Fall Detection via Inaudible Acoustic Sensing

JIE LIAN, University of Louisiana at Lafayette, USA
XU YUAN*, University of Louisiana at Lafayette, USA
MING LI, The University of Texas at Arlington, USA
NIAN-FENG TZENG, University of Louisiana at Lafayette, USA

The fall detection system is of critical importance in protecting elders through promptly discovering fall accidents to provide immediate medical assistance, potentially saving elders' lives. This paper aims to develop a novel and lightweight fall detection system by relying solely on a home audio device via inaudible acoustic sensing, to recognize fall occurrences for wide home deployment. In particular, we program the audio device to let its speaker emit 20kHz continuous wave, while utilizing a microphone to record reflected signals for capturing the Doppler shift caused by the fall. Considering interferences from different factors, we first develop a set of solutions for their removal to get clean spectrograms and then apply the power burst curve to locate the time points at which human motions happen. A set of effective features is then extracted from the spectrograms for representing the fall patterns, distinguishable from normal activities. We further apply the Singular Value Decomposition (SVD) and K-mean algorithms to reduce the data feature dimensions and to cluster the data, respectively, before input them to a Hidden Markov Model for training and classification. In the end, our system is implemented and deployed in various environments for evaluation. The experimental results demonstrate that our system can achieve superior performance for detecting fall accidents and is robust to environment changes, i.e., transferable to other environments after training in one environment.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing**.

Additional Key Words and Phrases: Ultrasonic, Fall Detection, Device-free, Hidden Markov Model

ACM Reference Format:
Jie Lian, Xu Yuan, Ming Li, and Nian-Feng Tzeng. 2021. Fall Detection via Inaudible Acoustic Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 114 (September 2021), 21 pages. https://doi.org/10.1145/3478094

1 INTRODUCTION

Fall is one of the most common cause of serious injury for elders. Each year, 3 million older people would be treated in hospitals due to falls [7], in which fractured bones and soft tissue injuries caused can lead to the death. The report from [5] has shown that lying on the floor for a long time after the fall may aggravate the injury. Given this consideration, the development of effective fall detection systems has attracted growing research attention in recent years, aiming to promptly discover the fall events of elders for needed immediate medical

*Corresponding author.

Authors' addresses: Jie Lian, University of Louisiana at Lafayette, Louisiana, Lafayette, 70504, USA; Xu Yuan, University of Louisiana at Lafayette, Louisiana, Lafayette, 70504, USA; Ming Li, The University of Texas at Arlington, Arlington, Texas, 76019, USA; Nian-Feng Tzeng, University of Louisiana at Lafayette, Louisiana, Lafayette, 70504, USA.

2474-9567/2021/9-ART114 $15.00
https://doi.org/10.1145/3478094

assistance. Such a system is important to enable rapid intervention and mitigate possible serious consequences resulting from falls, desirable for home deployment to protect elders' safety.

Diverse fall detection systems have been developed to serve this purpose for years, and they can be categorized into two types: wearable detection systems and non-wearable detection systems. Among wearable systems, a range of solutions have been proposed, based on accelerometers and gyroscopes [18, 34], RFID [10, 38], and smartphones [1, 9]. However, this line of solutions require the elders to wear the sensors or carry a smartphone using its built-in sensors to identify the fall events with high accuracy, albeit inconvenient. Given the elders may forget or be reluctant to wear/carry the sensors/devices, the non-wearable based solutions are more attractive, with a set of solutions being developed based on Wi-Fi, radars, and cameras. However, several limitations in existence restrict the wide deployment of these solutions. Specifically, WiFi-based solutions [32, 43, 51, 55] typically need to analyze CSI signal to extract features for the fall activities, usually calling for hardware level support to incur high cost and development overhead. In addition, this category of solutions occupies the communication channels on 2.4GHz or 5GHz, which competes for the channel resources with other Wi-Fi devices to degrade the quality of data communication. The radar-based solutions [5, 12, 36] is promising, but they require to purchase an often expensive radar device. The vision-based solutions [8, 19, 41], could also achieve high accuracy, but it would raise serious privacy concerns. The acoustic-based solutions have been proposed in [21, 35, 39], and they detect the fall by extracting Mel-Frequency Cepstral Coefficients (MFCC) [44] through the fall sound. Unfortunately, such solutions could be affected adversely by other acoustic sounds in the environment.

In contrast to existing solutions, we aim to propose a lightweight and convenient solution by leveraging the home audio device to emit the ultrasonic sound for sensing fall occurrence events, applicable for home use. Specifically, we control the speaker of an audio device to generate 20kHz ultrasonic signals inaudible by humans while using the equipped microphone to receive reflected signals for analysis. Since motions of a human in general will cause the Dopper effect of reflected signals, i.e., the frequency changes corresponding to different human activities to yield certain enclosed patterns, this allows us to detect the fall event through analyzing patterns in the reflected signals. Indeed, existing work based on Doppler radar [5, 12, 36], has demonstrated that human falls can be detected via analyzing the Doppler signals. This inspires us to investigate the fall patterns out of the Doppler effect of acoustic signals, distinguishable from other daily activities.

Our design relies only on one speaker and one microphone equipped in a typical audio device to achieve our goal. The speaker would generate a sequence of 20kHz continuous wave, and the microphone would record the Doppler signals for analysis. One key challenge here is that the signals include responses not only from human falls but also from other human activities or object reflection. Thus, it is necessary while challenging to develop a series of signal processing solutions and perform the fine-grained analyses to extract the desirable signal patterns corresponding to falls and the features representing their characteristics. We first apply Short Time Fourier Transform (STFT) for computing the spectrogram of received signals and then develop a collection of solutions to remove the interferences caused by different factors, i.e., direct transmission, environmental reflection, and system imperfection, resulting in the pure fall signals. The Power Burst Curve (PBC) is then applied to the spectrogram for automatically detecting human motions from a sequence of received signals, where the start and the end points of the falls, if any, can be identified to indicate fall duration time. After that, the effective features are extracted so as to represent the fall patterns. We continue to apply the Singular Value Decomposition (SVD) to reduce the feature dimension to 1, so as to input to a Hidden Markov Model (HMM) for training. Notably, the selection of SVD and HMM methods is mainly due to their low computational complexity and few parameters involved, since we aim to deploy our system in commercial smart home devices with limited processing capabilities. For example, the SVD is a simple and quick way for the feature reduction, while HMM has a few parameters but is sufficient to tackle our binary classification problem. Our system is implemented and deployed in practical house environments for evaluation. The experimental results demonstrate that our system can achieve superior performance for fall detection and it is robust to environmental changes.

Our contribution can be summarized as follows:

- To the best of our knowledge, we are the first to explore fall detection via ultrasonic sensing, by leveraging an existing home audio device, requiring no dedicated devices or the specialized changes on hardware. Our system is operated at the 20kHz ultrasonic band, imperceptible to humans and also interference-free to Wi-Fi devices.
- We design a series of solutions for effectively processing the acoustic signals and extracting the salient patterns. Specifically, a set of interference cancellation solutions are designed to eliminate various noise or interference, for obtaining the desired signals that include only clear human activity patterns. We extract a set of effective features and then apply SVD to further reduce their dimensions so as to retain the useful information, for inputting to a HMM model to train.
- Extensive experiments are conducted in different environments, demonstrating that our system can achieve superior performance in detecting the fall and distinguish it from other normal activities. In addition, we show the transferability of our system, i.e., the model trained in one environment (or frequency) can be applied directly to other environments (or other inaudible frequencies) without noticeable performance degradation.

2 RELATED WORK

Our work closely relates to two research directions: 1) fall detection and 2) acoustic-based sensing. We review the state of the arts as follows.

2.1 Fall Detection

Fall detection has attracted considerable interests in healthcare in the past decade. Roughly, the fall detection systems can be classified into two categories: wearable sensor-based systems and non-wearable systems.

Wearable solutions are typically based on accelerometers and gyroscopes [18, 34], smartphones [1, 9], RFID [10, 38], etc. These systems can work when users who wear/carry specific devices. In addition to their inconvenience, wearable-based solutions tend to have limited deployment due to the memory decline of the elders that make them forget to wear/carry the sensors/devices. In addition, some older people feel uncomfortable to wear the devices and will be reluctant to use them at home [61].

On the other hand, non-wearable technologies overcome the aforementioned limitations, enabling continuous fall monitoring without the need to wear/carry devices. Different categories of non-wearable solutions have been designed, including camera-based, radar-based, ambient-based, and Wi-Fi-based approaches. The camera-based system [8, 19, 41] leveraged a camera to capture the photos or video sequences of human activities and developed activity classification algorithms for discovering the fall events. However, it has been well known that this line of solutions will invade people's privacy, suffer from occlusion, and require intensive computation cost for real-time processing. The Doppler radar-based approaches [5, 12, 36] can directly measure motion velocity by leveraging the Doppler frequency relationship, but they require dedicated radar devices that work in high frequency bandwidth to achieve high resolutions. Meanwhile, ambient-based approaches have been proposed to monitor the sound of fall for detection. In [21, 35, 39], the MFCC features of fall sound were extracted and then processed by the machine learning classifier. However, these solutions rely on audible sounds, which could result from many daily activities, such as playing music, talking, etc., to yield degraded performance and be environment-specific.

