Opportunities for Activity Recognition using Ultrasound Doppler Sensing on Unmodified Mobile Phones

Biying Fu[1]**, Jakob Karolus**[2]**, Tobias Grosse-Puppendahl**[1]**, Jonathan Hermann**[1]**, Arjan Kuijper**[1,2]

[1]Fraunhofer IGD, Fraunhoferstr. 5, 64283 Darmstadt, Germany, {firstname.lastname}@igd.fraunhofer.de
[2]Technische Universität Darmstadt, Karolinenplatz 5, 64289 Darmstadt, Germany
jakob.karolus@stud.tu-darmstadt.de, arjan.kuijper@gris.tu-darmstadt.de

ABSTRACT

Nowadays activity recognition on smartphones is ubiquitously applied, for example to monitor personal health. The smartphone's sensors act as a foundation to provide information on movements, the user's location or direction. Incorporating ultrasound sensing using the smartphone's native speaker and microphone provides additional means for perceiving the environment and humans. In this paper, we outline possible usage scenarios for this new and promising sensing modality. Based on a custom implementation, we provide results on various experiments to assess the opportunities for activity recognition systems. We discuss various limitations and possibilities when wearing the smartphone on the human body. In stationary deployments, e.g. while placed on a night desk, our implementation is able to detect movements in proximities up to 2 m as well as discern several gestures performed above the phone.

Author Keywords

ultrasound sensing; activity recognition; acoustic sensing

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation: User Interfaces - Graphical user interfaces; Input devices & strategies

INTRODUCTION

Streaming classification techniques are widely used in mobile devices to recognize human behaviours and contexts. This is extremely valuable to realize implicit interaction systems, for example to support healthy and independent living. The most important parameters to sense include indoor location, gestures, heart-rate, or emergencies like falls.

Mobile devices bring along many sensors that can be used for this purpose. However, interaction is mostly centered around the device and information from the distant environment is harder to capture with integrated sensors. Besides that, the applicability of certain sensors can be limited by the
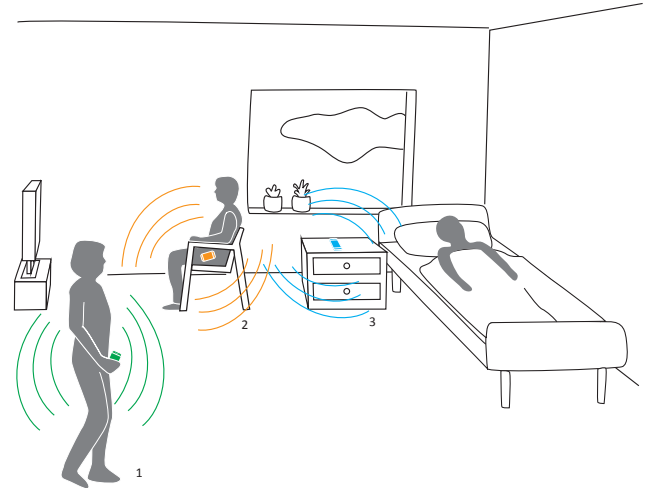
Figure 1: Ultrasound sensing using the smartphone's native speaker and microphone can be used to capture information in various scenarios: (1) holding the phone, (2) carrying the phone on the body, and (3) stationary deployments.

use-case. For example, camera-based systems induce privacy issues and body-worn systems are sometimes inconvenient to wear over long periods.

In this paper, we investigate the use of ultrasound to support new, unobtrusive, sensing possibilities on the mobile phone to overcome these issues. We present related work and initial experiments that outline opportunities for activity recognition in wearable and stationary settings. In our exemplary implementation we emit an ultrasound wave with the mobile phone's speaker. The phone's microphone picks up the reflected signals and enables to derive information on objects at distances up to 2 m. In order to demonstrate the capabilities of ultrasound sensing, we use continuous-wave with a center frequency at 20 KHz.

In the targeted scenario, a consumer-grade smartphone could potentially nullify the need for additional hardware equipment. It could realize new use-cases, such as detecting falls while placing the smartphone on the night desk. This provides the consumer with a simple-to-execute solution that can be installed conveniently as a single mobile application. However, there are numerous challenges which are mainly

induced by the heterogeneity of target systems. Firstly, this comprises the use of low-cost to high quality microphones and speakers. Secondly, the target system's software and hardware capabilities are very different in terms of processing power and battery runtime.

RELATED WORK
The use of ultrasound as a sensing modality has been widely investigated in the research literature. In [8] the feasibility and limitations of ultrasound sensing using mobile phones is explored.

However, most previous works focus on custom-built systems, instead of consumer-grade devices without hardware modifications. The signal analysis can be based on measuring timespans until the signal reflects at an object. Exploiting the Doppler shift caused by reflections at moving objects can be a suitable indicator which is less sensitive to timing constraints [23]. This enables to recognize body parts moving away or towards a sound pickup [11]. Ultrasound signals from other devices allow for implementing active messaging systems with the possibility to localize objects.

Gesture Recognition
Detecting finger and hand movements in free air is often achieved by analyzing a backscattered ultrasound signal. [15] apply ultrasonic waves to unobtrusively recognize one-handed gestures on custom hardware. The authors employ the Doppler effect caused by a moving human body part. Here, a single transmitter and three receiver microphones are sufficient to determine a 3D movement. Due to the availability of ultrasound capabilities in consumer hardware, ultrasound approaches have also been ported to consumer smartphones and laptops. SoundWave enables to recognize gestures in front of the screen of an unmodified consumer-grade laptop [11]. Very similarly, Dolphin detects various gestures performed above a consumer smartphone using the system's speaker and microphone [22]. Although the Doppler measurements are not discriminant in theory, the authors also argue that left-and right-swipe gestures can be classified on a per-user basis.

Activity Recognition, Localization & Context Awareness
In order to detect whole-body movements, worn nodes as well as passively backscattered signals can be used. Multiple microphones allow for ranging and localization based on the angle or time-difference of arrival. Tarzia et al. [30] use a similar technique to determine user presence near an unmodified laptop. Using a measurement window of only 10 s results in an accuracy of approximately 96 %, discriminating the two classes of absence and presence. Extending the windows to 25 s almost exceeds perfect accuracy.

In WALRUS by Borriello et. al [5], an indoor localization service was realized by leveraging the properties of ultrasound and radio frequencies. In each room a pc emits wifi messages and ultrasound pulses on a regular basis. As the mobile phone only picks up the sound signal from the respective room the user resides in, it can be related to the wifi message that was used to activate the recording. A similar approach with contrary execution is presented in ASSIST [13]. Here, the phone

serves as the transmitter emitting a high frequency chirp signal, while several custom made receivers pick up the signal. These are time synced via a Wifi network, which enables the authors to use the difference in arrival time of the chirps to estimate the smartphone location up to an error margin of 30 cm.

