# Exploring Smartwatch-based Deep Learning Approaches To Support Sound Awareness for Deaf and Hard of Hearing Users

Dhruv Jain, Hung Ngo, Pratyush Patel, Steven Goodman, Leah Findlater, Jon Froehlich




ASSETS 2020 Paper

Sound recognition is at the heart of many modern AI systems.

Our past work examined sound recognition to support d/Deaf and hard of hearing (DHH) users in the home.

However, the sensing and classification was done on non-portable devices.

Jain *et al.*, HomeSound, CHI 2020

Recent iOS 14 update introduced sound recognition in consumer smartphones.

But this release is closed-source and the implementation details are unknown.

# Two Studies

**Study 1** A **quantitative** examination of four lightweight deep-learning models to classify sounds.

**Study 2** A **qualitative** evaluation of a smartwatch-based sound awareness app with 8 DHH participants.

A recent study with 201 DHH users showed that smartwatch was the most preferred device for sound feedback.

# Two Studies

**Study 1** A **quantitative** examination of four lightweight deep-learning models to classify sounds.

**Study 2** A **qualitative** evaluation of a smartwatch-based sound awareness app with 8 DHH participants.

# Two Studies

**Study 1**

A **quantitative** examination of four lightweight deep-learning models to classify sounds.

# Study 1


## Goal

- Performance evaluation of four deep learning sound classification models across four architectures.

## Models

- Three recently released TensorFlow-Lite models: MobileNet (3.4MB), Inception (41MB), ResNet-Lite (178.3MB) and a quantized version of our model: VGG-Lite (281.8MB).
- Also, a comparison with state-of-the-art full-VGG model (845.5MB) running on a laptop.

## Architectures

- Watch-only, watch+phone, watch+cloud, and watch+phone+cloud.
- A commercially available smartwatch (Tickwatch Pro) and smartphone (Honor 7x) were used.

# Study 1 Findings


Models

Architectures

# Study 1 Findings


## Models

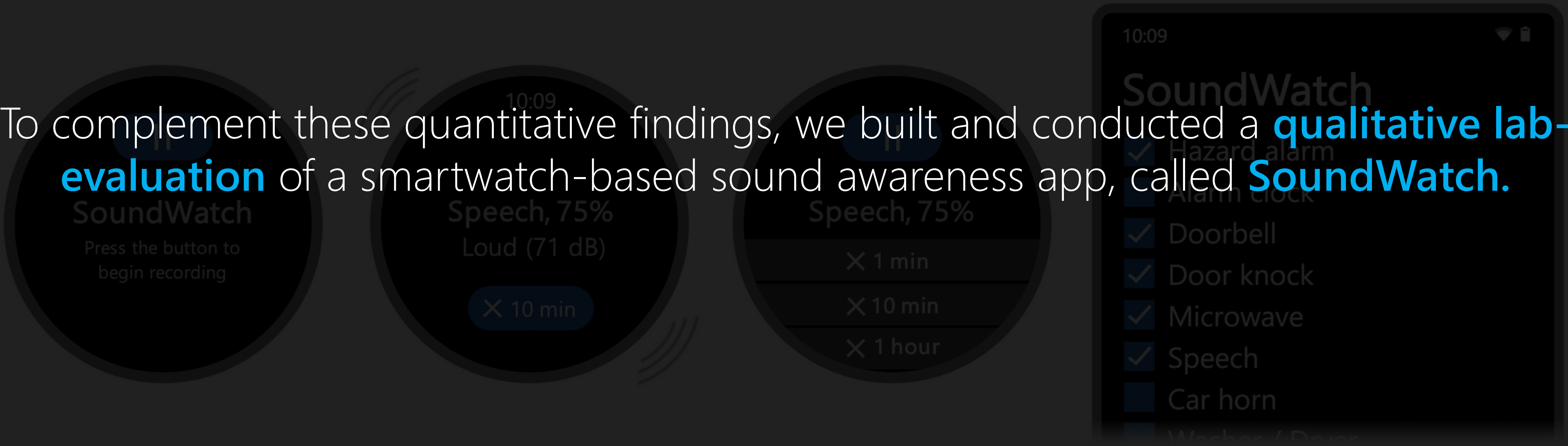
- The best classification model (VGG-lite) **had similar accuracy as the state-of-the-art** for non-portable (VGG) but required substantially **less memory (~1/3rd).**
- Accuracy of best model was **81.2%** (*SD*=5.8%) for 20 sound classes and **97.6%** (*SD*=1.7%) for three high-priority sounds, when evaluated on our dataset of field sound recordings.
- Among our four models, we also observed a **strict accuracy-latency trade-off:** the most accurate model was also the slowest (avg. acc=81.2%, avg. latency=3.4s).

