

## Detecting middle ear fluid using smartphones

**Authors:** Justin Chan,<sup>1\*†</sup> Sharat Raju,<sup>2,3\*†</sup> Rajalakshmi Nandakumar,<sup>1</sup> Randall Bly<sup>2,3</sup> and Shyamnath Gollakota<sup>1†</sup>

### Affiliations:

<sup>1</sup>Paul G. Allen School of Computer Science and Engineering, University of Washington, WA, (98195)

<sup>2</sup>Department of Otolaryngology – Head and Neck Surgery, University of Washington, WA, (98195)

<sup>3</sup>Seattle Children's Hospital and Research Institute, Seattle, WA, (98105)

\* Equal contribution

†Corresponding authors. Email: jucha@uw.edu (J.C.); sraju@uw.edu (S.R.); gshyam@uw.edu (S.G.)

**One-sentence summary:** To diagnose otitis media, the authors detect the presence of middle ear fluid using the built-in microphone and speaker of smartphones.

**Abstract:** The presence of middle ear fluid is a key diagnostic marker for two of the most common pediatric ear diseases: acute otitis media and otitis media with effusion. We present an accessible solution that uses speakers and microphones within existing smartphones to detect middle ear fluid by assessing eardrum mobility. We conducted a clinical study on 98 patient ears at a pediatric surgical center. Using leave-one-out cross-validation to estimate performance on unseen data, we obtained an area under the curve (AUC) of 0.898 for the smartphone-based machine learning algorithm. In comparison, commercial acoustic reflectometry, which requires custom hardware, achieved an AUC of 0.776. Further, we achieved 85% sensitivity and 82% specificity, comparable to published performance measures for tympanometry and pneumatic otoscopy. Similar results were obtained when testing across multiple smartphone platforms. Parents of pediatric patients ( $n = 25$  ears) demonstrated similar performance to trained clinicians when using the smartphone-based system. These results demonstrate the potential for a smartphone to be a low-barrier and effective screening tool for detecting the presence of middle ear fluid.

### Main text:

#### Introduction

The presence of middle ear fluid is the key diagnostic marker for the two most common pediatric ear diseases, acute otitis media and otitis media with effusion (1). Acute otitis media (AOM), known commonly as an “ear infection,” is characterized by the presence of infected fluid in the middle ear and results in symptoms of fever and ear pain. It is a leading cause of pediatric

healthcare visits and, although many cases can resolve without antibiotics, complications may include eardrum perforation, mastoiditis, facial nerve palsy, or meningitis (2–4). Otitis media with effusion (OME) is the presence of middle ear fluid without signs of an acute infection and affects up to 80% of children (5, 6). Although OME has few overt symptoms, making diagnosis more difficult, it is associated with speech delay, sleep disruption, poor school performance, balance issues, and a higher likelihood of developing AOM (5).

Diagnosis of OME or AOM requires detecting middle ear fluid using either pneumatic otoscopy or tympanometry (1). Pneumatic otoscopy is used by only 7–33% of primary care providers and is not designed for home screening purposes (5). Tympanometry necessitates a referral to an audiologist and the use of expensive equipment (7,8). For these reasons, in 2016 the American Academy of Otolaryngology called for research into a brief, reliable, and objective method to detect middle ear fluid as well as new in-home strategies to help parents and caregivers monitor fluid after initial physician evaluation (5).

Here, we describe a system that uses the microphone and speaker of existing smartphones to detect middle ear fluid by assessing eardrum mobility. The system sends a soft acoustic chirp into the ear canal using the smartphone speaker, detects reflected sound from the eardrum using the smartphone microphone, and uses a logistic regression machine learning model to classify these reflections and predict middle ear fluid status. No additional attachments are required beyond a paper funnel, which acts as a speculum and can be constructed from printer paper, scissors, and tape. Real-time implementation and data processing are performed entirely on the smartphone, compatible with both iPhone and Android devices. The system demonstrated comparable performance across multiple smartphone platforms and when used by parents versus clinicians. Given the ubiquity of smartphones, this system may hold potential as a middle ear screening tool for parents as well as health care providers in resource-limited regions.

## Results

### *Concept and prototype*

Our system uses the smartphone speaker to play audible, 150 ms frequency-modulated continuous wave (FMCW) chirps from 1.8 – 4.4 kHz into the patient's ear canal. The microphone remains active during the chirp, collecting both incident waves from the speaker and reflected waves from the eardrum. Sound reflected from the eardrum will destructively interfere with the incident chirp, causing a dip in sound pressure along a range of frequencies. A normal eardrum resonates well at multiple sound frequencies, creating a broad-spectrum, soft echo; as a result, the shape of the resulting acoustic dip is broad and shallow in the frequency domain. In contrast, a fluid or pus-filled middle ear, as found in OME and AOM, restricts the vibrational capacity of the eardrum; sound energy that would have vibrated the eardrum is instead reflected back along the ear canal, creating more destructive interference and resulting in a narrower and deeper 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 (9). Thus, although individual differences in ear canal length affect the location of the dip along the frequency domain, the shape of the dip primarily depends on eardrum mobility.

Our system builds upon existing acoustic reflectometry methods in three ways (10–14). First, it is a predominantly software-based solution that takes advantage of existing smartphone hardware rather than requiring a separate device. Current acoustic reflectometers require a microphone and speaker in close proximity to produce and measure sound waves along the ear canal. Many modern smartphones have a similar configuration, with a co-located speaker and microphone on their bottom edge for noise cancellation (Fig. 1A). This includes all versions of iPhone, Samsung Galaxy phones after the S5, and other Android phones, including the Google Pixel. Second, we use a paper funnel as a speculum to direct sound into the ear canal. The funnel (Fig. 1B) can be assembled using a printed paper template (fig. S1), scissors, and tape. Without this attachment, the resulting waveform can be highly variable because sound could reflect off of different structures of the pinnae. Third, the system uses a logistic regression machine learning model to classify the waveforms received by the microphone. To identify whether a patient has middle ear fluid, we first preprocessed the raw waveform to locate and isolate the acoustic dip (Fig. 1C, D). We then used logistic regression to determine if the shape of the dip was more indicative of a normal or fluid-filled ear. A text-based message is presented to the user indicating a result: ‘suggestive of middle ear fluid’ or ‘middle ear fluid unlikely’ (fig. S2). On an iPhone 5s and Galaxy S6, data processing and classification took 771.98 ms and 1.2 s, respectively.

