33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Improving Acoustic Fall Recognition by Adaptive Signal Windowing

Yun Li, Mihail Popescu, *senior IEEE member,* K. C. Ho, *IEEE fellow*, Daniel P. Nabelek

Abstract— Each year more than a third of elderly fall in the United States. To address this problem we are developing an acoustic fall detection system based on a microphone array. The main task of the acoustic system is to detect all the falls that occur in an indoor environment while producing as few as possible false alarms. One of the challenges of this task is to accurately locate where the fall signal comes from so that beamforming can be applied to improve the recognition of fall signals. In this paper we describe a simple fall signal location procedure that proved effective in preliminary testing.

I. Introduction

Falls are one of the most serious concerns of older adults. In spite of extensive fall prevention programs [1], about 13 million older adults fall each year in the United States, which results in an estimated hospitalization cost of about $9 billion [2]. About a third of the people who fall suffer severe injuries, such as fractures and head trauma [1], which can render them unable to rise or to ask for help. If the person lives alone, a fall might result in lying on the floor for a prolonged period of time which can cause hypothermia, dehydration, pressure sores, or rhabdomyolysis (destruction of skeletal muscle) [3]. By alerting the care giving personnel early using an unobtrusive automated fall detection system, we reduce the recovery time and save lives [4].

The number of falls in older adults is most likely higher than the one mentioned above, due to unreported falls [5]. An increasing frequency of uneventful, hence unreported, falls might be an indicator of physical decline, and of an imminent serious fall [5]. By recording the unreported falls using an unobtrusive automatic fall detection system, we can inform the caregiver about the necessity of a fall prevention intervention. As stated by Robinson et al. ([5], pg. 687) "detecting falls is likely to reduce the likelihood of future falls."

The fall detection methods found in the literature are based on two types of devices: wearable and non-wearable. The wearable devices tend to be easier to deploy, while the non-wearable ones tend to be less obtrusive [6]. The wearable devices are, in general, rejected by older people who perceive them as obtrusive [7]. Moreover, they are not effective in situations where they can't be worn, such as taking a shower or getting out of bed at night time.

Y.L. is with the ECE Dept., University of Missouri (corresponding author, e-mail: yl874@mail.mizzou.edu).

M.P. is with the Health Mngmt. & Inform. Department, University of Missouri, Columbia, MO 65211, USA (email: popescum@missouri.edu)

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

D.N is with the ECE Dept., University of Missouri.

Acoustic sensors have previously been used in habitat monitoring [8-13]. In previous papers [14-17] we tried various configurations and algorithms of an acoustic human fall detection system (FADE) based on a linear array of microphones. In [17] we investigated a circular microphone array configuration that proved to greatly increase the fall detection rate. However, noise and reverberation still present a great challenge for sound recognition. To better deal with environmental challenges in source recognition, we need to identify the source location more accuracy to increase the effectiveness of beamforming from the microphone array signals. In this paper we propose to use a simple adaptive windowing procedure that will detect the part of the signal for providing better localization accuracy. The fall classification results of using the proposed localization procedure in fall data measurements will be demonstrated.

II. System Architecture

The acoustic 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 25cm radius, which is determined based on the simulation results presented in [17]. The microphone array board was hanged vertically on a wall about 1.5m above the floor with the microphone sides pointing away from the wall. A data acquisition card is used for converting the analog signals into digital data. The microprocessor board works as a computer to process and analyze the digital data in real-time. The system architecture is shown in Fig. 1.

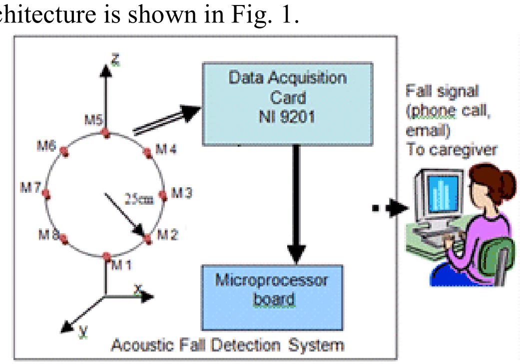

Fig. 1. The acoustic fall detector architecture

The working hypothesis for the fall detection is that the person is alone in the apartment room hence only moving person has to be tracked. A motion detector is used to detect if there is motion performed during a given interval (one minute) after a fall event is computed as likely. If there is motion after a fall is detected, then this event is cataloged as a false alarm and will not trigger the alarm to the caregiver. To preserve the privacy of the resident, the recorded sounds will be internally processed on the microprocessor and only an external fall signal (email or pager) will be sent to the caregiver.