# Study 1 Findings

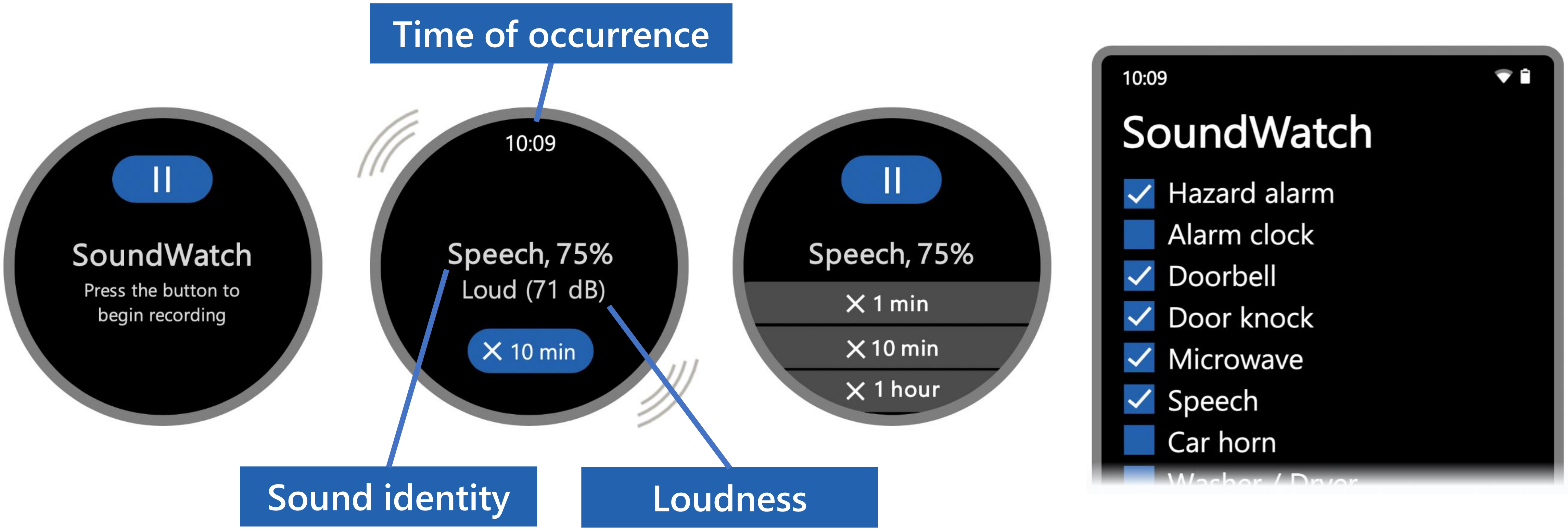

## Architectures

- The two phone-based architectures (*watch+phone*, *watch+phone+cloud*) **outperformed the watch-centric designs** (*watch-only*, *watch+cloud*) in terms of CPU, memory, battery usage, and end-to-end latency.

To complement these quantitative findings, we built and conducted a **qualitative lab-evaluation** of a smartwatch-based sound awareness app, called **SoundWatch.**