#### *Clinical testing*

We tested system performance for detecting middle ear fluid in two separate cohorts. First, we conducted a clinical study on patients between 18 months and 17 years of age. We used this population to train the algorithm and obtain cross-validated performance measures. Second, we recruited a separate cohort of patients under 18 months of age and evaluated performance using the algorithm trained in the first clinical study.

The first clinical study was conducted at Seattle Children’s Hospital surgical centers using a cohort of 98 patient ears between 18 months and 17 years of age from two different subgroups: patients undergoing ear tube placement, a common surgery performed on patients with chronic OME or recurrent AOM ( $n = 48$  ears); and patients undergoing a different surgery, such as a tonsillectomy, with no recent symptoms of AOM or OME and no signs of middle ear fluid by physical examination ( $n = 50$  ears). The median age of recruited patients was 5.0 [interquartile range (IQR): 2.0] years, height was 113.2 (IQR: 19.0) cm, weight was 20.0 (IQR: 9.2) kg, and the female-to-male ratio was 0.6 (table S1).

A trained clinician performed all patient testing in a private waiting room just prior to surgery and with the patient awake and held or sitting upright (Movie S1). Soft chirps were played into

the ear canal using multiple smartphone models and a new paper funnel for each patient. All patients were also tested in parallel with commercial acoustic reflectometry hardware (7, 12, 15). After surgery, we prospectively assigned each ear its actual middle ear fluid status. A patient was considered positive for middle ear fluid if, during ear tube placement, an incision into the eardrum (myringotomy) yielded fluid ( $n = 24$ ), or if the patient had a red, bulging eardrum consistent with AOM ( $n = 2$ ). A patient was considered negative for middle ear fluid if, during ear tube placement, myringotomy yielded no fluid ( $n = 24$ ), or if the patient did not receive ear tubes, did not have ear-related symptoms, and was negative for fluid on pneumatic otoscopy performed by the otolaryngologist ( $n = 48$ ).

For classification, we used a logistic regression algorithm on preprocessed microphone acoustic data. The sound intensity (in decibels) of each frequency along the acoustic dip was inputted as a separate feature. The algorithm was trained with iPhone 5s data collected from patients. Its classification accuracy was evaluated with leave-one-out cross validation (LOOCV), a rigorous method to validate machine learning models (16). During each iteration of LOOCV, 97 of 98 patient ears are used to train a model which is then used to output a prediction for the remaining 1 patient ear. This process is repeated for all 98 ears to estimate the accuracy of a model trained on all 98 ears when tested on unseen data. A receiver-operating characteristic (ROC) curve was generated from the cross-validation step with an area under the curve (AUC) of 0.898 (Fig. 2A). The operating point was chosen to have an overall sensitivity and specificity of 84.6% (95% CI: 65.1 - 95.6%) and 81.9% (95% CI: 71.1- 90.0%), respectively. With  $k$ -fold ( $k = 10$ ) cross-validation (17), we obtained a comparable AUC of 0.906. To address potential bias from training on the same patient's opposite ear, we repeated LOOCV but during each iteration we also excluded the contralateral ear from the training set, achieving an AUC of 0.899. We also downsampled the frequency response curve to 100 samples and obtained a similar AUC of 0.888. The fluid type was recorded as either being serous ( $n = 7$ ), mucoid ( $n = 11$ ), or purulent ( $n = 4$ ) for 22 of the 36 ears that had middle ear fluid. The algorithm correctly classified 86% (6 of 7) of ears that had serous fluid, 91% (10 of 11) of ears with mucoid fluid, and 100% (4 of 4) of ears with purulent fluid. These estimates predict the real-world clinical performance on unseen data of our final algorithm, which is trained on all 98 ears from the iPhone 5s dataset (Fig. 2B). Across male patients, for the iPhone 5s, 15 out of 17 positive ears and 31 out of 40 negative ears were classified correctly. Across female patients 7 out of 9 positive ears and 25 out of 30 negative ears were classified correctly. For two ears, we did not record gender.

Post hoc, we examined how the algorithm classified acoustic waveforms. Fig. 2C and 2D plot the mean sound intensities at each frequency for all ears classified by the model. The algorithm predicted that ears with narrower and deeper acoustic dips were more likely to have middle ear fluid. Similarly, on univariate analysis, sound intensities at the top and bottom of the waveform, which determine the depth of an acoustic dip, were given the most weight by the predictive model (Fig. 2E). This result indicates that the algorithm can independently identify an acoustic

pattern for middle ear fluid that is consistent with the known acoustic response of the eardrum (7, 12, 15).

The smartphone-based system demonstrated improved clinical performance compared to acoustic reflectometry (18), which uses custom hardware to assess middle ear fluid status. Head-to-head testing (Fig. 2B) across the 98 patient ears demonstrated an AUC of 0.898 for the smartphone-based approach compared to an AUC of 0.776 for commercial acoustic reflectometry (EarCheck Middle Ear Monitor, Innova Medical). The smartphone algorithm's improved clinical performance may be the result of applying machine learning over the waveform rather than relying on the hand- selected features used by acoustic reflectometers (7). When classifying patient waveforms obtained from smartphones, we found that using only the spectral-angle, as described in prior literature (7), reduces the AUC to 0.687.

