# A Microphone Array System for Automatic Fall Detection

Yun Li, *Student Member, IEEE*, K. C. Ho*, *Fellow, IEEE*, and Mihail Popescu, *Senior Member, IEEE*

***Abstract*—More than a third of elderly fall each year in the United States. It has been shown that the longer the lie on the floor, the poorer is the outcome of the medical intervention. To reduce delay of the medical intervention, we have developed an acoustic fall detection system (acoustic-FADE) that automatically detects a fall and reports it promptly to the caregiver. Acoustic-FADE consists of a circular microphone array that captures the sounds in a room. When a sound is detected, acoustic-FADE locates the source, enhances the signal, and classifies it as "fall" or "nonfall." The sound source is located using the steered response power with phase transform technique, which has been shown to be robust under noisy environments and resilient to reverberation effects. Signal enhancement is performed by the beamforming technique based on the estimated sound source location. Height information is used to increase the specificity. The mel-frequency cepstral coefficient features computed from the enhanced signal are utilized in the classification process. We have evaluated the performance of acoustic-FADE using simulated fall and nonfall sounds performed by three stunt actors trained to behave like elderly under different environmental conditions. Using a dataset consisting of 120 falls and 120 nonfalls, the acoustic-FADE achieves 100% sensitivity at a specificity of 97%.**

***Index Terms*—Beamforming (BF) and height information, fall detection, microphone arrays, sound source localization.**

## I. Introduction

ACCORDING to a Center for Disease Control report [1], more than one third of older Americans fall each year. For elderly, falls are the leading cause of death [1]. In 2007 alone, over 18,000 older adults died from fall injuries [1]. Falls are also the most common cause of nonfatal injuries such as lacerations, hip fractures, or head traumas [2]. In these cases, it has been found that the longer the lie on the floor, the poorer is the outcome of the medical intervention [3]. Among entirely elderly population, those living alone have the greatest risk of delayed intervention. As a consequence, each year, about 3% [4] of them are found helpless or dead at home by paramedics.

To address the problem of medical intervention delay, it is imperative to detect the falls as soon as they occur so that immediate assistance can be provided. A variety of fall detection methods have been published in the recent scientific literature. There are two main types of fall detection devices: wearable and nonwearable. Wearable devices, like accelerometer-based ones [5]–[8], detect falls by measuring the applied acceleration along the vertical axis. Although the wearable devices are versatile and effective in both indoor and outdoor environments, they often have power management problems and potential inconvenience for carrying them all the time during daily living activities. Some nonwearable devices include floor vibration sensors [9]–[11], video cameras [12], [13], IR cameras [14], and bed sensors [15]. The floor vibration sensors [9] are inexpensive and privacy preserving, but their performance is dependent on the floor type. Video cameras, IR cameras, and bed sensors are promising technologies that are still trying to address challenges related to low light, field of view, privacy, and high cost. In this paper, we describe an acoustic fall detection system (acoustic-FADE) that detects a fall as soon as it occurred based on its sound signature since the acoustic sensors are practically easy to use, able to handle privacy issues, reliable, and inexpensive.

Several previous works [10], [11], [16], [17] have described acoustic-FADEs based on a linear array of microphones. In addition, we investigated several fall detection algorithms for use with the microphone array such as fuzzy rule methods [16] and one-class classifiers [17]. However, previous acoustic systems had limited success in increasing the specificity due to the environmental noise and interference, in part because they were not considering the entire spatial information related to the sound source. The study in [18] showed that a vertical linear microphone array can greatly increase the specificity by using the estimated height of the sound source and passing the near-ground signals (e.g., with detected height $<0.5$ m) to the classification algorithm. This approach not only improves the specificity, but also increases the computation efficiency. However, the height estimation accuracy using a linear array is very sensitive to the acoustic properties such as background noise, reverberation, and interference. In addition, the height estimation itself was not reliable due to the conical shape localization ambiguity of the linear microphone array. To deal with these challenges, we proposed a new version of acoustic-FADE [19] that employs an eight-microphone circular array for person tracking and fall detection.

Beamforming (BF) can enhance the desired signal and reduce interferences such as from TV, radio, or phone ringing. In this

Manuscript received October 21, 2011; revised December 6, 2011 and January 19, 2010; accepted January 21, 2012. Date of publication February 3, 2012; date of current version April 20, 2012. This work was supported in part by the National Science Foundation under Grant CNS-0931607. *Asterisk indicates corresponding author.*

Y. Li is with the Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO 65211 USA (e-mail: yl874@mail.missouri.edu).

*K. C. Ho is with the Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO 65211 USA (e-mail: hod@missouri.edu).

M. Popescu is with the Department of Health Management and Informatics, University of Missouri, Columbia, MO 65211 USA (e-mail: popescum@missouri.edu).

Digital Object Identifier 10.1109/TBME.2012.2186449

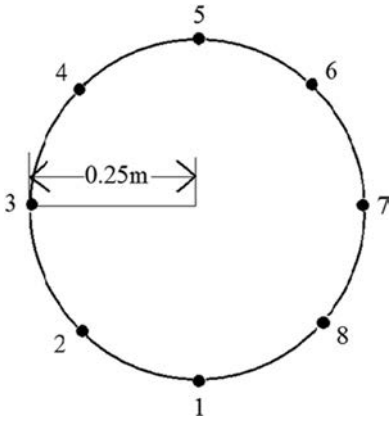


Fig. 1. Acoustic circular microphone array configuration.

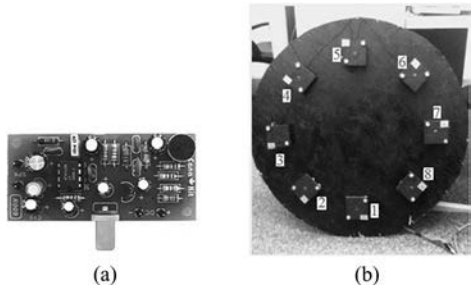


Fig. 2. (a) Front view of an amplifier-integrated microphone. (b) Front view of the microphone array (numbers mark the indices of microphones).

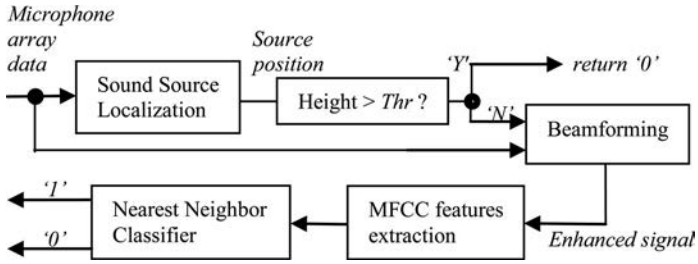


Fig. 3. Processing block diagram of the proposed acoustic-FADE (*Thr* = positive scalar height threshold, "*Y*": yes, "*N*": no, "*1*": classified as a fall, "*0*": classified as a nonfall).

paper, we use BF to improve the classification performance in realistic environments that often have large background noise and strong interferences. The BF technique is widely used in a variety of applications, such as videoconferencing [20], human–robot interaction [21], and bird monitoring [22] for increasing sound quality. BF requires sound source localization and we use steered response power with phase transform (SRP-PHAT) method [19] to locate the source. For sound classification, we apply the mel-frequency cepstral coefficients (MFCCs) features with a nearest neighbor (NN) approach, as previously used in [18] and [19].

The acoustic-FADEs proposed in [16]–[18] use a single microphone and the specificity is low. In [19], although a circular array is employed, it only focuses on locating the sound source and does not use any method on improving the signals for better classification. The cross-validation results show that the acoustic-FADE in [19] has a low specificity of about 90% in order to reach at a sensitivity of 87%. In this paper, in continuation of our preliminary work in [19], we propose the acoustic-FADE which improves the sound source localization accuracy, utilizes the estimated source height to increase the specificity, applies BF to enhance the received signals, and employs feature-based classification to improve the fall detection performance.