Using custom hardware, Kalgaonkar et al. [14] are able to detect whether a person talks in front of an ultrasound Doppler sensor. Sound propagations on human skin can be applied to measure touches on gestures with combined hardware worn on the finger and arm [18]. Attaching ultrasound emitters to various parts of the body and a transceiver worn around the neck enables the authors of [32] to derive various activities like sitting and cleaning. Rossi et al. [26] measure the impulse response of different environments with a mobile phone to infer an indoor location. The authors are able to distinguish between more than 20 rooms. Similarly on a smaller scale, emitting sound and vibrations can be used to excite the underlying surface of a device and thus enable localization [17]. Very recently, the authors of [19] presented an approach to recognize sleeping apnoea using unmodified mobile phones.

Multi-Device Interaction
Ultrasound messaging and ranging represents a very convenient way to identify and track multiple devices. Techniques which use simple backscattering are not able to identify objects and thus leave a certain amount of ambiguity in their results. Embedding information in emitted ultrasound signals can thus enable for object and person localization [3]. Relying on a secondary modality with faster wave propagation enables to reliably calculate the absolute time-of-flight of the ultrasound signal between two devices [5, 27, 12].

Approaches for device selection on mobile phones, e.g. by exploiting the Doppler effect when pointing at a device, has been investigate in [24, 29]. BeepBeep [20] solely relies on ultrasound generated by smartphones, which supports ranging other smartphones. When more than three objects are involved, it is possible to determine a relative position based on the information of multiple devices. Sensing device location in a car was investigated by [33]: By classifying the mobile phone's position it is possible to differentiate between the driver using the phone and a passenger. Reynolds et al. [25] introduce an ultrasound position sensing system for tangible objects above LCD-screens. The authors present interactive 'pucks' that communicate by emitting and receiving ultrasound and reconstruct their position.

Alternative Technologies
The area of ultrasound sensing is closely related to radio-frequency based techniques, as both are based on a wave propagation [23]. The use-cases are very similar but require a higher temporal resolution to analyze Doppler shifts and time-of-flight measurements. Zhao et al. [34] and Kellogg et al. [16] use the reflections of existing RF signals to recognize various types of gestures. In stationary installations, gesture recognition and indoor localizations was realized using reflections of radio waves by the human body [1, 21]. Refining
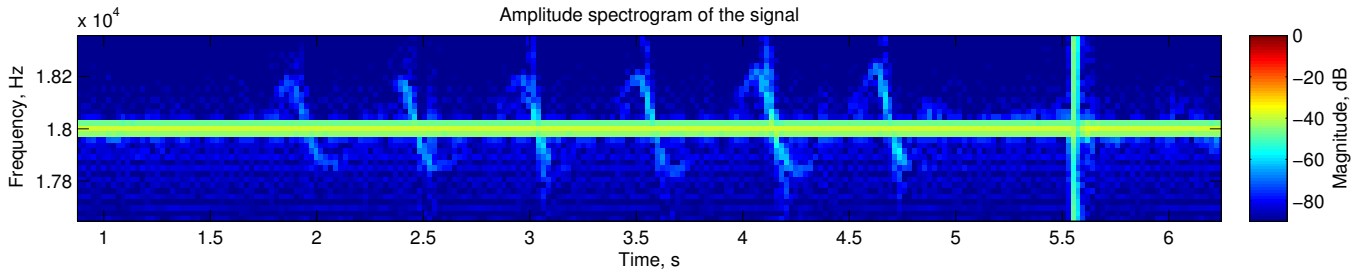
Figure 2: An exemplary spectrogram: The Doppler-shift is caused by a waving hand in distances between 0.1 and 0.3 m atop the smartphone. At second 5.5, the user claps both hands, resulting in a large noise overlay.

these techniques even enable to recognize speech or breathing rate from a distance [2, 31]. Other techniques which are suitable for recognizing the presence and proximity of objects may also comprise quasi-electrostatic field measurements [9].

Summary
From a use-case perspective, ultrasound sensing has been applied to measure a great variety of environmental parameters. The physical sensing setups are often based on custom hardware, with a recent trend towards existing mobile devices like laptops and smartphones. In contrast to previous works on unmodified mobile phones, we provide an explorative view on activity recognition in this paper. We conduct several experiments to outline the method's current opportunities and particularly its limitations. We identified three major setups for ultrasound sensing on mobile phones: (1) Stationary deployments (e.g. on a table), (2) holding the phone in a hand, and (3) carrying the phone on the body. In the following, we will describe the experiments to investigate the different deployments.

PHYSICAL PRINCIPLES OF DOPPLER SENSING
Currently established range and movement measurements are often based on pulsed radar. Here, a short pulse or a burst of short pulses is generated and the reflections are measured. This technique allows for inferring distance measurements to nearby objects. Other methods emit a continuous wave (CW) with a fixed frequency or a frequency modulated continuous wave (FMCW). The first technique easily allows for recognizing Doppler effects caused by movements, while FMCW also provides static distance measurements. In contrast to FMCW and pulsed methods, simple CW does not induce demanding timing constraints and is easily realizable on a wide range of smartphones. For this reason, we use continuous wave modulation in our experiments.

A continuous wave, which can be a pure sine-wave with a frequency f_0 is sent out from the system's speaker. We use the low ultrasound range at 20 kHz, which is about the maximum frequency that can be achieved with the phone's hardware components. A reflection from a moving target broadens the frequency spectrum around the central carrier frequency, which is called the Doppler frequency shift. An approaching target induces a shortening of the received wave-front, which means the frequency increases and a positive Doppler shift is observed. Respectively, a departing target leads to an expanded wave-front with a negative frequency shift. In order

to measure the Doppler frequency, Fast Fourier Transform (FFT) is used. The number of samples N_{FFT} used by the Fast Fourier Transform (FFT) determines the individual bin width of Δf. The bin width determines further the resolution of the measurable Doppler frequency.

Sampling an audio signal using the in-built microphone of a smartphone can usually be conducted at a maximum frequency of $f_s = 44.1$ KHz. Regarding the Nyquist theorem, the highest retrievable frequency for an ultrasound measurement is $\frac{f_s}{2} = 22.05$ KHz. Since the human ear is typically able to perceive sounds at frequencies up to 18 KHz [7], this leaves us with an effective frequency range of 4.05 KHz. In order to obtain a reasonable temporal resolution for gestures and movements, we divide the signal into overlapping windows (75 %) with a length of 93 ms. This corresponds to a 4096-point FFT with a bin resolution of 10.75 Hz. With a carrier frequency of 20 kHz, Doppler-shifts are observable with a resolution of $0.09 \, ms^{-1}$. In comparison, gestures in front of laptops were observed with a speed up to $3.9 \, ms^{-1}$ [11].