The Wi-Fi-based sensing has become popular in recent years, with many solutions being proposed for fall detection. For example, some systems measure the fall based on the changes of the received signal strength indicator (RSSI) [14, 16, 23]. However, such RSSI-based solutions are not easy to deploy due to the requirement of sensors and the detailed fingerprinting of environment. In addition, the CSI-based solutions [32, 43, 51, 55] have

been proposed, utilizing the Short Time Fourier Transform (STFT) or the wavelet transformation to estimate fast changes in the RF signals. For example, WiFall [55] leveraged the CSI amplitude related time domain features to characterize falls of a single person. RTFall [51] took into account both CSI amplitude and CSI phase while Falldefi [32] extracted a set of features from the CSI, to characterize the fall activity. However, Wi-Fi-based solutions require certain hardware changes for receiving the CSI signals. They also occupy the data communication channels, inevitably causing interference to home Wi-Fi devices.

2.2 Acoustic Sensing

Recently, acoustic sensing has also attracted wide attention, and it can be categorized into two lines of research. The first research line is device-dependent, requiring users to hold a device for sensing. Solutions [24, 48, 59, 60] have been developed for motion tracking. Relying on the low propagation speed of acoustic signals, they could achieve high accuracy. The core idea is to estimate the phase change of the signals from the transceiver to calculate the distance. But the nature of these systems is to track the device movement via acoustic signals, rather than a human activity.

The other research line is device-free, freeing the user from carrying a device for sensing. Recent works [26, 27, 40] have shown that acoustic sensing with the FMCW signals can achieve the high accuracy in terms of respiration sensing with a limited range (i.e., about 1 meter). However, the FMCW signal is unsuitable for use in our system. Notably, our system relies on analyzing the Doppler shift signals for detecting the fall occurrence. Since the FMCW signal continuously swipes its frequency, the multi-path reflection signals in the room environments will somewhat overlap with the Doppler signals caused by the fall in the frequency domain. Hence, it is difficult for our system to remove the reflection signals for acquiring the Doppler signals due to fall events.

Acoustic signal has also been used for gesture recognition [13, 37, 54], in which the Doppler shifts are leveraged from inaudible acoustic transmissions for recognizing different hands movement patterns and gestures. However, such designs require hands to be close to the mobile devices. Some systems [28, 42, 52] also have been developed to perform contactless tracking via acoustic signals. The basic idea of these systems is to generate the special modulated signal such as the OFDM symbols for measuring the distance between the target and the transmitter by the Time of Flight (TOF) or the phase changes of reflected signals. However, their sensing ranges are usually limited within one meter, unsuitable for home applications. A recent work RTrack [25] enables the room-scale hand motion tracking by combining the signal from a microphone array with a series of signal processing techniques and an RNN. Requiring a microphone array, it is not pervasive at home for AOA measurement, and its performance tends to be affected by the microphone layout.

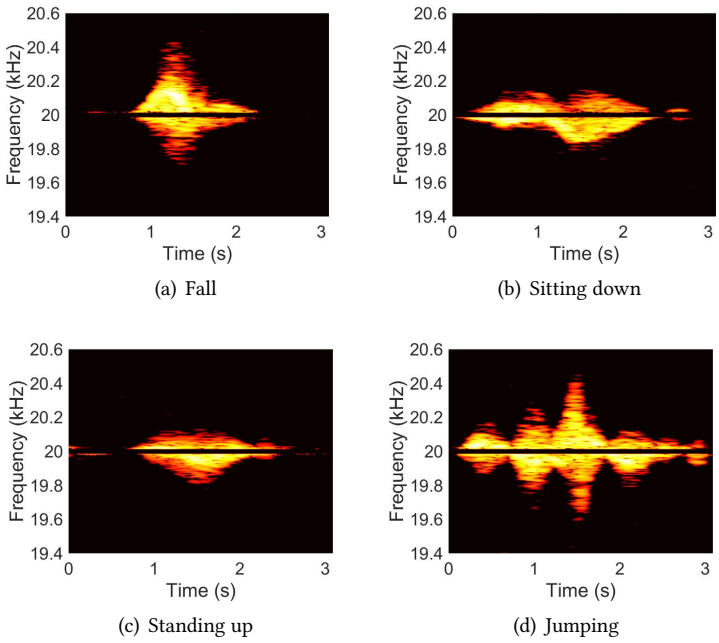
In addition, some systems focus on human activity recognition. For example, covertband [29] enables activity recognition by the inaudible OFDM symbols. It employed a smart speaker as the transceiver and two microphones as the receiver for identifying rhythmic motions such as jumping, pumping arms, or pelvic tilts, by analyzing OFDM symbols' correlation profile. However, fall detection is not considered in its design. Separately, other systems [4, 53, 57] can identify human's gait patterns from Doppler signals via extracting a set of features to train a machine learning classifier. Our design of effective fall detection via acoustic sensing is inspired in part by them.

3 PRELIMINARIES

We briefly present the preliminary knowledge and necessary background below.

3.1 Acoustic Sensing

In acoustic sensing, acoustic signals reflected from a moving object yield a frequency change, called the Doppler effect. The frequency shift is determined by the source frequency and the velocity of a moving object. Denote f_t



(a) Fall (b) Sitting down

(c) Standing up (d) Jumping

Fig. 1. Spectrogram of different activity

as the source frequency emitted from the speaker and f_r as the perceived frequency at the microphone, then the Doppler shift phenomenon can be modeled as

$$f_r = f_t \cdot \left(\frac{c + v \cdot \cos\theta}{c - v \cdot \cos\theta} \right), \tag{1}$$

where c is the sound speed in air, v is the human motion velocity, and θ is the angle between the motion and the beam of signals.

3.2 Disparate Patterns between Falls and Other Activities

Since we rely on the Doppler signal to detect falls, we should examine if there are disparate patterns between the falls and other activities. We leverage a speaker on an audio device to emit continuous wave at 20kHz and use its microphones to receive the reflected signals. One participant is asked to perform different actions in front of the audio device, like falls, walking, sitting, jumping. Figure 1 shows the spectrograms corresponding to four activities. Note that these figures have been processed by our interference cancellation methods, to be introduced in Section 4.2. From this figure, we can see different activities will result in total disparate patterns, making them distinguishable through analysis. Specifically, we observe the fall usually happens with a sharp burst, comparing to walking and sitting. This is due to the relatively high body moving speed caused by fall. For both walking and sitting, their resulting peaks are slower and wider. Although a jump also results in a sharp peak, it is followed by a set of small peaks, corresponding to a series of actions from landing to restoring balance. This experiment confirms the possibility of designing a solution for correctly distinguishing the fall from other normal activities via acoustic sensing, making our fall detection system work. It also gives insights for the nature of our system design, as detailed in Section 4.

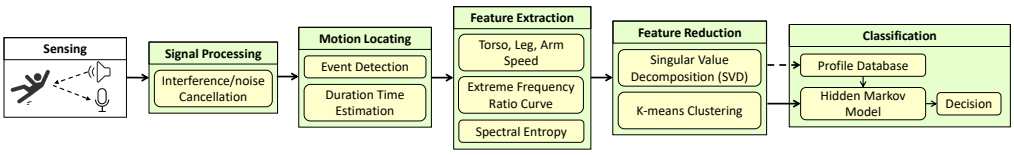


Fig. 2. The workflow of our fall detection system.

3.3 Hidden Markov Model

Hidden Markov Model (HMM) [15, 22, 45] is a double random process, in which each hidden state (human motion state) has a translation probability to the following hidden state and a probability distribution related to the possible observable state. HMM provides an approach to infer hidden states from the observation sequence. In our context, fall events can be modeled as transitions of human motion states in HMM. The hidden states related to states in a human fall include balance, losing balance, impact with floor or other lower object, and stabilization after impact. The observation sequence refers to the features extracted from the fall's doppler signal. Based on signals, HMMs are able to infer the hidden states during the fall, providing a better description of the fall and yielding more accurate recognition.

4 SYSTEM DESIGN

This section illustrates our design of the fall detection system via inaudible acoustic sensing. We control the speaker to emit continuous ultrasound signals while letting the microphone record the reflected signals. By conducting a series of fine-grained processing and analyses on the reflected signals, our system can identify the signal of interest and is capable of extracting the desirable features, realizing the fall detection goal. Many technical challenges exist in the design of our system, including but not limited to:

- The received signals contain intensive interference coming from the object reflection, multi-path and/or system defection, which can substantially distort the desirable signals reflected from the human fall events. Sometimes, the energy level of such interference even exceeds that of the signal of interest. It is necessary to develop a collection of solutions to eliminate them for gaining the pure signals resulting from fall events.
- To recognize fall events, the effective features should be extracted from the signals. However, the fall events are diverse which can happen at different directions, resulting in disparate patterns. How to extract a set of features for characterizing all different fall actions with disparate patterns is a challenging task. Essentially, we aim to make our detection system be robust to different fall direction changes. On the other hand, the fast attenuation of acoustic signals will result in fall's pattern unapparent, further elevating difficulty in the feature extraction task and thus hindering the development of an effective detection system.
- The received signals also contain some from other irrelevant activities, such as walking. Typically, the fall mostly happens when a person is walking. This brings a high requirement for our detection system, which should be able to correctly detect the onset of a fall in a sequence of received signals. In addition, some actions such as sitting down, standing up or jumping also have similar patterns as the fall. How to distinguish them from the fall remains challenging.