The structure of the paper is as follow. In Section II, we describe the proposed acoustic-FADE. Section III presents the sound source localization and BF techniques. In Section IV, we elaborate the methodology of fall detection. Section V describes the data and performance evaluations. Section VI provides the results and Section VII is the discussion. Finally, Section VIII concludes the paper.

## II. Proposed Acoustic-FADE

The proposed acoustic-FADE has two components: the acoustic sensor hardware and the data processing software. The acoustic sensor is a circular microphone array, as shown in Fig. 1, that consists of eight omnidirectional microphones uniformly distributed along a circle having a radius of 0.25 m. From array signal processing literatures, wider separation of array elements gives better localization accuracy. However, the separation between microphones should not be larger than the half signal wavelength to avoid spatial aliasing [20]. The fall signal has highest frequency component around 1000 Hz, which corresponds to a wavelength of about 35 cm. The separation of microphones should therefore be about 18 cm. This will be satisfied when choosing the radius of the array as 25 cm [19]. Each microphone component has an amplifier and a gain adjuster, as depicted in Fig. 2(a). Fig. 2(b) depicts the actual microphone array assembled. The array has enough coverage for a typical living room.

The acoustic signals from the microphones of the array are sampled synchronously at a rate of 20 KHz and the data are quantized to 12-bit resolution by data acquisition board mounted on the back side of the array.

Depending on the manufacturing variations, the gain and phase delay of each microphone element may differ from one another which can degrade the performance of sound source localization and BF. The microphone array is carefully calibrated using a white noise signal emitted at a known location to ensure each microphone has the same gain and phase characteristics [19].

For the software component, the processing block diagram of the proposed acoustic-FADE is shown in Fig. 3. From the sound data collected by the microphone array, we first apply localization to determine the position of the sound source.

If the source is found to be located above ground when the estimated height of the source is larger than a threshold *Thr*, it will be considered as from a nonfall and no further processing is needed. Otherwise, BF is used to enhance the sound signal using the estimated source position. MFCC features are extracted and a NN classifier determines if the sound is from a fall. We shall

explain the individual blocks in different sections in the following (the utilization of the estimated source height is considered as a component for classification).

## III. Sound Source Localization and BF

In practical environments, the fall sound may have low SNR and could be corrupted by interference. It is necessary to enhance the acoustic signal in order to achieve better detection accuracy. Improving an acoustic signal using a microphone array requires the location of the source. Given the signal measurements observed by a microphone array, the first step is to estimate the source location and BF can be followed to enhance the acoustic signal.

In the near-field scenario, the signal wavefront is curved, and traditional direction of arrival methods such as Multiple Signal Classification (MUSIC) [23], which assumes a planner wavefront, are not appropriate to determine the source location. Furthermore, reverberation effect could be significant in typical home environments, where the microphone array is deployed. We shall use a different localization method than MUSIC that is robust and resilient to reverberations.

The steered response power (SRP) [24] technique provides reasonably accurate near-field source location estimate under very noisy conditions. Its variation, SRP-PHAT [25], is able to yield a better source location estimate than PHAT under reverberation environments. We shall describe below the signal model, the SRP and SRP-PHAT methods for the localization of a sound source.

### A. Measurement Model

Let $\boldsymbol{u^o}$ be the true 3-D position in the Cartesian coordinate of a sound source to be located. The source radiates a signal and it is received by a sensor array that has $M$ microphones located at $\boldsymbol{s_i}$, $i = 0, 1, \ldots, M-1$. The observed signal at microphone $i$ is modeled as follows:

$$x_i(n) = s(n - \tau_i^o) + \varepsilon(n), \qquad n = 0, 1, \ldots, N-1 \quad (1)$$

where the sampling time interval is normalized to unity and $N$ is the total number of data points. $s(n)$ is the source signal, $\varepsilon(n)$ is the random noise, and $\tau_i^o$ is the propagation delay for the source signal to reach microphone $i$. To minimize the reverberation effect, our previous study [26] shows that the processing window for sound source localization should be around 25 ms ($N =$ 25 ms $\times$ 20 KHz $=$ 500 data points) started at the beginning of the signal.

Without loss of generality, we choose sensor 0 as the reference sensor for representing the relative propagation delays of the signals at different sensors. Hence, $\tau_0^o = 0$ and

$$\tau_i^o = \frac{1}{c}(|\boldsymbol{u^o} - \boldsymbol{s_i}| - |\boldsymbol{u^o} - \boldsymbol{s_0}|), \quad i = 1, \ldots, M-1 \quad (2)$$

where $c$ is the signal propagation speed. The symbol $|\cdot|$ denotes the Euclidean norm and $|\boldsymbol{u^o} - \boldsymbol{s_i}|$ is the distance between the source and the $i$th sensor. We would like to estimate the source position $\boldsymbol{u^o}$ using the signal measurements $x_i(n), i = 1, \ldots, M-1$, from the microphone array.

### B. Steered Response Power

If we assume that the source is at certain position $\boldsymbol{u}$, the delays for the source signal to reach the microphones can be regenerated based on (2):

$$\tau_i(\boldsymbol{u}) = \frac{1}{c}(|\boldsymbol{u} - \boldsymbol{s_i}| - |\boldsymbol{u} - \boldsymbol{s_0}|), \quad i = 1, \ldots, M-1. \quad (3)$$

After correcting the delays from the microphone outputs and adding them together, we have

$$z(n) = \sum_{i=0}^{M-1} x_i(n + \tau_i(\boldsymbol{u})). \quad (4)$$

The energy of $z(n)$ is expected to be largest when the assumed location is equal to the true location $\boldsymbol{u^o}$ in the absence of noise because the delay corrected microphone outputs are aligned exactly and added coherently to form $z(n)$. It is based on this idea that the SRP method estimates $\boldsymbol{u^o}$ by maximizing

$$J_{\mathrm{SRP}}(\boldsymbol{u}) = \sum_{n=0}^{N-1} z^2(n) = \sum_{n=0}^{N-1} \sum_{i,j=0}^{M-1} x_i(n+\tau_i(\boldsymbol{u}))x_j(n+\tau_j(\boldsymbol{u}))$$
$$\approx \sum_{i,j=0}^{M-1} R_{i,j}(\tau_i(\boldsymbol{u}) - \tau_j(\boldsymbol{u})) \quad (5)$$

where

$$R_{i,j}(\lambda) = \sum_{n=0}^{N-1} x_i(n)x_j(n-\lambda) \quad (6)$$

is the cross correlation between $x_i(n)$ and $x_j(n)$ at relative delay $\lambda$. The approximation in (5) is valid when $N$ is large, which can be easily satisfied for our application. The relative delay $\lambda$ can be any real value and the cross correlation is computed via the Fourier transforms of $x_i(n)$ and $x_j(n)$ as follows:

$$R_{i,j}(\lambda) = \frac{1}{N} \sum_{k=0}^{N-1} X_i^*(k)X_j(k)e^{-j(2\pi/N)k\lambda} \quad (7)$$

where $X_i(k), i = 0, 1, \ldots, M-1$ is the fast Fourier transform (FFT) of $x_i(n)$ and $(\cdot)^*$ is the complex conjugate of $(\cdot)$. Since the SRP function is highly nonlinear with respect to $\boldsymbol{u}$, numerical search is needed to maximize $J_{\mathrm{SRP}}(\boldsymbol{u})$.

The SRP method can be summarized as follows. For each hypothesized source position $\boldsymbol{u}$,

1) obtain $\tau_i(\boldsymbol{u})$ and $\tau_j(\boldsymbol{u})$ using (3), for $i, j = 0, 1, \ldots, M-1$;
2) find $R_{i,j}(\tau_i(\boldsymbol{u}) - \tau_j(\boldsymbol{u}))$ via (7), for $i, j = 0, 1, \ldots, M-1$;
3) compute SRP function $J_{\mathrm{SRP}}(\boldsymbol{u})$ using the second line of (5).