EXPERIMENTS
In this paper, we present various experiments leveraging the above mentioned technique and discuss their benefit and limitations. We analyze what kind of experiments can be conducted using a stationary deployment of a mobile phone. These include the recognition of hand gestures during which the user actively interacts with the phone, as well as passive interaction when determining activities in the vicinity of the device. Additionally, we investigate the performance of our approach when executing gestures while holding the phone as well as while wearing it on the body, e.g. in the trouser pocket.

The first part of experiments shows how the Doppler shift can be used in a controlled setup and whether it is feasible for gesture and activity recognition. The second part aims to test real life scenarios where we expect a noisy signal. These tests will show if our approach can overcome those obstacles. We conducted the experiments on a ZTE Blade, an Asus Nexus 7, and a Samsung Galaxy S3 running Android 4/5.

Stationary Deployments - Gesture Recognition
For the first set of experiments, the phone is situated on a table, while the user performs gestures above the phone. Due to

(a) Downward motion. **(b)** Upward motion. **(c)** Strong swipe motion. Right-to-left for right hand; Left-to-right for left hand. **(d)** Weak swipe motion. Right-to-left for left hand; Left-to-right for right hand.
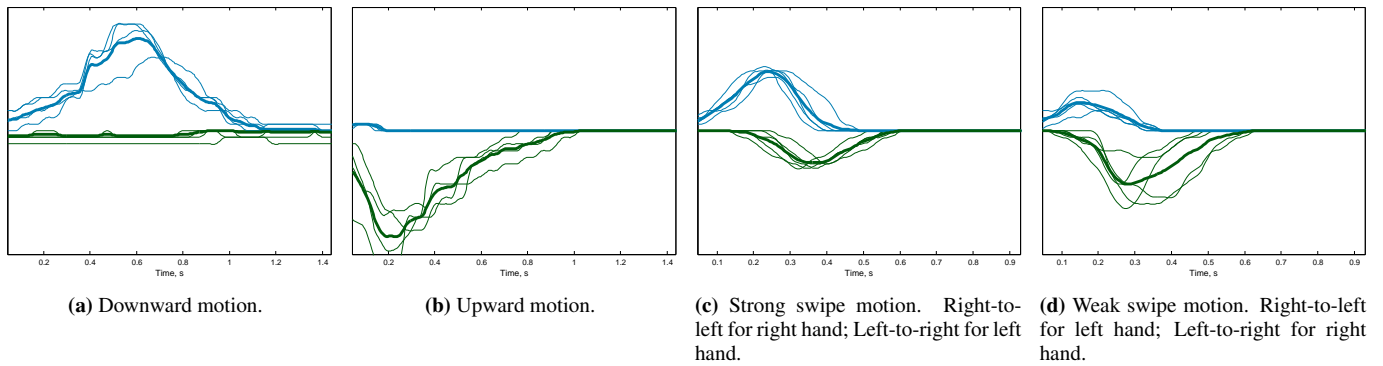
Figure 3: Broadening of the spectrum around the central carrier frequency. Thin lines indicate individual experiment runs, while the thick line represents the mean of all runs. In addition, blue colored curves depict the broadening due to a positive Doppler shift while the green curves represent the negative shifts.

the Doppler frequency shift caused by the user's hand movement, we can determine whether the hand approaches or withdraws from the device. For the downward motion, we expect a positive shift in frequency as we approach the phone, while an upward motion will cause a negative shift. This effect can be qualitatively measured by thresholding the frequency amplitude and calculating the remaining bandwidth around the central carrier frequency for each time step. A simple example for this effect in shown in figure 3 displaying the broadening of the spectrum around the central carrier frequency for various gestures. Note that for these kind of diagrams, we deliberately abstained to quantize the vertical axis to direct the focus on a qualitative interpretation of the results.

A central limitation of recognizing gesture via the Doppler Frequency shift in combination with consumer mobile phone is the fact that we can only determine the absolute change in distance of an object to the device. For example approaching the phone with the user's hand can be done in any arbitrary angle between the table and the velocity vector of the hand. As long as the euclidean distance between the hand and the phone changes at the same rate, one cannot distinguish these motions by means of the Doppler Frequency shift. Figure 4a illustrates this challenge. At the start of the experiment the user places both hands slightly above the phone and then does a zoom out and zoom in movement by moving the hands into opposite directions and back together again. The resulting shape is similar to a combination of up- and downward motions, although we used two hands and moved them horizontally with regard to the table.

However, slight variations in auxiliary movements can be picked up by the system. For example, the movements of the arm when doing a swipe gesture above the phone. Due to the nature of the Doppler effect swiping from left-to-right should be equal to swiping from right-to-left, hence these motions should not be distinguishable. This holds true for the hand movement, but not for the motion of the arm. Doing a right-to-left swipe with the right arm is more natural than doing the opposite swipe direction with the same arm. We performed several experiments using swipe motion with both arms and concluded that it is possible to distinguish the strong swiping motion from the weak one. In this scenario, a strong swipe motion would be the natural movements for the respective
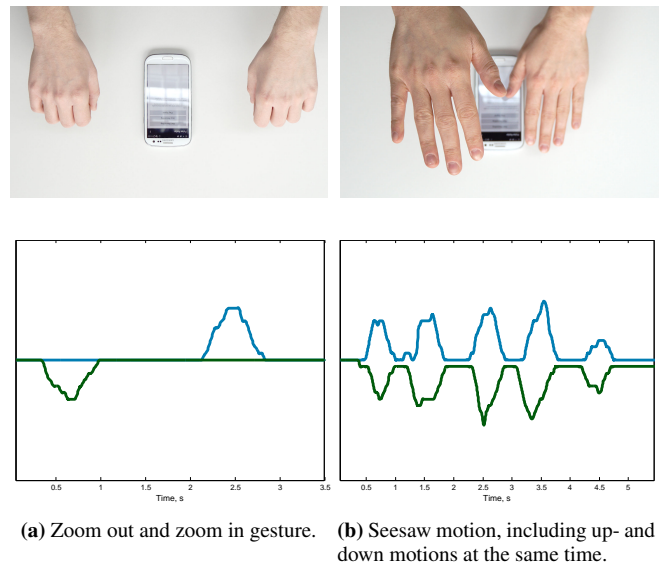

(a) Zoom out and zoom in gesture. **(b)** Seesaw motion, including up- and down motions at the same time.

