HeadsUp: Mobile Collision Warnings through Ultrasound Doppler Sensing

Negin Yaghoubisharif
Bauhaus-Universität Weimar
Weimar, Germany
negysharif@gmail.com

Christopher Getschmann
Aalborg University
Aalborg, Denmark
cget@cs.aau.dk

Florian Echtler
Aalborg University
Aalborg, Denmark
floech@cs.aau.dk

Figure 1: HeadsUp: mobile collision warnings using Doppler-shifted ultrasound signals.

ABSTRACT

Smartphone-using pedestrians are often distracted, leading to frequent accidents of varying severity with a risk of both awkwardness and injury. We introduce HeadsUp, a mobile app designed to warn the user of imminent collisions with solid obstacles. HeadsUp runs on unmodified commodity smartphones without additional hardware and uses active ultrasound sensing based on the Doppler effect. We contribute an analysis of the ultrasound audio characteristics of six different smartphone models to verify the feasibility of our approach across vendors and device classes, and a description of two implementation variants of our signal processing pipeline. We evaluate our system both in a lab environment and under real-world conditions, and we conclude that HeadsUp can effectively work at a range of up to 3 meters, even though overall performance is heavily dependent on both the individual user and the environment characteristics.

CCS CONCEPTS

• **Human-centered computing** → **Smartphones**; **Field studies**; **Sound-based input / output**.

KEYWORDS

smartphones, ultrasound, Doppler, collision warning, environment, sensing

MUM 2022, November 27–30, 2022, Lisbon, Portugal

ACM ISBN 978-1-4503-9820-6/22/11.
https://doi.org/10.1145/3568444.3568458

ACM Reference Format:
Negin Yaghoubisharif, Christopher Getschmann, and Florian Echtler. 2022. HeadsUp: Mobile Collision Warnings through Ultrasound Doppler Sensing. In *21th International Conference on Mobile and Ubiquitous Multimedia (MUM 2022), November 27–30, 2022, Lisbon, Portugal*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3568444.3568458

1 INTRODUCTION

Pedestrians who use their smartphones while walking are usually distracted and therefore prone to accidents. Shortly after the introduction of the iPhone in 2009, this phenomenon was already sarcastically described as "Death by iPhone" [12], and the British Automobile Association predicted in 2010 that a growing proportion of traffic deaths and casualties would result from distracted traffic participants [25]. More recent results from the United States support this theory, as pedestrian fatalities have increased at 4 x the rate of overall deaths in traffic [14], and a study from Korea also found a correlation between smartphone addiction and accident rate [7]. To counter this issue, some cities (e.g. Augsburg [11], Bodegraven [22], or Seoul [27]) have recently experimented with embedding additional traffic lights into the pavement at critical intersections to capture the attention of downward-looking pedestrians, while Ilsan in South Korea has seen the installation of warning projectors which show alerts on the ground near road crossings [15].

To address this issue, we explore collision sensing that happens unobtrusively in the background. HeadsUp, our collision-warning system runs on unmodified commodity smartphones using the devices' audio system to emit an inaudible ultrasound signal. The simultaneously recorded audio data is analyzed for the presence of the reflected and Doppler-shifted signal that indicates an approaching object. When a signal is detected in the correct frequency range that matches the walking speed of the phone-using pedestrian and the relative speed of common obstacles in urban environments (cars, other pedestrians, stationary objects), a notification is shown to warn the user.

2 RELATED WORK

An alternative approach to solve the problem of increasing pedestrian accidents is to design applications that are less distracting for the user. Various research projects have looked into limiting the use of addictive apps (AppDetox [10]), into using wearables and non-visual feedback to improve attention on the environment (Shoe Me The Way [19]), into showing a live camera view as background for the keyboard (Type'n'Walk [9]), into better management of notifications [26], or even into using drones as mobile guides for urban navigation [2].

However, the fact remains that users can be expected to look at their devices while on the move, as this is one of the defining features of *mobile* devices. We, therefore, do not consider HeadsUp as a replacement for the approaches described above, but rather as a complementary feature that has the potential to further reduce the risk of accidents in urban environments.

A related approach is BumpAlert by Tung and Shin [23], which also attempts to warn distracted smartphone users about obstacles by using a combination of acoustic signal runtime analysis and camera video analysis. However, as opposed to our approach, the system has to deal with multipath reflections from both stationary and moving objects, and requires a specific posture of the user holding the mobile device for camera data. In addition, users found the audible 11 kHz signals used here to be distracting.

Another area of work relevant to our approach is support for visually impaired users. Various existing projects have looked into using ultrasound collision detection for this task, using e.g. sensors embedded in a vest [20], in an overall [16], in a bracelet [17], in a belt [18], or in a cane [1]. However, these approaches all have in common that they use dedicated extra sensor hardware that has to be carried by users in addition to their smartphone, and which can be expected to lower user acceptance noticeably.

Regarding the usage of ultrasound on commodity hardware, an early example is SoundWave by Gupta et al. [5] which used the built-in speaker and microphone of a laptop to recognize in-air hand gestures. Ultrasound is emitted from the laptop's speaker, and any signals reflected by a nearby moving object are recorded by the microphone. The recorded data is then analyzed for Doppler-shifted signatures of predefined gestures, which include single and double tap, scrolling, and "two-handed seesaw". While the fundamental approach is similar to our work, SoundWave only recognizes motion in front of a stationary device at a maximum distance of one meter.