Before we elaborate our detailed design, we give an overview of our system, consisting of six key modules: *Sensing, Signal Processing, Motion Locating, Feature Extraction, Feature Reduction* and *Classification*, as shown in Figure 2. In the *Sensing* module, the system programs the built-in speaker in an audio device to emit inaudible acoustic signals and uses its microphone to collect reflected signal for analyses. In the *Interference Cancellation* module, we develop solutions to cancel the interference from environment reflections and system defects to get a clean spectrogram for analysis. The *Fall Locating* module will apply the Power Burst Curve (PBC) to locate the

fall event or other motion events from the signals. In the *Feature Extraction* module, a set of effective features that characterize the fall events will be identified and extracted from the spectrogram for classification. Then, new solutions based on the SVD decomposition and the K-means algorithm will be designed to reduce the feature dimension and cluster the data, respectively, for getting rid of the redundant features and determining the hidden state numbers in the Hidden Markov model (HMM), required by the *Feature Reduction* module. In the end, the *Recognition* module takes the data to train a HMM for classifying the events as falls or non-falls.

4.1 Sensing

We program the built-in speaker in an audio device to emit continuous inaudible acoustic signals at 20kHz. Once the signal is reflected from the human motion, the Doppler effect will be generated for making the frequency shift. The microphone is used to collect Doppler effect signals with a sampling frequency of 48kHz, which ensures it is twice higher than the highest Doppler shift frequency. Notably, it is not necessary to let the speaker keep sensing. We can program the speaker to periodically send low power 20kHz ultrasound when there is no human motion in the room. Once the microphone senses a certain sound pressure level between 19kHz and 21kHz, our fall detection system will be triggered to emit the ultrasonic signals with normal power.

After receiving the reflected signals at the microphone, we generate its spectrogram for analysis, via applying the short-time Fourier transform (STFT). In particular, the sequence of original signals is sliced by a set of small windows, each with a length of 0.4s while any two consecutive windows have 95% overlapping. As such, a 1s signal sequence would be sliced into 50 frames. Then we multiply each frame by a Hamming window, and apply an 8192 point Fast Fourier transform (FFT) on each frame. That is, we divide each frame into 8192 sub-bands, which produce a frequency resolution about 2.5Hz. After the above process, we construct a spectrogram for further analysis.

4.2 Interference Cancellation

Since the speaker is omnidirectional, the received signal will contain not only the desirable signals, but also many interferences. In our experiments, we observe three types of interferences resulted from 1) direct transmission, 2) environment reflection, and 3) system defects. The first interference type represents the signals directly transmitted from the speaker to the microphone. Since the speaker and the microphone are co-located, such an interference will dominate the spectrogram. The second interference type is caused by the reflection from the environment objects. The last one is the noise generated due to the imperfection of devices.

We notice that the direct transmission and the environmental reflection are static over time. So, to remove the direct transmission noise, we can simply remove the energy on the spectrogram between 19.99kHz and 20.01kHz. While the environmental reflection can be considered as the stationary noise, we can remove it by using spectral subtraction. To eliminate it, we enable the microphone periodically to record environmental reflection when no people is moving in the room. We also notice that such a reflection varies only slightly with time, so we could apply it to the whole signal in the same environment. We perform the STFT on the recorded reflections to get the noise spectrogram, and then subtract it from our spectrogram. Through the two steps, the direct transmission and the reflection of static objects could be completely removed.

The system defects will cause a certain level of noise, with one example shown in Figure 3, from which we can see noise points spreading on the spectrogram. This type of noise appears even when there is no activity happens. Our experiments show that the energy level of the system defect can be modeled as a Gaussian distribution. We record the amplitude of noise at high frequencies (higher than 20.6kHz) of the spectrogram and construct a noise histogram of amplitudes. A threshold N_t needs to be set for removing the noise. Assume μ and σ are the mean and the standard deviation of the Gaussian approximation of the noise histogram, respectively, then we can

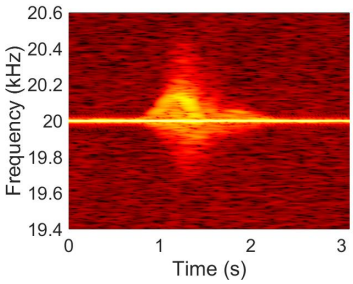


Fig. 3. Noisy spectrogram.

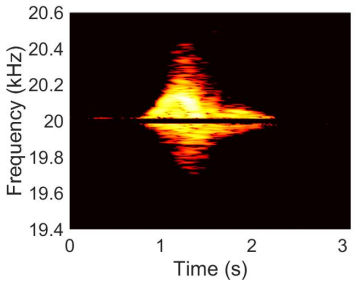


Fig. 4. Denoised spectrogram.

calculate them directly from this histogram. Hence, the threshold N_t can be calculated by:

$$N_t = \mu + \sigma \tag{2}$$

We then apply the threshold N_t to the spectrogram, yielding:

$$S(f,t) = \begin{cases} S(f,t) & \text{if } S(f,t) \geq N_t \\ 0 & \text{if } S(f,t) < N_t \end{cases} \tag{3}$$

where $S(f,t)$ represents the power at frequency f and time t on the spectrogram. This step will retain the points on the spectrogram with their energy levels greater than the threshold.

Figure 4 shows the spectrogram after the above three-step process, observed to retain only the Doppler signals and demonstrating the effectiveness of our interference cancellation methods.

4.3 Motion Locating

After canceling the interference, we obtain a clean spectrogram. Since the spectrogram in general contains information about a sequence of events, with fall occurrences possibly accounting for a small fraction, it is highly desirable to locate the onset of any human motion, if present, from the spectrogram, instead of processing the spectrogram as a whole, to lower the processing time. Note that a fall event belongs to one kind of human motions, which may also include other activities such as walks, jumps, sit-downs, etc. Thus, the next key step is to locate the starting and end points of a human motion in the spectrogram. Considering that the motion involves a set of limb movements, a series of high frequency will result due to the Doppler effect. This can be observed as a power burst on the spectrogram. The power burst curve (PBC) is then applied to locate the motion, where PBC represents the summation of signal power within a specific frequency range between frequencies f_l and f_u. Taking the positive frequency (frequency higher than 20kHz) as an example, the Power Burst Curve (PBC) could be represented as:

$$PBC(t) = \sum_{f=f_l}^{f_u} |S(f,t)|^2 \ . \tag{4}$$

Here $f_l = 20.01kHz$, is the lower frequency band after eliminating the direct transmission noise and the f_u is the upper frequency band which can be set to $20.6kHz$ or higher. The energy components in this frequency range include the large reflection from the body. On the other hand, we set up a threshold value, i.e., PBC_{th}, in this frequency range, calculated by

$$PBC_{th} = \sum_{f=f_l}^{f_u} N_t \ . \tag{5}$$

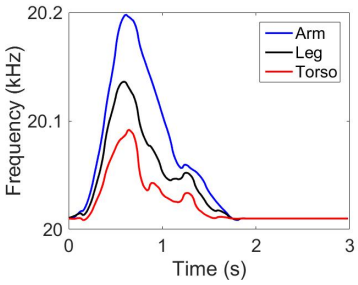


Fig. 5. Motion curve caused by fall.

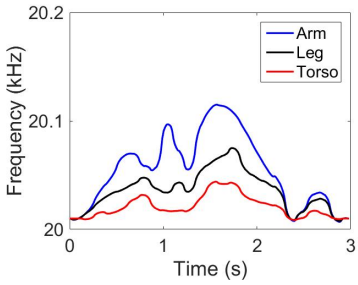


Fig. 6. Motion curve caused by sitting.

We sum up the signals in this range and compare the result with PBC_{th}; if the summation exceeds this threshold, we consider the presence of a motion. The start and end points of the motion can be determined accordingly, by identifying the two intersections between PBC and the noise threshold. We truncate signals between the start point and the end point for further analysis.

4.4 Feature Extraction for Fall Detection

Once a motion event is identified from the spectrogram, we next identify and extract a set of features representing the fall's characteristics. We observe that the fall spectrogram includes rich information such as the reflection of different body parts. In particular, the unique movement of a fall yields a frequency distribution different from those of other activities. This allows us to extract a set of features from the spectrogram for characterizing the body movement, in order to detect the fall. Collectively, we extract three sets of features.