Figure 4: Broadening of the spectrum for two-handed gestures.

arm, right-to-left for the right arm and left-to-right for the left arm. The weak swipe motion follows analog: The results are presented in Figure 3c,3d and depict the motion of the hand coming closer to the phone and going away again. A swipe motion from right-to-left with the right hand can be split into a fast (larger positive frequency shift) approaching motion and a slower (smaller negative frequency shift) withdraw motion. Generally speaking, the motion part that happens near the respective arm is executed faster and allows us to distinguish strong and weak motion. In an application scenario where swipes are executed using a predetermined hand, one can in fact discern between different swiping directions.

An important advantage of our approach is the possibility to recognize multiple different motions within the same time frame. Executing a motion towards and away from the phone at the same time results in two separate frequency shifts in opposite directions. In this experiment two hands are used to execute a downward and an upward motion at the same time. Upon reaching the phone with one hand the respective
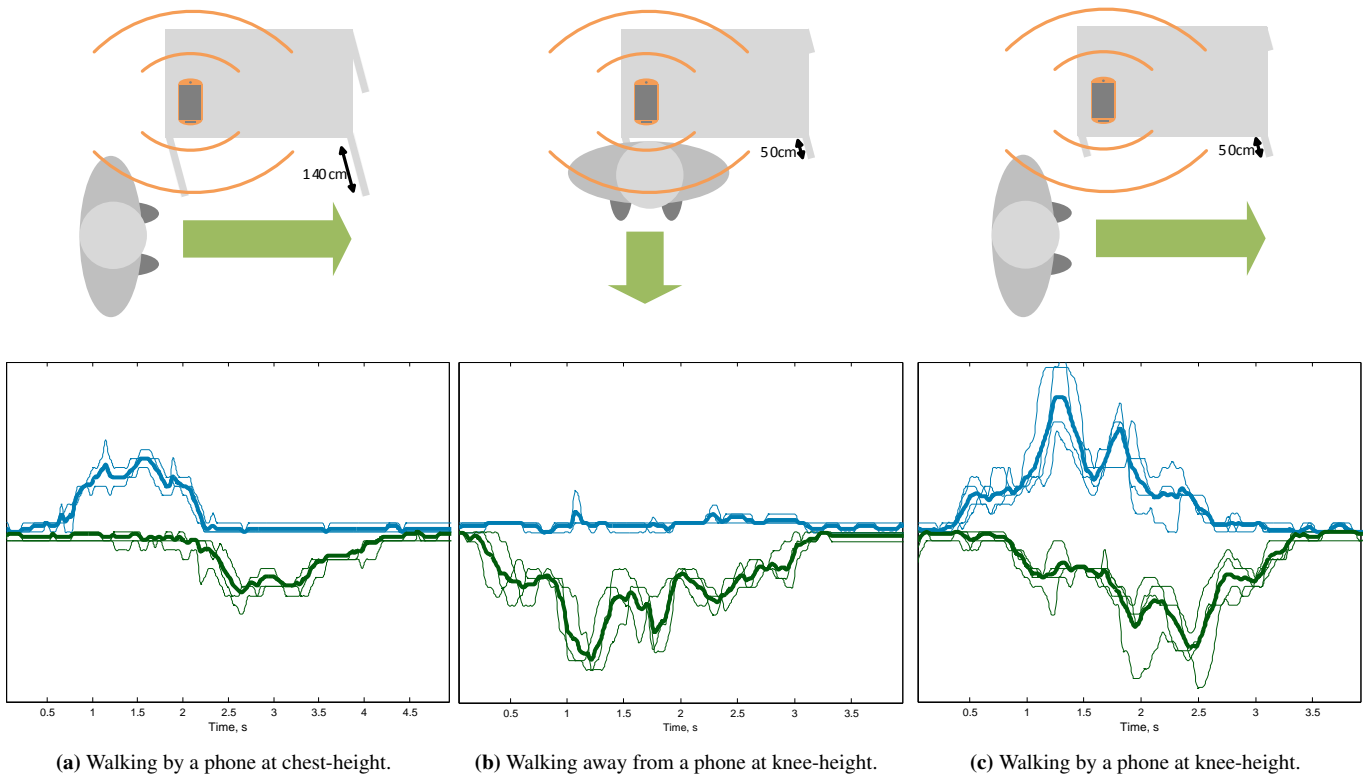
(a) Walking by a phone at chest-height.   **(b)** Walking away from a phone at knee-height.   **(c)** Walking by a phone at knee-height.

Figure 5: Broadening of the spectrum for different motions.

motion are reversed. We call this the seesaw motion. Figure 4b illustrates this scenario and shows the respective positive frequency shift for the hand going downwards as well as the negative frequency shift for the hand doing the upward motion. In the end both hand come to rest in a middle position at the same height resulting in a lower frequency shift due to the slower movement.

Stationary Deployments - Motion Recognition

Besides recognizing hand gestures, we believe that the system is able to perceive general motion or activities in the vicinity of the mobile phone. To test this hypothesis we mimicked our swipe motion during gesture recognition. In this experiment, the phone is deployed on a table at chest-height and the subject simply walks by the table. The resulting graph, shown in figure 5a, clearly depicts a motion towards and away from the phone. This experiment was designed to mostly pick up the motion of the upper body of a person. Although having a similar appearance as Figure 3c (strong swipe gesture), both can be clearly distinguished by including the duration of the movement. In particular, walking by the phone takes almost five seconds, while a swipe is generally executed way faster at around one second.

To test whether the motion of individual body parts can be detected as well we conducted another test run where the phone is deployed on a lower table, approximately at knee-height, while the subject walks away from the phone. Since the person increasing the distance to the phone, the broadening of the spectrum should show an overall negative Doppler shift,

which can be seen in figure 5b. It also shows single peak shifts corresponding to the movement of each leg. These are stronger the closer a subject is to the device, as the reflections are easier to be picked up in close proximity.

Mimicking our first test in this series, we redid the walk-by experiment with the phone placed on the lower table. We expect to pick up a stronger factor for the movement of individual body parts, like the legs, contrary to our first test, where the phone was deployed on a table at chest-height. The graph is shown in figure 5c and confirms our presumption. Clearly visible are spikes of Doppler shifts caused by the - in comparison to the overall - fast movements of each leg. Surprisingly, it seems that each part of the motion (approaching and withdrawing) has been shifted closer together. We conclude that the device registers the motion of the upper body just as in the previous experiment. However, this motion may get smeared by individual leg motion, e.g. when the body still approaches the phone, but one leg already past it.

