ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Enhanced Voice-Activity Home Care System

Perla Krishnakanth
Department of Embedded Systems,
Nova College of Engineering and Technology,
Hyderabad, Telangana – 501512, India.

Abstract:
This work proposes a voice-activity home care system which can construct a life log associated with voices at home. Accordingly, the techniques of sound-pressure-level calculation, abnormal sound detection, noise reduction, text-independent speaker recognition and keyword spotting are developed. In abnormal sound detection and speaker recognition, we adopt the two-stage recognition processes of Gaussian Mixture Model (GMM) for sound rejection, and Support Vector Machine (SVM) for sound classification. The experimental results reveal that the proposed abnormal sound detection, speaker recognition, and word spotting can reach accuracy rates above 82%, 90%, and 87%, respectively. Based on the recognized abnormal sounds or special words, an emergent event can be identified for home care where a speaker is known as well. Finally, the abovementioned recognition results versus time scales can fairly build a daily life log for home care.

INTRODUCTION:

Due to the increase of senior population, home care is a critical topic in recent years. With increasing age, the majority of older people take most time at home, so that home care and accident prevention grow into an area of concern to everyone. Accordingly, how does a family caregiver remotely and effectively monitor the elder has become an important research direction. In the prior art, many researchers have proposed various voice processing techniques to realize home care. For example, the event model was established to determine whether there was an emergency occurred [1]. Unusual sounds like cough, groan, wheeze, and cry for help were detected to understand the health condition of a subject [2].

Some special words of an utterance like “help” were recognized with location perception to provide necessary assistance [3]. In addition to the context of voices for help, abnormal sounds, such as screaming and glass broken, were discriminated from normal sounds where Gaussian Mixture Model (GMM) and Support Vector Machine (SVM) were commonly used for recognitions [4]-[6]. Thanks to the advance of technologies, IP cameras are usually used to capture life activities of elderly persons. The recorded video can be further processed to identify when a special event occurs. However, a camera has a limited angle of view and cannot be deployed at all areas of home, like bathroom. Hence, IP microphones can aid cameras to overcome these shortcomings [7]. Besides, video-based surveillance will make users' lives no privacy. This work explores how to build a daily life log from the recorded sounds for home care. Accordingly, the identification and context awareness of speaker's and environmental voices at home are developed to construct a life log.

PROPOSED HOME CARE SYSTEM:

This work employs multiple IP microphones to obtain voice data. First, the silence detection is performed by using the technique of energy ratio to avoid unnecessary computing where Sound Pressure Level (SPL) in a decibel unit is calculated at a voice frame of 256 points (32ms). The SPLs computed and recorded in a whole day present energy variations of environmental sounds. This parameter can reveal activity time of family members. Restated, sounds in day time and midnight tend to have high and low SPLs. If the situation is not consistent with the previous one in a daily log, the unusual event may occur. Once a sound is acquired, it is further classified into a speech or non-speech sound.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Second, our system is to identify whether there exists a special sound of siren, scream, sob, crash, glass broken, or crying. Third, speaker(s) are distinguished and their speech is translated to texts via Google web speech API. Meanwhile, the far-field recorded sounds go through noise cancellation to raise speech clarity. Additionally, special words like asking for help are recognized. Fourth, once an emergency event comes out, the proposed system will issue a call to a family member or hospital. Figure 1 shows the block diagram of the proposed voice-activity home care system.

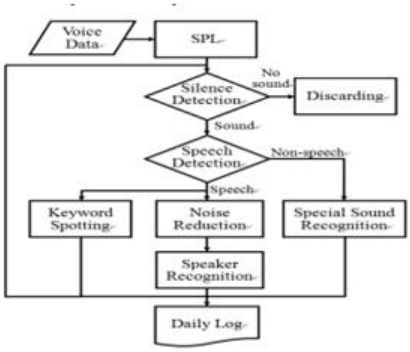

Figure 1. Proposed voice-activity home care system

A. Special Sound Recognition:

The recognition of special sounds adopts a two-stage process in which the first stage is the 16 mixture GMM to reject normal sounds, and the second stage is the linear-kernel SVM to identify six anomalies. Particularly, the feature selection scheme of Sequential Floating Backward Selection (SFBS) is employed to choose six features: zero-crossing rate, spectral kurtosis, spectral flatness, spectrum spread, spectral roll-off, mel-frequency cepstral coefficients. Figure 2 shows the flowchart of special sound recognition.

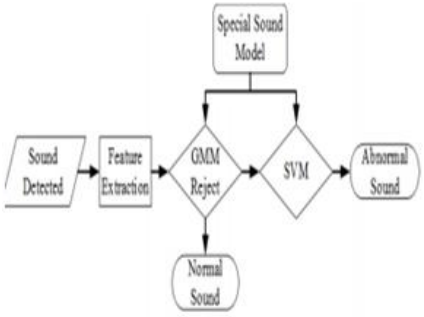

Figure 2. Flowchart of special sound recognition

B. Noise Reduction:

Due to far-field recording, a microphone receives many sound sources which may include noises of fan, air conditioner, and range hood as well as speakers' sounds. To make speaker's voice clear, the noise reduction scheme is applied to minimize noise. Based on silence frames of a dialogue, the ambient noise is modeled and estimated. Waveforms of a voice are partitioned into many frames which go through the Fourier transform. Based on a priori SNR and a posteriori SNR , a spectral gain function is estimated, and then used to multiply with spectrum signals of a sound frame. Afterwards, the normalized spectrum signals are inversely transformed to attain a time-domain sound frame with noise lessening. In our system, only speaker recognition takes the pre-process of noise reduction. In order to lower the computation complexity, the sound context associated with human speech is estimated where the autocorrelation scheme is performed frame by frame. Such an approach can prevent our system from doing noise reduction for special and environmental sounds. The autocorrelation scheme can be formulated as

$$R(\delta) = \sum_{n=0}^{P-1-\delta} x(n)x(n+\delta), \qquad (1)$$

Where x(n) is a signal of a voice frame, n is a time index, P is a is the number of delay points. Due to theδframe size, and can beδshort-term relationship among speech signals, = 8 yields fairly good estimation. δaround 2 to 20. In our tests, The larger R(δ) is, the higher the probability of a voice frame including speech signals is.