We evaluated test-retest reliability in the pediatric patients enrolled in our clinical study. Each ear was tested twice per smartphone; between each attempt, the funnel was fully removed from the ear and reinserted. Of the 66 ears tested twice, 94% of the ears were classified the same between each attempt. When a discrepancy occurred, the algorithm used the positive result to minimize false negatives. Each testing attempt consisted of ten chirps and we tested the consistency across these chirps. In the clinical study, 93 of 98 ears showed no difference in classification among all ten chirps. When doing a majority vote across the first three chirps, 96 of 98 ears showed no difference in classification compared to using a single chirp. Across all 98 ears, there was no difference when considering the classification result to be the majority (more than five) of the ten chirps (table S2).

Finally, using the algorithm trained in the first study, we evaluated the system's performance in a separate cohort of patients under 18 months of age to assess accuracy in a younger population. We again recruited surgical patients at Seattle Children's Hospital, using the same criteria described in the first study with the exception of age. This cohort included 15 patient ears and had a median age of 1.1 (IQR: 0.3) years, height of 76.0 (IQR: 7.8) cm, weight of 9.3 (IQR: 2.1) kg, and female-to-male ratio of 1.5 (Fig. 3A). The lowest age amongst this cohort was 9 months. All 5 ears that were positive for fluid and 9 of 10 ears that were negative for fluid were classified correctly (Fig. 3B). The shape of the acoustic dips paralleled those in the first clinical study: ears with fluid had a deeper and narrower acoustic dip compared to ears without fluid (Fig. 3C, D and fig. S3). This shows that an algorithm trained on patients over 18 months of age can properly classify patients under 18 months. We also trained and tested our algorithm's performance when using patients under 18 months as the training cohort. When running LOOCV across the 15 patient ears that were under 18 months of age, 5 of 5 ears positive for fluid and 10 of 10 ears negative for fluid were correctly classified. This is similar to the performance of the algorithm that is trained on the 98 patient ears over 18 months of age.

### *Performance across other mobile platforms*

All patients in the first cohort ( $n = 98$  ears) were tested in parallel with both the iPhone 5s and the Samsung Galaxy S6. Using LOOCV, we estimated performance of the iPhone 5s-trained system on unseen Galaxy S6 data. Specifically, the entire iPhone 5s dataset was used for training except for one patient ear, which was “held out” for testing. The trained algorithm was then tested on Galaxy S6 data from the held out ear. This was repeated for all patient ears in the cohort to generate an AUC of 0.851, as shown in Fig. 4A. In the same manner, we also tested a subset of this cohort using an iPhone 6s ( $n = 10$  ears), Samsung Galaxy S7 ( $n = 12$ ), and Google Pixel ( $n = 8$ ). The algorithm correctly classified 80% (8 of 10) of iPhone 6s data, 91.7% (11 of 12) of Galaxy S7 data, and 83.3% (7 of 8) of Pixel data (Fig. 4B). The low sample size in these subgroups precluded generation of meaningful AUC values. Processed waveforms for a given test ear across phone models is shown in fig. S4, waveforms for the remaining test ears are shown in fig. S5.

### *Performance testing with non-clinicians*

In a clinical setting, we evaluated the system’s performance when used by parents. Trained clinicians briefly demonstrated proper technique for testing, and the parent of a pediatric study participant subsequently performed unaided testing on their child. The parent’s results were then compared to those of the trained clinician. This cohort included 25 patient ears and had a median age of 4.0 (IQR: 6.0) years, height of 105.0 (IQR: 38.1) cm, weight of 16.4 (IQR: 13.9) kg, and female-to-male ratio of 1.1 (Fig. 5A). All 6 ears positive for fluid were classified the same by clinicians and parents, and 18 of 19 ears negative for fluid were classified the same (Fig. 5B). Additionally, the mean acoustic dip was similar between clinicians (red) and parents (black) (Fig. 5C, D). Individual curves for each patient are shown in fig. S6.

We tested the usability of funnel construction with a separate cohort of ten untrained adults. After playing a short instructional video (see Movie S2), we first measured the time it took participants to create and mount the funnel using a paper template, tape, and scissors. The average time was 2.8 ( $\pm 0.93$ ) minutes. We then queried participants about the usability of the entire system; they gave an average usability rating of 8.9 ( $\pm 1.1$ ) on a scale of 1 (unusable) to 10 (extremely usable) (table S3).

### *Effect of confounding ear pathologies*

In the above studies, we exclude patients with ear pathologies that affect eardrum mobility. Next, we evaluated the algorithm’s performance in the presence of ear pathologies such as cholesteatoma ( $n = 1$ ), ossicular chain discontinuity ( $n = 1$ ), acute eardrum inflammation ( $n = 1$ ) and prior tympanoplasty surgery ( $n = 3$ ) (fig. S7). The algorithm produced false positives for middle ear fluid in all these patients.

Similarly, patients undergoing myringotomy but lacking middle ear fluid may have abnormal middle ear pressure which can affect eardrum mobility. In our cohort, there were 7 patients reported by the surgeon as having acutely inflamed eardrums. Only 1 of these patients presented without fluid on myringotomy. This patient's ear was classified as positive by the algorithm (fig. S7). Thus, in the event that a patient presents with an inflamed eardrum but has not yet developed middle ear fluid, the algorithm would likely test positive and appropriately prompt further evaluation. The other acutely inflamed eardrums ( $n = 6$  ears) had middle ear fluid and were appropriately classified as positive by the algorithm.

#### *Benchmark testing*

Fig. 6 demonstrates benchmark performance of our smartphone-based system across various design and environmental conditions. We identified design and environmental conditions that could impact system accuracy, including background noise, incident angle of the smartphone, alterations to the funnel, and changes in chirp volume. Testing was performed with a 2.5 cm closed-ended plastic tube used to calibrate existing acoustic reflectometers as a positive control (33). The hard-backed end of the tube reflects sound to produce a narrow and deep acoustic dip that mimics an ear with middle ear effusion.