Stationary Deployments - Activity Recognition

In this series of experiments, we aim to analyze more complex activities with the gained knowledge. For the first test, the user is laying on a bed and the mobile phone is deployed on the nearby bedside table (at knee-height). We want to know whether our approach can perceive subtle movements of a person while asleep. By doing so, one could analyze the sleeping rhythm of the user which could be beneficial for estimating the best time to sound an alarm clock. Figure 6a shows the broadening of the spectrum for a person moving
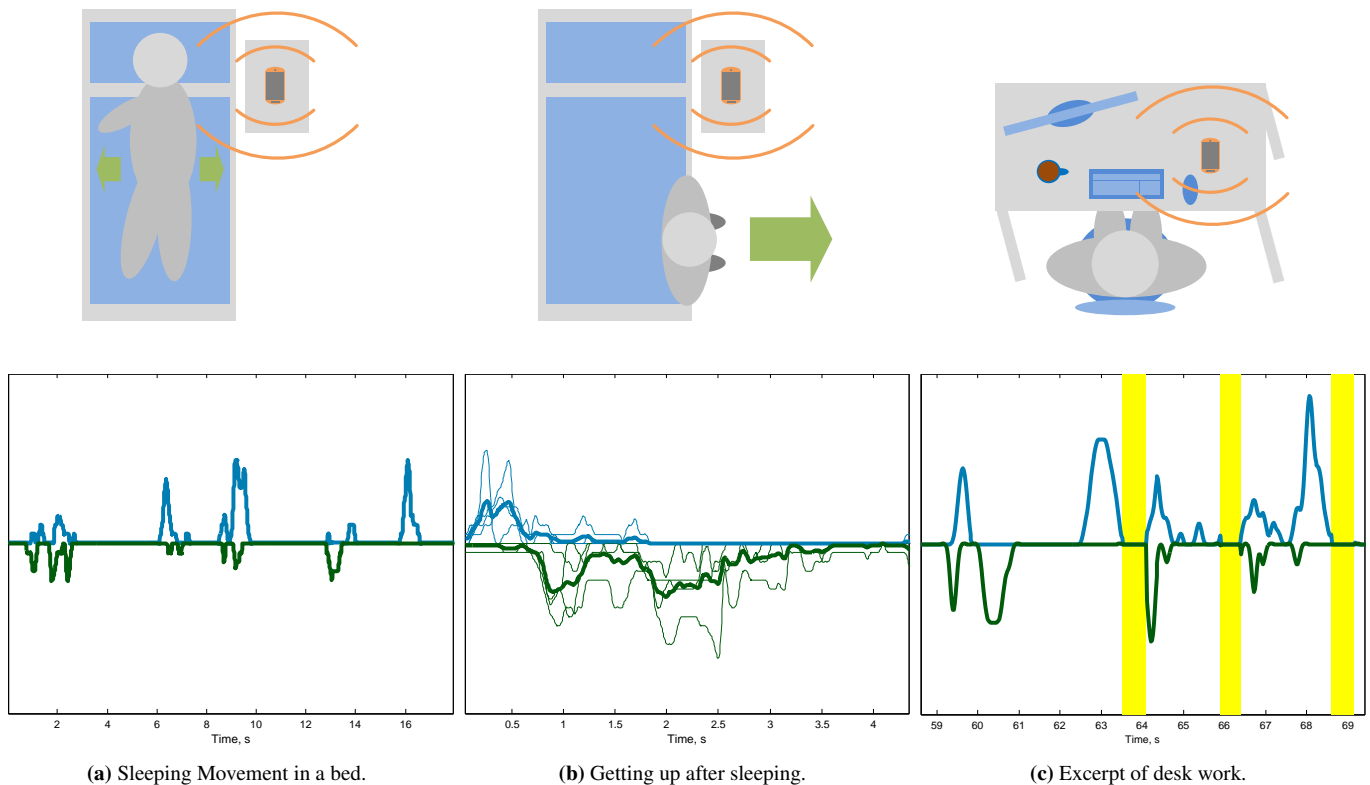
(a) Sleeping Movement in a bed. **(b)** Getting up after sleeping. **(c)** Excerpt of desk work.

Figure 6: Broadening of the spectrum for different user activities.

around on the bed, e.g. the subject changes from a face-down lying position into a dorsal position. These motions are being picked up by the device as Doppler shifts.

We want to combine this experiment with our motion experiment by letting the subject get up from the bed and walk away from it. The broadening of the spectrum for this experiment is shown in figure 6b and is consistent with the executed motion. At first, we observe a positive shift due to the subject standing up, hence decreasing distance to the phone. When walking away, we register negative frequency shifts and several peaks when the subjects moves each leg. Additionally exaggerate arm movements may also be picked up.

In a last test run, we investigate our approach in the context of everyday desk work. The user is instructed to work with a computer at the desk as usual, while we observe the signal received by the mobile phone, which is also placed on the table. For simplicity, we only show a part of the whole recording in figure 6c. The graph depicts various positive and negative Doppler shifts corresponding to hand and/or mouse movements. In the spectrum itself, we observed large peaks over the whole frequency spectrum indicating some sort of noise produced by the user. In particular, pressing keys on the keyboard may result in such sounds. To better differentiate these from the measured Doppler shifts, we replaced these peaks with yellow boxes in the figure. In some cases the peaks come paired with positive and negative Doppler shifts. In our scenario, the subject was drinking from a cup of coffee. The movement of picking up the cup and placing it on the table

can be related to such a signal response.

This series of tests indicate that the system is able to perceive general motion in the proximity as well as the movement of individual body parts. Given an appropriate data set and calibration, previously trained activities can be recognized.

Stationary Deployments - Range Limitation

This leads us to the question: What is the maximum distance to the smartphone at which we can reliable detect a Doppler shift, considering the noise from the received echo signal? The setup sees a ASUS Nexus 7 placed on top of a shelf with a height of 140 cm. This corresponds to the chest height of an average man. In the experiment, test subjects were asked to approach the device from different distances. We start at 250 cm away from the device and instruct the subject to get closer in each following step. The step width is chosen to be 50 cm which is equivalent to a normal step width. We stop the recording until we are close to the device. The result is depicted in Figure 8 and shows the spectrum of the received signal over time. However, it should be noted that the signal is filtered with a digital notch filter at the center frequency of 20 kHz, so that the main component at this center frequency coupled back from the device is reduced. The first peak appears at a distance of 200 cm away from the device and corresponds to the first step taken from 250 cm. The following peaks are separated 50 cm from each other. Therefore, the last peak is taken directly in front of the device. It should also be mentioned that in order to increase the clarity of the plot, only frequencies above 20 kHz are shown in Figure 8. Since
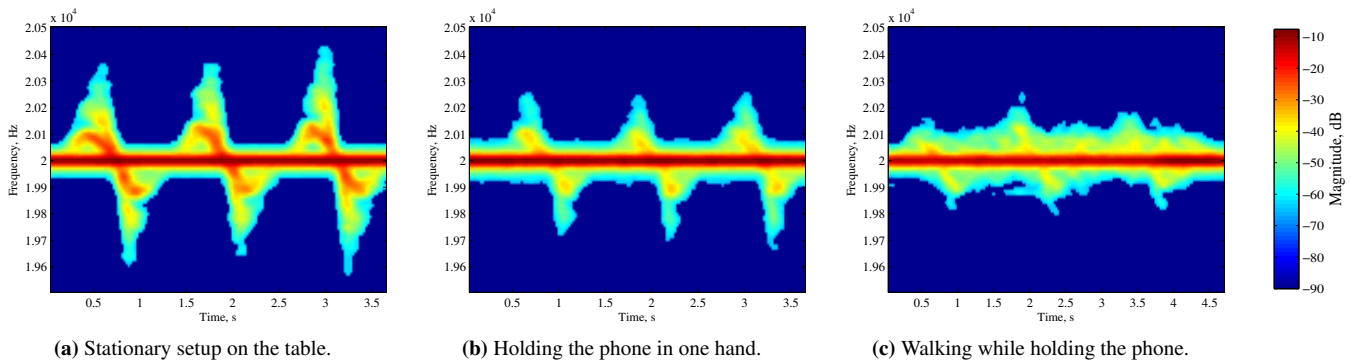
(a) Stationary setup on the table.   **(b)** Holding the phone in one hand.   **(c)** Walking while holding the phone.

