# Preliminary validation of Hyfe cough detection & classification

The Hyfe Team[1]*

**Abstract**

- In live indoor trials, 97% of coughs were captured using Hyfe smartphone apps.

- Performance was above 93% across all smartphones tested (max = 98%).

- Specificity was perfect (no false positives).

- Experimental playback of pre-recorded coughs also yielded high sensitivity (91%).

[1] *www.hyfe.ai*
*Correspondence: research@hyfe.ai

## Introduction

Hyfe develops smartphone apps that use A.I. and passive listening to unobtrusively measure patterns and trends in a user's cough. Hyfe currently offers two apps: a public-facing consumer version entitled Hyfe Cough Tracker (hereafter, the Consumer app) and a research-oriented version optimized for clinical studies (the Research app). The former is available on both iPhone and Android; the latter is available only on Android.

We use multiple approaches to evaluate the accuracy of Hyfe's cough detection algorithms and the performance of its apps in various real-world scenarios. In a clinical context, Hyfe has optimized its apps for high performance in quiet indoor settings, such as a bedroom or hospital room, for any user, and for all common smartphone models. This is the context for which performance tests are designed to evaluate.

Multiple factors interact to govern the performance of any acoustic detector and classifier: microphone specifications, microphone handling patterns, phone software stability, user vocal characteristics, environmental noise levels, interaction with other apps, etc. Rigorous performance tests must account for all of these factors without becoming overly complex or sacrificing experimental reproducibility, and these competing interests typically require strategic trade-offs.

To guide subsequent validation efforts, here we present preliminary results from two approaches to Hyfe performance evaluation: (1) live trials in which a small group of volunteers produce solicited (ie, forced) coughs, and (2) playback trials in which pre-recorded (solicited) coughs from a large number of volunteers were played over high-quality speakers. The live trials emulate the real world more accu-

rately while sacrificing the variability inherent to a larger population of users and the reproducibility of using pre-recorded sounds, while the playback trials reflect the opposite trade-off.

## Methods

Coughs were produced in front of an array of devices from various manufacturers, which were provisioned with a mixture of Consumer and Research versions of the Hyfe app (Tables 1 - 2). Devices were positioned such that all were an equal distance from the cough source, and were either oriented with their microphone inputs facing the cough source or equipped with an omni-directional lapel lavalier microphone (www.slintshop.com).

*Live participant trials*
We recruited volunteers (n=10, 5 females, 5 males, ages 28 to 37) to contribute 30 solicited coughs in a quiet bedroom without air conditioning or other ambient noise. Participants stood such that their mouths were approximately 50 cm from the device array (9 devices; 1 iPhone iOS, 8 Android; 3 with Consumer app, 6 with Research; 3 with lapel microphone, 6 without; Table 1). Instructions were provided to participants on a computer screen that logged the timestamp of each instructed cough. Coughs were separated by random intervals that ranged between 5 and 15 seconds.

*Playback trials*
Pre-recorded coughs were drawn from a screening study at Sewanee: the University of the South (Sewanee, TN, USA) in which coughs from students, staff, and faculty (approx. 900 participants over 10 weeks) were recorded in 5-second

**Table 1. Results of live trials** (10 volunteers, 30 coughs each) by device. Sensitivity (fraction of ground-truth coughs detected) and specificity (false positive detections per user hour) are reported as the mean + SD across volunteers.