Biying et al. [4] investigated the feasibility of gesture and activity recognition around a smartphone as well. For recognizing hand gestures, similar to SoundWave, they performed different hand gestures over a stationary device and used the Doppler shift to recognize hand approaching, hand withdrawal, swipe motion of right-to-left and left-to-right, and seesaw motion. The other purpose of their study was activity recognition from three predefined categories of sleeping in a bed, getting up from bed and walking away from it, and everyday desk work. They showed that their methods for hand gesture and activity recognition were feasible for a limited number of predefined gestures and activities in case the device was stationary, but when the device was held in hand, carried on the body, or covered with clothes, the received signal was strongly attenuated, and the system was exposed to strong noise.

Sun et al. presented VSkin [21], a novel way of interacting with mobile devices by sensing touch gestures on surfaces of the device. VSkin detects touch gestures with high accuracy using both structure-borne sounds (propagation of sound through the structure of the device), and air-borne sounds, the propagation of sound through the air. This was achieved by measuring the amplitude and the phase of each path of sound signals. The system detected tapping events with an accuracy of 99.65% and captured finger movements with an accuracy of 3.59 millimeters. MilliSonic [24] by Wang and Gollakota reaches even higher positional accuracy in the sub-millimeter range by using an external microphone array.

Apart from the usage of ultrasound in gesture recognition and motion detection for interaction, in medical science, Nandakumar et al. [13] proposed a contactless system for detecting restricted breathing due to opioid overdose using inaudible acoustic signals from the smartphone's speaker and the measured frequency shift. They produced a short-range active sonar system by a frequency-modulated continuous waveform and tracked breathing in people for the diagnosis of opioid overdose by motions of the chest. They showed that by providing the proper connectivity to emergency calls in case of a detected overdose, their system prevented fatal outcomes.

3 BACKGROUND

HeadsUp is based on the Doppler effect, which describes the frequency shift experienced by a receiver when recording signals from a source which is moving relative to the receiver. The base equation for calculating this shift is:

$$f = f_0 * \left(\frac{C \pm V_r}{C \pm V_s} \right) \tag{1}$$

with f_0 base frequency, C wave velocity (343 m/s for sound under standard conditions), and V_r/V_s speed of receiver and source, respectively (signs depend on relative direction of motion). However, in our scenario, we are analyzing *reflected* signals, with source and receiver being coincident in the smartphone. This case can be modeled as two separate Doppler shifts, one before and one after reflection, resulting in the following equation:

$$f = f_0 * \left(\frac{C \pm V_{obj}}{C \pm V_{usr}} \right) * \left(\frac{C \mp V_{usr}}{C \mp V_{obj}} \right) \tag{2}$$

As above, the individual signs of the inner fractions depend on the relative direction of movement (object approaching user or user following object). Based on this equation, we can now calculate the expected Doppler shift for plausible scenarios once we select a suitable base frequency.

Since our approach works exclusively in the *frequency* domain (i.e. detection of a specific frequency in the recorded signal), it is unaffected by multipath reflections and other signal degradations which manifest in the *time* domain. Also, since we do not need to differentiate between several types of motion as e.g. [5] does, our signal processing is less sensitive to background noise and can therefore work at longer ranges.

As a standard Android smartphone is capable of playing back and recording audio data with a maximum sample rate of 48 kHz, the Nyquist limit places the highest possible frequency that can still be processed at 24 kHz. However, as the audio hardware of smartphones is primarily designed to operate in the audible frequency range up to at most 18 kHz, we cannot assume that these high frequencies can be reliably used. We, therefore, select a base frequency of 20 kHz as a "sweet spot" that is well above the hearing limit even for children and young adults, but still below the Nyquist limit and well within the capabilities of most commodity smartphones. This is in line with the frequency used by other ultrasound-based systems such as Biying et al. [4].

4 FEASIBILITY STUDY

To confirm that our assumptions about smartphone audio capabilities are correct, we conducted an initial feasibility study with external sound equipment to characterize a variety of devices from the last 5 years: Motorola Moto E and Sony Xperia XZ1 Compact (low-end), LGE Nexus 5 and LGE Q6 M700 (mid-range), ZTE Axon 7 and Motorola Moto Z (high-end). Our goal was to verify whether this sample of devices would be capable of a) emitting suitable ultrasound signals, b) recording the reflection with sufficient sensitivity, and c) providing satisfactory noise characteristics. We assume that this sample across various model years, manufacturers, and price classes is representative of at least 90% of smartphones currently in use.

4.1 Playback

In the first test phase, we used a Behringer U-Phoria UMC404 audio interface with a sampling rate of 192 kHz at 16 bit resolution and an Audix TM-1 measurement microphone with a calibrated 25 kHz upper-frequency limit. All data was analyzed using the open-source audio workstation software Audacity[1]. Using this setup, we recorded a continuous 20 kHz ultrasound sinewave signal played back from the individual devices with maximum volume, at distances of 0 m, 3 m, and 6 m. All devices were placed horizontally on a flat surface, with the bottom of the device facing towards the microphone. Note that the LG Q6 and the Moto Z produced distinctly audible subharmonics in this scenario.