Figure 7: Received spectrogram after thresholding of waving gestures (up- and downward motion) for different scenarios
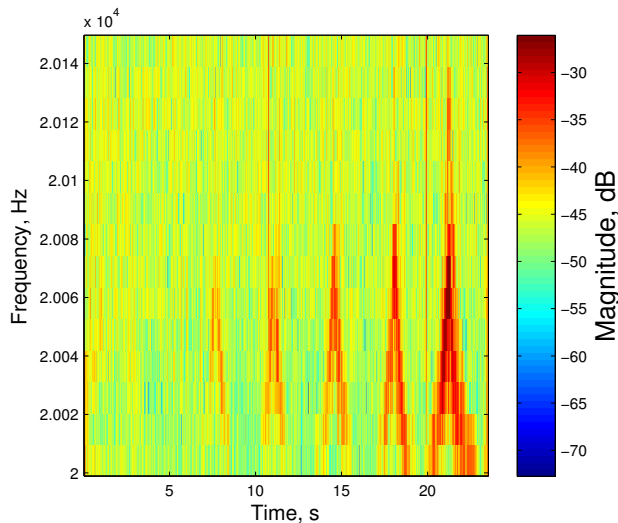

Figure 8: Spectrum of a user approaching the ASUS Nexus 7 from certain distance is depicted here. The first peak at time instance of 7 s corresponds to a distance of 200 cm away from the mobile device. The following peaks are separated 50 cm from each other till the person is standing right in front of the device.

we approach the device, only positive Doppler shift have to be evaluated.

Holding the Phone

Contrary to stationary deployment of the device, holding it can provide an auxiliary input method via simple gestures. For example up-, downward motions of the other hand in front of the phone can be used to control the volume of a music track or zooming into a map view. Swipe gestures prove useful when sorting through some kind of catalog, like a photo gallery or the mail folder. However, movements are limited to one-handed gestures, while stationary deployment can make use of two-handed gestures or whole body movement. We executed test runs using a wave-like (up- and downward motion) and a swiping gesture and successfully extracted the spectrum broadening as shown in stationary experiment. Nevertheless, while holding the mobile phone in one hand, we detected an increase in noise around the central carrier frequency and the shifted signal. This is due to the unavoidable movement of the hand carrying the phone, causing

very small, irregular frequency shifts and thereby deteriorating the received signal.

We want to find out to what extend the movement of the phone itself influences the received signal and prevents a reliable detection of gestures. In this series of experiments we vary the amount of 'noise movement' while performing a wave-like gesture (up- and downward motion). As a baseline we perform the gesture while the phone rests on top of a table (figure 7a). Additionally we execute the same gesture while holding the phone in one hand (figure 7b) and while walking around (figure 7c). These experiments show that moving the phone itself while recording gestures induces a lot of frequency shift noise and hinders gesture recognition quite a lot. Possible solutions would include the estimation of 'walking noise' and subtract it later on to clean the signal. However, this is only suitable while walking in wide areas. Narrow corridors, for example, may cause unintentional frequency shift when walking by an open door.

Carrying the Phone on the Body

In this section we conduct experiments with mobile devices carried on the body. In the first stage, we placed the mobile phone inside the clothing. The mobile device was carried in the pocket of a thin trouser. The device was put into the pocket after the recording was started. Figure 9 shows an extraction of the recording, where a hand is approaching the mobile device and departing again. The same action was repeated for three times. Afterwards the recording was stopped. However, it can be seen in figure 9, that through the thin material, noise covers most of the echo signal. This effect hinders the reliable recognition of gestures. For thicker trousers, like jeans, no frequency shifts can be registered at all.

In the second stage of the experiment, we strap the mobile device to a runner's arm to measure his surroundings while jogging. We hope to explore possibilities of distinguishing the users' environment, like e.g. urban area, woods or fields, via the reflected signal. Results however show that a clear signal cannot be extracted due to the weak echo and high 'noise movement', as discussed earlier. Furthermore, commercial products to strap the device on the runner's arm often cover either the transmitter or the receiver, which weakens the outgoing or incoming signal. Here, noise estimation based on the proximity of certain entities, as thoroughly researched in
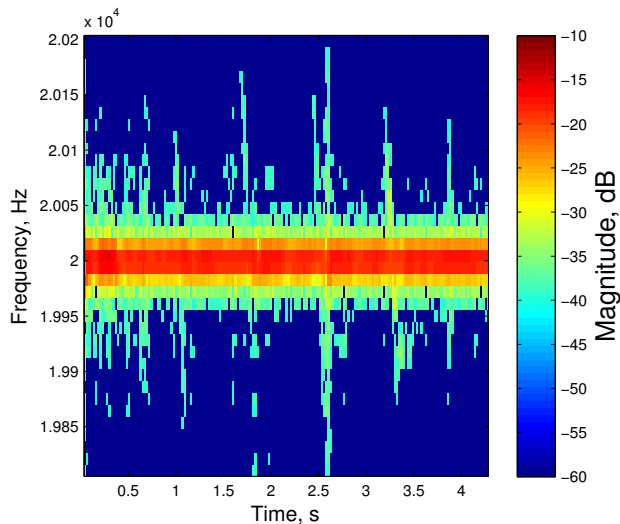
Figure 9: Spectrogram of gestures performed on a mobile device put inside a trouser pocket. The Doppler shifts have a very low magnitude but can be recognized.

[28], might be applied to clean the incoming signal as well as apply some form of calibration.

