Sound environment analysis for ADL detection from Living Lab to Medical Nursing

Dan Ovidiu Andrei, Dan Istrate, Jérôme Boudy, Said Mammar

SOUND ENVIRONMENT ANALYSIS IN MEDICAL NURSING HOMES

Dan-Ovidiu ANDREI[1,2,3], Dan ISTRATE[1,3], Jérôme BOUDY[2], Said MAMMAR[3]

ABSTRACT

In this paper we present the application of sound environment analysis algorithms previously developed for ADL recognition on real recordings made in a medical nursing home. Initially the sound algorithms were tested in laboratory, secondly in a living lab in Grenoble and lastly in a nursing home (EHPAD). Several days of audio signal have been recorded, but for the moment only 24h are labelized and used for evaluation. The proposed system is based on a combination of Wavelets Transform, Gaussian Mixture Models (GMM) and Suport Vector Machines (SVM).

Keywords: ADL identification, Sound Recognition, Real-Time, Audio Processing, GMM

1. Introduction

In this paper we shall present the utilisation of the sound environment as an information source for activities of daily life (ADL) identification. In the intelligent ambient systems the use of the sound extracted information has been more or less limited to speech recognition and speaker identification. During recent years, other everyday life sounds started to be investigated [1], [2], [3]. The sounds can have different sources: human (such as snorring, cries, screams, coughs), human generated (door clapping, dishes, object fall etc.) or nature generated (bird chirping, wind, thunder etc.).

2. Sound environment analysis

The sound environment analysis is difficult because: a very large sound classes, distant analysis using omnidirectional microphones, noise presence, large variability of the same sound.

[1] ESME SUDRIA, 38 Rue Molière, 94200 Ivry sur Seine, France; e-mail: ovidiu@esme.fr, istrate@esme.fr

[2] TSP/IMT, 9 Rue Charles Fourier, 91000 Evry, France; e-mail: jerome.boudy@telecom-sudparis.eu

[3] IBISC, 40 Rue de Pelvoux, 91000 Evry, France; e-mail: said.mammar@ibisc.univ-evry.fr

The already proposed architecture is composed by a real time step which continuously analyses the sound flow in order to detect the useful signals. This step is based on time-frequency analysis using Wavelet Transform.

Once a useful signal is detected, a hierarchical recognition stage is started. We first make the difference between speech and other sounds. Afterwards, if a sound has been identified, its membership to a sound class is analised; if speech was identified, a distress expression recognizer is started.

In this work the distress expressions recognition is not presented. The sound/speech classification and sound classification in the laboratory conditions and living lab was made using a combination between GMM and SVM [5] and the real recordings analysis is made using only GMM.

The architecture is presented in Figure 1.

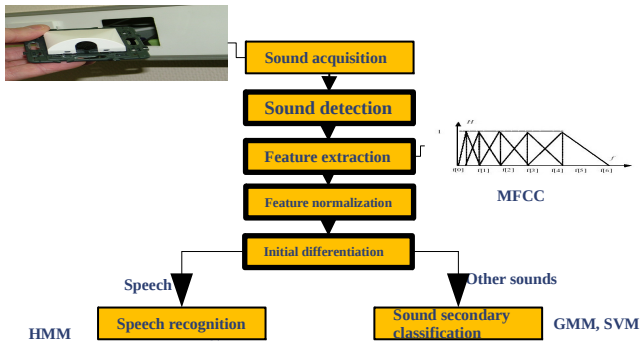

Figure 1: Sound environment analysis architecture

3. Sound database

The system evaluation was made initially using sounds from sound effects CDs, files from internet, and recordings in our laboratory; this step has allowed us to design the system architecture. In this step we have 18 sound classes with a total of 1049 files and a duration of 1 hour.

In a second step we have evaluated the system using recordings made in a Living Lab in Grenoble in the framework of SweetHome project. These recordings have been made by 21 persons and each scenario contains 2 hours of continuous recording using 7 microphones.

In the last step we have recorded files in a nursing home in the framework of EMonitor'age project.

We have recordings of 3 consecutive days. All recordings are in the format 16kHz/16 bits per sample/one channel (mono). For the moment only one day is used for evaluation.

The initial list of sound classes taken into consideration for the laboratory sounds and living lab recordings is presented in the Table 1.

Coughing	Person_Fall
Snoring	Brushing Teeth
Yawning	Object_Drop
Hands_Clapping	Radio_TV
Door_Clapping	Vacuum_Cleaner
Door_Opening	Kitchenware
Water	Window_Shutters
Ring	Speech
Paper	Unknown
Keys	

Table 1. List of sound classes used for laboratory and living lab recordings

In the framework of the 3 days continuous recordings in the nursing home, we have identified 13 sound classes and one unknown class, as presented in Table 2. The number of detected useful signals is about 6000. Newly added classes are presented in bold, the undetected classes were deleted.