The first set is the speed feature, including the torso speed curve and the arm speed curve, which can quantify the moving speed and moving pattern of each body part during the fall. To compute the speed, we first calculate the frequency shift caused by the movement of each body part. Inspired by the percentile method in Doppler radar [46], we extract the frequency curve as follows:

$$T(f,t) = \frac{\sum_{f_l}^{f} S(f,t)}{\sum_{f_l}^{f_u} S(f,t)}, \tag{6}$$

where $S(f,t)$ represents the Doppler effect energy on the frequency f at a certain time t on the spectrogram, $\sum_{f_l}^{f} S(f,t)$ denotes the cumulative energy associated with the frequency range from f_l to f, and $\sum_{f_l}^{f_u} S(f,t)$ represents the total cumulative energy of the Doppler signal at time t. Hence $T(f,t)$ represents the ratio of cumulative energy attributed by the frequency below f over the total Doppler effect energy needed at time t to identify a point in the torso contour. Our extensive experiments exhibit that we can set the threshold value of $T(f,t)$ as 30%, 75%, 95%, respectively, to roughly derive the frequency f corresponding to the torso movement, leg movement, and arm movement. The reason for choosing such three thresholds is that they could better describe the motion of each body part based on our experiments. We take a non-fall activity, i.e., sitting, for comparison, as shown in Figure 6. Comparing Figures 5 and 6, we observe that although the fall and sitting events have similar movement patterns, their motion curves differ significantly, i.e., maximum positive frequencies of the fall curve and the sitting curve are 200Hz and 100Hz, respectively. Since the fall occurrence will cause strong power appearing on both the positive frequency (higher than 20kHz) and the negative frequency (lower than 20kHz) sides, we shall calculate the frequency curve on both sides to characterize human motions comprehensively. As such, we can get a total of 6 frequency curves corresponding to arm movement, leg movement, and torso

movement on both positive and negative sides. With the curves, we can apply Eqn. (1) to get the moving speed curves of torso, leg, and arm.

The second set of features is the extreme ratio curve, calculated by

$$R(t) = \max\left(\left|\frac{f_{+\max}(t)}{f_{-\min}(t)}\right|, \quad \left|\frac{f_{-\min}(t)}{f_{+\max}(t)}\right|\right) \tag{7}$$

where $f_{+\max}$(t) represents the max frequency shift above 20kHz, deriving from the Eqn. (6) by setting the threshold to 95%. $f_{-\min}$(t) represents the minimum frequency shift below 20kHz, calculated similarly as $f_{+\max}$(t). Typically, the extreme ratio curve of a fall accident is sharper, having a larger peak value than those of other activities. This is due to no stationary movement of body parts in a fall event, so its energy amounts on both sides of the spectrogram are highly asymmetric, resulting in a sharper and highly extreme frequency ratio curve. Other types of motion such as sitting or standing, often have high and symmetric energy amounts in both positive and negative frequency bands, making the extreme frequency curve more stable.

The third set of features is the spectral entropy, which measures the randomness of the energy distribution on a spectrogram of the fall. Considering the quick and sudden motions of a fall event, it usually has a higher entropy than other events at both positive and negative frequency bands due to high energy fluctuation on the spectrogram. We calculate the spectral entropy, denoted by H, as follows:

$$H(t) = -\sum_{f=f_l}^{f_u} p(f,t) \ln p(f,t)$$

where f_l and f_u are upper and lower frequency bounds. $H(t)$ indicates the spectral entropy at a certain time t. $p(f,t)$ is the normalized power spectral density at frequency f and time t, calculated by:

$$p(f,t) = \frac{P(f,t)}{\sum_{f=f_l}^{f_u} P(f,t)}$$

where $P(f,t)$ is the Power Spectral Density, expressed by

$$P(f,t) = \frac{1}{f_u - f_l} \sum_{f=f_l}^{f_u} |S(f,t)|^2 ,$$

where $\sum_{f=f_l}^{f_u} |S(f,t)|^2$ represents the cumulative sum of square for energy attributed by the frequency between f_l and f_u.

4.5 SVD Algorithm

Due to the corresponding movement of the entire body, some features are highly correlated, so they may not contribute useful information for our fall detection task. We can further apply singular value decomposition (SVD) to get rid of the redundant features and reduce their dimensions. Here, SVD aims to reduce the feature dimension to 1 before inputted to HMM, which will be elaborated in Section 4.7.

We assume 0.8s data are inputted each time to HMM. We extract 8 feature series in the above step, and all of them have the same time resolution as our STFT, which is 0.02s. As such, we can get an 8×40 feature matrix as the input, denoted by F. The SVD decomposition algorithm can be represented by:

$$F = U\Sigma V^T, \tag{8}$$

where Σ is an 8×40 matrix, with all its elements off the main diagonal to be 0. Each element on the main diagonal is called a singular value. U and V are unitary matrices, with their sizes of 8×8 and 40×40, respectively. To

compute U, we multiply F and F^T, to get an 8×8 matrix. Then we apply the Eigenvalue Decomposition to FF^T:

$$\left(FF^T\right) u_i = \lambda_i u_i \,, \tag{9}$$

where we can get 8 eigenvalues of matrix FF^T and their corresponding 8 eigenvectors u. All the eigenvectors of FF^T are transformed into 8×8 matrices, which together denote the U matrix given by Eqn. (8).

To calculate V, we multiply F^T and F to obtain a 40×40 matrix. Then, we apply the eigenvalue decomposition to F^TF, yielding

$$\left(F^TF\right) v_i = \lambda_i v_i \,. \tag{10}$$

We can get 40 eigenvalues of matrix F^TF and their corresponding 40 eigenvectors v. The 40 eigenvectors v are combined with the V matrix expressed by Eqn. (8).

We next compute Σ, which has non-zero values only on its main diagonal, including the singular value σ. With the unitary matrices of U and V, we have

$$\begin{aligned} F = U\Sigma V^T \Rightarrow FV = U\Sigma V^TV \Rightarrow FV = U\Sigma \\ \Rightarrow Fv_i = \sigma_i u_i \Rightarrow \sigma_i = Fv_i/u_i \,. \end{aligned} \tag{11}$$

Note Σ is arranged in the descending order, so the reduction of singular values is extremely fast. In many cases, the top 10% or even 1% of singular value account for more than 99% of the sum from all singular values. In other words, we only need to use the largest k singular values and corresponding left and right singular vectors to approximate the matrix. So, we have:

$$F_{m\times n} = U_{m\times m}\Sigma_{m\times n}V^T_{n\times n} \approx U_{m\times k}\Sigma_{k\times k}V^T_{k\times n} \tag{12}$$

If $k = 2$, we can convert $F_{8\times 40}$ to $F_{2\times 40}$. By selecting the first line, we reduce the feature dimension to 1. Through this way, we have compressed the multi-dimension features to a 1-D sequence.

We divide F into 10 data units and then calculate the averaged value on each unit to get the observation sequence O with length 10. Then we input O to HMM for further classification.

4.6 Data Cluster

We next cluster the data, to determine how many hidden states contained in a fall, so as to further improve the performance of our detection system. The number of clusters is equal to the number of hidden states in HMM. In our design, we employ the K-means algorithm [17] to cluster the features extracted from the fall spectrogram analytic process. To be specific, we first randomly set a center for each cluster. Then the distances of each feature vector to those centers are calculated. We associate each vector with one cluster that has the nearest center and iteratively execute this step until a convergence. In the end, the number of clusters equals the number of hidden states in HMM. Our experiments found 4 clusters, representing the number of invisible states in human fall events, such as: balance state, losing balance state, impact state, and stable state after the impact.

4.7 HMM (Hidden Markov Model)

HMM had been widely applied to speech recognition [2, 31] and to the medical purpose [33, 58]. Its successful application in this field proves the ability of HMM to process biological sequence signals. Although the human motion-induced Doppler signal is very complex and dynamic with many different features, we can transform it to a simple feature sequence via our SVD algorithm. Then we apply HMM to process such a sequence to analyze the state transition indirectly through the feature sequence extracted from the Doppler signals.

We denote the HMM that represents the fall process as λ, yielding:

$$\lambda = (A, B, \pi) \,.$$

where A is the state transition matrix, B is the Emission matrix, and π is the initial state distribution. The number of observation values is 10, equal to the length of input O.

To train $\lambda = (A, B, \pi)$, the Baum-Welch algorithm [6] is applied. We first randomly initialize the parameters of A, B, π. Then the forward and backward procedure would calculate the expected hidden states according to the observed data O_{fall} and the parameters of A, B, π. We update the parameters to best fit the observed data O_{fall} and the expected hidden states. These steps repeat until the parameters are converged. After being trained, $\lambda = (A, B, \pi)$ can describe the features of a fall process. With the trained HMM (i.e., λ) and a new observation series O_{new} acquired from any motion process, the conditional probability $P(O_{new} \mid \lambda)$ could represent the marching degree of λ and O_{new}.