III. STUDY METHODOLOGY

A. Problem description

The acoustic fall detection system consists of 3 processing steps: 1) sound source localization (SSL), 2) beamforming and 3) fall recognition. The purpose of beamforming is to increase the signal-to-noise ratio (SNR) of the microphone signals so as to improve the fall recognition rate in highly noisy environments. However, the effectiveness of SNR improvement given by the beamformer is strongly dependent on the estimation accuracy of the SSL. Poor estimate of the source location can degrade the enhancement on the microphone signals, and even obscure them. Challenging acoustic environment that has high noise level and large reverberation effect often degrades the estimation accuracy of the SSL. The purpose of this paper is to investigate how we can improve the estimation accuracy of the SSL by processing a better segment of the signal. The motivation comes from the fact that the fall signal is non-stationary and using different portions of the signal could yield different results. Now, the problem is which portion of the data should be processed for the SSL for achieving better estimation accuracy.

B. The definition of the problem

The window formulation is used for the selection of the data segment. Let us denote a data window by $w(n;\tau,L)$ where n is the time index, τ is the center position of the window and L is the window size. In this study, we choose the commonly used Hamming window function. In this case $w(n;\tau,L)$ is given by

$$w(n;\tau,L) = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi n}{L-1}\right), & \tau-\frac{L}{2} \le n \le \tau+\frac{L}{2} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where L is assumed to be even. Let us represent the signal from the i^{th} microphone by $x_i(n)$, $n = 0,1,\dots,N-1$, $i = 0,1,\dots,M-1$, where N is the number of data points in a microphone channel of a sound file and M is the number of microphones in the array. The windowed signal is the product of the microphone signal and the window function defined by (1), i.e.

$$\tilde{x}_i(n;\tau,L) = \begin{cases} x_i(n)\cdot w(n;\tau,L), & \tau-\frac{L}{2} \le n \le \tau+\frac{L}{2} \\ 0, \ 0 \le n < \tau-\frac{L}{2} \ and \ \tau+\frac{L}{2} < n \le N-1 \end{cases} \quad (2)$$

$\tilde{x}_i(n;\tau,L)$ is the windowed signal from the i^{th} microphone with the windowing parameters τ and L. Since a sound signal may occur at any time position in a sound file (assuming only one sound even is contained in each file), it is necessary to set a time reference for each file so that the window position can be measured with respect to the reference. The time reference, τ_s, is found by the proposed method described below.

Suppose we have acoustic signals from M microphone channels of a fall. Fig.2 shows a 500ms-long acoustic signal of a typical fall in channel 1.

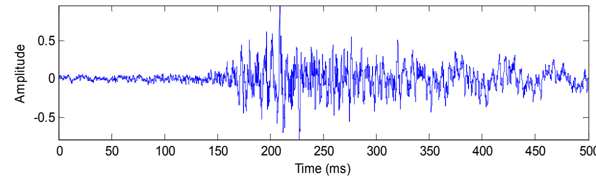

Fig.2. A typical acoustic fall waveform

The time reference is selected as the position where the largest amplitude occurs in the signal envelope of the strongest channel. The strongest channel is determined by comparing the signal energies among the channels and the waveform envelope is obtained by the Hilbert transform. The energy of the i^{th} microphone signal $x_i(n)$ is

$$E_i = \sum_{n=0}^{N-1} x_i(n)^2 \quad i = 0,1,\dots,M-1 . \quad (3)$$

The strategy for determining the strongest channel I is

$$I = arg \max_i E_i . \quad (4)$$

The envelope $e(n)$ of $x_I(n)$ is given by the magnitude of the analytic signal of $x_I(n)$, which is

$$e(n) = \sqrt{x_I(n)^2 + \hat{x}_I(n)^2} . \quad (5)$$

In (5), $\hat{x}_I(n)$ is the Hilbert transform of $x_I(n)$. The Hilbert transform is implemented by using a 32-point Parks-McClellan FIR filter [18]. Therefore, the time reference τ_s is chosen as

$$\tau_s = arg \max_n e(n) \quad (6)$$

In this study, we shall examine the performance of P window positions relative to the reference: $\tau_s^{(p)} = \tau_s + \Delta T \cdot f_o \cdot p$, $p = 0,\dots,P-1$. Note $\tau_s^{(0)}$ is the same as the time reference. ΔT is the time interval increment in second and f_o is the sampling frequency.