TECHNICAL CHALLENGES

Our experiments disclose that using the Doppler frequency shift is a feasible technique to recognize gestures as well as activities and general movement in the proximity of the device. In a constraint setup, e.g. a stationary deployment of the mobile phone, the received signal is adequate to detect a variety of gestures. Especially two-handed ones, where the motion of each hand is diverse, are a domain where our approach has its advantages. On the other hand, real-life applications require some sort of noise handling depending on the circumstances. Yet the technique is still usable if the phone movement is not too excessive. Wearable applications on the other hand induce greater challenges in signal processing. The main limitation is due to strong noise via movement or from user's environment. As long as the mobile device is placed inside the clothing, the signal strength is attenuated through the clothing material. Furthermore, additional noise from user's own movements, like in case of the application on a jogger's arm, hinders the detection of additional Doppler shifts from the environment.

Recursive Doppler Reflections

In addition to noise handling, some scenarios may require the algorithm to consider and handle multiple reflections of the emitted peak signal. To illustrate this effect to a higher degree, we repeated the wave gesture using a large sheet of paper. In this scenario, the original signal, emitted by the mobile phone, gets reflected once upon reaching the paper. This is the signal we receive via the device and it exhibits a Doppler shift. However, it is reflected once more by the table, making it seem like a weaker version of the original peak signal, though including a Doppler shift. Again this signal is subject to a Doppler frequency shift upon reaching the paper

making it look like there is another Doppler shift with double the velocity of the first. The effects of this phenomena as shown in Figure 10 can be observed more intensely when using large reflective surfaces. In a setup using hand gestures it is more commonly witnessed as a second flare being present at times when the hand is in close proximity of the phone (cf. the visible flares on every second zero-crossing in Figure 7a). For known movements, this issue can be avoided by excluding multiples of the first frequency shift. However, if there is in fact a second movement twice as fast, things can get ambiguous and hinder reliable gesture recognition.
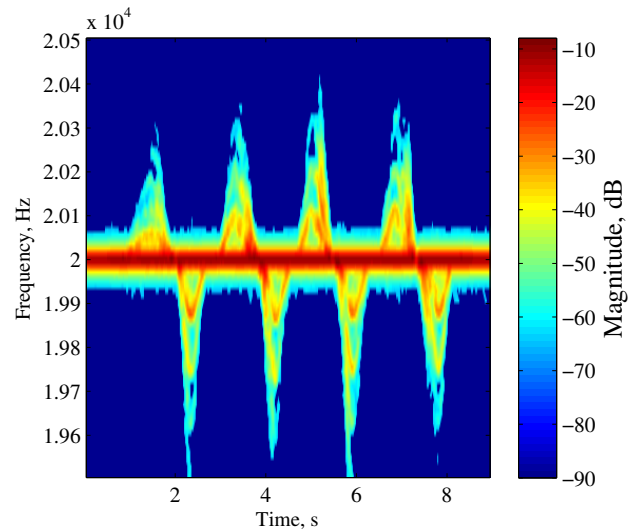

Figure 10: Received spectrogram showing multiple reflections of the original Doppler shift. For instance at 6 seconds, there are three distinct doppler shifts visible. A fourth can be adumbrated. The maximum extend of the first shift at this timestep is just lower that 1.99×10^4 Hz. Every follow up shift (around 1.975×10^4 Hz and 1.96×10^4 Hz) occurs due to multiple reflections of the sent signal.

Missing Directivity

Newer smartphones include two or even three microphones. It is difficult to determine the exact position of each microphone for each device and address the right one. This issue also applies to the position of the loudspeaker, leading to a lack of generalizability between multiple devices. Using a setup with more than one receiver or transmitter, e.g. as shown in [4], could yield directivity by combining the registered measurements.

Quality of Hardware

Another negative aspect concerning a smartphone's hardware is their quality. Usually, the modules for microphone and speaker have to be as small as possible for design purposes, which restricts audio quality in both the listening and recording domain. To compensate for this, a possible algorithm must be robust to noisy data, especially since data from one phone might be different in magnitude than from another even when performing the same activity or gesture.

Occasionally, speaker and microphone are co-located, leading to unpredictable amplitude variations in the recording, as

pointed out by [19]. For experiments that require a stable amplitude, these kind of phones, e.g. the Galaxy Nexus, cannot be utilized.

Furthermore, using the phone's speakers at maximum output increases the amount of eigenfrequency excitations, which may fall into the human range of audibility and disturb the user.

SUMMARY & FUTURE WORK
In this paper, we illustrated the feasibility of a smartphone using its native microphone and speaker to recognize nearby movements. We presented a number of experiments using ultrasound Doppler frequency shifts to detect gestures, general motion and activities of a user. We showed that our approach can be successfully applied in stationary setups or when holding the phone. When the phone is worn on the body, overlying clothes may strongly attenuate the send and received signal. Although we have not tested our approach on a wide range of target platforms and with many users, our results represent a starting point for other researchers.

In the future, we will investigate machine learning approaches for detailed gesture or activity recognition. Using multiple microphones, we aim to overcome some limitations in directivity and noise rejection. Especially extracting suitable feature vectors from the received signal would be hugely beneficial and help generalizing the approach.

Moreover, we aim to integrate FMCW instead of just relying on Doppler frequency shifts, which allow for absolute distance measurements but induces more timing constraints. Recent research demonstrated promising results to detect sleeping apnoea based on this technique [19]. Our next challenge will be to recognize falls in front of a bed while the mobile phone is placed on the night desk. This use-case has already been investigated with acoustic features, which can be enriched by ultrasound sensing [6]. We also plan to develop new generalizable object recognition and tracking algorithms, similar to [10].

REFERENCES
1. Adib, F., Kabelac, Z., Katabi, D., and Miller, R. C. 3d tracking via body radio reflections. In *Proceedings of the 11th USENIX Conference on Networked Systems Design and Implementation*, NSDI'14, USENIX Association (2014), 317–329.

2. Adib, F., Mao, H., Kabelac, Z., Katabi, D., and Miller, R. C. Smart homes that monitor breathing and heart rate. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, ACM (2015), 837–846.

3. Aumi, M. T. I., Gupta, S., Goel, M., Larson, E., and Patel, S. Doplink: Using the doppler effect for multi-device interaction. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '13, ACM (2013), 583–586.

4. Bannis, A., Zhang, P., and Pan, S. Adding directional context to gestures using doppler effect. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, UbiComp '14 Adjunct, ACM (New York, NY, USA, 2014), 5–8.

5. Borriello, G., Liu, A., Offer, T., Palistrant, C., and Sharp, R. Walrus: Wireless acoustic location with room-level resolution using ultrasound. In *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services*, MobiSys '05, ACM (2005), 191–203.