First, we used 4 different paper types of varying thickness and consistency to construct the funnel: filler paper, inkjet paper, laserjet paper, and cardstock. Changing paper type did not affect the classification accuracy (Fig. 6A). We also tested whether changes to tip opening diameter impacted accuracy. The funnel is designed to have a tip opening diameter of 7 mm to approximate the diameter of the ear canal (32). Variations in tip diameter from 5 mm to 10 mm did not affect performance. However, diameters of 1 mm and 3 mm produced false negatives (Fig. 6B). This suggests that a tip opening diameter between 5 and 10 mm is required.

Second, we varied the incident angle of our smartphone with respect to the calibration tube to examine the system's performance with slight deviations from direct line of sight. The smartphone tolerated up to a 45° offset from line of sight (Fig. 6C). Offsets of 60° or 75° produced a less prominent dip and false negatives. This suggests that while a parallel orientation is ideal, the system has some tolerance for non-ideal positioning. To validate the benchmark testing, we evaluated the effect of angle of insertion in an upright patient (16 months of age) with middle ear fluid confirmed on myringotomy. To accurately assess angle of insertion, we used the built-in smartphone gyroscope to measure smartphone rotation. Initially, the smartphone was placed in line with the axis of the ear canal and began playing and recording chirps. Angular data was recorded for each chirp while the phone was rotated up (positive) and down (negative), up to 30 degrees off axis. All chirps were correctly classified as positive within this range (fig. S8). Different insertion angles are also accounted for during clinical testing, where there was natural variance in measurement angle.

Third, we examined whether changes in background sound affected device accuracy, particularly with a crying child. We used an external speaker to play an audio file of a baby crying, with an average volume from 80 – 110 dBA. We tested our system in the calibration tube when it was placed directly next to the speaker (within 2 cm). In a measurement attempt where five chirps were played, the majority of chirps were correctly classified across tested volume levels during three different measurement attempts (Fig. 6D). When the background volume was lower than 80 dBA, all chirps were correctly classified. In the clinical study, there was no significant background noise. We report data on a two year old patient ear who was crying and had partial motion of the head. fig. S9 shows the mean and the standard deviation of the processed chirps used by our algorithm; this validates our benchmark testing on the effect of environmental noise.

Fourth, we tested the effect of deforming the funnel on classification of waveforms when using the calibration tube as a positive control. No deformation and partial deformation (more severe than typical use), as demonstrated in fig. S10, were appropriately classified as positive. Full deformation resulted in a false negative. Our clinical study had variance in funnel deformation among different users and ears, and thus the clinical results account for slight variations during actual use. We also tested the effect of different funnel instances on classification accuracy. Five different untrained users were instructed to construct a funnel and test it on the positive control. All acoustic curves were correctly classified as a positive ear (Fig. 6E). We also varied the sound intensity of chirps from 55 to 68 dBA. We found no difference in classification of the positive control (Fig. 6F).

Finally we consider the presence of cerumen (ear wax) and its effects on system performance. Our patients had partial cerumen occlusion (range: 0% - 50%) as estimated by the surgeon; our data indicates that this did not impair the algorithm's performance. Since none of our patients had complete cerumen occlusion, we used a positive control calibration tube for further testing. As expected, playing chirps into the tube generated a deep and narrow acoustic dip. Using putty to mimic cerumen, we found that partially occluding wax (60–70%) had little effect on the shape or position of the dip, as shown in fig. S11A. This is consistent with previous observations that acoustic-based techniques are unaffected by less than 50% cerumen occlusion (35). In contrast, 100% cerumen occlusion, also known as impaction, occurs in 10% of children (36) and causes significant waveform changes. As the site of impaction moves closer to the entrance of the ear canal, the acoustic dip appears shallower and occurs at a higher frequency due to an effectively shorter canal and corresponding quarter-wavelength, shown in fig. S11B. In these cases, chirps can reflect off cerumen, generating a false acoustic dip that does not reflect middle ear status. For example, at a depth of 1 cm, which is the deepest point cerumen would naturally accumulate (37), our tests produced a false acoustic dip located approximately 1 kHz higher compared to a normal dip from eardrum reflections. At shallower depths of impaction, the false dip was even more right-shifted. Impaction at the entrance produced a waveform similar to calibration chirps played into open air (fig. S11C). Given that the mean acoustic dip in our patients was located at 3

kHz (range: 2.4 – 3.7 kHz), a cerumen detection system in future prototypes could include an error display if an acoustic dip is identified outside the normal range or if the waveform resembles an in-air calibration chirp. Such a system must acknowledge that cotton swab insertion or iatrogenic manipulation can result in cerumen impaction deeper than 1 cm. These findings suggest that partial occlusion does not affect results, and full occlusion can be readily identified.

## Discussion

Proper diagnosis of AOM and OME requires an examination of middle ear fluid status (1, 5). Currently, most assessments of middle ear fluid are made either in primary care clinics, urgent care centers, or, in recent years, remotely using smartphone-attached otoscopes (19–22). These assessments can be costly and time-consuming. Further, they usually rely on visual information without an assessment of eardrum mobility, which can compromise accuracy because middle ear fluid often produces only subtle changes in the eardrum’s appearance (23–25). Techniques that assess eardrum mobility, such as pneumatic otoscopy and tympanometry, have high sensitivity and specificity (90% and 80% for specialist-performed pneumatic otoscopy), but are infrequently performed outside of a specialist’s office (26). Other methodologies—such as air-coupled ultrasound, short-wave infrared imaging, and optical coherence tomography—hold promise in terms of accuracy, but they require additional specialized and expensive hardware (27–29). Thus, there is a need for a middle ear fluid screening technique that does not use costly equipment or attachments, can evaluate eardrum mobility, and requires minimal expertise.