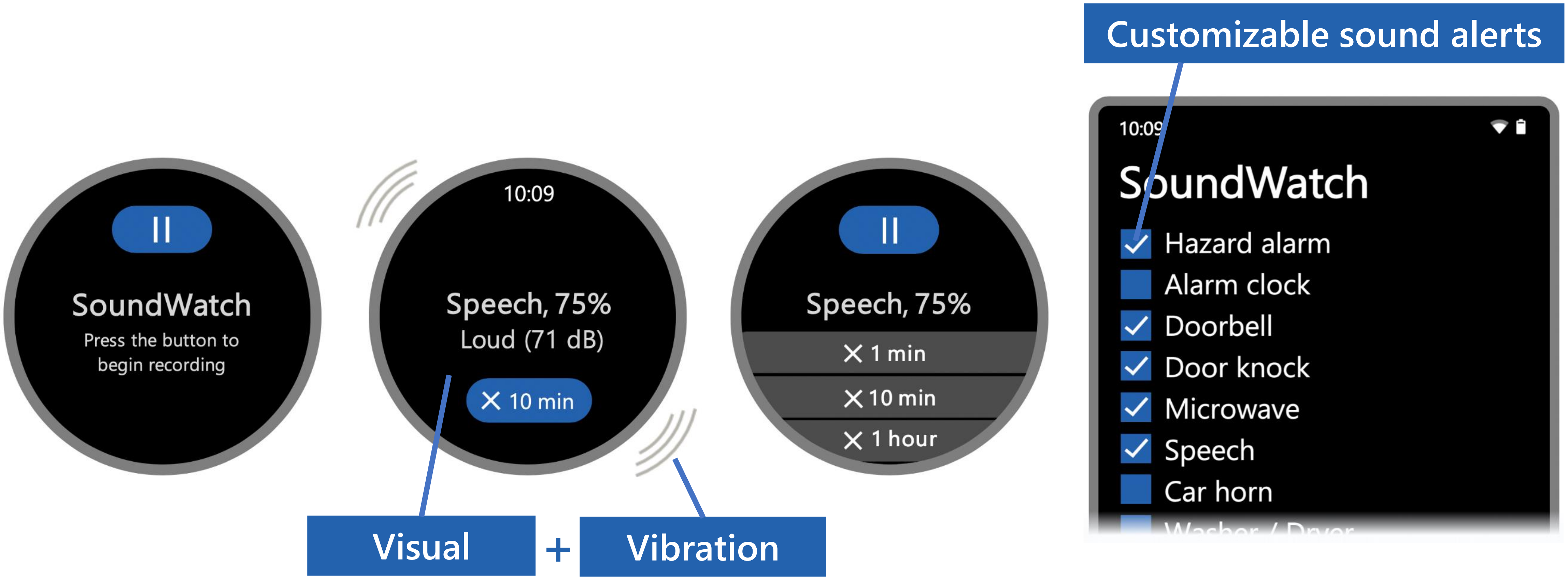


# SoundWatch



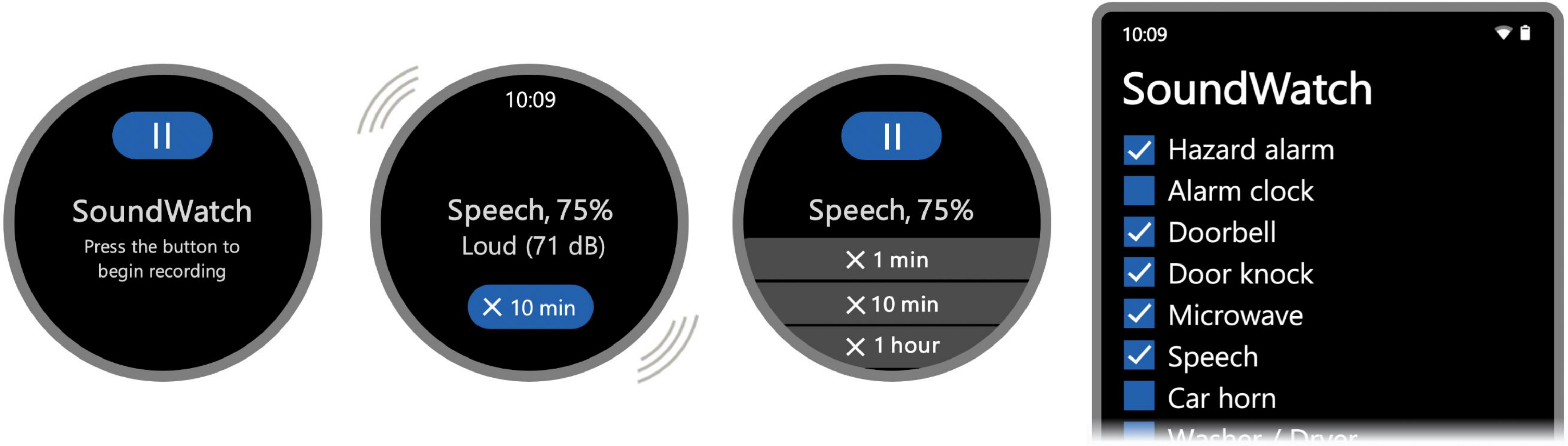
Time of occurrence
10:09
SoundWatch
Press the button to begin recording
Speech, 75%
Loud (71 dB)
10 min
1 min
10 min
1 hour
Sound identity
Loudness
SoundWatch
Hazard alarm
Alarm clock
Doorbell
Door knock
Microwave
Speech
Car horn

# SoundWatch


Customizable sound alerts
SoundWatch
Press the button to begin recording
10:09
Speech, 75%
Loud (71 dB)
10 min
Speech, 75%
1 min
10 min
1 hour
Visual
+
Vibration
10:09
SoundWatch
Hazard alarm
Alarm clock
Doorbell
Door knock
Microwave
Speech
Car horn

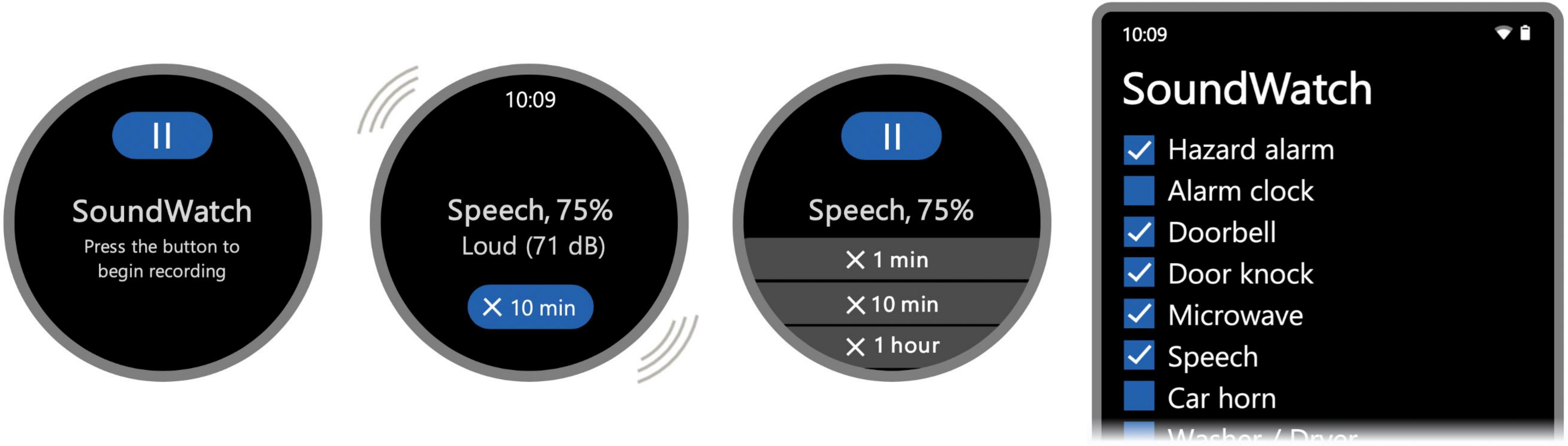
# SoundWatch

Support for four architectures with deep-learning model running on either watch (watch-only), phone (watch+phone), or cloud (watch+cloud, watch+phone+cloud).



# SoundWatch

**SoundWatch processes the sound locally on the watch or phone and, in the case of the cloud-based architectures, only uploads non-reconstructable mel-spectrogram features.**