The source location estimate is the position where $J_{\mathrm{SRP}}(\boldsymbol{u})$ has the largest value.

### C. Steered Response Power With Phase Transform

The SRP-PHAT applies a prefilter with frequency characteristics

$$H(k) = \frac{1}{\sqrt{|X_i(k)X_j(k)|^\beta}} \tag{8}$$

to $x_i(n)$ and $x_j(n)$ before cross correlating them. The parameter $\beta$ is a user-defined constant that is between 0 and 1. The resulting cross correlation becomes

$$\tilde{R}_{i,j}(\lambda) = \frac{1}{N}\sum_{k=0}^{N-1} \frac{X_i^*(k)X_j(k)}{|X_i(k)X_j(k)|^\beta} e^{-j(2\pi/N)k\lambda}. \tag{9}$$

The SRP-PHAT scoring function is

$$J_{\text{SRP-PHAT}}(\boldsymbol{u}) = \sum_{i,j=0}^{M-1} \tilde{R}_{i,j}(\tau_i(\boldsymbol{u}) - \tau_j(\boldsymbol{u})). \tag{10}$$

The source location estimate $\hat{\boldsymbol{u}}$ is the value of $\boldsymbol{u}$ that maximizes $J_{\text{SRP-PHAT}}(\boldsymbol{u})$. Again, numerical search is needed due to the complicated nature of the $J_{\text{SRP-PHAT}}(\boldsymbol{u})$ surface. Under reverberation environments, it has been shown that SRP-PHAT yields more accurate and reliable location estimate than SRP.

### D. Beamforming

BF is the process of improving the signal reception by properly combining a source signal received at several spatially separated microphones. BF not only reduces the background noise to increase the SNR, it also attenuates any interference coming from a direction other than that of the intended source signal, even if the noise and interferences occupy the same frequency band as the intended signal.

There are many BF techniques available such as LCMV, MVDR, etc. [20]. Perhaps the simplest one is delay and-sum [20]. The delay-and-sum beamformer output is constructed using

$$\hat{s}(n) = \frac{1}{M}\sum_{i=0}^{M-1} x_i(n + \tau_i(\hat{\boldsymbol{u}})) \tag{11}$$

where $\hat{\boldsymbol{u}}$ is the estimated source position. Note that $\tau_i(\hat{\boldsymbol{u}})$ is not an integer in general. The correction of delay in $x_i(n)$ is implemented in the frequency domain through the use of FFT by introducing linear phase shift with slope equal to $\tau_i(\hat{\boldsymbol{u}})$.

## IV. Fall Detection Methodology

### A. Utilizing Height Information

The previous study [18] proposed an approach to increase the specificity by using the height estimate of the sound source. The motivation is that given a height threshold, a fall, which comes from ground, can be separated from a nonfall, which comes above ground. In the proposed acoustic-FADE, the height information is directly available from sound source localization without additional cost. The criterion for determining the height threshold value is to filter out the above-ground nonfalls as many as possible while keeping all the falls. The height threshold was determined manually over a small set of data in the previous work [18].

To improve the robustness of fall detection in different acoustic environments, we propose here an approach to automatically set the height threshold using Otsu's method [27] that is originally used for automatic thresholding in image processing. Otsu's method determines the threshold to separate two classes by minimizing their intraclass variances. Since Otsu's method is designed for image, we first generate a grayscale image of which the intensity of each pixel represents the height of a sound signal in the training dataset. Then Otsu's method can be applied to the image and generate the threshold.

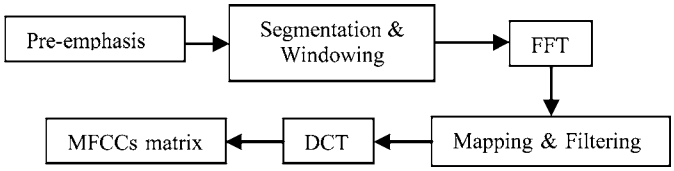


Fig. 4. Steps of computing MFCCs.

### B. MFCC Features

Mel-scale frequency cepstral coefficients (MFCCs) are the most commonly used acoustic features for speech/sound recognition. MFCCs take into consideration human perception sensitivity with respect to frequencies and therefore they are often considered best for speech/sound recognition. In fall detection, we use MFCC features to identify the sounds of falls and nonfalls. The computation of MFCCs is commonly performed in six steps [28], which are shown in the blocks diagram of Fig. 4.

The descriptions of the blocks are as follows.

1) *Preemphasis:* It is to compensate the HF signal component that was attenuated during sound production. It is implemented by passing the input through a first-order high-pass filter expressed as follows:
$$y(n) = x(n) - \eta x(n-1) \tag{12}$$
and typically $\eta \in [0.95, 1]$. We set it to 0.96 in this study.
2) *Segmentation and Windowing:* The input sound sample is segmented into a number of frames with 32% overlap. The frame size is set to 256 data points and the data length is 500 ms $\times$ 20 KHz $= 10^4$ points. The number of frames in a sound sample is $[(10^4 - 0.32 \times 256)/(0.32 \times 256)] = 121$. Each frame is multiplied by a Hamming window in order to minimize the boundary effect due to segmentation.
3) *FFT:* It is used to convert the time-domain points into the frequency domain to obtain the frequency spectral features of a sound sample. It is applied to each frame separately.
4) *Mapping and Filtering:* For each windowed frame, the mel-scale mapping is performed using the relation [29]
$$\text{Mel}(f) = 1125 \ln\left(1 + \frac{f}{700}\right) \tag{13}$$
where $f$ denotes the linear frequency ranging from 0 to 10 KHz. A bank of 30 triangular bandpass filters (BPFs) is assigned over the mel-frequency Mel($f$) in equal interval

spacings and they are multiplied to the FFT values. The energy $E_l$ of the $l$th BPF output, $l = 1, 2, \ldots, 30$, is generated.

5) *Discrete cosine transform:* It is applied to $E_l$ to create $J = 6$ MFCCs for frame $i$, i.e.,

$$C_{j,i} = \sum_{l=1}^{L} \cos\left[\frac{(j+1)(l-0.5)\pi}{L}\right] E_l$$
$$L = 30, j = 1, \ldots, J. \quad (14)$$

6) Repeat steps 4 and 5 for all 121 frames to obtain the MFCC matrix $\boldsymbol{C}$ of one sound sample as follows:

$$\boldsymbol{C} = \begin{bmatrix} C_{1,1} & \cdots & C_{1,121} \\ \vdots & \ddots & \vdots \\ C_{6,1} & \cdots & C_{6,121} \end{bmatrix}. \quad (15)$$

### C. NN Classifier

The NN classifier is a special case of the $k$-NN classifier when k is equal to 1. The idea of NN is to assign an unknown test sample to the class that has a sample closest to it. In our case, the closeness is measured by Euclidean distance based on the sample features. The MFCC features of a sample are represented by a $J \times I$ matrix ($J = 6$, $I = 121$ in our case). We use the Frobenius norm as a measure of the distance between a pair of $J \times I$ feature matrices $\boldsymbol{p}$ and $\boldsymbol{q}$

$$\text{dist}_{\boldsymbol{p}-\boldsymbol{q}} = |\boldsymbol{p} - \boldsymbol{q}|_{J\times I} = \sqrt{\sum_{j=1}^{J}\sum_{i=1}^{I}(p_{ji} - q_{ji})^2}$$
$$j = 1, \ldots, J, i = 1, \ldots, I. \quad (16)$$

The null and alternate hypotheses of a test sample for fall detection are

$$\text{H}_0\text{: Nonfall}$$
$$\text{H}_1\text{: Fall.}$$

$\text{H}_1$ is chosen if

$$\frac{\min(\text{dist}_{\text{test-nonfall}})}{\min(\text{dist}_{\text{test-fall}})} > 1 \quad (17)$$

otherwise $\text{H}_0$ is chosen. $\min(\text{dist}_{\text{test-fall}})$ represents the minimum Frobenius norm between a test feature matrix and a fall-labeled feature matrix. The same definition applies to $\min(\text{dist}_{\text{test-nonfall}})$.

