A Fuzzy Logic System for Acoustic Fall Detection

Mihail Popescu[(1)], Simon Coupland[(2)]

(1) Health Management and Informatics Dept., University of Missouri, Columbia, Missouri, USA
(2) Centre for Computational Intelligence, De Monfort University, Leicester, UK

Abstract

More than one third (13 million) of adults aged 65 and above fall each year in the United States. Developing automated systems that detect falls is an important goal for those working in the field of eldercare technology. We developed an acoustic fall detection system (FADE) that automatically recognizes falls using purely acoustic (sound) information. The main challenge of building a fall detection system is providing testing data, since, no matter how realistic the falls for training the system are, they can not fully replicate the real elder falls. To address this challenge, we developed a knowledge based system rather than a data driven one. The system uses fuzzy rules based on knowledge of the specific frequency fingerprint of a fall and on the height of the origin of the sound. The rules were implemented in a Mamdani fuzzy rule system. We tested our system in a pilot study that consisted of a set of 23 falls performed by a stunt actor during six sessions of about 15 minutes each (1.3 hours in total). We compared the results of the fuzzy rule system to the results obtained using a K-nearest neighbor (KNN) approach with cepstral features. While the fuzzy rule system did not perform as well as the KNN one in the low false alarm region, it had the advantage that it reached 100% detection rate.

Introduction

More than one third (13 million) of adults aged 65 and above fall each year in the United States (CDC 2006). In spite of extensive fall prevention programs (CDC 2006), in 2006 there were about 400,000 fall related hospitalizations with an estimated direct cost of about $19 billion (Stevens et al. 2006). About 30% of people who fall suffer severe injuries such as fractures and head trauma (CDC 2006) that can render them unable to raise themselves or to ask for help. If the person lives alone in the apartment, a fall might result in a prolong period of laying on the floor which can cause hypothermia, dehydration, pressure sores or rhabdomyolosis (destruction of skeletal muscle) (Ratclife et al. 1984). Moreover, the delay in hospitalization can increase the mortality risk in some clinical conditions (Gurley et al. 1996). For example, a one day delay in hip fracture surgery may increase the 30 day mortality risk from 7.3% to 8.7% (Moran et al. 2005). Hence, it is imperative that the falls are detected and the necessary help is provided as soon as possible.

The fall detection methods found in the literature (Noury et al. 2007) 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. The wearable devices are in general rejected by older, more frail, people (Noury et al. 2006). Among the wearable devices we cite accelerometers, gyroscopes, mercury tilt switches and velocity sensors. Among the non-wearable fall detection devices we mention floor vibration sensors (Alwan et al. 2006), video cameras (Rougier et al. 2007; Anderson et al. 2008), infrared cameras (Sixsmith et al. 2005), and smart carpets. The floor vibration sensor (Alwan et al. 2006) was proposed to be deployed in a motion sensor network that helped reduce the false alarm rate if motion is sensed in a given time window after the fall signal (Dalal et al. 2005). In our experience, the floor vibration sensors do not perform well on the floors made of concrete covered by carpet that are typical for the nursing homes currently built in the United States. The use of video cameras is promising, although the computational requirements for processing represent a challenge. The infrared cameras and the smart carpet technologies are still under development.

Any fall detection system faces two challenges. The first challenge is represented by the false alarms which may lead to its rejection by the user (Noury et al. 2006). The second challenge is the lack of adequate training data which can lead to falls not being detected. An ideal fall detection system will have a demonstrable detection rate of 100%. In this context, a knowledge based system has two advantages over a data-driven system: it is less dependent on training data and it has a transparent decision process where the reasons for classifying or not classifying a sound as a fall can be clearly understood.

In (Popescu et al. 2008) we introduced a dedicated acoustic **fa**ll **de**tection system (FADE) based on an array of acoustic sensors. The system is inexpensive and built from off-the-shelf components. In that system, we used a data driven approach to fall detection based on cepstral features and K-nearest neighbor algorithm. In this paper, we are investigating a knowledge based approach to fall recognition by implementing a fuzzy rule system in the FADE system.

Acoustic sensors have been previously used in habitat monitoring (Castelli et al. 2003; Vacher et al. 2003; Istrate et al. 2006; Ladyrus et al. 2007; Schmandt and Vallejo 2003). In (Istrate et al. 2006) a set of acoustic sensors was used to differentiate between several sound classes such as breaking glass, screams, steps, door sound and human sound. A microphone was placed in each room of the apartment to identify the location of the sound. The acoustic sensor used in the ListenIn system (Schmandt and Vallejo 2003) was designed for activity monitoring (baby noise or loud noise). The alarm, together with the encrypted sound, was send to a mobile device held by a caregiver. Human falls were not included in the sound classes detected in any of the above acoustic systems.