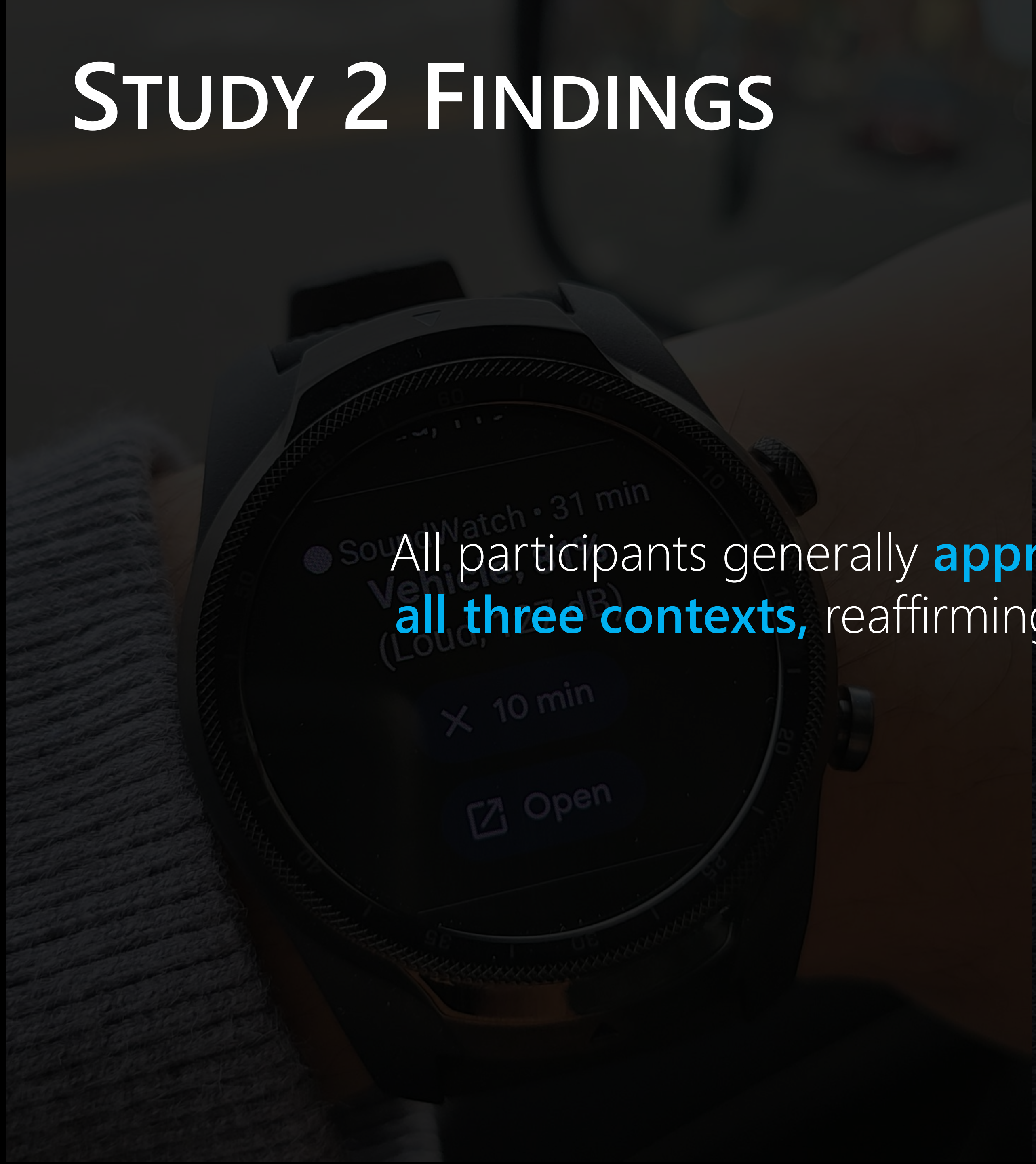
# STUDY 2

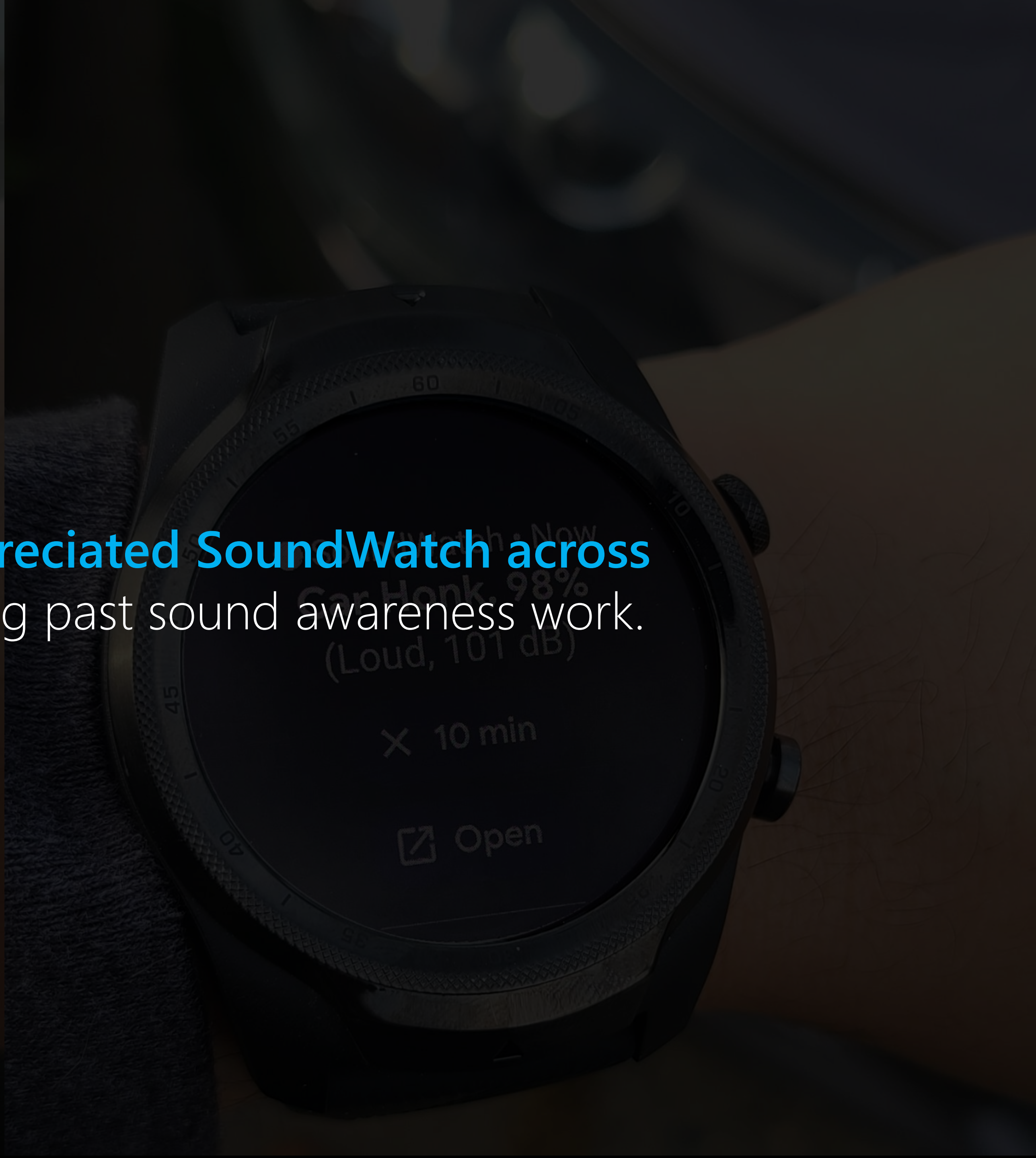
**Goal**
- Gather user feedback on our system results and the SoundWatch app.