We then analyzed a continuous 5-second stretch of recorded audio with a 4096-sample Fast Fourier Transform (FFT) using a Hann window function. At close range, the signal from the Moto Z was loud enough to cause clipping of the audio signal, resulting in a positive dB value. As shown in figure 2, all tested devices achieve

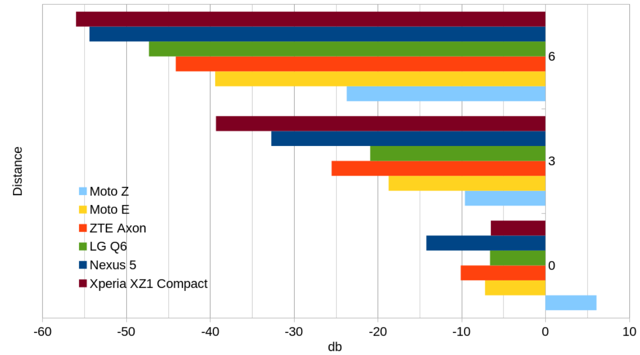

Figure 2: Amplitude of 20 kHz ultrasound emitted from smartphone speakers. Measurements on the x-axis are in dB relative to the maximum recording amplitude of the external UMC404 audio interface (higher values are louder), distance on the y-axis is in meters.

a signal strength of at least -40 dB at a distance of 3 m. Note that all signal strength measurements are relative to the maximum amplitude which the system can record, i.e. 0 dB = $2^{16} - 1$.

4.2 Recording

In the second test phase, we used a microcontroller (Arduino Micro) driving a piezo transducer as an ultrasound source at a frequency of 20 kHz, as consumer audio hardware often contains additional bandpass filters that heavily attenuate signals above 18 kHz. Using the UMC404 interface and Audix TM-1 microphone as above, we confirmed that this setup emits 20 kHz ultrasound with a signal strength of about -12.3 dB at close range, similar to the devices tested previously. We then recorded this signal via the individual devices' microphones, again at distances of 0 m, 3 m, and 6 m, with the devices placed on a flat surface and the bottom facing the emitter. Signal analysis was performed as before.

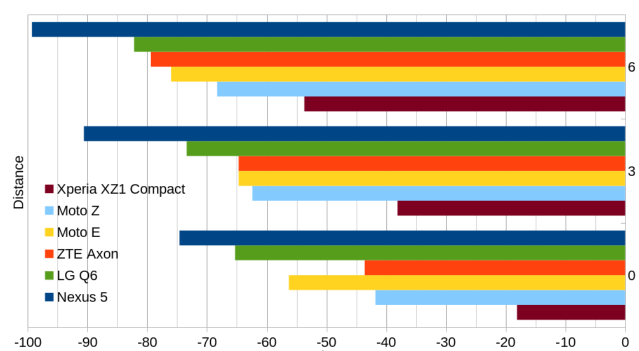

Figure 3: Amplitude of 20 kHz ultrasound recorded by smartphone microphones. Measurements on the x-axis are in dB relative to the maximum recording amplitude of the individual smartphones (higher values are louder), distance on y-axis is in meters.

[1] https://www.audacityteam.org/

As shown in Figure 3, all phones are capable of recording the signal at all tested distances, however at greatly reduced signal strength. Based on the first phase of the test, we can assume that an object at a distance of 3 meters will receive a signal strength of at least -40 dB when the signal is emitted from a smartphone (cf. Figure 2, center group). As an emitter with comparable signal strength was used in the second test phase, even taking potential reflection losses into account, we can conclude that our approach will work at a distance between user and object of at least 3 meters.

4.3 Signal Bandwidth

For the third test phase, we analyzed the bandwidth of the recorded signal when the phone is both emitter and receiver in a quiet environment. The emitted sound (pilot tone) will also be directly recorded by the devices' own microphone at a very high volume. Consequently, any reflected sound that is too close to the pilot tone's base frequency of 20 kHz will be drowned out in the resulting frequency spectrum, particularly as any reflection will also be at least 1-2 orders of magnitude less powerful than the pilot tone.

For signal analysis, we again used a FFT with a window size of 4096 samples (equivalent to 85.3 ms window duration) and a Hann window function. Using the `kiss-fft` library[2], processing time for one window is about 25 ms on a mid-range smartphone, so real-time processing of each buffer is easily possible while the next one is being recorded.

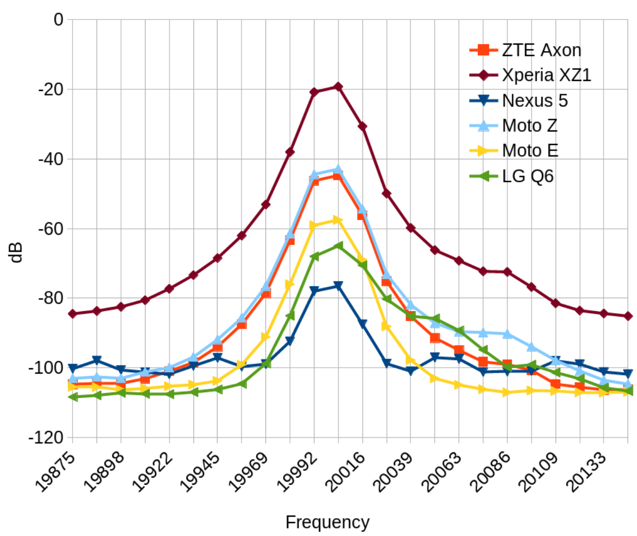

Figure 4: Relative bandwidth of 20 kHz ultrasound pilot tone, re-recorded by smartphone microphones. The x-axis shows the center frequency of each FFT bin, y-axis shows measurements in dB relative to maximum recording amplitude of the individual smartphones.

