32nd Annual International Conference of the IEEE EMBS
Buenos Aires, Argentina, August 31 - September 4, 2010

Acoustic Fall Detection Using a Circular Microphone Array

Yun Li, Zhiling Zeng, Mihail Popescu *Senior Member*, *IEEE*, K.C. Ho *Fellow, IEEE*

Abstract— **Falling is a common health problem for elderly. It is reported that more than one third of adults 65 and older fall each year in the United States. To address the problem, we are currently developing an acoustic fall detection system, FADE, which automatically detects a fall and reports it to the caregiver. In a previous version, FADE used a 3-microphone linear array to eliminate the false alarms produced by sounds produced well above the floor level. To improve the fall detection in noisy and reverberant environments, we replaced the linear array by an 8-microphone circular array that can provide a better 3-D estimation of the sound location. Preliminary experiments show that the sound location estimation performed by the circular array is reliable and robust to interference. We obtained encouraging classification results on a pilot dataset with 55 falls and 120 non-fall sounds.**

I. Introduction

More than one third of about 38 million adults of 65 and older fall each year in the United States [1]. A fall can cause serious health problems such as head injuries and hip fractures [1]. Moreover, older people who live alone have a further increase in early death risk due to the likely inability to ask for assistance after a fall occurs [2]. The annual risk for a person living alone of being found helpless or dead at home by paramedics is about 3.2% [3]. The delay in hospitalization can increase mortality risk in some conditions, such as hip fracture or traumatic brain injuries [3]. Other studies have shown that the longer the lie on the floor, the poorer is the outcome of the medical intervention [3-4]. To address the problem of medical intervention delay, it is imperative to detect the falls as soon as they occur such that immediate assistance 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 non-wearable. Wearable devices, like accelerometer-based ones, detect falls by measuring the applied acceleration along the vertical axis. The wearable devices are versatile and effective both indoor and outdoor but they are, in general, rejected by older, more frail people [5]. Among the non-wearable devices, we mentioned floor vibration sensors [6], video cameras [7,8], infrared cameras [9] and smart carpets. The floor vibration sensors [6] are inexpensive and privacy preserving but their performance is questionable in most of the US nursing homes due to the ground-level concrete floors covered by carpet, that vibrate little on impact. Video cameras, infrared cameras and smart carpets are promising technologies that are still trying to address challenges related to low light, field of view and privacy. Ideally, the goal of a fall detection system is to have as few false alarms as possible while detecting all falls. In order to achieve this goal, we believe that several different sensors have to be integrated in a smart network architecture. Consequently, the developing of different fall sensing modalities is a necessity for a successful sensor fusion approach.

In previous papers [10-12] we described an acoustic human fall detection system (FADE) based on a linear array of microphones. We investigated several fall detection algorithms such as fuzzy rule systems [10] and one-class classifiers [11]. However, both approaches had limited success in reducing the false alarms due to the environmental interference, in part, because neither of them took into account spatial information related to the sound source. In another version of FADE [12], we showed that using sound source height information can greatly reduce the false alarm rate. That is, sound sources located above a certain height (e.g. 1 m) above the floor, like talking or preparing food, are directly filtered out. Only sounds located at the floor level, such as falls or steps, are passed to the classification algorithm. This approach not only reduces the false alarm rate but it also increases the computation efficiency. However, the height estimation accuracy is influenced by the acoustic properties of the environment. For instance, noise and reverberation present a great challenge for sound localization. To better deal with environmental challenges in source localization, in this paper we propose to use of an 8-microphone array for person tracking and fall detection.

The sound localization technique has been applied to many applications such as videoconferencing [13], human-robot interaction [14] and bird monitoring. Multiple acoustic sensors are required in order to suppress the noise and improve the sound quality for recognition purpose [13]. Among the algorithms for sound localization, the time delay of arrival (TDOA) [12,14] and phase transform (PHAT) [13] are often used.