In speech recognition, knowledge based systems were designed to address variation in environment (acoustic variability) and speaker (within-speaker and across speaker variablity) (Zue 1985). As in the case of fall detection, one can not train a speech recognition system for all possible speech conditions and accents. Many features in the speech knowledge systems were based on the link between the speech sound spectrogram (the distribution of the energy across frequency) and the speech phonetics. The most relevant phonetic features for speech (Alexin et al. 1997), the fundamental frequencies (called formants), can not be used for fall detection since the fall sound it is not produced by a specialized instrument such as the human vocal tract. Other possible sound features are (Zue 1985) the zero crossing rate and total energy in a given frequency band.

We found two previous attempts using fuzzy rule systems (FRS) in speech recognition. Awais and Rehman (2003) used the average energy in two different frequency bands as inputs of a FRS for classification of Arabic phonemes. Hsieh et al. (1997) used the value of the first formant and the zero crossing rate as inputs of a FRS designed to perform Chinese speech segmentation.

In the next section we describe the architecture of the acoustic fall detection system used in this paper.

FADE System Architecture

The architecture of the FADE system used in this paper is shown in Figure 1.

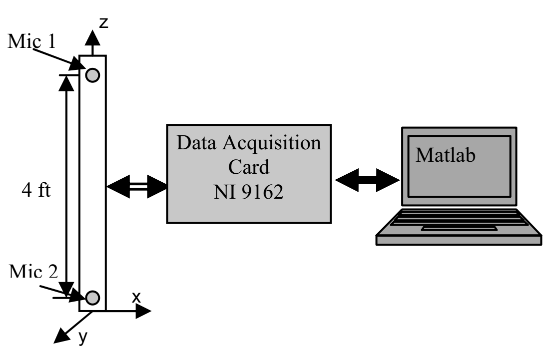

Figure 1. The architecture of the FADE acoustic fall detection system.

The fall detector consists of a linear array of electret condenser acoustic sensors (two shown: Mic 1 and Mic 2) mounted on pre-amplifier boards Cana Kit CK495 (about $20 each). The acoustic sensor array was mounted vertically (along z-axis) in order to be able to capture information about the height of the origin of the sound. The working hypothesis for FADE is that the person is alone in the apartment. In order to preserve the privacy of the patient, 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 work, we were mainly interested in the nature of the algorithm employed to detect the falls. For this reason, we only used a laptop computer to perform the signal processing instead of a microprocessor board.

The sound was recorded on a laptop computer using a National Instruments data acquisition card NI 9162 with 8 differential analog inputs. The recorded sound was later processed using Matlab (http://www.mathworks.com). In the experiments presented here we used 2 sound sensors (Mic 1 and Mic 2) mounted vertically, 4 feet apart.

Data Set Used in the Experiment

The falls used in this experiment were performed by a stunt actor (Rantz et al. 2008). The actor was instructed by our nursing collaborators to fall as an elderly person would fall. There were 5 types of falls: forward (Figure 2.a), backward, toward-left (Figure 2.b), toward-right and fall from a chair.

a. Forward fall b. Left-side fall

Figure 2. Stunt actor performing falls

A typical fall session was about 10-15 minutes long and it contained 3-5 falls one of each type mentioned above. A nurse directed the actor during the fall session, instructing her when and how to fall and when to get up. Additionally, various other sounds such as moving chairs, table knocks, feet stomping were produced by the actor and by the team members. We recorded 6 fall sessions with a total of 23 falls and a total recording time of 1.3 hours. A special 20 minute long session with 17 falls and noises was recorded and used for training. From this training session, 17 fall files and 30 false alarm files (1 second long) were extracted and used in the feature defining process and in the KNN algorithm. The sound was sampled at f_s=1000 Hz. During the feature selection process it was obvious that f_s=1000 Hz was not sufficient to capture the subtle sounds involved in falls. In the future, we intend to increase the sampling frequency to 10 kHz.

Algorithm description

In this section we will first present the sound features used to build the fuzzy rule system, then we describe the rules and, in the end, we briefly describe the data driven procedure (Popescu et al. 2008) that we used for comparison.

Sound feature selection

Energy in the 0-200 Frequency Band

Based on laboratory testing and on inspecting the available fall data we concluded that the majority of the fall energy is concentrated in the frequency interval 0-200 Hz. This can be seen in Figure 3. While the non-falls have a somewhat even energy distribution across all frequencies, the falls have the energy concentrated in the 0-200 Hz range.