We also examine the performance of Q window sizes: $L^{(q)}, q = 0,1,\dots,Q-1$. Now we can describe the procedures for evaluating the fall recognition performance with respect to the window parameters $\tau_s^{(p)}$ and $L^{(q)}$ below:

1) For each sound event, use $\tilde{x}_i\left(n;\tau_s^{(p)},L^{(q)}\right)$, $i = 0,1,\dots,M-1$ to obtain the 3D position estimate of the sound source $\hat{\boldsymbol{u}}(\tau_s^{(p)},L^{(q)})$ by the SRP-PHAT algorithm [17].
2) Use $x_i(n)$, $i = 0,1,\dots,M-1$ to compute the delay-and-sum beamformer output $\hat{s}(n;\tau_s^{(p)},L^{(q)})$ based on $\hat{\boldsymbol{u}}(\tau_s^{(p)},L^{(q)})$.
3) Perform the Mel-frequency cepstral coefficients (MFCCs)-based nearest neighbor classifier [17] with varying thresholds on $\hat{s}(n;\tau_s^{(p)},L^{(q)})$ for all sound events to compute a set of pairs of detection rate $dr(\tau_s^{(p)},L^{(q)})$ and corresponding false alarm rate $fa(\tau_s^{(p)},L^{(q)})$.
4) Obtain the receiver operating characteristic (ROC) curve, ROC($\tau_s^{(p)},L^{(q)}$), by drawing the dr with respect to the fa.

The criteria to find the optimum window parameter is area under the ROC curve, denoted by AUROC$(\cdot)$. Hence the optimum window parameter $(\hat{p},\hat{q})$ is

$$(\hat{p},\hat{q}) = arg\max_{p,q} AUROC\left(\tau_s^{(p)}, L^{(q)}\right), \quad (7)$$

$$p = 0,1,\dots,P-1;\ q = 0,1,\dots,Q-1\ .$$

Due to the large number of combinations of window positions and sizes to optimize from, we assume that the performance is dominated by the window position. Hence,

$$\hat{p} = arg\max_{p} AUROC\left(\tau_s^{(p)}, L\right),\ p = 0,1,\dots,P-1\ . \quad (8)$$

where L is a nominal window size.

$$\hat{q} = arg\max_{q} AUROC\left(\tau_s^{(\hat{p})}, L^{(q)}\right),\ q = 0,1,\dots,Q-1\ . \quad (9)$$

We denote the optimized window size $L^{(\hat{q})}$ by $\hat{L}$.

C. Data description and experimental procedures

The dataset for this study consists of 12 falls and 12 non-falls. They were acquired by a trained stunt actor under the instructions of a geriatric nurse in our laboratory. The dataset contains many possible fall types such as a forward fall, a backward fall, a sideway fall, etc. and typical daily non-fall activities such as knocking, clapping, dropping an object, etc. Each type of fall or non-fall is measured in a 500ms-long data record at a sampling rate f_o =20 KHz using the microphone array.

The 8-folded cross-validation is performed on the dataset to generate the ROC curves for different window positions and window sizes. The fall recognition performance is examined through the ROC curves.

IV. Results

A. The effect of window position

We investigate the performance of fall recognition with different window positions based on the experimental results on the real data measurements. The nominal window size L is set to 100ms. The ROC curves for 5 different window positions: $\tau_s^{(p)}, p = 0,1,2,3,4$ with a time interval ΔT=50ms are plotted in Fig. 3.