Due to spectral smearing and aliasing errors, both during playback and recording, the re-recorded signal will no longer result in a single sharp frequency peak. As shown in figure 4, the re-recorded pilot tone now covers a total of 9 FFT bins before dropping to at least -50 dB below the peak. Based on this result and the FFT parameters, we calculate a bandwidth of 9 bins * 11.72 Hz = 105.5 Hz and therefore a minimum detectable frequency shift of 52.8 Hz. Referring back to equation 2, we can now calculate the minimum detectable speed difference as $\Delta_v = 0.44$ m/s, which is well below the average walking speed of 1.4 m/s [8]. As we are only interested in detecting objects which exhibit relative motion *towards* the user, we only need to look for signals shifted to *higher* frequencies and can ignore any signals at frequencies which are lower than the pilot tone.

5 IMPLEMENTATION

From our previous experiments, we concluded that our approach is indeed feasible, and therefore proceeded to develop an Android app for further testing and field experiments. To collect data about the noise floor, we first recorded 140 samples of FFT windows from four devices (Moto E, ZTE Axon, Nexus 5, LG Q6) in static indoor environments. For each individual FFT window, we then determined the ratio between the pilot tone amplitude in the center frequency bin and any higher-frequency local maxima, i.e. potential collision indicators.

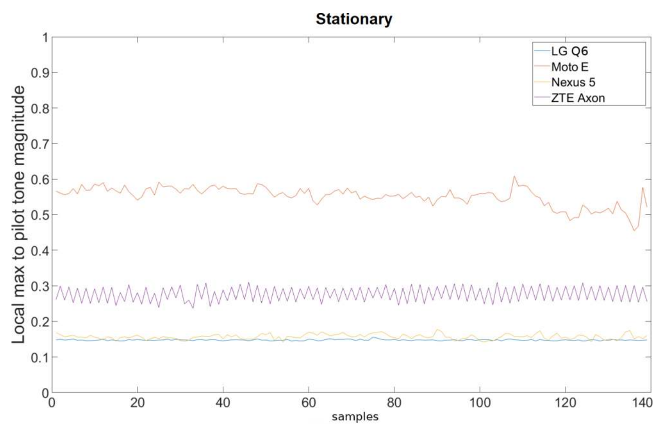

Figure 5: Signal-to-noise ratio between pilot tone amplitude and higher-frequency local maxima (stationary device).

As illustrated in Figure 5, the low-cost Motorola device has a relatively high noise floor, with higher-frequency local maxima that consistently reach about 55% of the pilot tone's amplitude. On the other hand, the mid-range and high-end devices have at most between 15% and 30% of noise floor. We, therefore, selected the LG Q6 for further experimentation, as it provided the lowest average signal-to-noise ratio.

However, a second test with 160 samples recorded during device usage while walking around in an open area (figure 6) showed that motion in the immediate environment of the device (e.g. the users' fingers) will raise the noise floor considerably, and also increase its randomness.

Based on these results, we implemented an initial calibration stage for our app, in which the user is asked to walk around in

[2] https://github.com/berndporr/kiss-fft

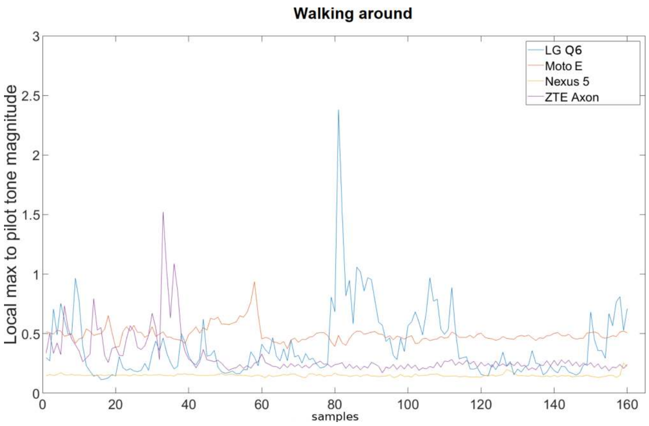

Figure 6: Signal-to-noise ratio between pilot tone amplitude and higher-frequency local maxima (walking around).

an open area and touch a series of random targets on the screen for 30 seconds (to simulate typical usage patterns). During this period, the pilot tone is emitted and the recorded data analyzed to determine the noise floor, which is then subsequently used as a base for setting the warning threshold. We use Tukey's upper fence (k = 3) of the local maxima collected during the calibration stage, and multiply the result with an empirically-determined factor of 1.35 to calculate the final threshold. This factor was determined based on a series of indoor and outdoor tests in which a person carrying the smartphone walked towards various obstacles (wall, other persons) while the device emitted the pilot tone and simultaneously recorded the reflected signal.

Once the threshold has been determined, the app will run in the background and continuously play the pilot tone while analyzing the recorded data. We currently do not have exact data on power consumption, but we assume a CPU utilization of roughly 30% on mid-range devices and therefore a corresponding runtime reduction by about 1/3rd. The sound playback and recording itself does not have a noticeable impact on battery lifetime.

If an FFT window shows a peak higher than the pilot tone's frequency which is a) above the pilot tone band and b) has a magnitude higher than the threshold for this specific frequency bin, a notification as shown in figure 7 is displayed. We intentionally kept the interface very simple and limited to a notification popup to immediately catch the users' attention, regardless of the foreground app currently in use. Users also have the option to give immediate feedback on whether the warning was a true or a false positive. To reduce the latter, the collision detection is disabled when the device's distance sensor registers an object in close proximity, e.g. the user's head during a phone call or the inside of a pocket or purse.