The value of a low cost, smartphone-based screening tool lies in its accessibility and user familiarity. Ninety-six percent of all parents we queried regarding potential participation in the study consented, and many were interested in learning more about the technology. Although some children were apprehensive before the study, the phone’s chirps (which sound like a small bird) had a calming effect, causing many children to respond with smiles or laughs.

Our system has several limitations. As with many screening tools, interpretation of the results requires appropriate clinical context such as symptoms and time-course, and positive results should prompt further clinical evaluation for potential misclassifications. This system also does not distinguish between different types of middle ear fluid (purulent, serous, or mucoid). Knowing fluid type could potentially be useful for identifying AOM versus OME, though this distinction is also made based on clinical history and symptoms. Furthermore, as with most middle ear assessment techniques, the system requires that a child not be agitated and remain relatively still for the duration of testing: the algorithm needs a minimum of three chirps for reliable results, which takes 1.2 seconds; significant head movement during this duration can cause inter-chirp inconsistency.

This smartphone-based screening tool relies on an evaluation of eardrum mobility to detect middle ear fluid. This is also true for tympanometry and pneumatic otoscopy, which are the

screening techniques recommended by the American Academy of Pediatrics and American Academy of Otolaryngology for middle ear fluid detection. Ear pathologies that affect eardrum mobility – such as cholesteatoma, ossicular chain discontinuity, acute eardrum inflammation and prior tympanoplasty surgery – can produce false positives for middle ear fluid. Similarly, patients undergoing myringotomy but lacking middle ear fluid may have abnormal middle ear pressure which can affect eardrum mobility. Despite the potential for false positives, use of eardrum mobility to predict middle ear effusion is well established in both active clinical practice and in prior large scale pediatric studies (30). Further, in the context of screening, these false positives would appropriately prompt additional evaluation.

In summary, we present a proof-of-concept screening tool that can be implemented on commodity smartphones to determine the presence of middle ear fluid. Given the ubiquity of smartphones, the system we describe may have clinically significant applications in developing countries and rural communities where smartphone availability is rapidly growing; in primary care settings as an adjunct to visual otoscopy; or for home-screening by parents as a platform to reduce health care costs. Further longitudinal clinical trials are required to determine the technology's impact in these and other potential scenarios.

## **Materials and methods**

### *Study design*

Our study was approved by the University of Washington and the Seattle Children's Hospital Institutional Review Boards. For our first clinical study ( $n = 98$  ears), we included otolaryngology patients undergoing surgery who were between the ages of 18 months and 17 years. Our second clinical study ( $n = 15$  ears) included otolaryngology surgical patients who were between 9 and 18 months of age. For both studies, we excluded patients with existing tympanostomy tubes, existing eardrum perforations, prior tympanoplasty, or known comorbid middle ear pathology, such as cholesteatoma or ossicular chain abnormalities. We stopped recruitment after a sufficient number of patient data was collected to demonstrate proof of concept. Randomization was not applicable. Surgeons were blinded to the results of preoperative smartphone testing.

### *Smartphone application*

Two custom applications — for iPhone and Android smartphones — were developed to emit chirps with a frequency range of 1.8 kHz to 4.4 kHz, each for a duration of 150 ms. Each chirp was interspersed with 250 ms of silence. The smartphone simultaneously recorded audio from the microphone at a sampling rate of 48 kHz, the highest sampling rate possible on the smartphones tested.

### *Data preprocessing*

Our processing pipeline was implemented on an iPhone 5s and a Samsung Galaxy S6. Each chirp had 7200 samples and was padded with trailing zeros so the total number of samples was 48,000. A 48,000 point Fast Fourier Transform (FFT) was then performed to capture the acoustic frequency response within a range of 0 – 24 kHz; we discarded frequencies outside the 1.8 – 4.4 kHz range of the transmitted chirp. This frequency range was chosen based on prior literature in acoustic reflectometry (31). Except when evaluating re-test and interchirp reliability, waveforms that were two or more standard deviations from the mean of other recorded chirp waveforms within a given attempt were excluded from analysis. Of the remaining chirps, we selected the second chirp for further processing. To reduce the variability caused by different funnels as well as microphone and speaker differences across devices, a calibration chirp was performed away from the ear and normalized to a unit frequency response to produce a set of weights. These weights were then used to normalize chirps captured in a subject's ear canal.

We then applied a moving average filter with a window size of 300 samples to smooth the waveform. Next, we used a peak-detection algorithm to identify the acoustic dip. Specifically, the algorithm identified the most prominent dip within a range of 2.3 – 3.8 kHz. Looking for dips within the full range of 1.8 – 4.4 kHz resulted in no change in AUC for the iPhone 5s. After the dip was identified, we selected frequencies within 500 Hz of the dip for further processing. This lets the machine learning algorithm focus only on portions of the acoustic response that are most predictive of middle ear effusion status. Selecting frequencies within 600 Hz, 700 Hz, 800 Hz, 900 Hz and 1000 Hz of the dip caused small ( $< 0.02$ ) AUC changes for the iPhone 5s.

#### *Machine learning classifier*

Our logistic regression classifier is computationally inexpensive and can run inferences on mobile devices. The frequency response for each ear was represented as an array of 1000 floating point values, where each element represents the amplitude for each of the 1000 selected frequencies around the acoustic dip. Each chirp was aggregated into a single matrix. Using a logistic regression machine learning model with an L2 penalty, we performed leave-one-out cross validation to validate the discriminative accuracy of the algorithm on our dataset of 98 ears collected with the iPhone 5s. We also performed  $k$ -fold ( $k = 10$ ) cross validation.

The classifier trained the model on the entire set of ears except for one. Testing was then performed on the omitted ear. This was repeated for all ears, and the overall accuracy was computed across all predictions. When computing predictions for the Samsung Galaxy S6 and other benchmark phones, we followed a similar LOOCV approach. We trained the algorithm on all ears from the iPhone 5s dataset except for one ear collected on the other device. The classifier then made a prediction on the omitted ear. This method evaluated whether the classifier can generalize to unseen ears on other devices.