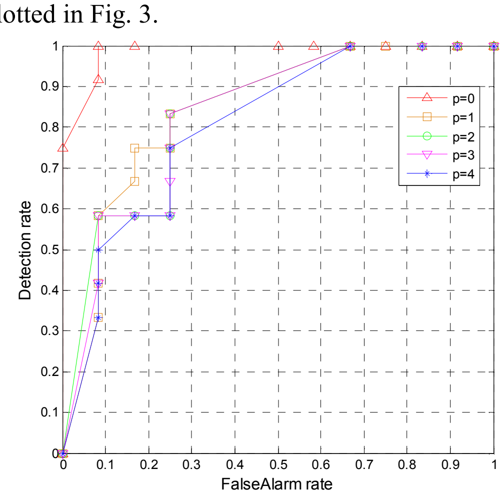

Fig. 3. ROC curves for different window positions

In Fig.3, it is interesting to observe that the performance is best when the window is located at the maximum signal envelope with its position at $\tau_s^{(0)}$. Using the data further away from the maximum gives worse and worse performance.

B. The effect of window size

Using the window position at $\tau_s^{(0)}$, the impact on the fall recognition with $Q = 8$ different window sizes $L^{(q)}$ of 2.5ms, 5ms, 10ms, 25ms, 50ms, 100ms, 200ms, 300ms are investigated and the ROC curves are shown in Fig. 4.

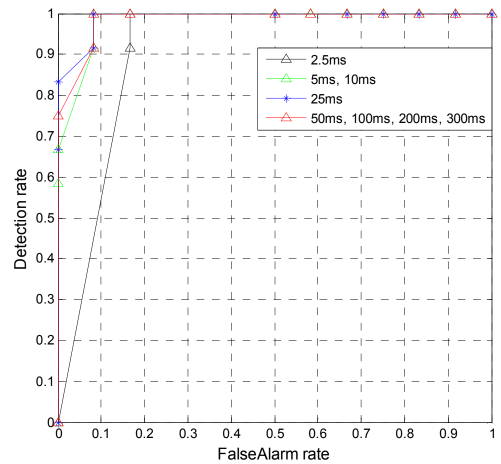

Fig.4. ROC curves with different window sizes

Fig.4 suggests that the window size smaller than 25 ms does not capture the entire fall signature. Using a window size of about 25ms provides best classification results. Further increasing the window size does not provide significant performance improvement.

C. Analysis and discussion

We shall provide some intuitive explanation of why positioning the data window at $\tau_s^{(0)}$ gives better results. In highly reverberating environments, the waveform of a sound event such as person falling, door knocking and object dropping has a particular pattern in which the very beginning part of the signal is 'cleaner' or louder than the following parts. This is because the energy contained in the very beginning portion of the signal is mostly generated by the direct propagation of the sound wave, which has the lowest attenuation on the amplitudes, compared to other signal components caused by the multiple-reflected propagations during the decay of the signal. This pattern leads us to believe that certain portion of the signal can provide better SSL accuracy. Fig. 5 shows the fall signal described in Fig. 2, separated into three portions indicated by three zones, Z1, Z2 and Z3. The zones are delimited by the two vertical dash lines at specified time points T1 and T2.

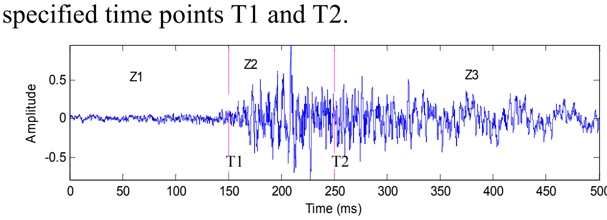

Fig.5. The fall waveform in Fig. 2 separated in 3 zones (Z1: noise-only zone; Z2: signal producing zone; Z3: potential reverberated zone)

Fig. 5 shows that the signal may encounter some reverberation effect since its amplitude remains largely different from zero even at 300ms away from T1 (fall time). Intuitively, the signal portion in zone 2 has higher amplitude and has the potential to achieve a better estimation accuracy of the SSL.

Fig. 6 shows the corresponding spectrogram of the fall waveform presented in Fig. 2, with the particular window we found as best ($\tau_s^{(0)} = 200$ ms $\times f_o = 4000$, $\hat{L}$ =25ms) indicated by the marked region in Fig. 6.