6 EVALUATION

We conducted an evaluation in an urban environment around the local university with 5 participants who walked along a predefined path with various obstacles (sculptures, doors, walls, pedestrians, passing traffic) to gain a realistic overview of the capabilities of our system. We used the LG Q6 for this test, as it provided the best

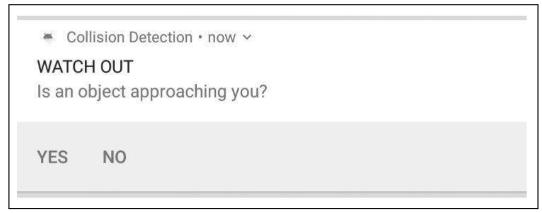

Figure 7: Collision warning popup.

overall ultrasound audio characteristics as described previously. Participants were provided with a news app and asked to read news articles during the test to simulate common usage patterns. An observer shadowed the participants during their walk and recorded all noteworthy events, but took care to keep a distance of at least 3 meters at all times (except for obstacle 3, where the observer intentionally simulated a pedestrian passing by the user at close range).

The evaluation path is shown in figure 12 and includes the following obstacles: (1) large wooden chair sculpture, (2) passing bus, (3) walking person (observer), (4) building entrance/exit doors, (5) brick wall, (6) outer wall of sculpture. See also figure 8 for a visual impression of the individual obstacles. In all cases, we asked participants to pass by the obstacles at a distance of at most 50 cm, ideally without slowing down (except for the doors, which participants had to open manually).

Figure 8: Obstacles 1 (left) to 6 (right) on evaluation path.

7 RESULTS

For our 5 participants and 6 planned obstacles, we expected a total of 30 collision warnings. Of these, 23 collisions (76.7%) were detected correctly, as opposed to 7 false negatives. Some of the obstacles such as the leg of the giant chair sculpture are apparently more difficult to detect (only for 2 out of 5 participants), likely due to smaller total cross-section. For the passing bus, we noticed that the two participants which did not receive a warning stood farther away from the curb than the others, so the combination of object speed, ambient noise, and distance is close to the detection limit in this case (at a speed of approximately 35 km/h, the bus will only be on approach within the 3 m detection range for about 0.3 seconds, corresponding to at most 2-3 FFT windows that can reliably be expected to detect a reflection at all). For the four other obstacles,

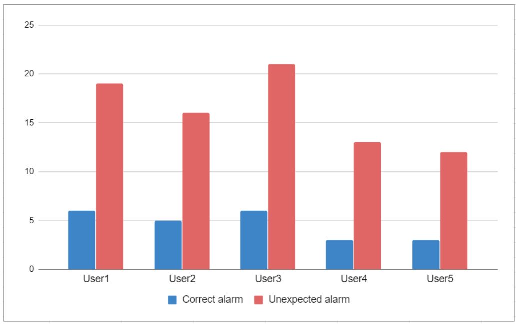

Figure 9: Expected and unexpected warnings per user. Expected warnings correspond to the 6 planned obstacles, unexpected warnings are both false positives and unplanned true positives (see text).

only two collisions in total were not detected.

When looking at the individual number of warnings for each test participant (figure 9), it is apparent that a) the total number of warnings is different for each user, and b) that those users that received less warnings in total also received less true positives. At the moment, we can only speculate about the factors influencing this difference; it is likely to depend on a combination of individual posture and way of holding the mobile as well as clothing worn by the user (e.g. a thick coat will reflect less sound energy, particularly if the speaker on the mobile device is oriented towards the user).

While the total number of 81 unexpected warnings may appear unacceptably high at first glance, the picture changes if we take our observations of the environment during the evaluation into account. In 38.3% of cases, random pedestrians passed the test persons at the moment of the alert and in 16.0% of cases, cars on the road triggered a warning. 9.9% and 3.7% of warnings were triggered when approaching walls or poles, respectively, and 18.5% of unexpected warnings happened while passing through the building (obstacle 4). This leaves us with 13.6% of instances where an unexpected warning happened without any recognizable external trigger; these are the false positives which can only be mitigated by improvements of the signal processing.

8 IMPROVED SIGNAL PROCESSING

The main limitation of our system is the high number of false positives, even if taking external triggers into account. In our evaluation, 13.6% of warnings could not be assigned to a category and were triggered spuriously, which would likely desensitize users to the warnings over the long term. One possible reason may be the continuously emitted pilot tone, which can easily cause short-term reflections that trigger a warning, even if the signal already vanishes again after 1-2 consecutive FFT frames.

To mitigate the high number of spurious warnings, we implemented a second variant of our system which uses a modified approach to collision detection. Instead of a continuous ultrasound signal, we now emit short "chirps" of 50 ms duration as illustrated in figure 10, with pauses of approximately 150 ms in-between signals.

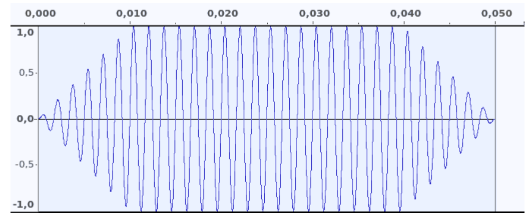

Figure 10: Pulse shape of ultrasound chirp (frequency not to scale).