## V. Experimental Data and Performance Evaluation

### A. Data Measurements

We have obtained approval by the Institutional Review Board, University of Missouri, for this research project. The experimental data consists of falls and nonfalls performed by three stunt actors (two females, age: 32 and 46, height: 5′3″ and 5′4″, weight: 135 and 117 lbs; one male, age: 30, height: 5′8″, weight: 170 lbs) and please refer to Table I for more details. They have been trained and instructed by our nurse collaborators to fall like an elderly where the training details have been documented in [30].

TABLE I
Stunt Actor Information

| | Gender | Age | Height | Weight | # falls | # nonfalls |
|---|---|---|---|---|---|---|
| A | Female | 32 | 5'3" | 135 lbs | 40 | 40 |
| B | Female | 46 | 5'4" | 117 lbs | 40 | 40 |
| C | Male | 30 | 5'8" | 170 lbs | 40 | 40 |


Fig. 5. (a) Front camera view of the experiment. (b) Side camera view of the experiment.

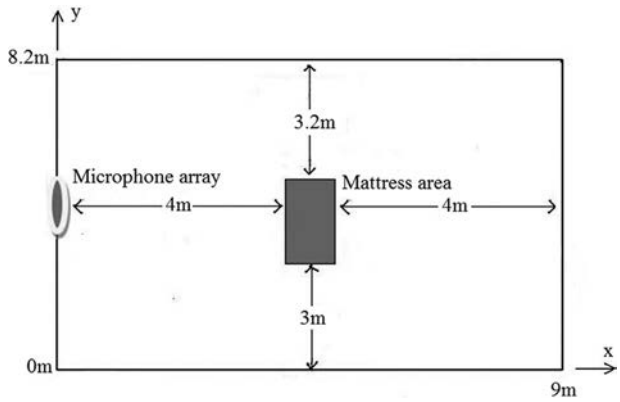


Fig. 6. Top view of the configuration for collecting dataset DAT1. The microphone array center is at $x$–$y$ coordinate of (0 m, 4.1 m) with a height of 2.46 m. The mattress occupies an area of 2 m × 1 m.

The experimental data has two sets—DAT1 and DAT2. DAT1 is a dataset collected in a laboratory room and DAT2 is a dataset collected in realistic living environments in TigerPlace.[1]

Both of them are collected using the proposed acoustic-FADE.

*1) Description of DAT1:* The size of the laboratory room is 9 m × 8.2 m × 3 m (L × W × H) and the reverberation time is $\text{RT}_{60} \approx 0.9$ s, which is calculated based on Sabine's reverberation equation for room acoustics [31]. Fig. 5 shows two camera views of a stunt actor who is ready to fall on the mattress. Fig. 6 shows the relative positions of the microphone array and the mattress, where the fall sounds come from. Higher sensitivity and specificity could be expected if the microphone array is closer to the sound source because of better SNR.

The stunt actors fall on a mattress and generate a fall sound in each trial. Apart from fall signals, the stunt actors also produce

[1]TigerPlace is an assisted living facility in Columbia, Missouri, MO. See website at http://eldertech.missouri.edu/overview.htm

TABLE II
DESCRIPTION OF DAT1

| Falls (format: 'type'-'trend') | #files[2] | Nonfalls[3] | #files[2] |
|---|---|---|---|
| Balance-Forward | 6 | Closing window | 6 |
| Balance-Backwards | 6 | Typing keyboard | 6 |
| Balance-Left | 6 | Key shaking | 6 |
| Balance-Right | 6 | Machine noise | 6 |
| Lose consciousness-Forward | 6 | Phone ringing | 6 |
| Lose consciousness-Backwards | 6 | Knocking door | 6 |
| Lose consciousness-Left | 6 | Talking | 6 |
| Lose consciousness-Right | 6 | Lying on a bed | 6 |
| Trip and fall-Forward | 6 | Lying on a couch | 6 |
| Trip and fall-Sideways | 6 | Sitting on a chair | 6 |
| Slip and fall-Forward | 6 | Normal walking | 6 |
| Slip and fall-Sideways | 6 | Slow walking | 6 |
| Slip and fall-Backwards | 6 | Rubbing | 6 |
| Reach fall (chair)-Forward | 6 | Dropping books | 6 |
| Reach fall (chair)-Left | 6 | Dropping balls | 6 |
| Reach fall (chair)-Right | 6 | Dropping cans | 6 |
| Slide fall-Forward | 6 | Dropping wood cases | 6 |
| Slide fall-Backwards | 6 | Dropping plastic bottles | 6 |
| Couch fall-Upper body first | 6 | Rolling a can | 6 |
| Couch fall-Hips first | 6 | Rolling a plastic bottle | 6 |
| Total # nonfall files | 120 | Total # nonfall files | 120 |

[2]Each file has eight-channel signals of 500 ms duration and is recorded at 20 KHz sampling rate.
[3]The first 60 nonfalls are from sound sources above the ground and the last 60 are from sound sources on the ground.

TABLE III
DESCRIPTION OF DAT2

| | Stunt actor | Room size (L×W×H) | $RT_{60}$ | Surface type | # fall files[2] |
|---|---|---|---|---|---|
| Subset 1 | B,C | 5m×4m×3m | ≈0.77s | Carpet | 12 |
| Subset 2 | B,C | 5m×3m×3m | ≈0.73s | Carpet | 12 |
| Subset 3 | B,C | 6m×5m×3m | ≈0.81s | Carpet | 12 |
| Subset 4 | B,C | 5m×4m×3m | ≈0.77s | Mattress | 12 |

[2]Each file has eight-channel signals of 500 ms duration and is recorded at 20 KHz sampling rate.

nonfall sounds from different actions and they can come from any possible locations in the room. DAT1 contains 120 files of falls and 120 files of nonfalls. Half of the nonfall sounds are above ground and the other half are on ground. There are 20 types of falls and 20 types of nonfalls. Table II gives the description of DAT1.

*2) Description of DAT2:* DAT2 consists of four subsets that were collected in four different apartments in TigerPlace. The four experiments are conducted independently and the descriptions of the experiments including the stunt actors, the acoustic characteristics (room size and $RT_{60}$), the type of floor surface, and the number of falls are shown in Table III.

The stunt actor B is a female of age 46, height 5′4″, and weight 117 lbs, and the stunt actor C is a male of age 30, height 5′8″, and weight 170 lbs, as tabulated in Table I. Each of them performed six falls for each subset. The mattresses used in these experiments are all the same and are similar to the one used in DAT1. The carpet was typical for assisted living homes.

### B. Performance Evaluation

The acoustic-FADE performance was assessed through receiver operating characteristic (ROC) curve, which depicts the classifier sensitivity versus 1-specificity at different detection thresholds. Since an ROC curve cannot be obtained by a binary decision rule like the one of NN, (17) is modified by replacing 1 with a threshold variable $r$ [32] and the decision rule becomes:

$H_1$ (fall) is chosen if

$$\frac{\min(\text{dist}_{\text{test-nonfall}})}{\min(\text{dist}_{\text{test-fall}})} > r, \quad r \in (0,2) \tag{18}$$

otherwise, $H_0$ (nonfall) is selected. To quantify the ROC curve for better comparison, in addition to sensitivity and specificity, we use area under the ROC curve (AUC) and accuracy as well.