| Device | Ground-truth | | Sensitivity | | Specificity | |
| --- | --- | --- | --- | --- | --- | --- |
| | *Volunteers* | *Coughs each* | *Mean* | *SD* | *Mean* | *SD* |
| iPhone SE (4) (iOS, Consumer) | 10 | 30 | 0.93 | 0.06 | 0 | 0 |
| Ulephone X8 (10) (Android, Consumer) | 10 | 30 | 0.96 | 0.03 | 0 | 0 |
| Ulephone X8 (13) (Android, Research) | 10 | 30 | 0.97 | 0.04 | 0 | 0 |
| Ulephone X8 (11) (Android, Research) | 10 | 30 | 0.98 | 0.02 | 0 | 0 |
| Ulephone X8 (2) (Android, Research) | 10 | 30 | 0.99 | 0.02 | 0 | 0 |
| Motorola G7 (8) (Android, Research) | 10 | 30 | 0.96 | 0.06 | 0 | 0 |
| Motorola G7 (9) (Android, Consumer + mic) | 10 | 30 | 0.97 | 0.04 | 0 | 0 |
| Motorola G7 (16) (Android, Research + mic) | 10 | 30 | 0.98 | 0.02 | 0 | 0 |
| Motorola G7 (5) (Android, Research + mic) | 10 | 30 | 0.97 | 0.04 | 0 | 0 |
| **Aggregate means** | | | | | | |
| Overall | - | - | 0.97 | 0.02 | 0 | 0 |
| Consumer mean | - | - | 0.95 | 0.05 | 0 | 0 |
| Research mean | - | - | 0.97 | 0.04 | 0 | 0 |
| With mic | - | - | 0.97 | 0.04 | 0 | 0 |
| Without mic | - | - | 0.97 | 0.04 | 0 | 0 |

files (44.1 kHz, wav format) using a professional-quality microphone (Movo PC-M6, www.movophoto.com) in February - May 2021. From this archive of de-identified recordings (n=3,918), 314 were selected randomly. Each was manually verified to contain a single cough. For each cough in this set, the following data were drawn from the Hyfe database: cough prediction score (based upon April 2021 version of classification models), and start time of cough in the parent recording file.

During playback trials, these 314 pre-recorded coughs were played in sequence in a quiet room without air conditioning or other ambient noise. Sounds were played over a Kali Audio-6 6.5" desktop speaker (https://www.kaliaudio.com; 87 Hz - 25 kHz frequency response). The speaker was placed 50 cm from the array of devices (7 devices; 1 iPhone iOS, 6 Android; 2 Consumer app, 5 Research app; no lapel microphones used). The volume of the speaker was set to approximate that of a real person's cough, and received levels were captured to test this setting post hoc. Each cough was separated by a random interval ranging from 5 to 15 seconds. For each sound file, the exact playback time of its constituent cough was recorded to a log file along with the cough's associated metadata.

*Analysis*
Cough detection data generated from the app were downloaded and synchronized to the "ground-truth" logs. Cough detections (defined as peak sounds with a cough prediction score greater than 0.50) that occurred within two seconds of a ground-truth timestamp were classified as a successful capture. All analyses were conducted in R 4.0.2.

## Results

In live trials, devices performed with a mean sensitivity of 0.97 across users (sd = 0.02, min=0.93, max=0.98) and perfect specificity (no false positives) (Table 1, Figure 1). App versions (Consumer, n=3, and Research, n=6) performed similarly (0.95 and 0.97, respectively), as did devices with and without mic (0.97 each) (Table 1). The single iOS device tested had the lowest sensitivity (mean = 0.93, sd = 0.06).