In this paper, we describe a robust microphone-array-based fall detection system (FADE) that is able to estimate the 3-D sound source location, in terms of not only the height but also the DOA (Direction of Arrival) and range. The specific sound localization algorithm used in this paper is called spectrum response power-PHAT (SRP-PHAT) [15-17]. For the sound classification task, we use mel-frequency

M. P. is with the Health Mngmt. & Inform. Department, University of Missouri, Columbia, MO 65211, USA (corresponding author, phone: 573-882-1266; e-mail: popescum@missouri.edu).

D.H. is with Elect. and Comp. Eng. Dept., University of Missouri, Columbia, MO 65211, USA (email: hod@missouri.edu).

Y.L. and Z.Z. are with the ECE Dept., University of Missouri.

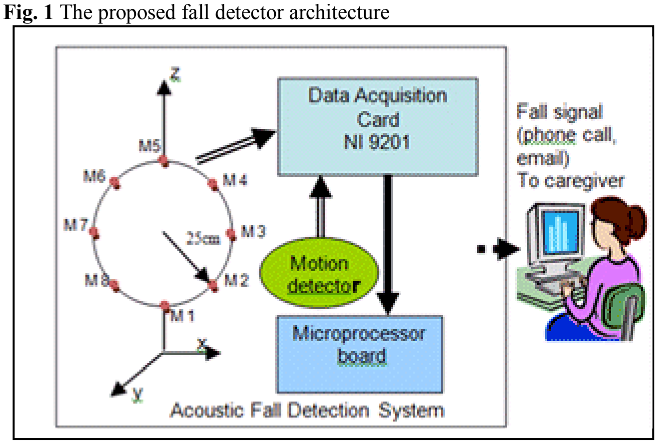
cepstral coefficients (MFCC) together with a nearest neighbor approach, as proposed in [12].

The structure of the paper is as follows: in section II we present the architecture of FADE, in section III we briefly describe the steered response power estimation using phase transform (SRP-PHAT) and the algorithm for sound localization, in section IV we show the results of our preliminary testing and in section V we conclude the paper.

II. System Architecture

The architecture of the FADE system is shown in Fig. 1.

Fig. 1 The proposed fall detector architecture



The fall detector consists of a circular array of 8 microphones. Each microphone has a mini amplifier and is mounted on a Cana Kit UK009 board. The microphones are installed on a plywood board in a circular pattern with a 25 cm radius. While a greater radius would improve the detection performance, it also limits the deployment options in an apartment setting. The array radius, R, was chosen based on the simulation results presented in Section IV.A. The microphone array board was hanged vertically on a wall in our lab at about 1.5 m above the floor, with the microphone side pointing away from the wall. The working hypothesis for FADE is that the person is alone in the apartment (room) hence only moving person has to be tracked. If motion is detected during a given interval (one minute) after a fall event is computed as likely, the caregiver alert is not issued. Instead, the event that provoked the alarm is cataloged as a false alarm and used to retrain the classifier(s). In order to preserve the privacy of the resident, the sound will be internally processed on a microprocessor board and only an external fall signal (email or pager) will be sent to the caregiver.

In this paper, we are mainly interested in investigating the sound localization and classification methods for fall detection. That is, we do not consider the motion detector and the communication processing with the caregiver. The sound is recorded using a National Instruments data acquisition card NI 9201 with 8-channel analog inputs. Currently, the localization and classification of the recorded sound are performed using Matlab (http://www.mathworks.com) installed on desktop computer instead of using a microprocessor board.

III. Algorithm Description

The main steps of the signal processing algorithm were: sound localization and fall recognition.