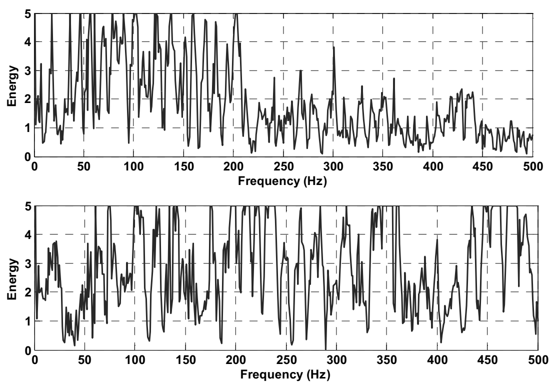

Figure 3. The typical fall frequency signature (above) and the typical non-fall frequency signature (below).

However, note that this energy concentration seems to be characteristic not only to a body fall but also to any object dropping on the floor. In the fuzzy rule system we used as an input the energy density in the 300-500 Hz band calculated as:

$$E_{d,300-500} = \frac{\sum_{i=301}^{500} S_t(i)}{200} \qquad (1)$$

where S_t ={1 if S>1;S if S<1}, S=abs(fft(s(w))), s is the fall signal, and w is a window (of size 1000 in our case). The signal will be classified as a fall if $E_{d,300-500}$ is low.

Zero Crossing Rate (ZCR)

The ZCR was calculated as the number of sign changes in a given widow divided by the length of the window. Although ZCR was mentioned in the literature (Zue 1985) as a reasonable feature to use in speech recognition application, we did not find it very useful for fall detection. We plotted in Figure 4 the energy density and ZCR for the 17 falls and 30 non-falls available in our training data. While the non-falls seem to have a mostly low ZCR, the falls do not exhibit any trend. However, experiments on our test data did not show ZCR to be a useful feature.

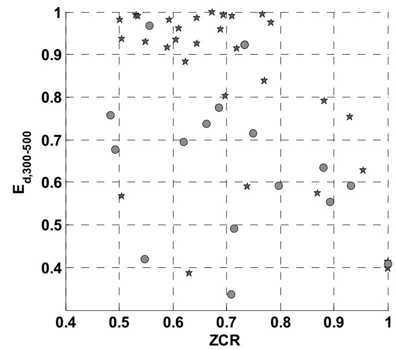

Figure 4. The 17 falls (circles) and 30 non-falls (stars) plotted in the ZCR-$E_{d,300-500}$ space.

The Height of the Sound

The height of the sound is computed using the correlation between the sound recorded by the two sound sensors. The signal correlation was computed using the whitened spectrum cross-correlation (rather than common time domain cross-correlation) (Valin et al. 2003), that is:

$$R_{12}(t) = \sum_{n=0}^{N-1} \frac{S_1(n)S_2(n)^*}{|S_1(n)||S_2(n)|} e^{i2\pi nt/N}, \qquad (2)$$

where $S_i(n)$ is the Fourier transform of the signal $s_i(n)$ received by the i^{th}, $i\in\{1,2\}$, sound sensor and $t\in[-N,N]$. Then, we computed the delay, δ_{12}, between the signals received by two sensors as:

$$\delta_{12} = \arg\max_{t\in[-N,N]}\{R_{12}(t)\}. \qquad (3)$$

To make the search more efficient the maximum of R_{12} was searched in the $t\in$ [-10, 10] interval. The signal in a window will be classified as a fall if the delay, δ_{12}, is low (that is, the signal reached sensor 1 faster than sensor 2, which means that the sound is coming from somewhere lower than 2ft). We used the delay as the second input to our FRS.

The fuzzy rule system architecture

The Mamdani fuzzy rule system used in this paper had the following four rules:

IF $E_{d,300-500}$ is LOW_E AND δ_{12} is LOW_δ THEN “fall”
IF $E_{d,300-500}$ is $HIGH_E$ AND δ_{12} is LOW_δ THEN “no-fall”
IF $E_{d,300-500}$ is LOW_E AND δ_{12} is $HIGH_\delta$ THEN “no-fall”
IF $E_{d,300-500}$ is $HIGH_E$ AND δ_{12} is $HIGH_\delta$ THEN “no-fall”

The above membership functions (LOW and HIGH) are shown in Figure 5.