4.8 Classification

In the feature series, we set the input length of HMM as 0.8s and the moving step length of the window as 0.1s. We slide a 0.8s data window on the signals, according to the method described in Section 4.4 for feature series extraction. Then, we apply SVD to the features for dimensional reduction to get a 1-D sequence, serving as the observation sequence O_{new} of HMM. O_{new} is inputted into the trained fall process model (Section 4.7) for calculating the output probability $P(O_{new} \mid \lambda)$. This probability is compared to a pre-set threshold P: if it is greater than P, this sequence represents a fall event; otherwise, it does not. Our empirical experiments show that P can be set to 0.156 for fall detection.

5 PERFORMANCE EVALUATION

We implement our fall detection system and deploy it in different room environments for performance evaluation. Our goal is twofold. First, we aim to examine the effective detection range and the accuracy of fall detection via acoustic sensing and also to show its capabilities of distinguishing fall accidents from other human activities. Second, we conduct various experiments under different environments in order to show the robustness of our system.

5.1 Experiment Setup

Our system is implemented on a speaker (Edifier R1280DB) and a microphone (SAMSON MeteorMic, 16 bit, 48 kHz), which are binding together as one device placed on the desk, with the speaker generating and emitting signals at 20kHz. The microphone's sampling frequency is set to 48kHz. The recorded Doppler signals are sent to a laptop for further analyses, with Matlab employed as the signal processing and machine learning tools. We set the transmission power to the 80% of speaker's maximum power and then measure the sound pressure at 1m away from the speaker, to get 45dB.

We conducted our experiment in four different environments: 1) a living room in an apartment with several tables and chairs, 2) a lab with tables and chairs, 3) a meeting room with a long table and dozens of chairs around, and 4) a long corridor.

5.2 Data Collection

Since the fall event may cause potential risks to elders, we only recruit 6 elders (5 males and 1 female) to participate. Besides, 17 young people are also recruited for mimicking the elders' behaviors. Hence, there are a total of 23 participants, whose ages range from 60 to 75 for elders and from 24 to 40 for young people. Among them, there are 17 males and 6 females. Notably, a mattress is provided for protection from being injured when they fall down. For each participant, we collect a continuous stream of activities, mixing the fall, fall-like, and other daily activities as listed in Table 1. We ask the elders to walk in their nature ways and then fall on the mattress, for collecting their fall events. All participants are asked to perform the fall actions according to their experiences

Table 1. Activities types

Fall activities	Fall-like activities	Daily activities
Lose balance-Forward	Jump	Walk
Lose balance-Backward	Sit-down on floor	Picking up
Lose balance-Left	Sit-down on chair	Open the door
Lose balance-Right	Stand-up from floor	Eat food
Trip-Forward	Stand-up from chair	Drink water
Trip-Backward	Raise hand	Working on computer
Trip-Left	Picking up	
Trip-Right	Large objects falling	
Slip-Forward		
Slip-Backward		
Slip-Left		
Slip-Right		
Lose consciousness-Forward		
Lose consciousness-Backward		
Lose consciousness-Left		
Lose consciousness-Right		

for simulating 1) the sudden loss of balance, including losing balance, losing consciousness, trip, and slip, and 2) fall forward, backward, and sideward. In addition, we also ask them to perform some non-fall activities and daily activities (see Table 1). For young people, we let them mimic elders' behaviors as follows. First, we ask them to walk at a slow speed (i.e., less than 1m/s) when collecting the continuous data, because [3] shows elders with a slow gait speed have a high risk of falling down, as expected due to the loss of muscle strength and mass [47]. Also, we ask them to try their best to mimic the falls of elders. Second, we ask them to do some fall-like activities and daily activities to mimic the elders' behaviors; for example, sitting down and standing up slowly, while supporting the body by hand. All collected data are labeled manually, serving as the ground truth.

5.3 Evaluation Metrics

We define the following evaluation metrics for measurement.

- *Accuracy:* which is defined as the ratio of correctly identified samples over all samples, i.e., $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$.
- *Precision:* which is the ratio of the correctly detected falls over all detected falls, i.e., $Precision = \frac{TP}{TP+FP}$.
- Recall: which means the ratio of correctly detected falls over the total falls. i.e., $Recall = \frac{TP}{TP+FN}$.
- *F1-score:*, which reflects the overall performance of the classifier, defined as F1 $score = \frac{2\times Recall \times Precision}{Recall + Precision}$.

Here, TP represents the true positive, which is the correctly detected falls. TN represents the true negative, which is the correctly detected non-falls. FP represents the false positive, which refers to non-falls but wrongly detected as falls. FN is the false negative, denoting falls but wrongly detected as non-falls.

5.4 Effective Detection Range

We test the working range of the system in this experiment, by letting a participant perform both fall and non-fall activities at different distances to the device. We collect 100 fall events from $1m$ to $5m$ as the training data to train a HMM. Then, another 100 fall events and 100 non-fall events are collected at $1m$, $2m$, $3m$, $4m$, and $5m$, with each

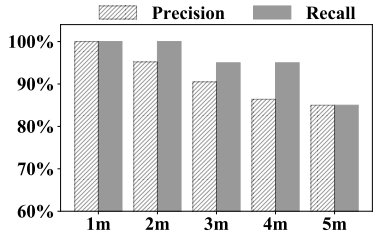


Fig. 7. Detection Range.

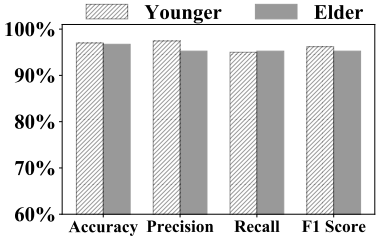


Fig. 8. Evaluation on Elders.

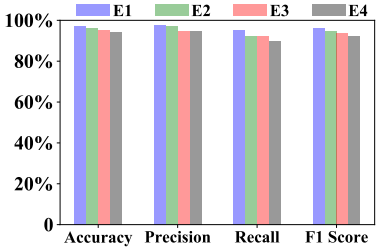


Fig. 9. Impact of different environments and different participants.

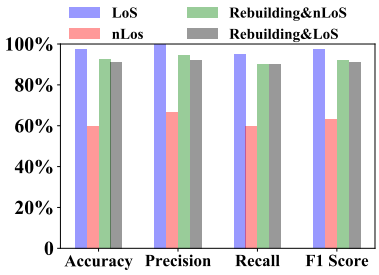


Fig. 10. Impact of the nLoS.

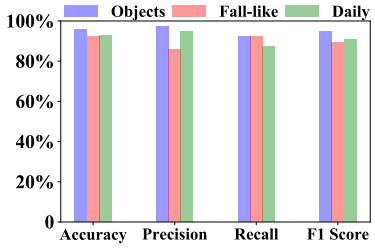


Fig. 11. Evaluation of the non-fall activities.

distance collecting 20 falls and 20 non-falls. Figure 7 shows the performance of our system in terms of precision and recall at various distances. Detection performance degrades when the distances from person to the sensing device increase. This is due to signal attenuation, where the signals reflected from the longer distance will result in the low SNR, making the feature extraction tasks much harder. At 5 meters, our precision can still maintain more than 80%. The recall stays higher than 95% up to 4 meters. Even when a person moves to 5 meters away, our system can still maintain 80% recall for fall detection. In the following experiments, we set the distance to 3 meters.

5.5 Performance on Elders

In this experiment, we evaluate our system on 6 old participants. We train a model based on the fall data collected from 2 young participants and then collect a set of fall, fall-like and daily activities for testing from the elders. In addition, we also collect data from the 2 young participants for testing. Figure 8 shows our system performance in terms of accuracy, precision, recall, and F1 score. From this figure, we observe our system can achieve very high performance for detecting the fall from elders, with the accuracy, precision, recall, and F1 score of 96%, 95%, 95%, and 95%, respectively. The performance on the young participants is even better, to be 97%, 98%,95%, and 96%, respectively. Such results show that our system could reliably detect falls from elders. In addition, it also demonstrates that our system does so, without collecting elders' data for training (in the purpose of protecting elders), amenable to real-world deployment.

5.6 Performance under Different Environments and Different People

We next evaluate our system performance under different environments and users. In particular, this experiment trains a HMM model only with the data collected from two participants in the apartment and then examine

the model on different participants under different environments. To be specific, we conduct four experiments on the 17 younger participants in four different environments: apartment, lab, meeting room, and corridor. For collection the testing data, the first experiment lets the same participants from the training set to perform activities in the apartment, indicated as E1. The second experiment asks the same participants from the training set to perform activities in other environments, denoted as E2. The third experiment lets participants other than the training set perform activities in the apartment, indicated as E3. The last experiment asks participants other than the training set to perform activities in other environments, indicated as E4. Figure 9 shows the averaged results in each experiment in terms of four metrics. We can observe our system to always achieve very high performance in the four scenarios (i.e., E1, E2, E3, and E4), with the the accuracy, precision, recall, and F1-Score of 97%, 97%, 95%, 96%, of 96%, 97%, 93%, 94%, of 95%, 94%, 92%, 93%, and of 94%, 94%, 90%, 92%, respectively. This set of experiments demonstrate that our trained model at one environment can be transferred to different environments and different people for use with negligible performance degradation.