The feature analysis in Fig. 3E was generated in sklearn using the SelectPercentile method, which calculates the ANOVA F-value between each feature and the ground truth label. The spectral gradient angle of the acoustic dip in Fig. 3B were generated using reference points 200 Hz around the dip. The angles were trained and validated using LOOCV using a logistic regression classifier to obtain an AUC. Briefly, spectral gradient angle is the measurement used by existing acoustic reflectometers to determine the presence of middle ear fluid. It is computed as the angle between the slopes of the acoustic dip.

#### *Funnel design*

The funnel was fabricated using printer paper and clear tape. A template for the funnel was designed using Adobe Photoshop and printed onto paper (fig. S1). The funnel was placed over a smartphone's speaker and microphone, and a small piece of tape attached it to the front and back. Each smartphone model required a customized funnel template optimized to envelope the speaker and microphone. The base of the funnel cone varied in size depending on the combined length and separation of the speaker and microphone. The funnel's base had a diameter of 45 mm for the Samsung Galaxy S6 and S7, 52 mm for the iPhone 5s, and 57 mm for the iPhone 6s and Google Pixel. For each template, the funnel was designed with a 7 mm diameter opening, which approximates the size of the opening into the ear canal (32). Not every smartphone has ideal positioning of the microphone and speaker. To increase accessibility, we began prototyping a detection system using earbud headphones with a paper funnel (fig. S12). Further patient studies are needed to validate this approach.

#### *Clinical study design*

We tested each patient's ear using a smartphone fitted with a paper funnel. Each phone had an installed application that played 10 identical chirps and simultaneously recorded the reflected echoes. We standardized the volumes for each phone to an average of 65 dBA. After attaching the funnel to our testing device, we played calibration chirps into the air, away from the patient. When the patient arrived, they were positioned upright for testing, either being held on their parent's lap or sitting in a chair.

The phone was placed near the child's ear, with the funnel positioned at the entrance to the ear canal. The ideal position for the funnel in most children was medial to the tragus, pointing medially and slightly anteriorly into the ear canal. In most cases, the canal was also straightened during testing by gently pulling the pinnae posteriorly. After playing the first set of chirps, the smartphone was withdrawn from the ear and repositioned in approximately the same location to produce a second set of chirps for reliability testing. This process was repeated for each phone and each ear. After device testing, we tested each ear with an FDA-approved acoustic reflectometry instrument (18, 33) and documented the ordinal output score on a scale from 1–5. We reclassified these outputs in accordance with prior studies (34): 1 and 2 results were classified as a normal ear, and 3–5 indicated an infected ear. Finally, a subset of patients,

particularly those not undergoing tympanostomy tube placement, underwent otoscopy to provide additional evidence regarding middle ear fluid status.

#### *Performance testing with non-clinicians*

Our usability study ( $n = 25$  ears) included both otolaryngology surgery and clinic patients between 9 months and 17 years of age. As with the other clinical studies, we excluded patients with existing tympanostomy tubes, existing eardrum perforations, prior tympanoplasty, or known comorbid middle ear pathology such as cholesteatoma or ossicular chain abnormalities.

Clinicians demonstrated proper ergonomic positioning and placement of the smartphone system with respect to the ear canal. Proper device use included first pressing the record button, then directing the funnel tip medially and slightly anteriorly at the entrance to the ear canal, and finally pulling the pinnae posteriorly with the opposite hand. Patients were directed to wait until the chirps were finished playing before removing the device. After the clinician performed testing, parents were allowed to experiment with device placement for approximately 30 seconds to 1 minute. Following this period, parents performed actual unaided testing, which was recorded. We performed all usability testing on the Samsung Galaxy S6.

For funnel construction testing, we made an instructional video. Subjects were instructed to watch the video and subsequently construct the funnel, unaided, using the provided paper template, scissors, and tape. After construction and smartphone mounting, subjects were asked to rate the overall usability.

#### *Run-time analysis*

Implementation on an iPhone was performed in Swift, and the Accelerate framework was used to perform FFTs. Android implementation was done in Java, and the JTransforms library was used to perform FFTs. On an iPhone 5s, data processing and classification took 771.98 ms, including 2.71 ms for dip detection, 0.06 ms for logistic regression, and 767.24 ms for 10 FFTs. On a Galaxy S6, the total runtime was 1.2 s, including 4.96 ms for dip detection, 0.67 ms for logistic regression, and 1200 ms for 10 FFTs.

#### *Statistical analysis*

To obtain clinical performance metrics and perform cross-validation, the FFT, filtering and peak detection were performed in Matlab. NumPy and scikit-learn were used to perform logistic regression, ROC analysis, and the calculation of accuracy, sensitivity, specificity and 95% confidence interval values. Line charts were created using matplotlib and seaborn.

### **Supplementary Materials**

Fig. S1. Funnel template for smartphone

Fig. S2. Conceptual diagram of the smartphone based system

Fig. S3. Processed waveforms of patients under 18 months of age

Fig. S4. Processed waveforms obtained from different smartphones for a single ear

Fig. S5. Individual patient waveforms obtained from different smartphones

Fig. S6. Processed waveforms of testing by parents

Fig. S7. Processed waveforms of confounding ear pathologies

Fig. S8. Effect of angle of insertion on system performance in a patient with middle ear fluid

Fig. S9. Processed waveforms for a crying two-year old patient with partial head movement