Every chirp has a ramp-up and ramp-down phase of 10 ms duration each to avoid audible clicking noise as a side effect of signals starting or stopping at full volume. Other ultrasound sensing applications using commodity hardware, such as e.g. [6] by Jin et al., have used a similar approach. We do not expect any impact on power consumption, as the sound playback itself only has negligible influence when compared to the CPU load from the continuously-running FFT.

In addition, we also shift the base frequency of every chirp randomly within a predefined range between 19.5 and 20.5 kHz in steps of 200 Hz, and use the known base frequency of the recorded chirp to only check for reflections within a narrower band directly above the pilot tone frequency, similar to the "frequency hopping spread spectrum" (FHSS) method used in various radio transmission standards such as Bluetooth. We select this approach to mitigate potential signal collisions between devices in the vicinity that also emit ultrasound, e.g. car parking sensors.

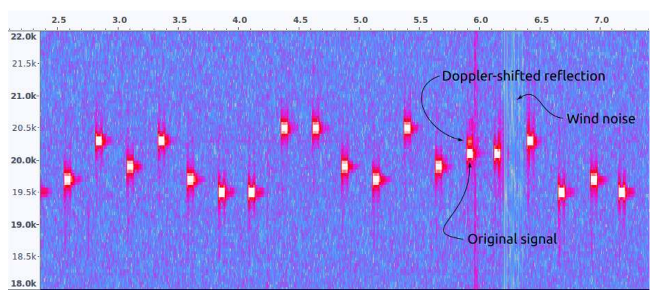

Figure 11: Spectrum plot of ultrasound chirps with simulated collision at 6.0 s.

In figure 11, a spectrum plot of a sample run recorded with our previously used external audio equipment (Behringer UMC404 + Audix TM-1 at 192 kHz sample rate and 16 bit resolution) is shown. Each frequency-shifted chirp is clearly visible. At 6.0 seconds in this recording, we simulated a collision by approaching the setup with a solid object (cardboard box, 15x20 cm). As expected, the chirp at 6.0 seconds shows an additional Doppler-shifted frequency line above the base frequency.

As the shape of each re-recorded chirp in the frequency domain is quite consistent (similar to a Gaussian distribution, see also figure 4), we can now use a template-matching approach to detect imminent collisions. Based on the known center frequency of each recorded chirp, we calculate an exponential running average for the

25 frequency bins on each side of the base frequency and subtract this signal template from the individual chirp signal. We then calculate the new center frequency of the chirp based on the difference values. If no Doppler shift has occurred, then the center frequency of the difference will be very close to the original base frequency; otherwise, it will shift to higher frequencies. If a shift above one standard deviation is detected in at least two consecutive chirps, we trigger a warning just like in the first prototype.

9 DISCUSSION & FUTURE WORK

When in use, our system will inevitably contribute to an increased ultrasound noise floor that might affect the environment, particularly animals. However, bats usually employ noticeably higher ultrasound frequencies in the 20-60 kHz range [3], so our system operates at the very low end of their hearing range and is therefore unlikely to have a large impact. We assume that other ultrasound-sensitive animals such as dogs are already used to similar signals in an urban environment, e.g. from car distance sensors, and therefore unlikely to be affected either.

One limitation of our setup is that the threshold multiplier of 1.35 has been empirically determined, but not in a strictly controlled environment and with a "convenience sample" of devices. This also applies to the evaluation, which contained various random encounters and unpredictable environmental conditions. Changes to any parameters might require a modified value for the multiplier; however, we assume that our automated running average calibration used in the improved implementation (cf. section 8) will address this issue.

Another notable limitation of our system is that HeadsUp will not work when the user is wearing headphones, as the default audio playback will then route the ultrasound output to the headphones, too. On some Android devices, it is possible to explicitly select the external speaker for a specific audio stream, however, this feature is hardware- and implementation-dependent. Moreover, HeadsUp is only able to warn about objects with a certain minimum cross-sectional area and will fail to detect e.g. a fence as the one near obstacle 3. In addition, if a system like HeadsUp were to gain wider traction, then the signals from other devices in the vicinity could potentially also trigger additional false positives, even if both devices employ random frequency hopping.

We see several options to mitigate these issues through future improvements of our system. To reduce the number of spurious warnings created in indoor environments, a straightforward approach would be to use GPS to only activate HeadsUp when walking outside. In addition, we will perform an additional in-the-wild evaluation of our improved signal processing approach. We will also investigate whether it is feasible to train a 1D convolutional neural network (CNN) using the data collected in the first and second test run to perform more accurate collision recognition and whether a frequency-modulated continuous wave (FMCW) approach (usually used for range detection) might also be suitable for our usage scenario. A future larger-scale deployment would also require a more elaborate user interface that allows to set parameters such as expected walking speed, audio routing, warning threshold, etc.

10 CONCLUSION

We have presented HeadsUp, a mobile collision warning system using unmodified commodity smartphones. HeadsUp emits ultrasound signals and simultaneously listens for Doppler-shifted reflections of these signals that indicate an approaching object, displaying a warning to the user who may be distracted by using their smartphone. To assess the feasibility of our approach, we first characterized the ultrasound playback and recording capabilities of six different smartphone models. In our real-world study, HeadsUp detected 76.7% of expected collisions (23 out of 30 staged situations). In addition, our system also produced 81 additional warnings in total, of which 13.6% could not be linked to any external trigger. To reduce spurious warnings, we implemented a frequency-hopping approach with pulsed ultrasound chirps. In the future, we will investigate whether the updated signal processing improves real-world results and whether a neural network or a FMCW-based approach can be used to further augment our system.