5.7 Impact of Non-Line of Sight (nLoS)

In the experiments above, we observe that the environment change will have a certain impact on our system performance. One possible reason is that the blocking of Doppler signals could somehow degrade system performance. Hence, we design another set of experiments to evaluate the impact of the nLoS path and show how to minimize its impact. We put a chair between the person and the device to simulate an nLoS scenario. Three experiments are conducted: 1) the chair is placed at the original place, so there is an LoS path between the person and the device; 2) the chair is placed on the LoS path, blocking the signals (i.e., nLoS scenario); 3) retraining the model by the nLoS data. In the first two experiments, the HMM model is trained under the LoS path scenario. Then, we ask a participant to perform 20 falls and 20 fall-like activities in the first two experiments. In the third experiment, we ask a student to conduct 60 falls to retrain the HMM model for the nLoS scenario and then evaluate the rebuilt model with data collected in the first and second scenarios to evaluate its performance. Figure 10 shows the results of three experiments. From this figure, a model trained under the LoS path scenario is observed to degrade its detection performance greatly when the chair is moved to the LOS path (from the first experiment to the second experiment), with all evaluation metrics declines to around 60% (see the red bar). This is due to the significant change of Doppler signals and the reduction of SNR caused by the obstacle. However, we found that if the model was reconstructed using the training data collected in the nLoS setting (i.e., the third experiment), the results restore to a comparable level, with all performance metrics recovering to around 90% (see the green bar). In addition, we also observe that when applying our model trained on the nLoS setting to test data from the first scenario (LoS sample), all metrics can still reach around 90% (see the gray bar). This experiment indicates that our system could eliminate the impact of nLoS path by retraining the model with the nLoS data. It gives us the insight that we can collect a rich set of training data to cover all scenarios for maintaining the robustness of our system.

5.8 Evaluation of Non-fall Activities

This section presents experiments to demonstrate system's performance in detecting different types of non-fall activities. Three types of non-fall activities are experimented: (a) objects (*i.e., ball, book, and box*) falling; (b) fall-like activities; and (c) some daily activities. In (a), an object is falling from a high place. In (b) and (c), the detailed fall-like and daily activities are shown in Table 1. For each activity type, we collect 60 samples for 40 fall activities in each experiment. We employ the model trained from two young people, as in Section 5.6. Our performance results are illustrated in Figure 11, where the purple, red, and green bars correspond respectively to the experiments for (a), (b), and (c). From this figure, we see our system to always achieve high accuracy in distinguishing these non-fall activities, with all metrics greater than 90%. Specifically, it has the highest

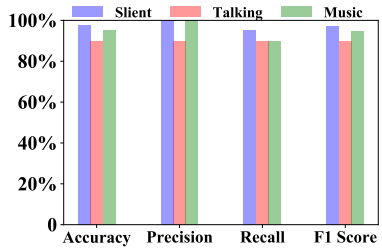


Fig. 12. Impact of the ambient sound.

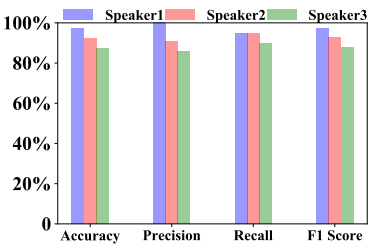


Fig. 13. Impact of the devices.

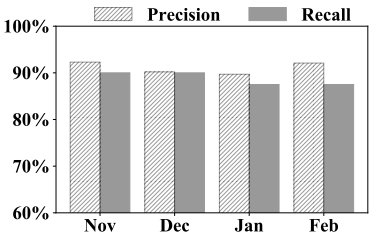


Fig. 14. Impact at different time.

performance in recognizing fall occurrences from the object falling. The reason is that the trajectory of any object falling is close to a straight line in the vertical direction, whereas that of a human's fall consists of a series of limbs' movements in different directions, causing a more complex Doppler signal. In addition, the sizes of these objects are relatively small comparing to that of a human body, resulting in their generated Doppler signals being much weaker. As a result, human fall accidents can be easily distinguished from the falls of those objects.

5.9 Impact of Time

We built a model based on the training data collected in November 2020, and then evaluated its performance for detecting human falls in November, December, January, and February. Notably, the worn clothes vary in different months due to the weather changes. We collect 40 fall events and 40 non-fall activities in each month for evaluation. Figure 14 shows the precision and recall at the four months. From this figure, we observe that the performance of our system changes only slightly across different months. This experiment demonstrate that the impact of clothes is negligible.

5.10 Impact of Ambient Sound

Since our system relies on acoustic signals, we evaluate the impact of ambient sound on our system. We conduct the experiments in the environment when playing music or having human talk. The sound sources are at $0.5m$ away from the device. The music volume is the same as that used for sending ultrasonic signals in our system, and people talk in the normal voice. Figure 12 depicts the performance of our system in the two scenarios for comparison with the scenario of no ambient sound (i.e., silent). Theoretically, the audible noise should not affect our system, since they are operating at different frequency bands. However, the figure reveals that the audible noises still have certain impacts on our system, despite slightly. This is due to the frequency leakage, which results in additional frequency components of the noises spreading to around 20kHz and mixing with the resulting Doppler signals of falls. Since the frequency leakage is not ubiquitous in the signals, its impact to our system is not severe.

5.11 Impact of Different Directions

We also ask one participant to fall in different directions for evaluating our system. Specifically, we make the participant fall in the 0°, 90°, -90°, and arbitrary degrees. The 0° means the participant falls along the line right in front of the speaker, whereas 90° and - 90° mean the fall directions are perpendicular to the sound propagation direction. Figure 15 shows the performance of our system in the four experiments. Our system is observed to perform the best along the 0° direction, with its accuracy, precision, recall, and F1 score equal to 98%, 97%, 97%, and 97%, respectively. When falling in an arbitrary direction, the four metrics drop to 92%, 92%, 91%, and 91%, respectively. Notably, when the fall direction is strictly perpendicular, one may assume no Doppler signals to be

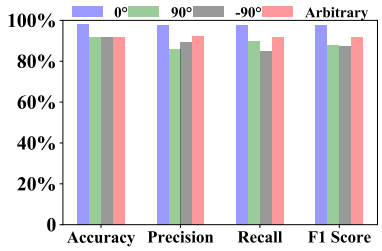


Fig. 15. Impact at different directions.

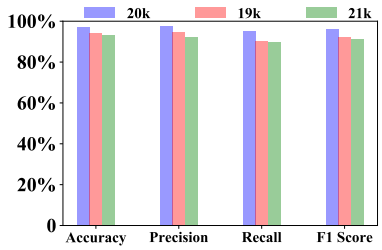


Fig. 16. Impact of the frequency.

generated so our system becomes inoperative. However, our experiment finds that strong Doppler signals are still generated, and our system can achieve the accuracy, precision, recall, and F1 score of around 91%, 85%, 90%, 87%, respectively. This could be due to the complex moving patterns of human falls. For example, the arm swing and the body movement would generate strong signals. Also, due to the relative short-range between the device and the participant (within a room), the sound waves are not confined to one plane. Instead, there is always an angle between the fall motion and the wave surface, resulting in a strong Doppler signal.

5.12 Transferability of Our System

We next evaluate our system with three different speakers, i.e., Edifier R1280DB, Logitech z200, and Amazon Echo, denoted as Speaker 1, Speaker 2, and Speaker 3. The first speaker is our default speaker, which is used for collecting the training data. We tune their volumes to the same transmission power and let them emit the signals at the same frequency, i.e., $20kHz$. The performance of our system with these three speakers are shown in Figure 13. We observe that Speaker 1 achieves the best performance whereas Speaker 3 achieves the worse on all four metrics. The reason is that Speaker 3, i.e., Amazon Echo, is a home speaker, which is not designed for generating the ultrasonic signal, emitting unstable signals on the high frequency band. However, our system with Amazon Echo still exhibits the accuracy, precision, recall, and F1 score of 87.5%, 85.7%, 90%, and 87.8%, respectively. This experiment demonstrates the transferability of our system with different devices.

In addition, we also conduct experiments to show model transferability across different frequencies. In particular, our experiment evaluates three working frequencies, i.e.,19kHz,20kHz, and 21kHz. We collect the training data only on the 20kHz, while conducting the detection tasks under 19kHz, 20kHz, and 21kHz. Figure 16 shows the performance of our detection tasks on the three frequencies. This figure reveals that, when comparing to that on the 20kHz, the performance outcomes of 19kHz and 21kHz degrade only slightly for the detection tasks. Specifically, the accuracy, precision, recall, and F1 score are 94%, 95%, 90%, and 93% (or 93%, 92%, 90%, and 91%), respectively, on 19kHz (or 21kHz). This experiment indicates that our system trained on one frequency can be generalized for use under other frequencies.