**Participants**
- Eight DHH participants (3 women, 3 men, 2 non-binary).

**Method**
- Campus walkthrough with the SoundWatch app in three contexts: a lounge, a lab, and a bus stop.
- Post-trial interview on the experience and other technical considerations—*e.g.*, desired accuracy-latency tradeoff, thoughts on the four SoundWatch architectures.

# Study 2 Findings

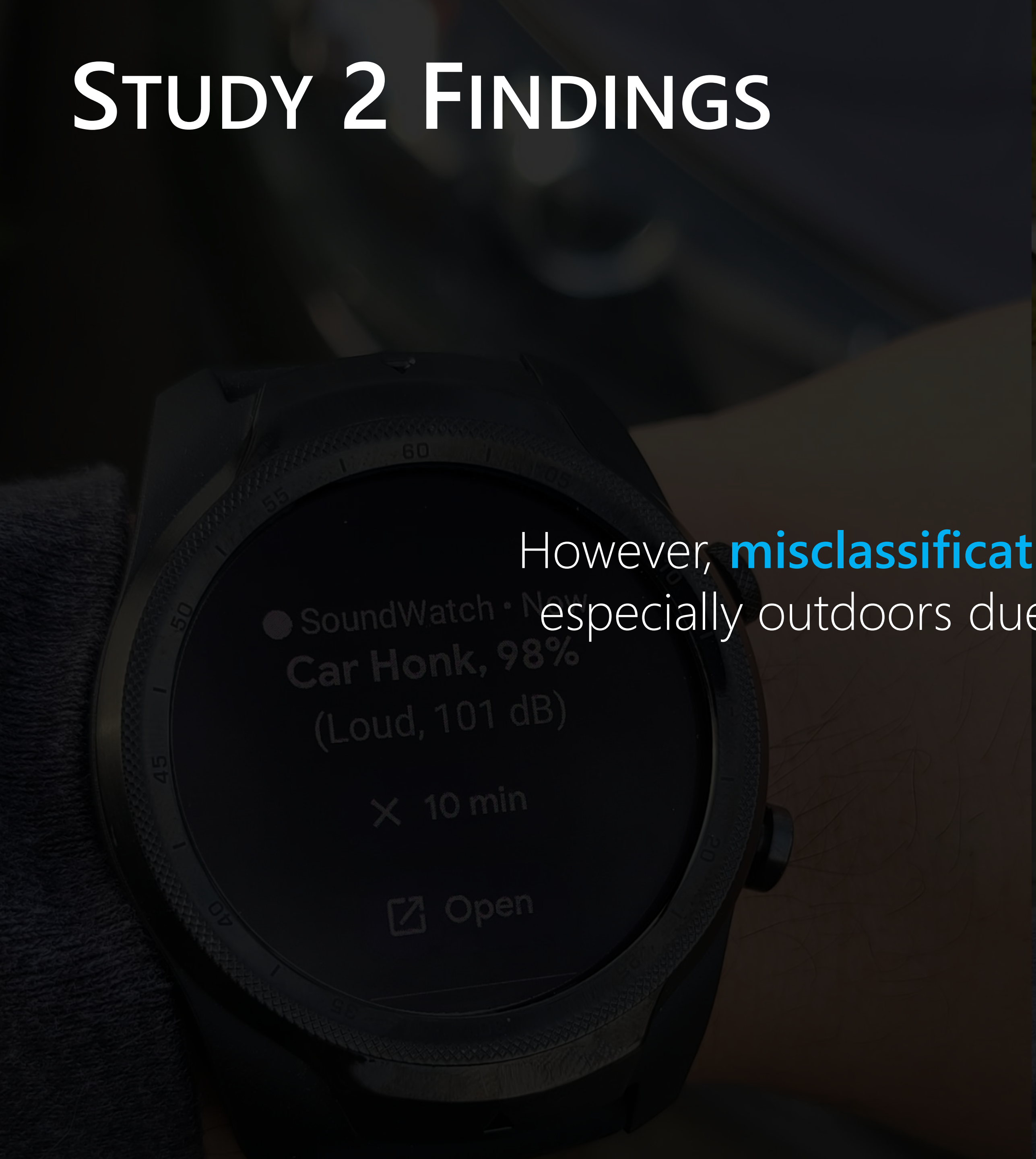


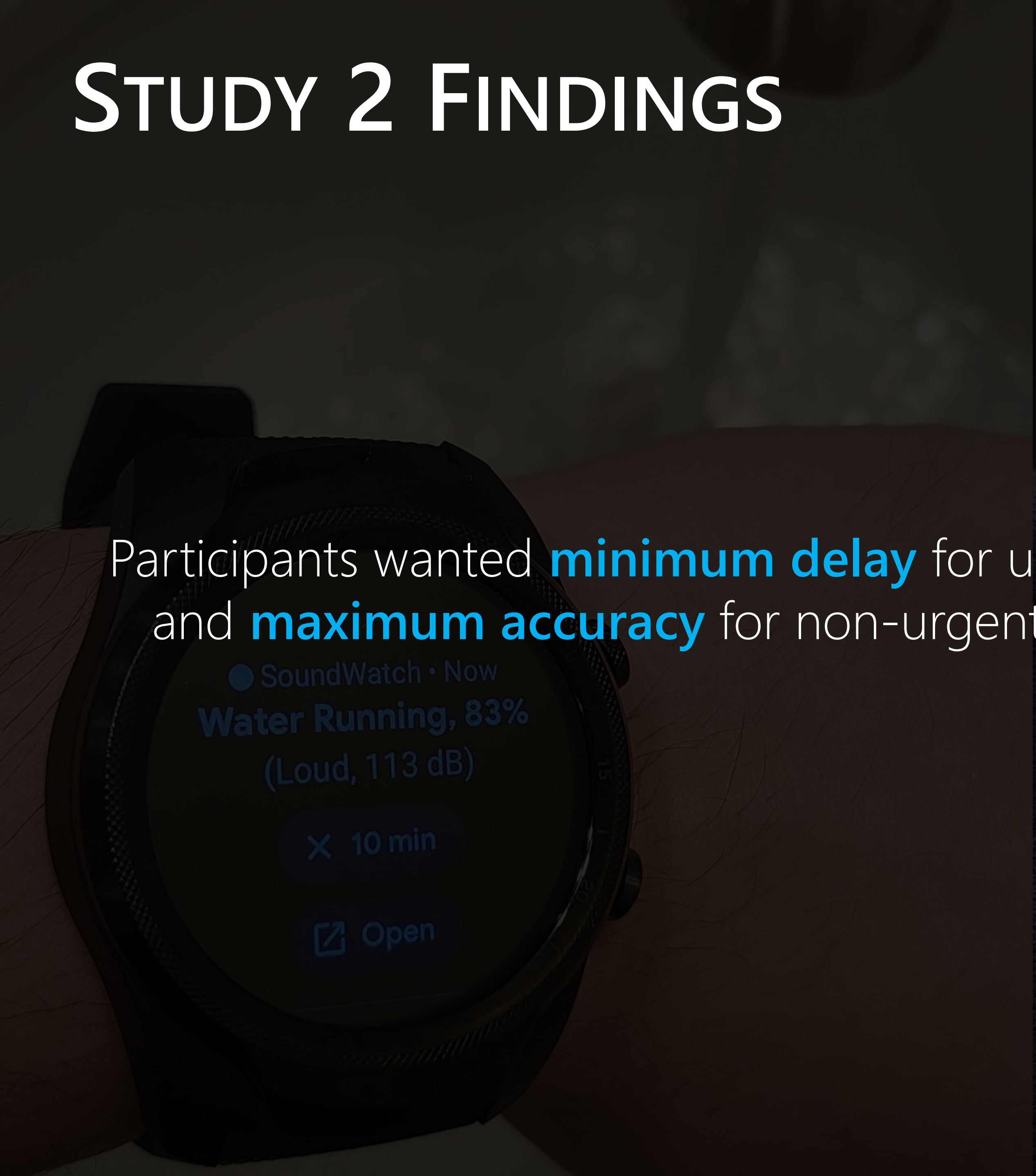

All participants generally **appreciated SoundWatch across all three contexts,** reaffirming past sound awareness work.