Fig. S10. Effect of funnel deformation on system performance

Fig. S11. Effect of cerumen on acoustic waveforms

Fig. S12. Design for earbud headphones

Table S1. Demographic summary for first clinical study

Table S2. Interchirp reliability testing

Table S3. Funnel construction times and usability ratings

Movie S1. Video illustrating proper technique for testing

Movie S2. Instructional video for funnel construction

## References

1. A. S. Lieberthal, A. E. Carroll, T. Chonmaitree, T. G. Ganiats, A. Hoberman, M. A. Jackson, M. D. Joffe, D. T. Miller, R. M. Rosenfeld, X. D. Sevilla, R. H. Schwartz, P. A. Thomas, D. E. Tunkel, The diagnosis and management of acute otitis media, *Pediatrics* pp. peds–2012 (2013).
2. D. Bowyer, *Complications of otitis media* (CRC Press, 2015).
3. J. Acuin, Chronic suppurative otitis media: Burden of illness and management options, *World Health Organization* (2004).
4. A. Qureishi, Y. Lee, K. Belfield, J. P. Birchall, M. Daniel, Update on otitis media–prevention and treatment, *Infection and Drug Resistance* **7**, 15 (2014).
5. R. M. Rosenfeld, J. J. Shin, S. R. Schwartz, R. Coggins, L. Gagnon, J. M. Hackell, D. Hoelting, L. L. Hunter, A. W. Kummer, S. C. Payne, D. S. Poe, M. Veling, P. M. Vila, S. A. Walsh, M. D. Corrigan, Clinical practice guideline: otitis media with effusion (update), *Otolaryngology–Head and Neck Surgery* **154**, S1 (2016).
6. I. Williamson, Otitis media with effusion in children, *BMJ Clinical Evidence* **2015** (2015).
7. H. Lindén, H. Teppo, M. Revonta, Spectral gradient acoustic reflectometry in the diagnosis of middle-ear fluid in children, *European Archives of Otorhinolaryngology* **264**, 477 (2007).
8. J. Lous, Why use tympanometry in general practice: A review, *World Journal of Otorhinolaryngology* **5**, 53 (2015).
9. E. Shaw, R. Teranishi, 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**, 240 (1968).

10. D. W. Teele, J. Teele, Detection of middle ear effusion by acoustic reflectometry, *The Journal of Pediatrics* **104**, 832 (1984).
11. R. M. Lampe, M. R. Weir, J. Spier, M. F. Rhodes, Acoustic reflectometry in the detection of middle ear effusion, *Pediatrics* **76**, 75 (1985).
12. J. T. Combs, M. K. Combs, Acoustic reflectometry: spectral analysis and the conductive hearing loss of otitis media, *The Pediatric Infectious Disease Journal* **15**, 683 (1996).
13. J. Chianese, A. Hoberman, J. L. Paradise, D. Colborn, K. Kathleen, R. Diana, E. Howard, M. Kurs-Lasky, Spectral gradient acoustic reflectometry compared with tympanometry in diagnosing middle ear effusion in children aged 6 to 24 months, *Archives of pediatrics & adolescent medicine* **161**, 884 (2007).
14. T. Muderris, A. Yazıcı, S. Bercin, G. Yalçınler, E. Sevil, M. Kırıs, Consumer acoustic reflectometry: accuracy in diagnosis of otitis media with effusion in children, *International Journal of Pediatric Otorhinolaryngology* **77**, 1771 (2013).
15. S. Kimball, Acoustic reflectometry: spectral gradient analysis for improved detection of middle ear effusion in children, *The Pediatric Infectious Disease Journal* **17**, 552 (1998).
16. M. Kearns, D. Ron, Algorithmic stability and sanity-check bounds for leave-one-out cross-validation, *Neural Computation* **11**, 1427 (1999).
17. S. Rose, Machine learning for prediction in electronic health data, *JAMA Internal Medicine* (2018).
18. Earcheck middle ear monitor, Innova Medical, Lenexa, KS (2018).
19. Cellscope, <https://www.cellscope.com/> (2018).
20. hearScope, <https://www.hearxgroup.com/hearscope/> (2018).
21. R. Mandavia, T. Lapa, M. Smith, M. Bhutta, A cross-sectional evaluation of the validity of a smartphone otoscopy device in screening for ear disease in Nepal, *Clinical Otolaryngology* **43**, 31 (2018).
22. N. Erkkola-Anttila, H. Irjala, M. K. Laine, P. Thtinen, E. Lyttyniemi, A. Ruohola, Smartphone otoscopy performed by parents, *Telemedicine and e-Health* (2018).
23. M. U. Shah, M. Sohal, T. A. Valdez, C. R. Grindle, iPhone otoscopes: Currently available, but reliable for tele-otoscopy in the hands of parents?, *International Journal of Pediatric Otorhinolaryngology* **106**, 59 (2018).
24. G. D. Hoople, Otitis media with effusion—a challenge to otolaryngology, *The Laryngoscope* **60**, 315 (1950).
25. M. E. Pichichero, Diagnostic accuracy, tympanocentesis training performance, and antibiotic selection by pediatric residents in management of otitis media, *Pediatrics* **110**, 1064 (2002).
26. A. Palmu, H. Puhakka, T. Rahko, A. K. Takala, Diagnostic value of tympanometry in infants in clinical practice, *International journal of pediatric otorhinolaryngology* **49**, 207 (1999).
27. R. Stoessel, N. Krohn, K. Pfleiderer, G. Busse, Air-coupled ultrasound inspection of various materials, *Ultrasonics* **40**, 159 (2002).