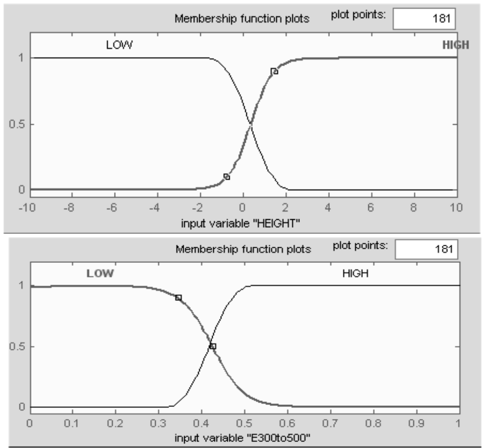

Figure 5. The memberships for the two input variables (δ_{12} or HEIGHT – above and $E_{d,300\text{-}500}$ – below) for the FSR used in paper.

The Cepstral Features Approach

We used the mel frequency cepstral coefficients (mffc) as features. The number of coefficients (features) used was C=7. The features were extracted using the Matlab function, mfcc, from (Slaney 2008). To make the system less dependent on the distance to the sound source, we did not use the first cepstral coefficient (proportional to the signal average) in the recognition procedure. The recognition was performed using the K-nearest neighbor (KNN) procedure with K=3. The "fall" and "no fall" training samples used in KNN were extracted from a fall session recorded by the same stunt actor but different from the six test sessions (see next section). A window was evaluated only if its energy E_w was greater than a threshold E_{THR} (=0.2 in our case). The energy is calculated as:

$$E_w(k) = \sum_{n=Nk-(N-1)/2}^{n=Nk+(N-1)/2} s^2(n) \qquad (4)$$

where N is the number of samples in the window. The considered window was 1 second of speech signal (N=1000 samples for a sampling frequency f_s=1000 Hz) with a 50% overlap between consecutive windows. A summary of for both algorithms ("CEPSTRAL" and "FRS") is given in algorithm 1. The low sampling frequency chosen (f_s=1000 Hz) was, in fact, due to an erroneous setting of the data acquisition software. The intended value was 10KHz. Preliminary results with a 20 KHz sampling frequency and using two new energy ratio subband features, ERSB (Liu et al. 1998), in the subbands 0-300 Hz and 300 Hz -2000 Hz, are significantly better.

Results

To describe the performance of the algorithm for all the fall sessions we used ROC curves. For the cepstral approach we varied the lag threshold L_{THR}={-10,…,10} and we obtained pairs of {detection rate, false alarm rate}. The detection rate for each threshold was computed as (# of falls detected)/23 and the false alarm rate as (#false alarm/minute). To get a ROC curve for the FRS we varied a threshold value in the interval [0,1], so at a given point say 0.7 all outputs above 0.7 were classified as a fall and all below 0.7 were classified as not being a fall. The two ROCs are shown in Figure 6. From Figure 6 we see that the cepstral approach is slightly better than the FRS in the low false alarm area.

```
FOR each window (k)
    -Compute energy E(k) given by (4)
    -IF E(k)> E_THR
        --Compute δ12 between channel 1 and 2 using (2)&(3)
       IF "CEPSTRAL"
        --IF δ12 <L_THR(sound height is 'Low')
                Extract mfcc features from signal in channel 1
                Do KNN to detect a fall: out(k)=1 if "fall", =0 else.
        --ELSE (sound height is 'High')
            "no fall", out(k)=0
        --END
       ELSE "FRS"
         -- compute the energy density E_d,300-500
         --out(k)=FRS(E_d,300-500 , δ12 ) ;
        END
    -ELSE (sound energy is low)
        "no fall", out(k)=0
    -END
END FOR
```

Algorithm 1. Overview of Fall Detection Algorithms.

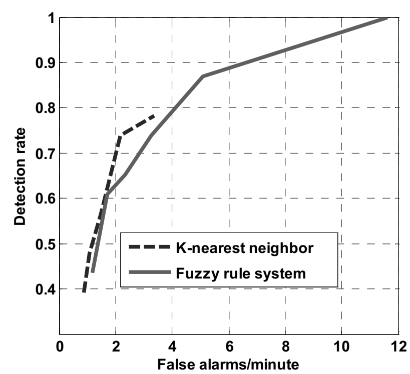

Figure 6. Comparison between the KNN-cepstral and the FRS algorithms.

The problem of the cepstral approach is that it only reaches 80% detection rate which comes from the insufficient training samples used in our experiment (which will always be the case, as we mentioned in the Section 1). On the other hand, the FRS approach achieves a 100% detection rate, albeit at a high false alarm rate. We mention that in this paper we were not concerned with reducing the false alarm rate but only with comparing the two approaches. The false alarm rate can be reduced once we can make sure we detect all the falls.