*1) Performance Metric Index:* Let us define four important evaluation factors, which are number of true positives (TP), number of false positives (FP), number of false negatives (FN), and number of true negatives (TN). Given a certain detection threshold $r$, then

$$\text{Sensitivity}|_r = \left.\frac{\text{TP}}{\text{TP}+\text{FN}}\right|_r \tag{19}$$

$$\text{Specificity}|_r = \left.\frac{\text{TN}}{\text{TN}+\text{FP}}\right|_r. \tag{20}$$

AUC is between 0 and 1. The larger the AUC, the better the classification performance. If the operating point of the ROC curve is chosen at detection threshold $r_o$, then the accuracy is

$$\text{Accuracy} = \left.\frac{\text{TP}+\text{TN}}{\text{TP}+\text{FN}+\text{TN}+\text{FP}}\right|_{r_o}. \tag{21}$$

Note that TP + FN is the total number of falls and TN + FP is the total number of nonfalls.

*2) Determination of ROC Operating Threshold:* In practice, we would like to achieve certain sensitivity and specificity from acoustic-FADE and the operating threshold must be determined. We choose the operating threshold $r_o$ by equalizing the total cost from false positives with that from false negatives when the penalty of a false positive is considered the same as a false negative. The operating point is the position on the ROC curve whose tangent has a slope of $(1/\rho) - 1$, where $\rho$ is the proportion of falls [33]. For DAT1, we have equal number of falls and nonfalls, $\rho = 0.5$, and the slope is 45°. Once the operating point is found, the corresponding classifier threshold $r_o$ is fixed.

### C. Performance Evaluation Using DAT1 and DAT2

To evaluate the performance of acoustic-FADE using DAT1 (see Table II), we use 10-fold cross validation. The 120 falls and 120 nonfalls are decomposed into 10 folds. Each fold has 12 falls and 12 nonfalls with even distribution of fall and nonfall types. Given a detection threshold $r$, one fold is used as testing dataset and the others as training dataset at each validation. Then the sensitivity and specificity are obtained for each validation. Taking the average of sensitivity and specificity over the 10 validations gives $\text{Sensitivity}|_r$ and $\text{Specificity}|_r$.

To examine the performance using DAT2 (see Table III), we use the following method. For each subset in DAT2, it is used

to replace the falls in the testing fold in DAT1 at each validation without performing retraining. Hence, we are testing each subset of DAT2 over the 10 trained classifiers from DAT1.

The algorithm to perform fall detection in the testing folds of DAT1 and DAT2 follows the block diagram in Fig. 3.

## VI. Experimental Results

The primary interest of the experiment is to validate the theory that localization and BF are able to improve fall detection by suppressing the background noise and interferences. In addition, we want to confirm that the specificity can be increased by utilizing the estimated height of the sound source. Finally, we would like to investigate the impact on the fall detection performance when the acoustic environments and the floor materials have changed.

### *A. Performance Evaluation Using DAT1*

The fall detection performance was evaluated under two cases. The first is with background noise only; the second is with background noise and TV interference. For both cases, we evaluated the performance of the proposed acoustic-FADE that consists of a circular microphone array with localization, BF and estimated source height (BFH) for fall detection. To examine better effect of using BF and height information, we also produce the results of using BF only and using a single microphone without BF and source height (SGL) for comparison purpose.

*1) Case I: Performance Under Background Noise Only:* The signal is modeled as follows:

$$\boldsymbol{x}(n) = \boldsymbol{s}(n) + \alpha\boldsymbol{\Phi}(n). \tag{22}$$

$\boldsymbol{x}(n)$ is the $M \times 1$ received signal vector from the $M = 8$ microphones at time $n$, $\boldsymbol{s}(n)$ is the original received signal, and the component $\boldsymbol{\Phi}(n)$ is white Gaussian noise of unity power. The noise level is adjusted by changing the scalar $\alpha(\alpha \geq 0)$. The original signal $\boldsymbol{s}(n)$ is from the experimental configuration shown in Fig. 6.

We choose four noise levels of $\alpha = \{0, 0.1, 0.2, 0.4\}$, where $\alpha = 0$ is referred to "clean" since no white noise is added. The other three noise levels correspond to three SNR[2] values of 10, 4, $-2$ dB. The 10-fold cross-validation ROC curves for the three processing cases (BFH, BF, SGL) at "clean" level are shown in Fig. 7. The AUC and the sensitivity, specificity, and accuracy computed using (19)–(21) at the operating threshold $r_o$ for each processing case under the SNR levels are listed in Table IV. The threshold is found by using the method described in Section V-B2.

Fig. 7 shows that the BF reduces the 1-specificity by about 30% at 100% sensitivity, compared to using a single microphone (SGL). The use of estimated source height (BFH) reduces the 1-specificity further by about 70% at 100% sensitivity. Overall,

[2]The SNR level is obtained by averaging the SNRs over all the fall and nonfall files in DAT1. The SNR of one file is calculated by $\mathrm{SNR} = P_s/\alpha^2$ where $P_s = (1/MN)\sum_{m=0}^{M-1}\sum_{m=0}^{N-1} s_m^2(n)$, which is the average power of the received signal $s(n)$ over $M$ microphones. $N$ is the duration of one file in samples.

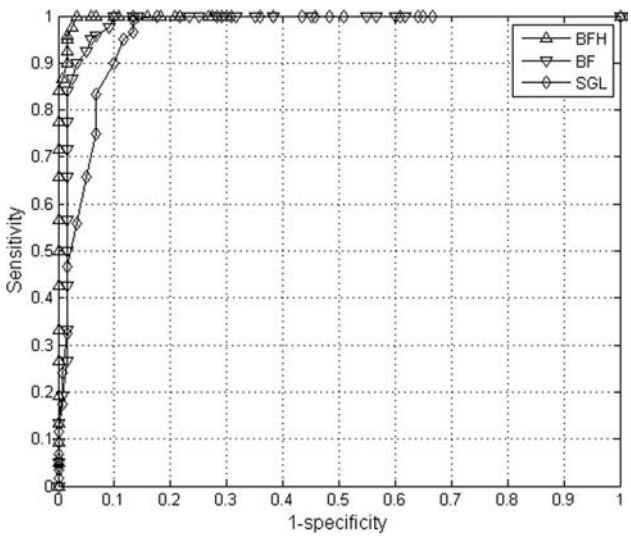


Fig. 7. Comparison of 10-fold cross-validation ROC curves in case I on DAT1 at "clean" condition.

TABLE IV
Ten-Fold Cross-Validation Results of Case I on DAT1 at Different SNR Levels (Sensitivity, Specificity, and Accuracy Values are in % at the Operating Thresholds)

| | | 'clean' | 10dB | 4dB | -2dB |
|---|---|---|---|---|---|
| BFH | AUC | 0.997 | 0.962 | 0.853 | 0.765 |
| | Sensitivity | 100 | 97 | 88 | 81 |
| | Specificity | 97 | 81 | 73 | 62 |
| | Accuracy | 98 | 89 | 81 | 72 |
| BF | AUC | 0.981 | 0.954 | 0.703 | 0.548 |
| | Sensitivity | 95 | 95 | 79 | 67 |
| | Specificity | 95 | 77 | 57 | 44 |
| | Accuracy | 95 | 86 | 68 | 55 |
| SGL | AUC | 0.961 | 0.892 | 0.653 | 0.498 |
| | Sensitivity | 95 | 91 | 64 | 60 |
| | Specificity | 88 | 75 | 60 | 42 |
| | Accuracy | 92 | 83 | 62 | 51 |

as shown in Table V for the "clean" condition, at the operating threshold $r_o$, the proposed acoustic-FADE achieves 100% sensitivity at 97% specificity, and the corresponding accuracy is 98%.