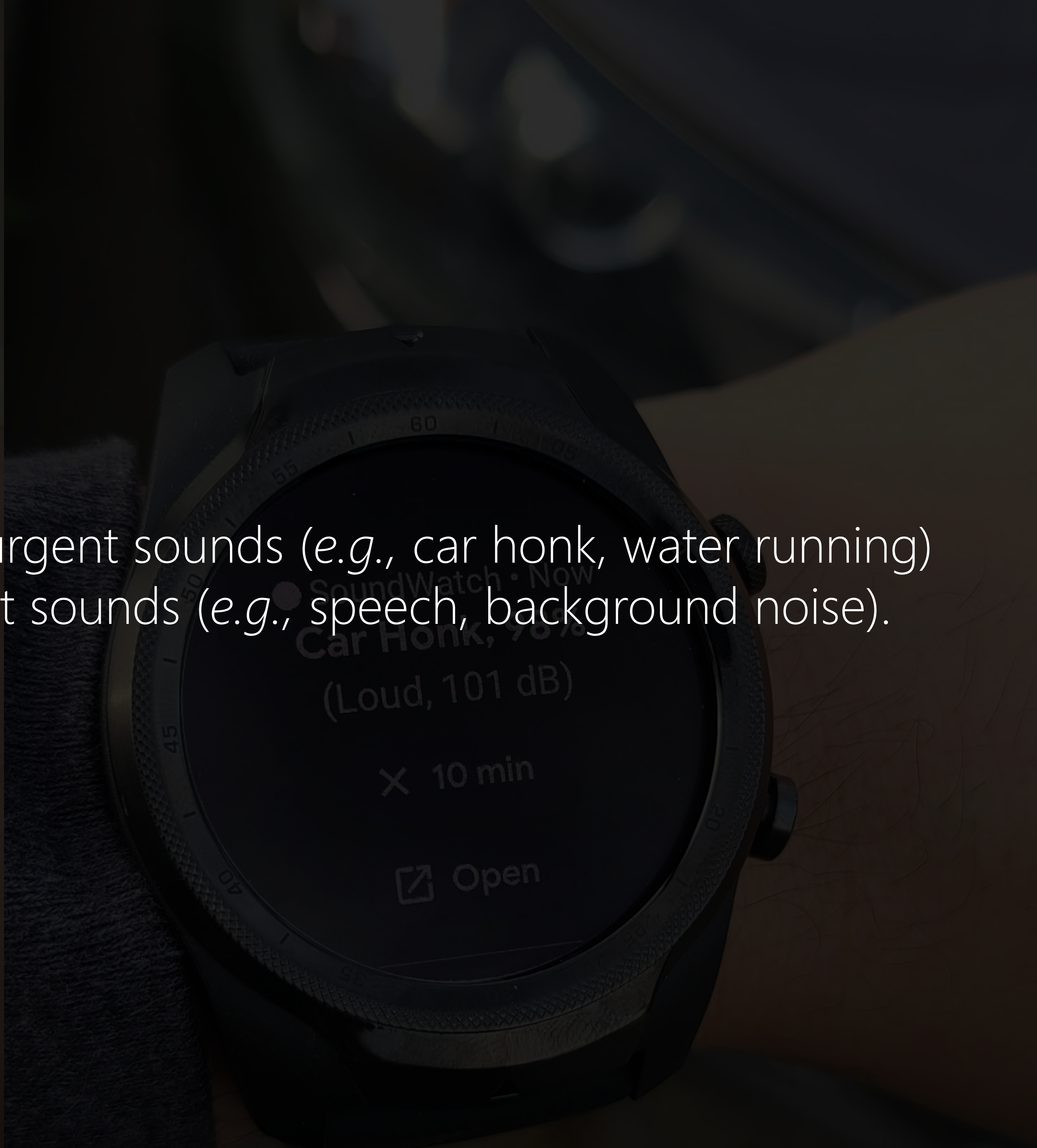
# Study 2 Findings

However, **misclassifications were concerning**, especially outdoors due to background noise.