A. Sound localization using SRP-PHAT algorithm

The signal received by the i^{th} microphone, $m_i(t)$, can be modeled as:

$$m_i(t) = \alpha_i \cdot s(t - \tau_i) + n_i(t) \quad , \qquad (1)$$

where $s(t)$ is a single sound source at an unknown position, τ_i is the time delay between the source and the i^{th} microphone, α_i is the attenuation factor of $m_i(t)$ propagated along the i^{th} path and $n_i(t)$ is an uncorrelated white noise process. The purpose of SRP for a single source is to search for a steering 3-D spatial vector $\hat{\boldsymbol{p}}$ that maximizes the likelihood function $f(.)$, as described by:

$$\{\hat{\varphi}|f(\hat{\varphi}) = J\} = \underset{\hat{\boldsymbol{p}}}{\operatorname{argmax}} f(\hat{\boldsymbol{p}}) \quad , \qquad (2)$$

where $\hat{\varphi}$ is the estimated location and J is the global maximum response power occurred at location $\hat{\varphi}$. The main idea of the likelihood function is to align in time the microphone signals $m_i(t)$ based on the measured delays $\Delta\tau_i = \tau_i - \tau_0$ (τ_0 is the reference time delay estimated as the minimum among all the delays), then sum the aligned signals to compute the response power. Note that for each $\hat{\boldsymbol{p}}$ in space, the delays $\Delta\tau_i$ are uniquely determined and denoted by $\Delta\tau_i = \rho_i(\hat{\boldsymbol{p}})$. As proposed by [15-16], the SRP is seen as the cross-correlation power over pairs of microphones. By subtracting the total energy at all microphones from the squared delay-and-sum term and taking the time average of the residue gives the standard SRP likelihood function for an M-microphone array as

$$f(\hat{\boldsymbol{p}}) = \frac{1}{t_l - t_0}\left[\int_{t_0}^{t_l}\left[\left(\sum_{i=1}^{M}\alpha_i m_i(t + \rho_i(\hat{\boldsymbol{p}}))\right)^2 - \sum_{i=1}^{M} m_i^2(t + \rho_i(\hat{\boldsymbol{p}}))\right] dt\right] \qquad (3)$$

in which the environment attenuation factor   is modeled as the normalized reciprocal of the i^{th} propagation distance d_i, that is, $\alpha_i = \frac{1/d_i}{\sum 1/d_i}$. The time interval (t_0, t_l) is the processing window which covers all aligned signals. The PHAT version of the SRP likelihood function is obtained by multiplying the cross-correlation spectrum with a weighting factor Φ_{ij}. Thus, by taking Parseval's theorem of the first term in $f(\hat{\boldsymbol{p}})$, we can derive the SRP-PHAT likelihood expression as:

$$f_{PHAT}(\hat{\boldsymbol{p}}) = \frac{1}{t_l - t_0}\left[\frac{1}{2\pi}\sum_{i=1}^{M}\sum_{j=1}^{M}\alpha_i\alpha_j\int_{t_0}^{t_l}\Phi_{ij}M_i(\omega)M_j^*(\omega)\,e^{j\omega(\rho_i(\hat{\boldsymbol{p}})-\rho_j(\hat{\boldsymbol{p}}))}\,d\omega - \int_{t_0}^{t_l}\left(\sum_{i=1}^{M} m_i^2(t + \rho_i(\hat{\boldsymbol{p}}))\right)dt\right] \qquad (4)$$

where the frequency-dependent weighting factor Φ_{ij} is generally defined as the inverse magnitude of the cross-spectrum, that is:

$$\Phi_{ij} = 1/\left|M_i(\omega)M_j^*(\omega)\right| \tag{5}$$

B. Fall recognition

Before extracting the sound features, an energy discriminator is used to select the frames with energy larger than a threshold value, E_T. For each selected frame, we extracted the MFCC features. We computed N MFCC features (coefficients) but we ignored the first one in the recognition procedure, as proposed in [12]. We chose N=7 based on fall detection experiments described in section IV.C. The recognition was performed using the nearest neighbor (NN) procedure. The "fall" and "non-fall" training samples used in the NN procedure were recorded by the same person that performed the test session. A fall has to be detected in at least two consecutive windows in order to be reported as such. The fall confidence, C, is calculated as $C = \frac{1}{8}\sum_{i=1}^{8} \delta_i$ where $\delta_i = 1$ if a fall has been detected in channel i, and 0 else. A threshold , $\in[0,1]$, is then used to declare a "fall" if $C >$. A flowchart of the fall recognition algorithm used in this paper is presented in Fig. 2. Source localization was not used in the fall recognition experiments presented in Section IV.C, but will be part of the final system.