*2) Case II (Performance With Background White Noise and TV Interference):* The signal model is

$$\boldsymbol{x}(n) = \boldsymbol{s}(n) + \boldsymbol{y}(n) + \alpha\boldsymbol{\Phi}(n). \tag{23}$$

The terms $\boldsymbol{x}(n)$, $\boldsymbol{s}(n)$, and $\boldsymbol{\Phi}(n)$ are defined as in case I and $\boldsymbol{y}(n)$ denotes the TV interference. The TV interference is created by putting a TV sound source near the mattress with height of 0 m (we set the height as zero to challenge the acoustic-FADE because the effect of interference will not be reduced by using the estimated source height) in the room. The experimental configuration for this case is also given in Fig. 6, where the TV sound source is located at *x*–*y* coordinate of (7.5 m, 3.5 m). The

TABLE V
TEN-FOLD CROSS-VALIDATION RESULTS OF CASE II ON DAT1 AT DIFFERENT SNR LEVELS (SENSITIVITY, SPECIFICITY, AND ACCURACY VALUES ARE IN % AT THE OPERATING THRESHOLDS)

| | | 'clean' | 10dB | 4dB | -2dB |
|---|---|---|---|---|---|
| BFH | AUC | 0.989 | 0.932 | 0.802 | 0.713 |
| | Sensitivity | 93 | 91 | 76 | 71 |
| | Specificity | 98 | 86 | 69 | 66 |
| | Accuracy | 95 | 89 | 73 | 68 |
| BF | AUC | 0.956 | 0.903 | 0.651 | 0.506 |
| | Sensitivity | 93 | 84 | 72 | 52 |
| | Specificity | 92 | 85 | 57 | 46 |
| | Accuracy | 93 | 85 | 64 | 49 |
| SGL | AUC | 0.902 | 0.815 | 0.577 | 0.381 |
| | Sensitivity | 84 | 83 | 67 | 41 |
| | Specificity | 92 | 78 | 41 | 37 |
| | Accuracy | 88 | 81 | 54 | 39 |

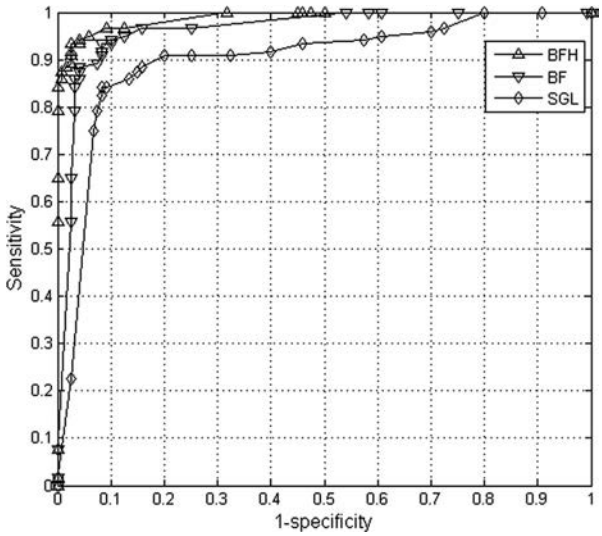


Fig. 8. Comparison of 10-fold cross-validation ROC curves in case II on DAT1 at "clean" condition.

level of the TV audio signal is constant. We set the signal-to-interference ratio[3] (SIR) of the falls and nonfalls in DAT1 to be 6 dB. Similar to case I, we show the 10-fold cross-validation ROC curves for the three processing cases at "clean" condition in Fig. 8. The AUC and the sensitivity, specificity, and accuracy at the operating threshold $r_o$ for each processing case under the SNR levels are listed in Table V. For example, the ROC curves in Fig. 8 show similar patterns to those in Fig. 7 except that the presence of the interference generally degrades the performance of the proposed acoustic-FADE for each processing case. However, at the operating threshold, acoustic-FADE can still reach to 93% sensitivity at 98% specificity, with the corresponding accuracy equal to 95%.

[3]The SIR is calculated in the similar way as calculating SNR except that $\alpha^2$ is replaced by $P_y$, which is the average power of the interference signal $\boldsymbol{y}(n)$.

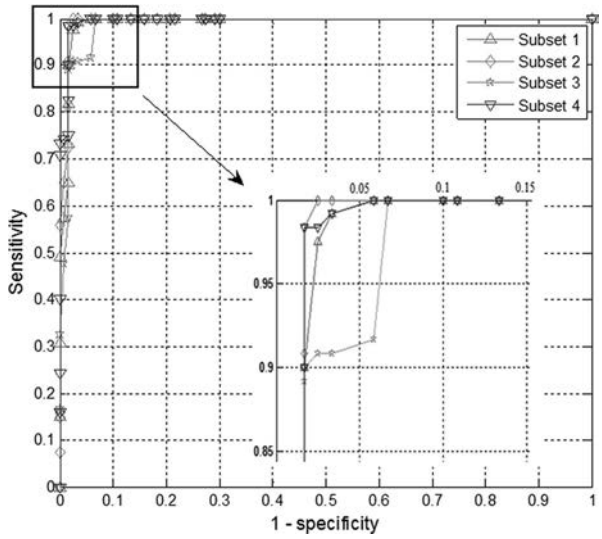


Fig. 9. Comparison of ROC curves using the proposed acoustic-FADE (BFH) among the four subsets of DAT2 at "clean" condition. The top-left-corner region is enlarged.

TABLE VI
VALIDATION RESULTS USING THE PROPOSED ACOUSTIC-FADE (BFH) ON DAT2 AT "CLEAN" CONDITION (SENSITIVITY, SPECIFICITY, AND ACCURACY VALUES ARE IN % AT THE OPERATING THRESHOLDS)

| | Subset 1 | Subset 2 | Subset 3 | Subset 4 | Average |
|---|---|---|---|---|---|
| AUC | 0.992 | 0.995 | 0.987 | 0.994 | 0.992 |
| Sensitivity | 99 | 100 | 91 | 99 | 97 |
| Specificity | 97 | 97 | 97 | 97 | 97 |
| Accuracy | 98 | 99 | 94 | 98 | 97 |

### B. Performance Evaluation Using DAT2

The ROC curves shown in Fig. 9 are generated using the proposed acoustic-FADE by the method described in Section V-C) for the four subsets at "clean" condition. Table VI gives the AUC and sensitivity, specificity, and accuracy at the detection threshold $r_o$ that is determined from the dataset DAT1.

## VII. DISCUSSION

By comparing BFH and BF with SGL in Fig. 7 and Table IV for case I, we observe that the fall detection performance using BF is always better than that of using a single microphone at all four SNR levels. Fig. 7 shows at "clean" condition that the 1-specificity is greatly reduced by about 70% using BFH compared to SGL when the sensitivity is at 100%.

Compared to linear arrays [18], circular arrays provide much better height estimation accuracy. The estimated height information of the source from localization is found to be very useful for improving the performance and it increases the specificity considerably. Although the detection performance becomes worse when the noise level increases, the improvement of using the proposed acoustic-FADE is more significant. These experimental results confirm our expectation that BF is able to enhance the signal and improve fall detection performance, and the height information helps to increase the specificity.

The results in Fig. 8 and Table V for case II are consistent with those for case I. At the same SNR level, the performance for case II is worse than that in case I because of the presence of TV interference. However, the improvement of using the proposed FADE remains very significant due to its ability to suppress the interference and enhance the signal before detection. Fig. 8 shows that at "clean" condition, at the operating threshold, the 1-specificity is greatly reduced by about 90% using BFH compared to SGL when the sensitivity is at 93%. Comparing the results in case I at the same sensitivity, the proposed acoustic-FADE provides more reduction in 1-specificity when directional interference is present in the environment. This is expected to be so because BF is known to have the advantage of reducing interference in addition to improving the SNR.

The results in Fig. 9 and Table VI show that the proposed acoustic-FADE performances on the four subsets in DAT2 are very similar except for subset 3, which has a little worse result than the others. This is possibly because subset 3 has carpet as the floor material and/or has different reverberation time. Another possibility is that the training uses DAT1 only. However, the performance reduction is not significant and it occurs at over 90% sensitivity, as shown in Fig. 9. Nevertheless, the results (AUC and accuracy) from DAT2 in Table VI are very close to those of BFH at "clean" condition from DAT1 in Table IV, indicating that the acoustic-FADE performance is not sensitive to room environments, and the difference in floor materials from mattress to carpet only results in marginal degradation in performance, although the classifiers were trained on mattress falls.