28. R. H. Wilson, K. P. Nadeau, F. B. Jaworski, B. J. Tromberg, A. J. Durkin, Review of shortwave infrared spectroscopy and imaging methods for biological tissue characterization, *Journal of Biomedical Optics* **20**, 030901 (2015).
29. A. F. Fercher, W. Drexler, C. K. Hitzenberger, T. Lasser, Optical coherence tomography principles and applications, *Reports on Progress in Physics* **66**, 239 (2003).
30. P. H. Karma, M. A. Penttilä, M. M. Sipilä, M. J. Kataja, Otoscopic diagnosis of middle ear effusion in acute and non-acute otitis media. i. the value of different otoscopic findings, *International journal of pediatric otorhinolaryngology* **17**, 37 (1989).
31. D. H. Keefe, J. C. Bulen, K. H. Arehart, E. M. Burns, Ear-canal impedance and reflection coefficient in human infants and adults, *The Journal of the Acoustical Society of America* **94**, 2617 (1993).
32. D. D. Mahoney, W. P. Sjursen, W. J. Staab, M. Leedom, One-size-fits-all uni-ear hearing instrument (2006). US Patent 7,092,543.
33. Earcheck middle ear monitor use & care manual, Innova Medical, Lenexa, KS.
34. H. Y. Lee, M. S. Choi, N. S. Park, C. S. Cho, Applicability of spectral gradient acoustic reflectometry (EarCheck): Screening patients who need surgical treatment for chronic otitis media with effusion, *International Journal of Pediatric Otorhinolaryngology* **78**, 107 (2014).
35. S. L. Block, E. Mandel, S. McLinn, M. E. Pichichero, S. Bernstein, S. Kimball, J. Kozikowski, Spectral gradient acoustic reflectometry for the detection of middle ear effusion by pediatricians and parents, *The Pediatric Infectious Disease Journal* **17**, 560 (1998).
36. D. F. McCarter, A. U. Courtney, S. M. Pollart, Cerumen impaction, *American Family Physician* **75** (2007).
37. L. S. Alvord, B. L. Farmer, Anatomy and orientation of the human external ear, *Journal of the American Academy of Audiology* **8**, 383 (1997).

**Acknowledgments.** We thank our participants and their families at Seattle Children's Hospital for their willingness to participate in our study. We are grateful for funding from the National Science Foundation, the National Institute of Health and the Seattle Children's Sie-Hatsukami Research Endowment for support. We would like to acknowledge S. Parikh, J. Perkins, H. Thomas, H. Ou, J. Dahl, D. Horn, and K. Johnson for permitting us to recruit their patients and contributing data to the study. We would like to thank the Department of Pediatric Otolaryngology at the University of Washington, the operating room staff and personnel at Seattle Children's Hospital, especially at Seattle Children's Bellevue Clinic and Surgery Center, for enabling and facilitating patient recruitment. We thank K. Sie, C. Heike, J. Sunshine, E. Bly, S. Kaplan, V. Iyer, A. Wang, H. Shen, E. Wang and S. Ainsworth for their critical and important feedback on the manuscript and methods.

**Author contributions.** SR, JC, RB, and SG designed the experiments; SR and JC conducted the experiments; JC and RN developed software and deployed the algorithm; SR and JC conducted

the analysis; SR, JC, and RN interpreted results; JC generated figures; SR, JC and SG wrote the manuscript; RN and RB edited the manuscript. Conceptualization: SR.

**Competing Interests.** All co-authors are inventors on US provisional patent application no. 62/72,8543, submitted by the University of Washington, which is related to this work. JC, SR, RB and SG have equity stakes in Edus Health, Inc., which is related to the technology presented in this manuscript. SG is a co-founder of Jeeva Wireless, Inc. and Sound Life Sciences, Inc. RB is a consultant for SpiWay, LLC and a co-founder of EigenHealth, Inc.

**Data and material availability.** All data necessary for interpreting the manuscript have been included. Additional information may be requested from the authors.

**Fig. 1. Using a smartphone to detect middle ear fluid.** (A) Location of speaker and microphone on the bottom of an iPhone 5s, with and without paper funnel attached. (B) Process of assembling smartphone funnel. (C) Proper placement of smartphone and funnel at ear canal entrance. (D) Raw acoustic waveform obtained when chirps are played into an ear with middle ear fluid (red) and without fluid (blue). The standard deviation (gray) is computed across 10 chirp instances on a patient's ear.

**Fig. 2. Classification of patient ears from clinical testing.** (A) ROC curve for our middle ear fluid detection algorithm, cross-validated on data collected from patients using an iPhone 5s ( $n = 98$ ) with operating point denoted by the red circle. (B) Comparison of performance for smartphone-based detection, acoustic reflectometry, and spectral angle-only classification during parallel clinical testing ( $n = 98$ ). (C, D) Mean acoustic dip classified by the algorithm as with middle ear fluid (red) and without middle ear fluid (blue). Shaded region represents one standard deviation from the mean. (E) Feature analysis indicating the weight that the classifier places on each frequency around the acoustic dip.

**Fig. 3. Classification of patient ears under 18 months.** (A) Demographic table of patients under 18 months. (B) Confusion matrix of the algorithm's performance for patients under 18 months. (C, D) Mean acoustic dip of ears of patients under 18 months ( $n = 15$ ) classified by the algorithm as with middle ear fluid (red) and without fluid (blue). Shaded region represents one standard deviation from the mean.

**Fig. 4. Classification performance across other mobile platforms.** (A) ROC curve for our middle ear fluid detection algorithm, cross-validated on data collected from patients using an Samsung Galaxy S6 ( $n = 98$ ). (B) Confusion matrices comparing performance on three other smartphones.

**Fig. 5. Performance testing with trained clinicians versus untrained parents.** (A) Demographic table of patients that were tested by parents. (B) Confusion matrix of the algorithm's performance for patients ears ( $n = 25$ ) tested by parents. (C, D) Mean acoustic dip of ears tested by parents (black) and clinicians classified by the algorithm as with middle ear fluid (red) and without fluid (blue).

**Fig. 6. Benchmark testing across different scenarios.** (A) Different paper types used to construct the funnel (B) Different tip diameters of the funnel (C) Different funnel placement angles (D) Different background noise (infant crying) intensities (E) Funnels created by different individuals (F) Different chirp volumes. Solid and dashed lines indicate conditions where the algorithm classifies the waveform correctly and incorrectly respectively. The figure shows the mean for each test and a standard deviation computed across 5 chirp instances.