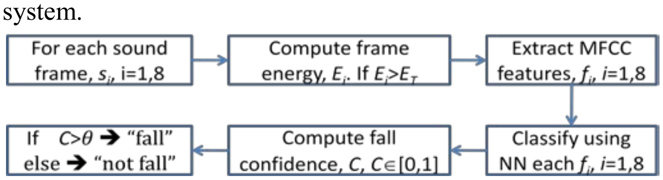


Fig. 2. The flowchart of fall detection algorithm used in this paper.

IV. RESULTS

In order to determine the behavior of the acoustic fall detector for various microphone array sizes and levels of background noise, we performed some initial experiments using a Matlab acoustic array simulation toolbox, ArrayToolbox [15, 18].

A. The choice of array radius, R

For determining the array radius, R, we radiated a 1-second duration of clean footstep signal at a specified location and simulated the received signals at all 8 microphone channels using the propagation attenuation and the time delay derived from the distances between the source and the microphone receivers. Omi-directional background noise (pre-recorded fan and computer noise) at 30 dB SNR was added to the microphone signals. The noise values at different microphone channels were independent of each other. The localization algorithm was then applied to the simulated signals to determine the source location. The room was modeled as a 7m×5m×2.5m parallelepiped. The sound source was placed at 5 locations in the room along an arc having a 3.5m radius from the array center (3.5m, 1.5m, 0m,) in the xy-plane (x, y, 0). At each location, we computed the root-mean-square error of DOA, range and height using 200 ensemble runs. The averaged DOA, height and range root-mean-square errors over the 5 locations are shown in Fig. 3. Fig. 3 indicates that height error is nearly 0 for R=25cm. Moreover, the DOA and range error curves are becoming closer to their minimum values for the same radius.

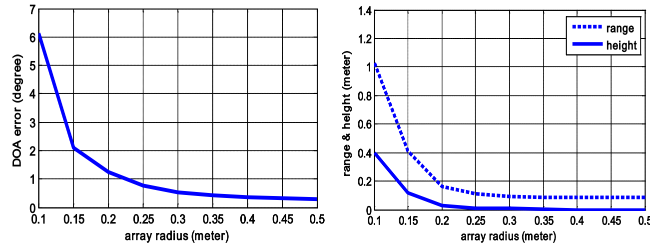


Fig. 3. Average DOA, range and height errors, at five locations.

While, theoretically, a larger array would result in better localization performance, a smaller array is often preferred in practice. Hence, we chose R=25 cm for the microphone array.

B. The influence of noise on the array tracking ability

For the tracking experiment, we used 12 possible locations of the sound source arranged in the (x,y) plane as in Fig. 4. We assumed that the source signal moved from one location to the next in a sequential manner and the microphone array tracked the location of the source.

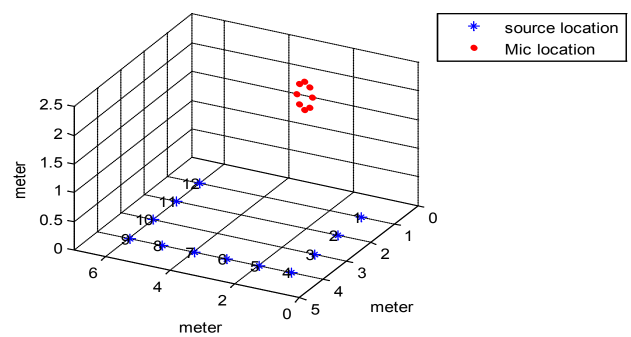


Fig. 4. Circular array (red dots) and 12 sound locations used in the footsteps tracking experiment

A foot-step signal was radiated at each location and collected together from the 12 locations as shown in Fig. 5.