# Study 2 Findings

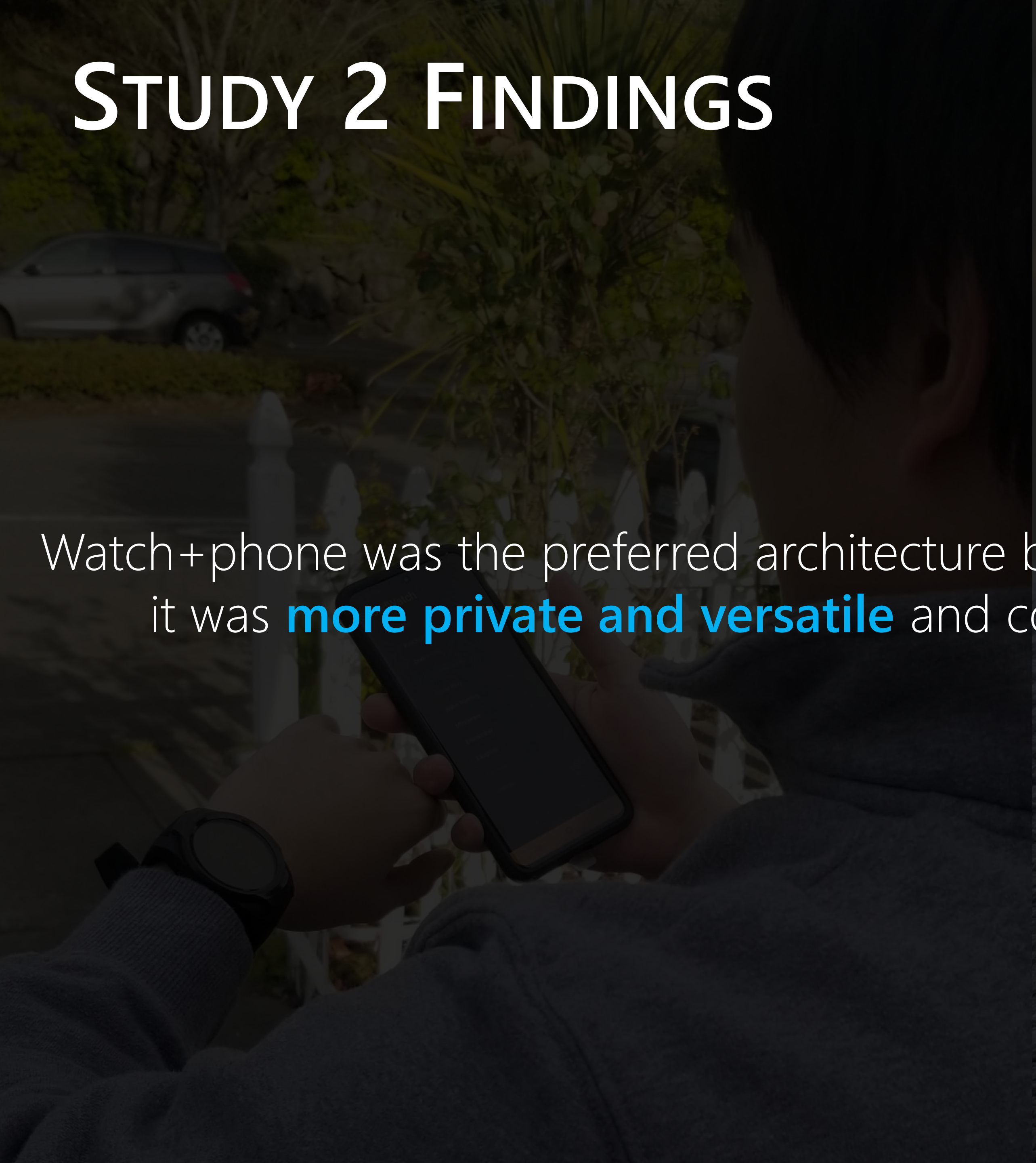


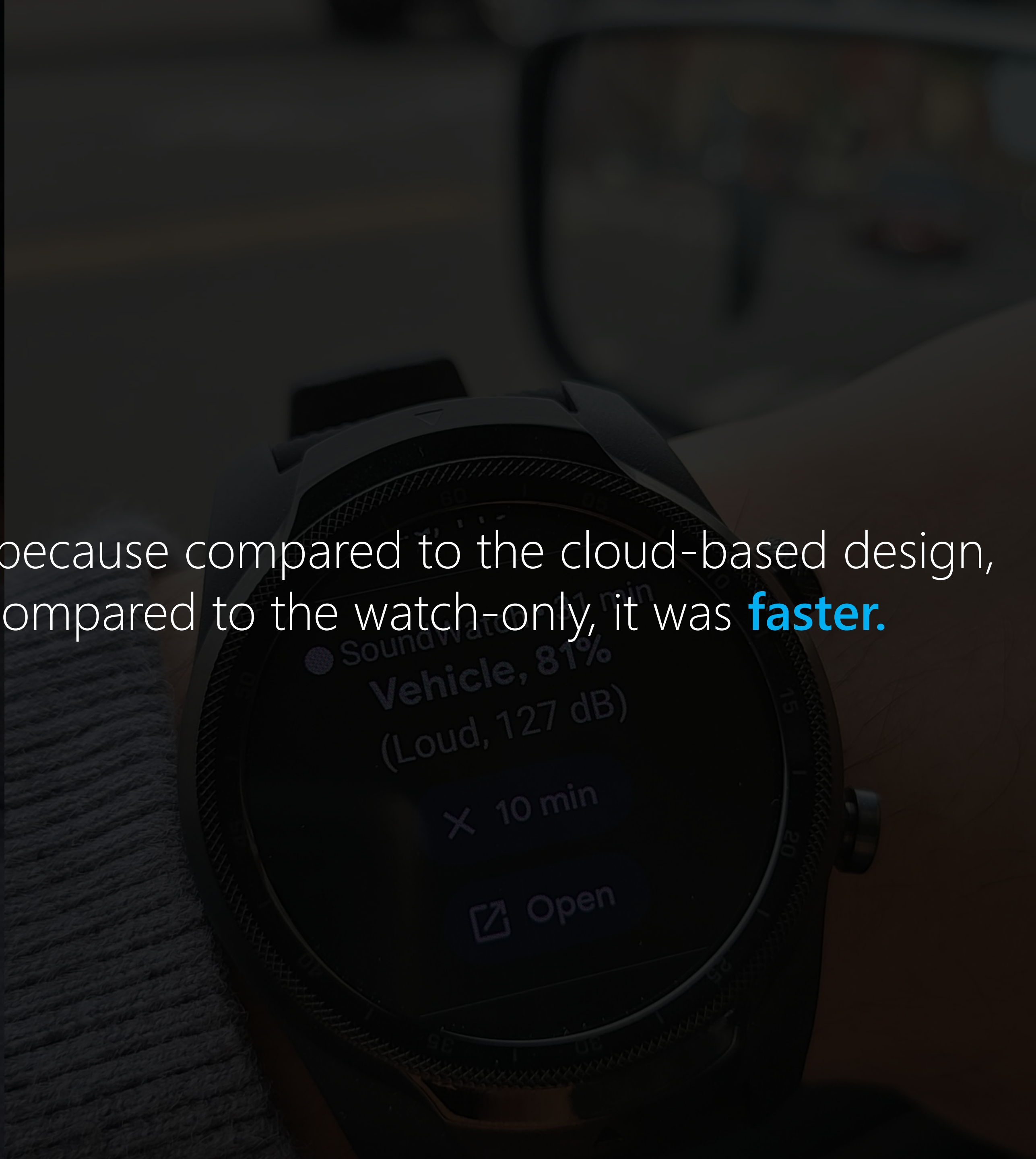




Participants wanted **minimum delay** for urgent sounds (*e.g.*, car honk, water running) and **maximum accuracy** for non-urgent sounds (*e.g.*, speech, background noise).

# Study 2 Findings

Watch+phone was the preferred architecture because compared to the cloud-based design, it was **more private and versatile** and compared to the watch-only, it was **faster.**





# Reflection

How well does a **smartwatch-based sound classification** tool need to perform?

Needs further study...

# Recommendations

1. Explore usage in the field.
   But this introduces ethical and safety concerns.
   Increasing transparency may help.

2. Explore showing multiple "possible" sounds.

3. Explore end-user customization.

4. Explore end-user interactive training of the model—*e.g.*, Wu, CHI '20.
   But this may be tedious if the sound is inaccessible to DHH users.

Smartwatch offers a myriad of possibilities for DHH users and beyond.

Please refer to the paper for more interesting ideas on **smartwatch + sound** feedback.