5.13 Computational Cost

We next measure the processing time of each main component in our system. Since our system aims to work in a real-time manner by processing the time sequence signals, we only need to take one sliced signal (over 0.8s) for measurement. The major computational costs come from the STFT and Interference Cancellation (Denoise) in signal processing, feature extraction, SVD decomposition, and HMM classification, which are $21ms$, $20ms$, $6.7ms$, $3.2ms$, and $1.2ms$, respectively. The computational costs from other components are negligible and can be omitted. Hence, the total processing time for each sliced signal is only around $52.1ms$, deemed to be

much less than the moving step length of the HMM window (= 0.1s). This experiment validates that our system could support the real-time processing for fall accident detection.

6 DISCUSSIONS

Our work successfully demonstrates the feasibility of employing an existing home device to conduct acoustic sensing for the fall detection purpose. It provides a proof of concept for validating the potential practice use of acoustic sensing in home applications. Yet, this system has some limitations that can be improved, leaving in our future work.

First, experimental results exhibit that our system is not perfect, unable to achieve 100% detection accuracy for surely identifying both fall and fall-like activities. This is due to many reasons, with one common reason being that some fall-like activities are very similar to falls under a certain context, misleading our system's detection. For example, when a person quickly sits down onto the chair or floor, its resulting pattern is very close to that of a fall. But, fortunately, these activities are not uniform for elders, and thus their information has little value for practice use. Nevertheless, we can investigate more salient and inconspicuous features as the input to help distinguish the fall and the fall-like activities. Another plausible solution is to collect more training data covering various fall-like activities to better training our model.

Second, our detection performance drops markedly in the nLoS scenario, where some objects block the signals. This makes the signal of interest to be much weaker and incurs more interference to degrade the performance of our system. In addition to retraining our model as discussed in Section 5.7, another plausible solution is to utilize multiple devices to cover all possible angles. Furthermore, the location information obtained by those devices can also be leveraged in our detection algorithm to improve its performance. In essence, we may first leverage the Doppler reflection to roughly estimate such a location information, and then train a location-based model for fall detection.

Third, the current effective detection range is still limited. For further practical deployments, we need to extend this range to be more than 5 meters. Several directions can be explored, such as developing more advanced interference cancellation and signal processing techniques to obtain more clear spectrogram, extracting more subtle and fine-grained features for event mining, among others.

Fourth, our work aims to demonstrate the possibility of developing a solution with only one speaker and one microphone for conducting fall detection, given that some homes do not possess any device with multiple microphones. Definitely, our developed system can also be directly migrated to multi-microphone devices with no changes. Thus, our solution is more general for home use. If we make a certain change (in the interference cancellation phase) by applying the beamforming algorithms [49, 50] with multiple microphones, our solution is expected to yield better performance.

Lastly, although ultrasound cannot be heard by the human being, it is reported in [30] that exposing one in the ultrasound environment would affect one's brain activity, known as the hypersonic effect. However, [11] shows no significant negative consequences observed when temporarily exposed in the environment even with the sound pressure level of 94dB. [20] continued to suggest a 70 dB guideline, since some people might be more sensitive to ultrasound. Hence, the safety guidelines for ultrasound are highly controversial. On the other hand, [56] exhibited that Commercial Off-the-Shelf (COTS) devices have the limited powers to produce ultrasound, so even the most sensitive people could hardly hear the ultrasound generated by them. Nonetheless, during our experiment, the transmission power is only set to 80% of the speaker's maximum power and the Sound Pressure Level (SPL) of the sensing signal is measured to be only 45dB at 1m away from the device. This SPL attenuates noticeably with an increase in distance, making us believe that our system is safe enough with little risks to the human being.

7 CONCLUSION

This paper has addressed the design and implementation of a non-wearable and accurate fall detection system, which leverages solely on an existing home audio device, widely applicable for home use. We control the speaker on the audio device to emit inaudible ultrasonic signals and let the co-located microphone to capture the reflected Doppler signals. Through developing a collection of solutions, including signal processing, interference cancellation, feature extraction, feature reduction, and model training, our system can get clear and rich patterns for fall events, distinguishable from other normal activities. We have evaluated our system with diverse fall types and different daily activities, demonstrating that it can achieve the precision and the recall of 92.6% and of 90.4% respectively in our default environment. When transferred to different environments, our system can still maintain its high accuracy, with over 90% for both precision and recall. In sharp contrast to the existing wearable and non-wearable solutions, our system possesses such salient features as convenient and easy for deployment, no additional cost to purchase dedicated devices, no resource competition with home Wi-Fi devices, among others, widely deployable for home use.

ACKNOWLEDGMENTS

This work was supported in part by NSF under Grants 1763620, 1948374, 1943509, and 2019511. Any opinion and findings expressed in the paper are those of the authors and do not necessarily reflect the view of funding agency.

REFERENCES

[1] Stefano Abbate, Marco Avvenuti, Francesco Bonatesta, Guglielmo Cola, Paolo Corsini, and Alessio Vecchio. 2012. A smartphone-based fall detection system. *Pervasive and Mobile Computing* 8, 6 (2012), 883–899.

[2] Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, and Gerald Penn. 2012. Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition. In *2012 IEEE international conference on Acoustics, speech and signal processing (ICASSP)*. IEEE, 4277–4280.

[3] Claire E Adam, Annette L Fitzpatrick, Cindy S Leary, Anjum Hajat, Elizabeth A Phelan, Christina Park, and Erin O Semmens. 2021. The Association between Gait Speed and Falls in Community Dwelling Older Adults with and without Mild Cognitive Impairment. *International Journal of Environmental Research and Public Health* 18, 7 (2021), 3712.

[4] M Umair Bin Altaf, Taras Butko, and Biing-Hwang Juang. 2015. Acoustic gaits: Gait analysis with footstep sounds. *IEEE Transactions on Biomedical Engineering* 62, 8 (2015), 2001–2011.

[5] Moeness G Amin, Yimin D Zhang, Fauzia Ahmad, and KC Dominic Ho. 2016. Radar signal processing for elderly fall detection: The future for in-home monitoring. *IEEE Signal Processing Magazine* 33, 2 (2016), 71–80.

[6] Paul M Baggenstoss. 2001. A modified Baum-Welch algorithm for hidden Markov models with multiple observation spaces. *IEEE Transactions on speech and audio processing* 9, 4 (2001), 411–416.

[7] Michael F Ballesteros, Kevin Webb, and Roderick J McClure. 2017. A review of CDC's Web-based Injury Statistics Query and Reporting System (WISQARS™): Planning for the future of injury surveillance. *Journal of safety research* 61 (2017), 211–215.

[8] Zhen-Peng Bian, Junhui Hou, Lap-Pui Chau, and Nadia Magnenat-Thalmann. 2014. Fall detection based on body part tracking using a depth camera. *IEEE journal of biomedical and health informatics* 19, 2 (2014), 430–439.

[9] Yabo Cao, Yujiu Yang, and WenHuang Liu. 2012. E-FallD: A fall detection system using android-based smartphone. In *2012 9th International Conference on Fuzzy Systems and Knowledge Discovery*. IEEE, 1509–1513.

[10] Yung-Chin Chen and Yi-Wen Lin. 2010. Indoor RFID gait monitoring system for fall detection. In *2010 2nd International Symposium on Aware Computing*. IEEE, 207–212.

[11] Mark D Fletcher, Sian Lloyd Jones, Paul R White, Craig N Dolder, Timothy G Leighton, and Benjamin Lineton. 2018. Effects of very high-frequency sound and ultrasound on humans. Part II: A double-blind randomized provocation study of inaudible 20-kHz ultrasound. *The Journal of the Acoustical Society of America* 144, 4 (2018), 2521–2531.

[12] Ajay Gadde, Moeness G Amin, Yimin D Zhang, and Fauzia Ahmad. 2014. Fall detection and classifications based on time-scale radar signal characteristics. In *Radar Sensor Technology XVIII*, Vol. 9077. International Society for Optics and Photonics, 907712.

[13] 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*. 1911–1914.