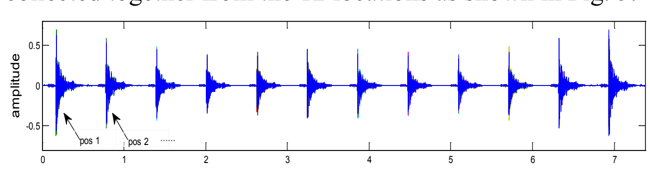


Fig. 5. Simulated signal used in the foot-steps tracking experiment

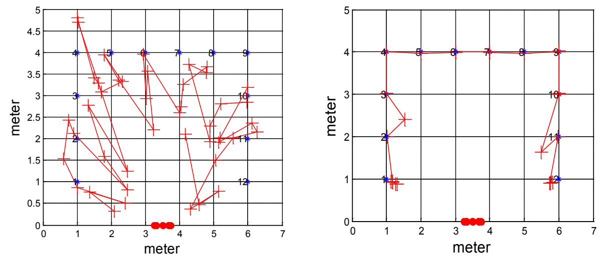
A sliding window of 1 sec. was used to segment out the microphone array data passed to the localization algorithm to estimate the source position. By varying the noise level added to the signal sequence, we obtained the localization results in height as shown in Table I. At a given SNR, the height error shown is the average of the absolute errors at the 12 locations.

TABLE I

AVERAGE HEIGHT ERROR VS. NOISE LEVEL

NOISE SNR (DB)	-10	0	10	20	30
HEIGHT ERROR (M)	0.47	0.26	0.14	0.09	0.07

A sample of the estimated accuracy projected in the x-y plane is illustrated in Fig. 6.



a) SNR=-10 DB b) SNR=30 dB

Fig. 6. Localization error at specific room locations for 2 SNR levels (o-original location, + =computed).

Table I shows that the maximum average height error was about 0.5 m when SNR=-10dB. The corresponding range error (see Fig. 6.a) is more severe, about 1-2 m. Depending on applications, the location error at -10dB SNR may still be acceptable. Otherwise, it is necessary to maintain a higher SNR to achieve better accuracy. It should be noted that as the SNR improves, the error reduces significantly.

C. *Pilot fall recognition experiment*

For the fall recognition experiment we recorded in our lab a training set of 25 falls (on a mat) and 50 false alarms (talking, key sounds, typing, phone ringing, soft and hard cover books dropped on the floor). The test set contains 30 falls and 120 false alarms. Each event had a duration of 0.5 s and was sampled at a frequency of 20,000Hz.

The sound features used were the MFCC coefficients. After training, the algorithm shown in Fig. 2 was applied to the test set. When varying the number of MFCC coefficients used in the classifier, a set of ROC curves are obtained as shown in Fig. 7.

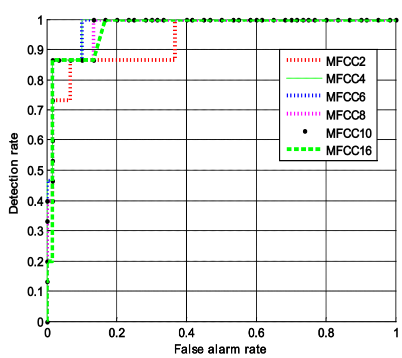


Fig. 7. The ROCs of the fall recognition experiment

The best result was observed when using 6 MFCC coefficients (i.e. 2th to 7th). The area under the curve in this case is 0.98. Only 6 false alarms were misclassified, while all 30 falls were detected.

V. CONCLUSIONS

In this paper we presented an audio sensor array designed for person tracking and fall detection. Key design features of the array such as the number of microphones, the preamp setting and radius, were determined using a Matlab array simulation toolbox. Due to the lack of space, we only presented the array radius choosing procedure, the other choices being done in a similar fashion.

While we showed that person tracking is a possibility, we have not implemented a complete audio tracking algorithm.

Preliminary fall recognition results are encouraging, although more realistic experiments such as fall in presence of noise (such as TV) are necessary.

