under Grant 2022YFE0137700, the Science and Technology Innovation Program of Hunan Province under Grant 2021RC4023, the Key Research and Development Program of Hunan Province under Grant 2021WK2001 and 2022GK2020, in part by the Hunan Natural Science Foundation of China under Grant 2022JJ30171, in part by the Open Research Fund from Guangdong Laboratory of Artificial Intelligence and Digital Economy [Shenzhen (SZ)] under Grants GML-KF-22-22 and GML-KF-22-23. The corresponding author of this paper is Hongbo Jiang.

References

[1] "Middle ear effusion," https://www.texaschildrens.org/departments/ear-nose-and-throat-otolaryngology/conditions-we-treat/fluid-ear-middle-ear-effusion, 2021.

[2] E. M. Sarrell, A. Mandelberg, and H. A. Cohen, "Efficacy of naturopathic extracts in the management of ear pain associated with acute otitis media," *Archives of pediatrics & adolescent medicine*, vol. 155, no. 7, pp. 796–799, 2001.

[3] M. E. Ravicz, J. J. Rosowski, and S. N. Merchant, "Mechanisms of hearing loss resulting from middle-ear fluid," *Hearing research*, vol. 195, no. 1-2, pp. 103–130, 2004.

[4] I. Williamson, "Otitis media with effusion in children," *BMJ clinical evidence*, vol. 2015, 2015.

[5] R. J. Nozza, C. D. Bluestone, D. Kardatzke, and R. Bachman, "Identification of middle ear effusion by aural acoustic admittance and otoscopy," *Foundations of pediatric audiology*, pp. 195–209, 2006.

[6] N. Shahnaz and L. Polka, "Standard and multifrequency tympanometry in normal and otosclerotic ears," *Ear and hearing*, vol. 18, no. 4, pp. 326–341, 1997.

[7] P. K. Harris, K. M. Hutchinson, and J. Moravec, "The use of tympanometry and pneumatic otoscopy for predicting middle ear disease," 2005.

[8] R. M. Rosenfeld, L. Culpepper, K. J. Doyle, K. M. Grundfast, A. Hoberman, M. A. Kenna, A. S. Lieberthal, M. Mahoney, R. A. Wahl, and W. C. Jr, "Clinical practice guideline: Otitis media with effusion." *American Family Physician*, vol. 130, no. 5, pp. S95–S118, 2004.

[9] Y. Jin, Y. Gao, X. Guo, J. Wen, Z. Li, and Z. Jin, "Earhealth: an earphone-based acoustic otoscope for detection of multiple ear diseases in daily life," in *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*, 2022, pp. 397–408.

[10] J. Chan, S. Raju, R. Nandakumar, R. Bly, and S. Gollakota, "Detecting middle ear fluid using smartphones," *Science translational medicine*, vol. 11, no. 492, 2019.

[11] J. B. Allen, "Measurement of eardrum acoustic impedance," in *Peripheral auditory mechanisms*, 1986, pp. 44–51.

[12] J. Mohd, Z. Rozli, A. Nowshad, H. F. Mohammad *et al.*, "Effect of different factors on the acoustic absorption of coir fiber." *Journal of Applied Sciences*, vol. 10, no. 22, pp. 2887–2892, 2010.

[13] Ludwig and D. George, "The velocity of sound through tissues and the acoustic impedance of tissues," *Journal of the Acoustical Society of America*, vol. 22, no. 6, p. 862, 1950.

TABLE III: Power consumption of EarSonar for different smartphone.

Smartphone	Power(mW)
Huawei	2100
Galaxy	2120
MI 10	2243

[14] Rozanov, Konstantin, and N., "Ultimate thickness to bandwidth ratio of radar absorbers." *IEEE Transactions on Antennas Propagation*, vol. 48, no. 8, pp. 1230–1230, 2000.

[15] M. H. Fouladi, M. J. M. Nor, M. Ayub, and Z. A. Leman, "Utilization of coir fiber in multilayer acoustic absorption panel," *Applied Acoustics*, vol. 71, no. 3, pp. 241–249, 2010.

[16] Keefe and H. Douglas, "Ear-canal impedance and reflection coefficient in human infants and adults." *Journal of the Acoustical Society of America*, vol. 94, no. 5, pp. 2617–38, 1993.

[17] E. C. Ifeachor and B. W. Jervis, "Digital signal processing: a practical approach," *Pearson Education*, 2002.

[18] A. Gnutti, F. Guerrini, and R. Leonardi, "Representation of signals by local symmetry decomposition," in *Signal Processing Conference*, 2015.

[19] C. C. Novak, J. La Lopa, and R. E. Novak, "Effects of sound pressure levels and sensitivity to noise on mood and behavioral intent in a controlled fine dining restaurant environment," *Journal of Culinary Science & Technology*, vol. 8, no. 4, pp. 191–218, 2010.

[20] Y. Gao, W. Wang, V. V. Phoha, W. Sun, and Z. Jin, "Earecho: Using ear canal echo for wearable authentication," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 3, pp. 1–24, 2019.

[21] J. Song and K. Hynynen, "Accurate assessment of middle ear effusion by monitoring ultrasound reflections from a tympanic membrane," *2009 IEEE International Ultrasonics Symposium*, pp. 193–195, 2009.

[22] N. Ozana, R. Califa, A. Schwarz, N. Lipschitz-Tayar, M. Wolf, and Z. Zalevsky, "Remote optical sensor for detection of middle ear effusion," in *2017 Conference on Lasers and Electro-Optics Europe European Quantum Electronics Conference (CLEO/Europe-EQEC)*, 2017, pp. 1–1.

[23] O. Clade, G. Palczewska, J. J. Lewandowski, P. Krakovitz, and D. Dinet, "P3q-2 development and evaluation of a 20mhz array for ultrasonic detection of middle ear effusion," in *2006 IEEE Ultrasonics Symposium*, 2006, pp. 2357–2360.

[24] J. Won, W. Hong, P. Khampang, D. R. Spillman, and S. A. Boppart, "Longitudinal optical coherence tomography to visualize the in vivo response of middle ear biofilms to antibiotic therapy," *Scientific Reports*, vol. 11, no. 1, 2021.

[25] F. Kawsar, C. Min, A. Mathur, and A. Montanari, "Earables for personal-scale behavior analytics," *IEEE Pervasive Computing*, vol. 17, no. 3, pp. 83–89, 2018.

[26] C. Min, A. Montanari, A. Mathur, S. Lee, and F. Kawsar, "Cross-modal approach for conversational well-being monitoring with multi-sensory earables," in *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, 2018, pp. 706–709.

[27] T. Hossain, M. S. Islam, M. A. R. Ahad, and S. Inoue, "Human activity recognition using earable device," in *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, 2019, pp. 81–84.

[28] X. Fan, L. Shangguan, S. Rupavatharam, Y. Zhang, J. Xiong, Y. Ma, and R. Howard, "Headfi: Bringing intelligence to all headphones," in *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '21, 2021, p. 147159.

[29] Z. Wang, S. Tan, L. Zhang, Y. Ren, Z. Wang, and J. Yang, "An ear canal deformation based continuous user authentication using earables," in *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '21, 2021, p. 819821.

[30] G. Cao, K. Yuan, J. Xiong, P. Yang, Y. Yan, H. Zhou, and X.-Y. Li, "Earphonetrack: Involving earphones into the ecosystem of acoustic motion tracking," in *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, ser. SenSys '20, 2020, p. 95108.