[14] Chih-Ning Huang and Chia-Tai Chan. 2014. A zigbee-based location-aware fall detection system for improving elderly telecare. *International journal of environmental research and public health* 11, 4 (2014), 4233–4248.
[15] Shehroz S Khan, Michelle E Karg, Dana Kulić, and Jesse Hoey. 2014. X-factor HMMs for detecting falls in the absence of fall-specific training data. In *International Workshop on Ambient Assisted Living*. Springer, 1–9.
[16] Sanaz Kianoush, Stefano Savazzi, Federico Vicentini, Vittorio Rampa, and Matteo Giussani. 2016. Device-free RF human body fall detection and localization in industrial workplaces. *IEEE Internet of Things Journal* 4, 2 (2016), 351–362.
[17] K Krishna and M Narasimha Murty. 1999. Genetic K-means algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 29, 3 (1999), 433–439.
[18] Bogdan Kwolek and Michal Kepski. 2015. Improving fall detection by the use of depth sensor and accelerometer. *Neurocomputing* 168 (2015), 637–645.
[19] Tracy Lee and Alex Mihailidis. 2005. An intelligent emergency response system: preliminary development and testing of automated fall detection. *Journal of telemedicine and telecare* 11, 4 (2005), 194–198.
[20] TG Leighton. 2016. Are some people suffering as a result of increasing mass exposure of the public to ultrasound in air? *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 472, 2185 (2016), 20150624.
[21] Yun Li, KC Ho, and Mihail Popescu. 2012. A microphone array system for automatic fall detection. *IEEE Transactions on Biomedical Engineering* 59, 5 (2012), 1291–1301.
[22] Dongha Lim, Chulho Park, Nam Ho Kim, Sang-Hoon Kim, and Yun Seop Yu. 2014. Fall-detection algorithm using 3-axis acceleration: combination with simple threshold and hidden Markov model. *Journal of Applied Mathematics* 2014 (2014).
[23] Brad Mager, Neal Patwari, and Maurizio Bocca. 2013. Fall detection using RF sensor networks. In *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. IEEE, 3472–3476.
[24] Wenguang Mao, Jian He, and Lili Qiu. 2016. CAT: high-precision acoustic motion tracking. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. 69–81.
[25] Wenguang Mao, Mei Wang, Wei Sun, Lili Qiu, Swadhin Pradhan, and Yi-Chao Chen. 2019. RNN-Based Room Scale Hand Motion Tracking. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–16.
[26] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Jacob E Sunshine. 2019. Opioid overdose detection using smartphones. *Science translational medicine* 11, 474 (2019).
[27] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaniel Watson. 2015. Contactless sleep apnea detection on smartphones. In *Proceedings of the 13th annual international conference on mobile systems, applications, and services*. 45–57.
[28] Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. 2016. Fingerio: Using active sonar for fine-grained finger tracking. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 1515–1525.
[29] Rajalakshmi Nandakumar, Alex Takakuwa, Tadayoshi Kohno, and Shyamnath Gollakota. 2017. Covertband: Activity information leakage using music. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–24.
[30] Tsutomu Oohashi, Emi Nishina, Manabu Honda, Yoshiharu Yonekura, Yoshitaka Fuwamoto, Norie Kawai, Tadao Maekawa, Satoshi Nakamura, Hidenao Fukuyama, and Hiroshi Shibasaki. 2000. Inaudible high-frequency sounds affect brain activity: hypersonic effect. *Journal of neurophysiology* (2000).
[31] Dimitri Palaz, Mathew Magimai-Doss, and Ronan Collobert. 2019. End-to-end acoustic modeling using convolutional neural networks for HMM-based automatic speech recognition. *Speech Communication* 108 (2019), 15–32.
[32] Sameera Palipana, David Rojas, Piyush Agrawal, and Dirk Pesch. 2018. FallDeFi: Ubiquitous fall detection using commodity Wi-Fi devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–25.
[33] Shing-Tai Pan, Yi-Heng Wu, Yi-Lan Kung, and Hung-Chin Chen. 2013. Heartbeat recognition from ECG signals using hidden Markov model with adaptive features. In *2013 14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*. IEEE, 586–591.
[34] Paola Pierleoni, Alberto Belli, Lorenzo Palma, Marco Pellegrini, Luca Pernini, and Simone Valenti. 2015. A high reliability wearable device for elderly fall detection. *IEEE Sensors Journal* 15, 8 (2015), 4544–4553.
[35] Mihail Popescu and Abhishek Mahnot. 2009. Acoustic fall detection using one-class classifiers. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 3505–3508.
[36] Luis Ramirez Rivera, Eric Ulmer, Yimin D Zhang, Wenbing Tao, and Moeness G Amin. 2014. Radar-based fall detection exploiting time-frequency features. In *2014 IEEE China Summit & International Conference on Signal and Information Processing (ChinaSIP)*. IEEE, 713–717.
[37] Wenjie Ruan, Quan Z Sheng, Lei Yang, Tao Gu, Peipei Xu, and Longfei Shangguan. 2016. AudioGest: enabling fine-grained hand gesture detection by decoding echo signal. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing*. 474–485.
[38] Wenjie Ruan, Lina Yao, Quan Z Sheng, Nickolas Falkner, Xue Li, and Tao Gu. 2015. Tagfall: Towards unobstructive fine-grained fall detection based on uhf passive rfid tags. In *proceedings of the 12th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services on 12th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and*

Services. 140–149.

[39] Arslan Shaukat, Muhammad Ahsan, Ali Hassan, and Farhan Riaz. 2014. Daily sound recognition for elderly people using ensemble methods. In *2014 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*. IEEE, 418–423.

[40] Xingzhe Song, Boyuan Yang, Ge Yang, Ruirong Chen, Erick Forno, Wei Chen, and Wei Gao. 2020. SpiroSonic: monitoring human lung function via acoustic sensing on commodity smartphones. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.

[41] Erik E Stone and Marjorie Skubic. 2014. Fall detection in homes of older adults using the Microsoft Kinect. *IEEE journal of biomedical and health informatics* 19, 1 (2014), 290–301.

[42] 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*. 591–605.

[43] Yonglong Tian, Guang-He Lee, Hao He, Chen-Yu Hsu, and Dina Katabi. 2018. RF-based fall monitoring using convolutional neural networks. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–24.

[44] Vibha Tiwari. 2010. MFCC and its applications in speaker recognition. *International journal on emerging technologies* 1, 1 (2010), 19–22.

[45] Lina Tong, Quanjun Song, Yunjian Ge, and Ming Liu. 2013. HMM-based human fall detection and prediction method using tri-axial accelerometer. *IEEE Sensors Journal* 13, 5 (2013), 1849–1856.

[46] Ph Van Dorp and FCA Groen. 2008. Feature-based human motion parameter estimation with radar. *IET Radar, Sonar & Navigation* 2, 2 (2008), 135–145.

[47] Stefano Volpato, Lara Bianchi, Fulvio Lauretani, Fabrizio Lauretani, Stefania Bandinelli, Jack M Guralnik, Giovanni Zuliani, and Luigi Ferrucci. 2012. Role of muscle mass and muscle quality in the association between diabetes and gait speed. *Diabetes care* 35, 8 (2012), 1672–1679.

[48] 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*. 1–11.

[49] Anran Wang, Dan Nguyen, Arun R Sridhar, and Shyamnath Gollakota. 2021. Using smart speakers to contactlessly monitor heart rhythms. *Communications biology* 4, 1 (2021), 1–12.

[50] Anran Wang, Jacob E Sunshine, and Shyamnath Gollakota. 2019. Contactless infant monitoring using white noise. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–16.

[51] Hao Wang, Daqing Zhang, Yasha Wang, Junyi Ma, Yuxiang Wang, and Shengjie Li. 2016. RT-Fall: A real-time and contactless fall detection system with commodity WiFi devices. *IEEE Transactions on Mobile Computing* 16, 2 (2016), 511–526.

[52] Wei Wang, Alex X Liu, and Ke Sun. 2016. Device-free gesture tracking using acoustic signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. 82–94.

[53] Yingxue Wang, Yanan Chen, Md Zakirul Alam Bhuiyan, Yu Han, Shenghui Zhao, and Jianxin Li. 2018. Gait-based human identification using acoustic sensor and deep neural network. *Future Generation Computer Systems* 86 (2018), 1228–1237.

[54] Yanwen Wang, Jiaxing Shen, and Yuanqing Zheng. 2020. Push the Limit of Acoustic Gesture Recognition. *IEEE Transactions on Mobile Computing* (2020).

[55] Yuxi Wang, Kaishun Wu, and Lionel M Ni. 2016. Wifall: Device-free fall detection by wireless networks. *IEEE Transactions on Mobile Computing* 16, 2 (2016), 581–594.

[56] Matt Wixey, Emiliano De Cristofaro, and Shane D Johnson. 2020. On the Feasibility of Acoustic Attacks Using Commodity Smart Devices. In *2020 IEEE Security and Privacy Workshops (SPW)*. IEEE, 88–97.

[57] Wei Xu, ZhiWen Yu, Zhu Wang, Bin Guo, and Qi Han. 2019. Acousticid: gait-based human identification using acoustic signal. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–25.

[58] Sen Yang, Moliang Zhou, Shuhong Chen, Xin Dong, Omar Ahmed, Randall S Burd, and Ivan Marsic. 2017. Medical workflow modeling using alignment-guided state-splitting HMM. In *2017 IEEE International Conference on Healthcare Informatics (ICHI)*. IEEE, 144–153.

[59] Sangki Yun, Yi-Chao Chen, and Lili Qiu. 2015. Turning a mobile device into a mouse in the air. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*. 15–29.

[60] Cheng Zhang, Qiuyue Xue, Anandghan Waghmare, Sumeet Jain, Yiming Pu, Sinan Hersek, Kent Lyons, Kenneth A Cunefare, Omer T Inan, and Gregory D Abowd. 2017. Soundtrak: Continuous 3d tracking of a finger using active acoustics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2 (2017), 1–25.

[61] Tong Zhang, Jue Wang, Ping Liu, and Jing Hou. 2006. Fall detection by embedding an accelerometer in cellphone and using KFD algorithm. *International Journal of Computer Science and Network Security* 6, 10 (2006), 277–284.