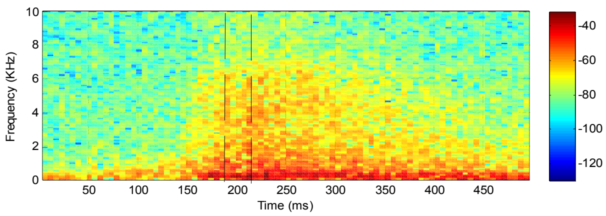

Fig. 6. Spectrogram of the fall waveform in Fig. 2.

It is commonly known that a SSL estimation of a signal richer in frequency content is more accurate than that with poorer frequency content. As indicated in Fig. 6, the frequency band of the signal in the particular window is much wider (0 ~ 6 KHz) which results in a higher SSL estimation accuracy. The frequency band in the noise-only part is very low (0~300 Hz). The frequency band in the reverberated part becomes narrow as the signal decays.

The window position $\tau_s^{(0)}$ in this particular signal is found using the procedures described in section III.B. This value is consistent with the time domain and the frequency domain analysis.

V. Conclusion

This paper presents a study for improving acoustic fall detection accuracy by choosing the data window position and duration. We find that positioning the window at the beginning of the signal where the amplitude is higher gives the best SSL results. The estimation accuracy of the SSL increases when the window size increases. However, beyond 25 ms window duration, no significant performance improvement is observed.

Acknowledgements

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

References

[1] Center for Disease Control (2006), http://www.cdc.gov/ncipc/factsheets/adultfalls.htm.

[2] Stevens J.A., Corso P.S., Finkelstein E.A., Miller T. R. (2006). "The costs of fatal and nonfatal falls among older adults", Injury Prevention, 12:290–5.

[3] Ratclife P. J., Ledingham J.G., Berman P., Wilcock G.K., Keenean J., "Rhabdomyolysis in elderly people after collapse", British Med. J., vol. 288, pp. 1877-8, 1984.

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

[5] Robinson Z.L., Powers C.M., MacLean C.H., "Quality indicators for the management and prevention of falls and mobility problems in vulnerable elders", Annals of Internal Medicine, vol. 135, No. 8, pp. 686-693, 2001.

[6] Noury N., Fleury A., Rumeau P., et al., "Fall detection-principles and methods", Proc. of the 29th IEEE EMBS, Lyon, France, Aug. 23-26, 2007.

[7] Demiris G., Rantz M., Aud M., Marek K., Skubic K., et al., "Older adults' attitude towards 'smart home' technologies", Medical Inf. and the Internet in Medicine, 29(2), pp 87-94, 2004.

[8] Castelli E., Vacher M., Istrate D., Besacier L., Serignat J.F., "Habitat telemonitoring system based on the sound surveillance", ICICTH (Int. Conference on Information Communication Technologies in Health), 11-13 July 2003, Samos Island, Greece.

[9] Vacher M., Istrate D., Besacier L., Castelli E., Serignat J-F., "Smart audio sensor for telemedicine" Smarts Objects Conference (SOC), 15-17 May, Grenoble, France, 2003.

[10] Istrate D., Castelli E., Vacher M., Besacier L., Serignat J-F., "Information extraction from sound for telemedicine" , IEEE Trans. on Tech. and Biomed., vol. 10, no. 2, pp 246-274, 2006.

[11] Schmandt C., Vallejo G., "'Listenin' to domestic environments from remote locations", Proc. of the 2003 Int. Conf. on Auditory Display, Boston, MA, 6-9 July 2003.

[12] Laydrus N. C., Ambikairajah E., Celler B., "Automated sound analysis system for home telemonitoring using shifted delta cepstral features", 15th International Conference on Digital Signal Processing, pp. 35-38, 2007.

[13] Y. Zigel, D. Litvak, 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, 56(12):2858-67, Dec, 2009.

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

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

[16] 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, pp. 4628-4631, Aug. 20-24, 2008.

[17] Y. Li, Z.L. Zeng, M. Popescu, K.C. Ho, "Acoustic Fall Detection Using a Circular Microphone Array," 32th Int. IEEE EMBS Conf., Buenos Aires, Argentina, Sept. 1-4, 2010.

[18] T.W. Parks and J.H. McClellan, "Chebyshev approximation for non-recursive digital filters with linear phase,'' IEEE Trans. Circuit Theory, vol. CT-19, no. 2, pp. 189–194, 1972.