To gain some insight about the difference of MFCC features between fall sounds and nonfall sounds, we display the patterns of the MFCC features of DAT1 (see Table II). For the MFCC matrix [see (15)], we concatenate the six rows to form a column vector of length $6 \times 121 = 726$. The vector is normalized to have unity norm for display purpose. Repeating this procedure for all 240 files forms a feature matrix of size $726 \times 240$. Fig. 10 shows the grayscale image of the features, where file indices 1 to 120 correspond to fall files, indices 121 to 180 correspond to on-ground nonfalls, and 181 to 240 for above-ground nonfalls.

Fig. 10 indicates that there is strong consistency for the features from falls. In addition, the fall features distinguish most from those of nonfalls in the first and last MFCC features. It is interesting that the above-ground nonfalls have large feature values in the second MFCC.

As described in the Section I, we have advanced the previous work [19] by employing and integrating sound source localization, estimated source height, and BF for fall detection. The work in [19] only achieves 87% sensitivity at a specificity of nearly 90%. Under the same acoustic conditions (data are collected in the same environment at "clean" condition without interferences) as those in [19], the results of the proposed acoustic-FADE in this paper greatly increase the specificity to 97% with the sensitivity reaching 100%. In addition, the results of this paper show a significant improvement as compared to those in [16]–[18], which use a single microphone for fall detection.

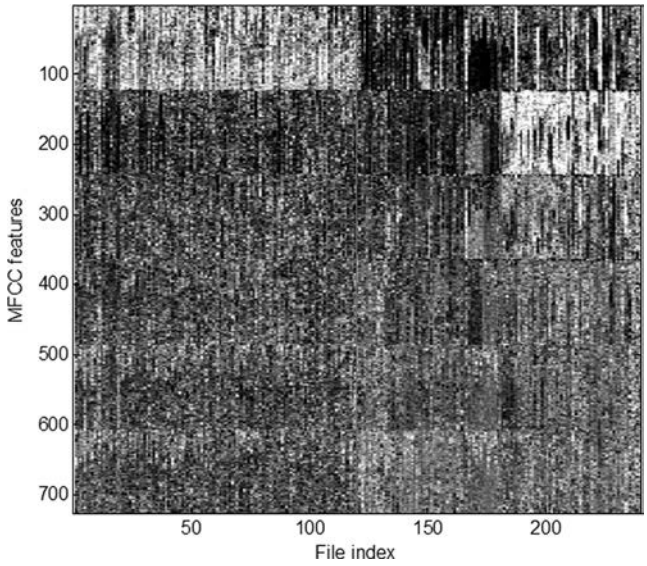


Fig. 10. Grayscale representation of the MFCC features of falls (1–120th files) and nonfalls (121–240th files) in Table II separated by a red solid line. For the nonfalls, the 121–180th files are on-ground nonfalls and the 181–240th files are above-ground nonfalls, and they are separated by a red dash line. The darker is the intensity and the lower is the coefficient value.

We would like to clarify that the current study in this paper is limited. First, the fall detection performance presented here has small sample size and uses simulated fall acoustic signals from stunt actors. Although the stunt actors have been well trained to fall like elderly people, there may still be some slight discrepancy with the actual falls from elderly. In order to increase the sample size and collect more realistic fall signals, we have deployed the proposed acoustic-FADE in the TigerPlace apartments of elderly to capture the falls. Additional study of the proposed approach and its performance will be conducted as our fall signal database becomes richer. Second, the false alarms are from simulated events. It is necessary to evaluate the proposed algorithm during actual daily living activities rather than simulated scenarios since they may not cover all potential sources of interference encountered during normal activities of elderly. The data we are currently collecting from the acoustic-FADE at the apartments of elderly will enable us to conduct this investigation.

## VIII. Conclusion

In this paper, we propose an acoustic-FADE and evaluate its performance. Acoustic-FADE uses a hardware that consists of a uniform circular microphone array that has eight omnidirectional microphones distributed along a circle of radius 25 cm. The processing of localization, height information, and BF are applied to locate the sound source, increase the specificity, and enhance the received signal. Feature extraction and classification are then performed for fall detection. Preliminary experimental results using the actual fall and nonfall data validate the promising performance of acoustic-FADE. In addition, the validation results from the external experiments confirm that the proposed

acoustic-FADE is not sensitive to the changing acoustic environments and floor materials.

Future work will focus on improving the localization and BF technique under more challenging situations such as multiple interferences and significant amount of reverberation. An improved classifier, which accounts for the temporal dynamic nature of the fall acoustic signal such as hidden Markov model, will also be investigated. Also, actual falls from elderly and larger dataset will be collected and analyzed to further evaluate and improve the performance of acoustic-FADE.

## ACKNOWLEDGMENT

The authors would like to thank the reviewers for providing very good comments to help us in improving the paper. The authors would also like to thank the staff in Center for Eldercare and Rehabilitation Technology, University of Missouri, for their efforts on the assistance in data collection.

## REFERENCES

[1] Center for Disease Control. (Sep. 2011). [Online]. Available: http://www.cdc.gov/HomeandRecreationalSafety/Falls/adultfalls.html

[2] D. A. Sterling, J. A. O'Connor, and J. Bonadies, "Geriatric falls: Injury severity is high and disproportionate to mechanism," *J. Trauma-Injury, Infection Crit. Care*, vol. 50, no. 1, pp. 116–119, 2001.

[3] C. G. Moran, R. T. Wenn, M. Sikand, and A. M. Taylor, "Early mortality after hip fracture: Is delay before surgery important," *J. Bone Joint Surg.*, vol. 87, no. 3, pp. 483–489, 2005.

[4] R. J. Gurley, N. Lum, M. Sande, B. Lo, and M. H. Katz, "Persons found in their homes helpless or dead," *N. Engl. J. Med.*, vol. 334, no. 26, pp. 1710–1716, 1996.

[5] N. Noury, A. Fleury, P. Rumeau, A. K. Bourke, G. O. Laighin, V. Rialle, and J. E. Lundy, "Fall detection—Principles and methods," in *Proc. 29th Annu. Int. IEEE Eng. Med. Biol. Soc. Conf.*, Lyon, France, Aug. 2007, pp. 1663–1666.

[6] A. K. Bourke, P. W. van de Ven, A. E. Chaya, G. M. OLaighin, and J. Nelson, "Testing of a long-term fall detection system incorporated into a custom vest for the elderly," in *Proc. 30th Annu. Int. IEEE Eng. Med. Biol. Soc. Conf.*, Aug. 2008, pp. 2844–2847.

[7] G. Wu and S. Xue, "Portable preimpact fall detector with inertial sensors," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 2, pp. 178–183, Apr. 2008.

[8] M. Prado-Velasco, M. G. del Rio-Cidoncha, and R. Ortiz-Marin, "The inescapable smart impact detection system (ISIS): An ubiquitous and personalized fall detector based on a distributed 'divide and conquer strategy'," in *Proc. 30th Annu. Int. IEEE Eng. Med. Biol. Soc. Conf.*, Aug. 2008, pp. 3332–3335.

[9] M. Alwan, P. J. Rajendran, S. Kell, D. Mack, S. Dalal, M. Wolfe, and R. Felder, "A smart and passive floor-vibration based fall detector for elderly," in *Proc. 2nd IEEE Int. Conf. Inf. Comm. Tech.*, Damascus, Syria, Apr. 2006, pp. 1003–1007.

[10] D. Litvak, Y. Zigel, and I. Gannot, "Fall detection of elderly through floor vibrations and sound," in *Proc. 30th Annu. Int. IEEE Eng. Med. Biol. Soc. Conf.*, Aug. 2008, pp. 4632–4635.