Conclusions

We presented a prototype of an acoustic fall detection system that uses a fuzzy rule system to detect the falls. The major problem with such systems is the lack of adequate training data to train automated systems on. This lack of training data led us to investigate several acoustic features that could be used in a knowledge based system. We identified the energy density in the 300-500 Hz band and the height of the origin of the sound as promising features. However, more features have to be identified to reduce the currently high false alarm rate.

In future work we intend to fuse the data from the acoustic sensor with data from a motion detector with a view to further reducing the false alarm rate. Such a system would only trigger an alarm should a fall be detected acoustically and little motion be sensed after the event. A motion detector could also be used to help the system learn new noises online by maintaining a false alarm library.

References

Alexin Z., Csirik J, Gyimothy T., Jelasity M., Toth L. 1997. Learning Phonetic Rules in a Speech Recognition System. Lecture Notes in Computer Science; Vol. 1297, Proceedings of the 7th International Workshop on Inductive Logic Programming, pp. 37 – 44.

Alwan M., Rajendran P. J., Kell S., Mack D., Dalal S., Wolfe M., Felder R., 2006. A smart and passive floor-vibration based fall detector for elderly. The 2nd IEEE International Conference on Information & Communication Technologies: from Theory to Applications - ICTTA'06, April 24 - 28, Damascus, Syria.

Anderson D., Luke R.H., Keller J., Skubic M., Rantz M., Aud M, 2008. Linguistic summarization of activities from video for fall detection using voxel person and fuzzy logic. Computer Vision and Image Understanding. (In Review)

Awais, M.M., Habib-ur-Rehman, 2003. Recognition of Arabic phonemes using fuzzy rule base system. 7th International Multi Topic Conference, INMIC 2003, pp. 367- 370.

Castelli E., Vacher M., Istrate D., Besacier L., Serignat J.F., 2003. Habitat telemonitoring system based on the sound surveillance. ICICTH (Int. Conf. on Inf. Communication Tech. in Health), Samos Island, Greece.

Center for Disease Control (CDC) 2006. http://www.cdc.gov/ncipc/factsheets/adultfalls.htm, accessed April 2008.

Dalal S., Alwan M., Seifrafi R., Kell S., Brown D., 2005. A Rule-Based Approach to the Analysis of Elders' Activity Data: Detection of Health and Possible Emergency Conditions. AAAI Fall 2005 Symposium (EMBC).

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

Hsieh C-T, Lai, E., Wang Y-C, 1997. Distributed fuzzy rules for preprocessing of speech segmentation with genetic algorithm. Proceedings of the Sixth IEEE International Conference on Fuzzy Systems, Vol. 1, pp. 427-431.

Istrate D., Castelli E., Vacher M., Besacier L., Serignat JF, 2006. Information extraction from sound for medical telemonitoring. *IEEE Trans. on Information Tech. in Biomedicine*, vol. 10, no.2.

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

Liu Z., Wang Y., Chen T., 1998, "Audio Feature Extraction and Analysis for SceneSegmentation and Classification", *Journal of VLSI Signal Processing*, no. 20, pp. 61–79.

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

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

Popescu M., Li Y., Skubic M., Rantz M., 2008. An acoustic fall detector system that uses sound height information to reduce the false alarm rate. IEEE EMBC08, Vancouver, British Columbia, Canada, August 20-24, 2008, pp 4628-4631.

Proakis J.G, Manolakis D.G, 2007. Digital signal processing. 4th edition, 2007.

Rantz M., Aud M., Alexander G., Wakefield B., Skubic M., Luke R.H., Anderson D., Keller J.K., 2008. Falls, Technology, and Stunt Actors: new Approaches to Fall Detection and Fall Risk Assessment. *Journal of Nursing Care Quality*, in press.

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

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

Schmandt C., Vallejo G., 2003. Listenin' to domestic environments from remote locations. Proc. of the 2003 Int. Conf. on Auditory Display, Boston, MA.

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

Slaney M., "Auditory toolbox 2.0", http://www.slaney.org/malcolm, accessed April 2008.

Stevens J. A., Corso PS, Finkelstein E. A., Miller T. R. 2006. The costs of fatal and nonfatal falls among older adults, *Injury Prevention*, 12:290–5.

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

Valin J.M., Michaud F., Rouat J., Letourneau D., 2003. Robust sound source localization using microphone array on a mobile robot. Proceedings International Conference on Intelligent Robots and Systems, pp. 1228-1233.

Zue V.W. 1985. The use of speech knowledge in automatic speech recognition. Proceedings of the IEEE, vol. 73, no. 11, pp. 1602-1615.