In playback trials , devices performed with slightly lower sensitivity to 0.91 (sd = 0.02, min = 0.88, max = 0.94) (Table 2, Figure 2). Sensitivity was similar across app versions (Consumer, n=2, mean = 0.91; Research, n=5, mean = 0.91), and the single iOS device tests fell in the middle of the array values (0.88). One Android device registered one false
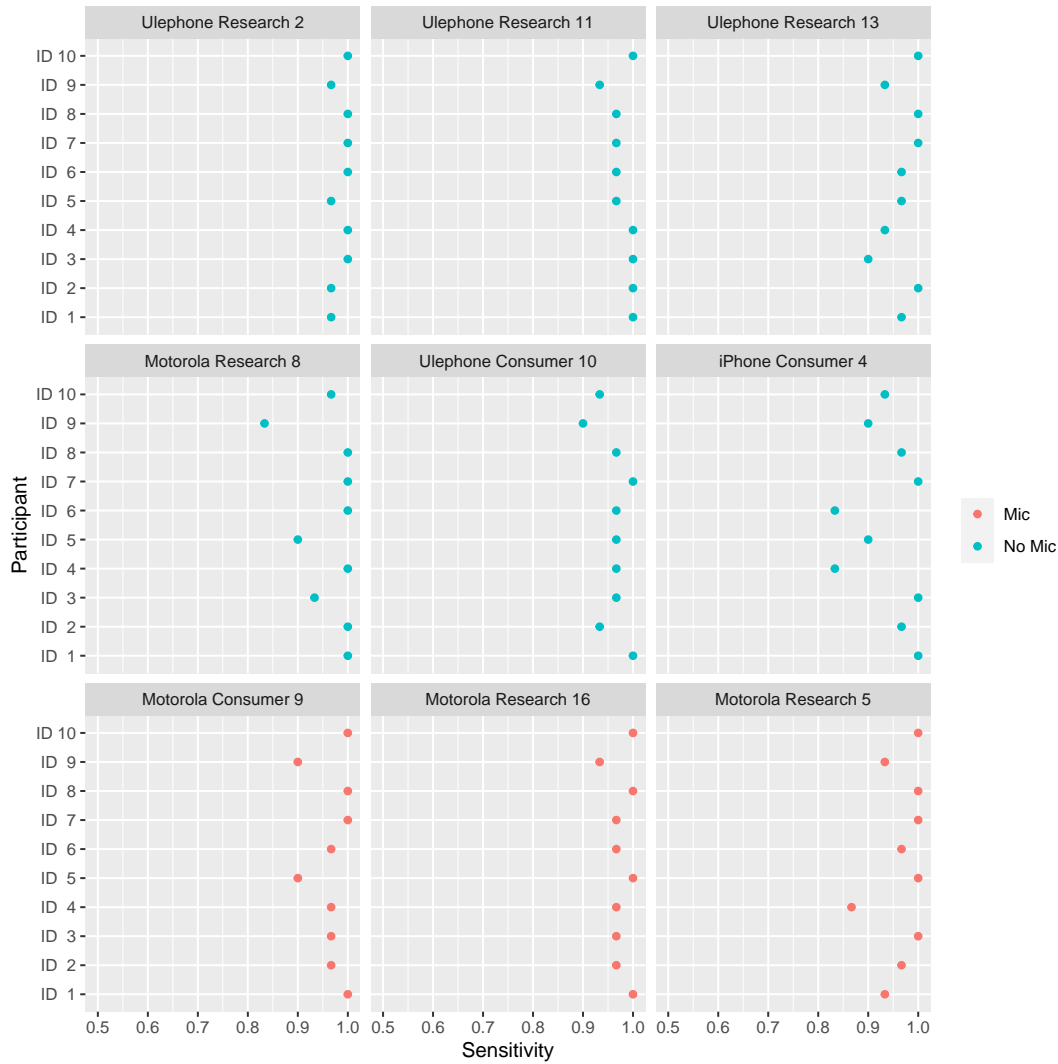
**Figure 1. Hyfe sensitivity as measured during live trials**, presented by for each participant (ID 1 - ID 10; 30 coughs each) on each device.

positive (one in 2,157 peak detections across the array), resulting in a per-device false positive rate of 0.14 per user hour (equivalent of 3 false positives per day).

## Discussion

Live trials demonstrated high performance in the ideal clinical context: a quiet indoor setting in which the device is positioned close to the user. This preliminary effort ought to provide a concrete example to inform the development of full-fledged clinical trials.

Playback trials of coughs also yielded high performance (above 90% sensitivity), but with a small loss (approx. 6% drop in sensitivity) which we attribute to artifacts from the process of digitally capturing and replaying sound through microphones and speakers. However, the fact that sensitivity loss is measurable and minor opens up opportunities for reproducible experiments using pre-recorded sounds, e.g., comparing performance among devices across multiple sessions; assessing the effect of sound source distance; assessing the effect of carrying devices and other material obstructions; and more rigorous assessments of the effect of auxiliary technology such as lapel microphones.