11 REPRODUCTION NOTE

All data and source code for this paper is available at Github via https://github.com/mmbuw/headsup.

ACKNOWLEDGMENTS

We thank our study participants for their help with our evaluation. This work was in part funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through project grant EC437/1-1.

REFERENCES

[1] Sonda Ammar Bouhamed, Imen Khanfir Kallel, and Dorra Sellami Masmoudi. 2013. Stair Case Detection and Recognition Using Ultrasonic Signal. In *2013 36th International Conference on Telecommunications and Signal Processing (TSP)*. 672–676. https://doi.org/10.1109/TSP.2013.6614021

[2] Ashley Colley, Lasse Virtanen, Pascal Knierim, and Jonna Häkkilä. 2017. Investigating Drone Motion as Pedestrian Guidance. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia* (Stuttgart, Germany) *(MUM '17)*. Association for Computing Machinery, New York, NY, USA, 143–150. https://doi.org/10.1145/3152832.3152837

[3] M B Fenton, C V Portfors, I L Rautenbach, and J M Waterman. 1998. Compromises: sound frequencies used in echolocation by aerial-feeding bats. *Canadian Journal of Zoology* 76, 6 (1998), 1174–1182. https://doi.org/10.1139/z98-043 arXiv:https://doi.org/10.1139/z98-043

[4] Biying Fu, Jakob Karolus, Tobias Grosse-Puppendahl, Jonathan Hermann, and Arjan Kuijper. 2015. Opportunities for Activity Recognition Using Ultrasound Doppler Sensing on Unmodified Mobile Phones. In *Proceedings of the 2nd International Workshop on Sensor-Based Activity Recognition and Interaction* (Rostock, Germany) *(iWOAR '15)*. Association for Computing Machinery, New York, NY, USA, Article 8, 10 pages. https://doi.org/10.1145/2790044.2790046

[5] Sidhant Gupta, Daniel Morris, Shwetak Patel, and Desney Tan. 2012. SoundWave: Using the Doppler Effect to Sense Gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1911–1914. https://doi.org/10.1145/2207676.2208331

[6] Haojian Jin, Christian Holz, and Kasper Hornbæk. 2015. Tracko: Ad-Hoc Mobile 3D Tracking Using Bluetooth Low Energy and Inaudible Signals for Cross-Device Interaction. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 147–156. https://doi.org/10.1145/2807442.2807475

[7] Hye-Jin Kim, Jin-Young Min, Hyun-Jin Kim, and Kyoung-Bok Min. 2017. Accident risk associated with smartphone addiction: A study on university students in

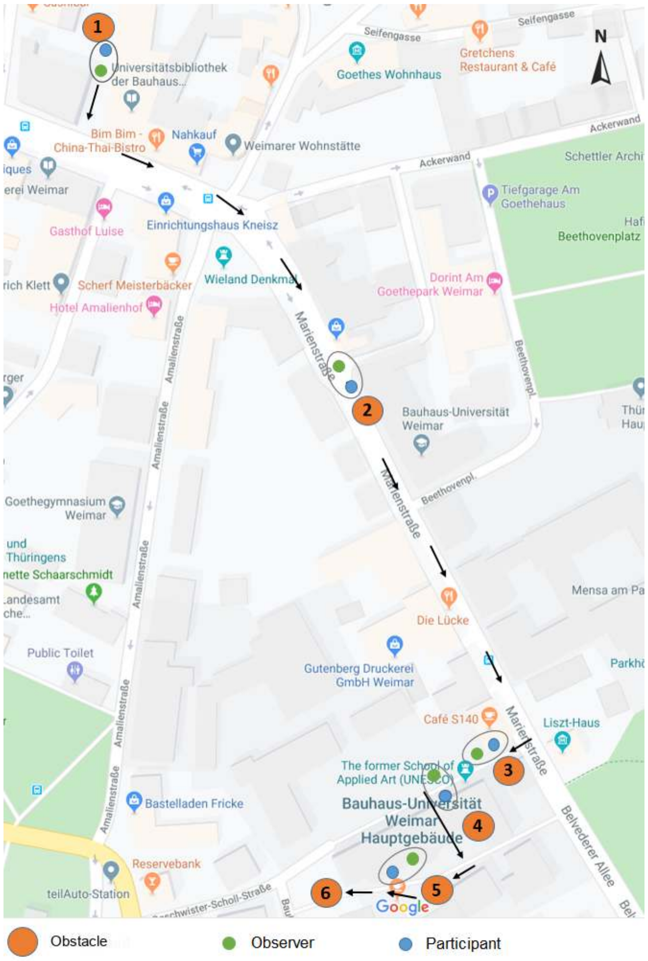

Figure 12: Evaluation walking path. Map data by Google. © 2020 GeoBasis-DE/BKG (© 2009)

Korea. *Journal of Behavioral Addictions* 6 (11 2017), 1–9. https://doi.org/10.1556/2006.6.2017.070

[8] Robert V. Levine and Ara Norenzayan. 1999. The Pace of Life in 31 Countries. *Journal of Cross-Cultural Psychology* 30, 2 (1999), 178–205. https://doi.org/10.1177/0022022199030002003

[9] "CGActive LLC". 2009. *Type 'n' Walk.* Retrieved Oct 17, 2022 from http://www.type-n-walk.com/