6. Cheffena, M. Fall detection using smartphone audio features. *Biomedical and Health Informatics, IEEE Journal of PP*, 99 (2015), 1–1.

7. Fastl, H., and Zwicker, E. *Psychoacoustics: Facts and Models*. Springer series in information sciences. Springer, 2007.

8. Filonenko, V., Cullen, C., and Carswell, J. Investigating ultrasonic positioning on mobile phones. In *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on* (Sept 2010), 1–8.

9. Grosse-Puppendahl, T., Berlin, E., and Borazio, M. Enhancing accelerometer-based activity recognition with capacitive proximity sensing. In *Ambient Intelligence*, vol. 7683 of *Lecture Notes in Computer Science*. Springer, 2012, 17–32.

10. Grosse-Puppendahl, T., Braun, A., Kamieth, F., and Kuijper, A. Swiss-cheese extended: An object recognition method for ubiquitous interfaces based on capacitive proximity sensing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, ACM (2013), 1401–1410.

11. Gupta, S., Morris, D., Patel, S., and Tan, D. Soundwave: Using the doppler effect to sense gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, ACM (2012), 1911–1914.

12. Helal, S., Winkler, B., Lee, C., Kaddoura, Y., Ran, L., Giraldo, C., Kuchibhotla, S., and Mann, W. Enabling location-aware pervasive computing applications for the elderly. In *Pervasive Computing and Communications, 2003. (PerCom 2003). Proceedings of the First IEEE International Conference on* (March 2003), 531–536.

13. Hoflinger, F., Zhang, R., Hoppe, J., Bannoura, A., Reindl, L., Wendeberg, J., Buhrer, M., and Schindelhauer, C. Acoustic self-calibrating system for indoor smartphone tracking (assist). In *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on* (Nov 2012), 1–9.

14. Kalgaonkar, K., and Raj, B. Recognizing talking faces from acoustic doppler reflections. In *Automatic Face Gesture Recognition, 2008. FG '08. 8th IEEE International Conference on* (Sept 2008), 1–6.

15. Kalgaonkar, K., and Raj, B. One-handed gesture recognition using ultrasonic doppler sonar. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on* (April 2009), 1889–1892.

16. Kellogg, B., Talla, V., and Gollakota, S. Bringing gesture recognition to all devices. In *Proceedings of the 11th USENIX Conference on Networked Systems Design and Implementation*, NSDI'14, USENIX Association (Berkeley, CA, USA, 2014), 303–316.

17. Kunze, K., and Lukowicz, P. Symbolic object localization through active sampling of acceleration and sound signatures. In *Proceedings of the 9th International Conference on Ubiquitous Computing*, UbiComp '07, Springer-Verlag (Berlin, Heidelberg, 2007), 163–180.

18. Mujibiya, A., Cao, X., Tan, D. S., Morris, D., Patel, S. N., and Rekimoto, J. The sound of touch: On-body touch and gesture sensing based on transdermal ultrasound propagation. In *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces*, ITS '13, ACM (2013), 189–198.

19. Nandakumar, R., Gollakota, S., and Watson, N. Contactless sleep apnea detection on smartphones. In *Proceeding of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '15 (2015).

20. Peng, C., Shen, G., and Zhang, Y. Beepbeep: A high-accuracy acoustic-based system for ranging and localization using cots devices. *ACM Trans. Embed. Comput. Syst. 11*, 1 (Apr. 2012), 4:1–4:29.

21. Pu, Q., Gupta, S., Gollakota, S., and Patel, S. Whole-home gesture recognition using wireless signals. In *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking*, MobiCom '13, ACM (New York, NY, USA, 2013), 27–38.

22. Qifan, Y., Hao, T., Xuebing, Z., Yin, L., and Sanfeng, Z. Dolphin: Ultrasonic-based gesture recognition on smartphone platform. In *Computational Science and Engineering (CSE), 2014 IEEE 17th International Conference on* (Dec 2014), 1461–1468.

23. Raj, B., Kalgaonkar, K., Harrison, C., and Dietz, P. Ultrasonic doppler sensing in hci. *Pervasive Computing, IEEE 11*, 2 (Feb 2012), 24–29.

24. Randell, C., and Muller, H. L. Low cost indoor positioning system. In *Proceedings of the 3rd International Conference on Ubiquitous Computing*, UbiComp '01, Springer-Verlag (2001), 42–48.

25. Reynolds, M., Mazalek, A., and Davenport, G. An acoustic position sensing system for large scale interactive displays. In *Sensors, 2007 IEEE* (Oct 2007), 1193–1196.

26. Rossi, M., Seiter, J., Amft, O., Buchmeier, S., and Tröster, G. Roomsense: An indoor positioning system for smartphones using active sound probing. In *Proceedings of the 4th Augmented Human International Conference*, AH '13, ACM (2013), 89–95.

27. Savvides, A., Han, C.-C., and Strivastava, M. B. Dynamic fine-grained localization in ad-hoc networks of sensors. In *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking*, MobiCom '01, ACM (2001), 166–179.

28. Schulz, A., Karolus, J., Janssen, F., and Schweizer, I. Accurate pollutant modeling and mapping: Applying machine learning to participatory sensing and urban topology data. In *International Conference on Networked Systems (NetSys2015)* (2015).

29. Sun, Z., Purohit, A., Bose, R., and Zhang, P. Spartacus: Spatially-aware interaction for mobile devices through energy-efficient audio sensing. In *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '13, ACM (2013), 263–276.

30. Tarzia, S. P., Dick, R. P., Dinda, P. A., and Memik, G. Sonar-based measurement of user presence and attention. In *Proceedings of the 11th International Conference on Ubiquitous Computing*, Ubicomp '09, ACM (2009), 89–92.

31. Wang, G., Zou, Y., Zhou, Z., Wu, K., and Ni, L. M. We can hear you with wi-fi! In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, MobiCom '14, ACM (2014), 593–604.

32. Watanabe, H., Terada, T., and Tsukamoto, M. Ultrasound-based movement sensing, gesture-, and context-recognition. In *Proceedings of the 2013 International Symposium on Wearable Computers*, ISWC '13, ACM (2013), 57–64.

33. Yang, J., Sidhom, S., Chandrasekaran, G., Vu, T., Liu, H., Cecan, N., Chen, Y., Gruteser, M., and Martin, R. P. Detecting driver phone use leveraging car speakers. In *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking*, MobiCom '11, ACM (2011), 97–108.

34. Zhao, C., Chen, K.-Y., Aumi, M. T. I., Patel, S., and Reynolds, M. S. Sideswipe: Detecting in-air gestures around mobile devices using actual gsm signal. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, ACM (2014), 527–534.