We suggest that live trials can be designed in tandem with playback experiments to optimize resources and expedite the Hyfe validation process. We suggest that live trials focus upon the objective validation of real-world use cases (e.g., unsolicited coughs recorded during deployments in hospital rooms) while playback trials focus upon the relative impacts of certain context factors (e.g., distance from device, effect of carrying device, effect of various forms of ambient noise, etc.).

**Table 2. Results of playback trials** (314 pre-recorded coughs played over speakers) by device. Sensitivity is defined as the fraction of ground-truth coughs detected; specificity is defined as false positive detections per user hour.

| Device | Playback coughs | Cough detections | Sensitivity | Specificity |
|---|---|---|---|---|
| iPhone SE (4) (iOS, Consumer) | 314 | 277 | 0.90 | 0.00 |
| Ulephone (10) (Android, Consumer) | 314 | 284 | 0.92 | 0.97 |
| Ulephone (13) (Android, Research) | 314 | 290 | 0.94 | 0.00 |
| Ulephone (2) (Android, Research) | 314 | 283 | 0.92 | 0.00 |
| Motorola (5) (Android, Research) | 314 | 270 | 0.88 | 0.00 |
| Motorola (8) (Android, Research) | 314 | 278 | 0.90 | 0.00 |
| Google Pixel (14) (Android, Research) | 314 | 283 | 0.92 | 0.00 |
| **Grand mean (SD)** | **-** | **-** | **0.91 (0.02)** | **0.14 (0.37)** |
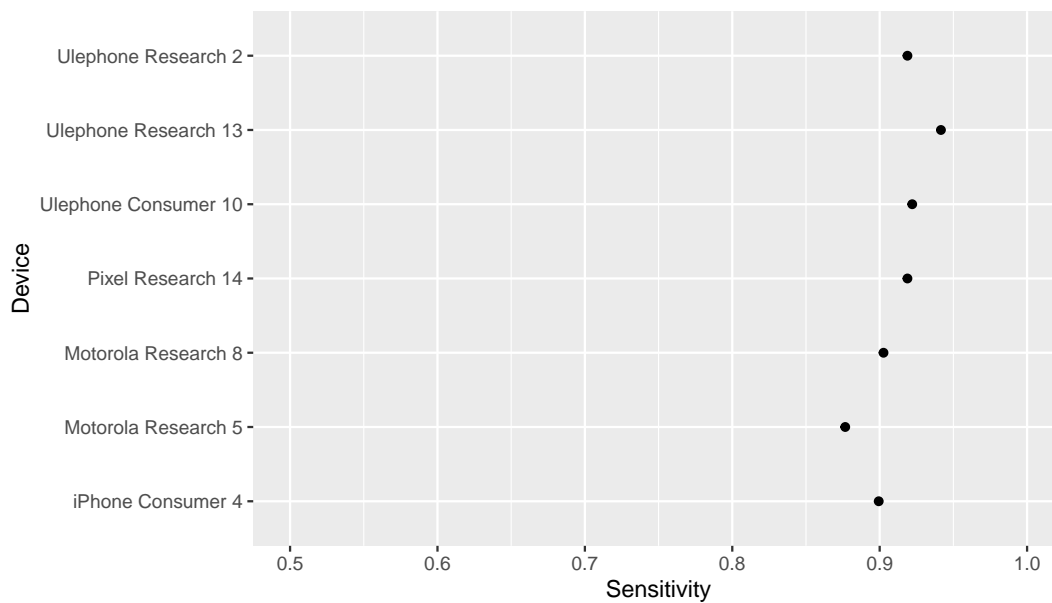


**Figure 2. Hyfe sensitivity as measured during playback trials**, presented by for each participant (ID 1 - ID 10; 30 coughs each) on each device.