[11] Y. Zigel, D. Litvak, and I. Gannot, "A method for automatic fall detection of elderly people using floor vibrations and sound—Proof of concept on human mimicking doll falls," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 12, pp. 2858–2867, Dec. 2009.

[12] C. Rougier, J. Meunier, A. St-Arnaud, and J. Russeau, "Fall detection from human shape and motion history using video surveillance," in *Proc. 21st Int. Conf. Adv. Inform. Netw. Appl. Workshops*, 2007, pp. 875–880.

[13] D. Anderson, R. H. Luke, J. Keller, M. Skubic, M. Rantz, and M. Aud, "Linguistic summarization of activities from video for fall detection using voxel person and fuzzy logic," *Comput. Vision Image Understanding*, vol. 113, no. 1, pp. 80–89, Jan. 2009.

[14] A. Sixsmith, N. Johnson, and R. Whatmore, "Pyrolitic IR sensor arrays for fall detection in the older population," *J. Phys. IV France*, vol. 128, pp. 153–160, 2005.

[15] J. Hilbe, E. Schulc, B. Linder, and C. Them, "Development and alarm threshold evaluation of a side rail integrated sensor technology for the prevention of falls," *Int. J. Med. Inform.*, vol. 79, no. 3, pp. 173–180, 2010.

[16] M. Popescu and S. Coupland, "A fuzzy logic system for fall detection," in *Proc. AAAI Fall Symp.*, Washington, DC, Nov. 2008, pp. 78–83.

[17] M. Popescu and A. Mahnot, "Acoustic fall detection using one-class classifier," in *Proc. 31th Annu. Int. IEEE Eng. Med. Biol. Soc. Conf.*, Sep. 2009, pp. 3505–3508.

[18] M. Popescu, Y. Li, M. Skubic, and M. Rantz, "An acoustic fall detector system that uses sound height information to reduce the false alarm rate," in *Proc. 30th Annu. Int. IEEE Eng. Med. Biol. Soc. Conf.*, Aug. 2008, pp. 4628–4631.

[19] Y. Li, Z. L. Zeng, M. Popescu, and K. C. Ho, "Acoustic fall detection using a circular microphone array," in *Proc. 32th Annu. Int. IEEE Eng. Med. Biol. Soc. Conf.*, Buenos Aires, Argentina, Sep. 2010, pp. 2242–2245.

[20] J. M. Valin, F. Michaud, J. Rouat, and D. Letourneau, "Robust 3D localization and tracking of sound sources using beam forming and particle filtering," in *Proc. IEEE ICASSP*, May 2006, vol. 4, pp. 841–844.

[21] J. M. Valin, F. Michaud, J. Rouat, and D. Letourneau, "Robust sound source localization using microphone array on a mobile robot," in *Proc. Int. Conf. Intell. Robots Syst.*, 2003, pp. 1228–1233.

[22] C. Kwan, K. C. Ho, G. Mei, Y. Li, Z. Ren, R. Xu, Y. Zhang, D. Lao, M. Stevenson, V. Stanford, and C. Rochet, "Automated acoustic system to monitor and classify birds," *EURASIP J. Appl. Signal Process.*, vol. 2006, pp. 1–19, 2006.

[23] H. L. Van Trees, *Optimum Array Processing*. New York: Wiley, 2002.

[24] J. P. Dmochowski, J. Benesty, and S. Affes, "A generalized steered response power method for computationally viable source localization," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 8, pp. 2510–2526, Nov. 2007.

[25] Y. Cho, D. Yook, S. Chang, and H. Ki, "Sound source localization for robot auditory systems," *IEEE Trans. Consumer Electron.*, vol. 55, no. 3, pp. 1663–1668, Aug. 2009.

[26] Y. Li, M. Popescu, K. C. Ho, and D. P. Nabelek, "Improving acoustic fall recognition by adaptive signal windowing," in *Proc. 33th Annu. Int. IEEE Eng. Med. Biol. Soc. Conf.*, Boston, MA, Aug. 30–Sep. 3, 2011, pp. 7589–7592.

[27] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man., Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[28] J. S. Jang, *Audio Signal Processing and Recognition*. [Online]. Available: http://www.cs.nthu.edu.tw/∼jang

[29] D. O'Shaughnessy, *Speech Communication: Human and Machine*. Reading, MA: Addison-Wesley, 1987, pp. 210–214.

[30] M. Rantz, M. Aud, G. Alexander, B. Wakefield, M. Skubic, R. H. Luke, D. Anderson, and J. Keller, "Falls, technology, and stunt actors: New approaches to fall detection and fall risk assessment," *J. Nursing Care Quality*, vol. 23, no. 3, pp. 195–201, 2008.

[31] W. C. Sabine, *Collected Papers on Acoustics*. Los Altos, CA: Peninsula, 1993 (Trade Cloth ISBN 0-932146-60-0).

[32] Y. Li, "The fall detection algorithm using sound sensors," Master's thesis, University of Missouri, Columbia, MO, Jul. 2009.

[33] M. H. Zweig and G. Campbell, "Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine," *Clin. Chem.*, vol. 39, no. 8, pp. 561–577, 1993.

**Yun Li** (S'08) was born in China. He received the B.S. degree in electrical engineering from the Harbin Engineering University, China, Harbin, in 2007, and the M.S. degree in electrical and computer engineering from the University of Missouri, Columbia, MO, in 2009, where he is currently working toward the Ph.D. degree in electrical and computer engineering.

He is a Research Assistant in the group of Center for Eldercare and Rehabilitation Technology, University of Missouri. His research interests include array signal processing, detection and localization, pattern recognition, and computational intelligence.

**K. C. Ho** (M'91–SM'00–F'09) was born in Hong Kong. He received the B.Sc. degree (first class Hons.) in electronics and the Ph.D. degree in electronic engineering from the Chinese University of Hong Kong, Hong Kong, in 1988 and 1991, respectively.

From 1991 to 1994, he was a Research Associate in the Royal Military College of Canada. In 1995, he was a member of scientific staff at the Bell-Northern Research, Montreal, PQ, Canada. From September 1996 to August 1997, he was a faculty in the Department of Electrical Engineering, University of Saskatchewan, Saskatoon, SK, Canada. Since September 1997, he has been with the University of Missouri, Columbia, MO, where he is currently a Professor in the Department of Electrical and Computer Engineering. He is an inventor of five United States patents, three Canadian patents, two patents in Europe, and five patents in Asia on geolocation, mobile communications, and signal processing. His research interests include sensor array processing, source localization, subsurface object detection, wireless communications, and the development of efficient adaptive signal processing algorithms for various applications including echo cancellation.

Dr. Ho is the Vice-Chair of the Sensor Array and Multichannel Technical Committee in the IEEE Signal Processing Society. He is also the Rapporteur of ITU-T Q15/SG16: Voice Gateway Signal Processing Functions and Circuit Multiplication Equipment/Systems. He has been serving his second Associate Editor term of the IEEE TRANSACTIONS ON SIGNAL PROCESSING since January 2009. He is the Editor of the International Telecommunication Union (ITU-T) Standard Recommendations G.160: Voice Enhancement Devices and G.168: Digital Network Echo Cancellers. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2003 to 2006, and the *IEEE Signal Processing Letters* from 2004 to 2008. He received the Junior Faculty Research Award in 2003 and the Senior Faculty Research Award in 2009 from the College of Engineering, University of Missouri.

**Mihail Popescu** (SM'08) received the M.S. degree in medical physics, M.S. degree in electrical engineering, and the Ph.D. degree in computer science from the University of Missouri, Columbia, MO, in 1995, 1997, and 2003, respectively.

He is currently an Assistant Professor in the Department of Health Management and Informatics, University of Missouri. His research interests include eldercare technologies, fuzzy logic, and ontological pattern recognition.