C. Speaker Recognition:

Text-independent speaker recognition is carried out in our System. Speaker recognition uses 16 mixtures GMM to reject speaker(s) who are not family members, and then linear-kernel SVM to identify whom they are. Due to text independent, the characteristics of speaker utterances rather than context must be well addressed. Accordingly, mean and standard deviation of the first-order differential mel-frequency cepstral coefficients, the first-order

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

differential linear prediction coefficients, and mean, standard deviation kurtosis and skewness of fundamental frequencies are considered to improve the recognition rate. Particularly, the feature selection scheme of SFBS is utilized to discover the adequate parameters which include low-frequency mel-frequency cepstral coefficients, high-frequency linear prediction coefficients, and fundamental frequencies. Figure 3 shows the flowchart of the speaker recognition.

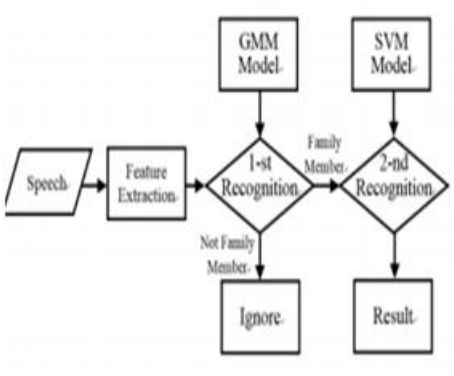

Figure 3. Flowchart of speaker recognition

D. Keyword Spotting:

Nowadays, Google web speech API is a quite convenient speech recognition engine which supports multiple languages. Hence, the proposed system sends the enhanced speech waveform file to Google for text generation. Additionally, we define some special words associated emergency (e.g. help, hurt, fire, police, thief), that are spotted from the text file for a further action. To consider the privacy, this function is only enabled by a special comment, like "Alexa" used by Echo from Amazon [10].

Conclusion:

This work develops a voice-activity home care system which consists of SPL calculation, noise reduction, special sound recognition, speaker identification, keyword spotting, and daily log establishment. In order to reduce computation complexity, SPL and speech intensity in a sound piece are analyzed to determine whether voice-activity recognition, and speech-related identification are activated, respectively.

In special sound and speaker recognitions, the two-layer recognition processes of rejection and classification are effectively employed. Experimental results exhibit that special sound, speaker, and keyword recognitions can attain 82%, 90%, and 87%, respectively. Based on these computed and recognition results, the voice-related daily log is effectively built for home care applications.

References:

[1] Danilo Hollosi, Jens Schr¨oder, Stefan Goetze, and Jens-E. Appell, "Voice activity detection driven acoustic event classification for monitoring in smart homes," Proc. of IEEE International Symposium on Applied Sciences in Biomedical and Communication Technologies, pp.1-5, 2010.

[2] Min-Quan Jing, Chao-Chun Wang, and Ling-Hwei Chen, "A real-time unusual voice detector based on nursing at home," Proc. of IEEE International Conference on Machine Learning and Cybernetics, vol. 4, pp. 2368-2373, 2009.

[3] Y.-W. Liu et al., "Developing "voice care": real-time methods for event recognition and localization based on acoustic cues," Proceedings of IEEE International Conference on Multimedia and Expo Workshops, pp. 1-6, July 2014.

[4] Jianzhao Qin, Jun Cheng, Xinyu Wu, and Yangsheng Xu, "A learning based approach to audio surveillance in household environment," International Journal of information Acquisition, vol. 3, no. 3, pp. 1-7, 2006.

[5] Huy Dat Tran, and Haizhou Li, "Sound event recognition with probabilistic distances SVM," IEEE Transaction on Audio, Speech, and Language Processing, pp. 1556-1568, vol. 19, no. 6, August 2011.

[6] M. A. Sehili, D. Istrate, B. Dorizzi, and J. Boudy, "Daily sound recognition using a combination of GMM and SVM for home automation," Proc. of 20th

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

European Signal Processing Conference, pp. 1673-1677,2012.

[7] O. T.-C. Chen, Yi-Heng Tsai, Che-Wei Su, Po-Chen Kuo, and Pin-Chih Chen, "Voice-activity recognition system for home care," Proc. of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Milan, Italy, August 25th -29th, 2015. (Late breaking research poster paper).

[8] Richard C. Hendriks, Richard Heusdens, and Jesper Jensen, "Forward-backward decision directed approach speech enhancement," Proc. of Int. Workshop Acoustics Echo and Noise Control, pp. 109-112, Sept. 2005.

[9] Tomi Kinnunen, and Haizhou Li, "An overview of text-independent speaker recognition: from features to supervectors," Speech Communication 52, pp. 12-40, 2010.

[10] Amazon Echo: Always Ready, Connected, and Fast.
http://www.amazon.com/Amazon-SK705DI-Echo/dp/B00X4WHP5E