ACKNOWLEDGEMENTS

This work has been supported in part by the NSF grant CNS-0931607.

REFERENCES

[1] Center for Disease Control (CDC), http://www.cdc.gov/ncipc/factsheets/adultfalls.htm.

[2] D. A. Sterling, J. A O'Connor., J. Bonadies, (2001). "Geriatric falls: injury severity is high and disproportionate to mechanism", Journal of Trauma–Injury, Infection and Critical Care (2001); 50(1):116–9.

[3] R. J. Gurley, N. Lum, M. Sande, B. Lo, M. H. Katz “Persons found in their homes helpless or dead”. N Engl J Med.; 334(26):1710-6, 1996.

[4] C. G. Moran, R.T. Wenn, M. Sikand, A.M. Taylor, Early mortality after hip fracture: is delay before surgery important", J. of Bone and Joint Surgery, pp. 483-9, 2005.

[5] N. Noury, A. Fleury, P. Rumeau, et al., "Fall detection-principles and methods", Proc. of the 29th IEEE EMBS, Lyon, France, Aug. 23-26,

[6] M. Alwan, P. J. Rajendran, S. Kell, D. Mack, S. Dalal, M. Wolfe, R.Felder, " A smart and passive floor-vibration based fall detector for elderly", 2nd IEEE Int. Conf. on Inf. & Comm. Tech.: from Theory to Applications - ICTTA’06, April 24 - 28, 2006, Damascus, Syria.

[7] C. Rougier, J. Meunier, A. St-Arnaud, J. Russeau, "Fall detection from human shape and motion history using video surveillance", 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07), 2007, pp. 875-880.

[8] D. Anderson, R.H. Luke, J. Keller, M. Skubic, M. Rantz , Aud M, "Linguistic summarization of activities from video for fall detection using voxel person and fuzzy logic. Computer Vision and Image Understanding, Vol. 113 (1), January 2009, pp. 80-89

[9] A. Sixsmith, N. Johnson, R. Whatmore, "Pyrolitic IR sensor arrays for fall detection in the older population", J. Phys. IV France, vol. 128.

[10] M. Popescu, S. Coupland, "A fuzzy logic system for fall detection", AAAI Fall Symposium, Washington DC, Nov 7-9 2008, pp. 78-83.

[11] M. Popescu, A. Mahnot ., "Acoustic fall detection using one-class classifier", 31th Int. IEEE EMBS Conf., Vancouver, BC, 2009

[12] M. Popescu, Y. Li, M. Skubic, M. Rantz, "An acoustic fall detector system that uses sound height information to reduce the false alarm rate", 30th Int. IEEE EMBS Conf., Vancouver, BC, Aug. 20-24, 2008, pp. 4628-4631.

[13] J. M. Valin, F. Michaud, J. Rouat, D. Letourneau, "Robust 3D localization and tracking of sound sources using beam forming and particle filtering", Proc. IEEE ICASP (2006), vol. 4, pp 841-844

[14] J. M. Valin, F. Michaud, J. Rouat, D. Letourneau, "Robust sound source localization using microphone array on a mobile robot", Proc. Int. Conference on Intelligent Robots and Syst. (2003), pp. 1228-1233.

[15] K. D. Donohue, J. Hannemann, H. G. Dietz, “Performance of phase transform for detecting sound sources with microphone arrays in reverberant and noisy environments”, Signal Processing, Vol. 87, no. 7, pp. 1677-1691, July 2007.

[16] B. Mungamuru, P. Aarabi, “Enhanced Sound Localization”, IEEE Trans. on systems, man, and cybernetics, Vol. 34. No.3, June 2004.

[17] H. Do, H. F. Silverman, “A fast microphone array SRP-PHAT source location implementation using coarse-to-fine region contraction (CFRC)”, IEEE Workshop on Application of Signal Processing to Audio and Acoustics, Oct. 21-24, 2007.

[18] Donohue, K.D., http://www.engr.uky.edu/~donohue/audio/Arrays/MAToolbox.htm