[10] Markus Löchtefeld, Matthias Böhmer, and Lyubomir Ganev. 2013. AppDetox: Helping Users with Mobile App Addiction. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia* (Luleå, Sweden) *(MUM '13)*. Association for Computing Machinery, New York, NY, USA, Article 43, 2 pages. https://doi.org/10.1145/2541831.2541870

[11] Cara McGoogan. 2016. *German city installs traffic lights in pavements to protect texting pedestrians.* Retrieved Oct 17, 2022 from https://www.telegraph.co.uk/technology/2016/04/26/german-city-installs-traffic-lights-in-pavements-to-protect-text/

[12] Damien Murphy. 2010. *Pedestrian death rise blamed on iPods.* Retrieved Oct 17, 2022 from https://www.smh.com.au/technology/pedestrian-death-rise-blamed-on-ipods-20100905-14w4d.html

[13] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Jacob E. Sunshine. 2019. Opioid overdose detection using smartphones. *Science Translational Medicine* 11, 474 (2019). https://doi.org/10.1126/scitranslmed.aau8914 arXiv:https://stm.sciencemag.org/content/11/474/eaau8914.full.pdf

[14] BBC News. 2017. *Smartphone use blamed for road deaths.* Retrieved Oct 17, 2022 from https://www.bbc.com/news/technology-39453497

[15] Jeff Parsons. 2019. *South Korea tackles smartphone 'zombies' with flickering lights.* Retrieved Oct 17, 2022 from https://metro.co.uk/2019/03/19/south-korea-tackles-smartphone-zombies-flickering-lights-8935321/

[16] António Pereira, Nelson Nunes, Daniel Vieira, Nuno Costa, Hugo Fernandes, and João Barroso. 2015. Blind Guide: An Ultrasound Sensor-based Body Area Network for Guiding Blind People. *Procedia Computer Science* 67 (Jan. 2015), 403–408. https://doi.org/10.1016/j.procs.2015.09.285

[17] Aliaksei L. Petsiuk and Joshua M. Pearce. 2019. Low-Cost Open Source Ultrasound-Sensing Based Navigational Support for the Visually Impaired. *Sensors* 19, 17 (Aug 2019), 3783. https://doi.org/10.3390/s19173783

[18] Andreas Riener and Harald Hartl. 2012. "Personal Radar": A Self-Governed Support System to Enhance Environmental Perception. In *Proceedings of the 26th Annual BCS Interaction Specialist Group Conference on People and Computers* (Birmingham, United Kingdom) *(BCS-HCI '12)*. BCS Learning & Development Ltd., Swindon, GBR, 147–156. https://doi.org/10.14236/ewic/HCI2012.20

[19] Maximilian Schirmer, Johannes Hartmann, Sven Bertel, and Florian Echtler. 2015. Shoe Me the Way: A Shoe-Based Tactile Interface for Eyes-Free Urban Navigation. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Copenhagen, Denmark) *(MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 327–336. https://doi.org/10.1145/2785830.2785832

[20] Byeong-Seok Shin and Cheol-Su Lim. 2007. Obstacle Detection and Avoidance System for Visually Impaired People. In *Haptic and Audio Interaction Design (Lecture Notes in Computer Science)*, Ian Oakley and Stephen Brewster (Eds.). Springer, Berlin, Heidelberg, 78–85. https://doi.org/10.1007/978-3-540-76702-2_9

[21] Ke Sun, Ting Zhao, Wei Wang, and Lei Xie. 2018. VSkin: Sensing Touch Gestures on Surfaces of Mobile Devices Using Acoustic Signals. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking* (New Delhi, India) *(MobiCom '18)*. Association for Computing Machinery, New York, NY, USA, 591–605. https://doi.org/10.1145/3241539.3241568

[22] James Titcomb. 2017. *Pavement lights installed near Dutch schools to guide smartphone-obsessed pedestrians.* Retrieved Oct 17, 2022 from https://www.telegraph.co.uk/technology/2017/02/17/pavement-lights-installed-near-dutch-schools-guide-smartphone/

[23] Yu-Chih Tung and Kang G. Shin. 2018. Use of Phone Sensors to Enhance Distracted Pedestrians' Safety. *IEEE Transactions on Mobile Computing* 17, 6 (June 2018), 1469–1482. https://doi.org/10.1109/TMC.2017.2764909

[24] Anran Wang and Shyamnath Gollakota. 2019. MilliSonic: Pushing the Limits of Acoustic Motion Tracking. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3290605.3300248

[25] Asavin Wattanajantra. 2010. *Zombie 'iPod pedestrians' endangered by mobile oblivion, says AA.* Retrieved Oct 17, 2022 from https://www.cnet.com/news/zombie-ipod-pedestrians-endangered-by-mobile-oblivion-says-aa/

[26] Dominik Weber, Alexandra Voit, Anja Exler, Svenja Schröder, Matthias Böhmer, and Tadashi Okoshi. 2017. Intelligent Notification and Attention Management on Mobile Devices. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia* (Stuttgart, Germany) *(MUM '17)*. Association for Computing Machinery, New York, NY, USA, 561–565. https://doi.org/10.1145/3152832.3152840

[27] Yonhap. 2022. *In-ground traffic lights installed across Seoul for 'smartphone zombies'.* Retrieved Oct 17, 2022 from https://www.koreaherald.com/view.php?ud=20220111000281