Class no.	Sount class	Occurences		Class no.	Sount class	Occurences
1	Cough	22		8	Water flow	36
2	Snoring	4065		9	**Object hit**	376
3	Yawn	24		10	Kitchenware	171
4	Door clapping	92		11	**Electric Razor**	66
5	Door opening	50		12	Speech	709
6	**Door knock**	3		13	**Sigh**	52
7	**Steps**	16		14	Unknown	328

Table 2. List of detected sound classes

4. Gradual evaluation

The system was developed and optimized using a gradual approach: started with testing in our laboratory, continued with a real test in a living lab and currently dealing with recordings from a nursing home.

4.1. Laboratory Evaluation

The proposed sound environment system was primarly designed and tested on a sound data base recorded in our laboratory and completed with files from sound effects CDs and Internet files. All results are presented in [5]. The average sound recognition rate using GMM was about 71% and using SVM/GSL method it was about 75%.

4.2. Living Lab evaluation

Once the sound system optimized in laboratory, a large test on recorded files in a living lab in Grenoble was realized. The results are presented in [4]. The average sound recognition rate was about 70% using SVM-GSL method.

4.3. Real Data Evaluation

We are currently evaluating the system on the nursing home recordings.

Different tests have been conducted using MFCC and LFCC. Confusion matrices for several tests using only EMonitor'age sounds, as well as sounds from the laboratory sound data base combined with sounds from the EMonitor'age project are presented in tables 3 to 6 (all values in the tables represent good recognition rates – as percentages).

Rec. class Class	01	02	03	04	05	06	07	08	09	10	11	12	13
01	**22.727**	4.545	4.545	4.545	13.636	0	13.636	0	9.091	0	0	4.545	22.727
02	0.319	**55.034**	5.329	1.645	1.866	0.368	2.308	15.373	3.143	0.074	3.954	1.179	9.406
03	0	34.783	**13.043**	4.348	8.696	0	0	17.391	4.348	0	0	4.348	13.043
04	1.087	3.261	2.174	**15.217**	19.565	1.087	15.217	8.696	18.478	1.087	2.174	1.087	10.870
05	0	12.000	2.000	10.000	**34.000**	0	20.000	4.000	8.000	4.000	0	2.000	4.000
06	0	0	0	0	0	**0.000**	0	0	0	0	33.333	66.667	0
07	0	0	0	6.250	18.750	0	**75.000**	0	0	0	0	0	0
08	0	5.556	2.778	0	8.333	0	0	**72.222**	0	0	11.111	0	0
09	2.128	3.191	3.457	11.170	9.574	1.862	3.723	2.128	**45.479**	0.532	3.723	2.926	10.106
10	0	1.170	0	1.754	15.789	0	3.509	0	7.018	**70.760**	0	0	0
11	0	3.030	1.515	1.515	0	0	3.030	25.758	0	0	**60.606**	3.030	1.515
12	1.958	1.818	4.476	0.699	3.077	0.420	2.238	3.636	2.797	0.280	1.119	**73.706**	3.776
13	1.923	17.308	3.846	1.923	5.769	0	5.769	13.462	5.769	0	1.923	3.846	**38.462**
Overall Good Recognition Percentage: **56.147%**													

Table 3. Confusion matrix for Leave-one-out method, GMM, MFCC, EMonitor'age only sounds

Rec. class Class	01	02	03	04	05	06	07	08	09	10	11	12	13
01	**4.545**	22.727	4.545	4.545	13.636	0	4.545	0	13.636	0	0	0	31.818
02	0.639	**76.621**	3.045	2.726	1.940	0.147	0.565	0.270	6.582	0.417	0.246	0.663	6.139
03	4.348	21.739	**26.087**	8.696	8.696	0	0	0	13.043	0	0	0	17.391
04	3.261	2.174	2.174	**53.261**	10.870	0	2.174	3.261	15.217	0	0	0	7.609
05	0	12.000	2.000	32.000	**24.000**	0	6.000	4.000	6.000	2.000	2.000	2.000	8.000
06	0	0	0	0	0	**0.000**	0	0	0	0	0	100.00	0
07	0	0	0	6.250	6.250	0	**68.750**	6.250	6.250	0	0	0	6.250
08	0	8.333	0	0	5.556	0	0	**77.778**	2.778	0	5.556	0	0
09	0.532	4.255	2.128	14.362	4.787	0.532	0.266	1.330	**65.160**	0.266	1.064	1.330	3.989
10	0	3.509	0.585	1.170	6.433	0	5.263	0	11.696	**67.836**	1.754	0	1.754
11	0	21.212	0	0	1.515	0	0	19.697	16.667	0	**36.364**	3.030	1.515
12	1.538	2.238	1.119	2.937	3.916	0.280	1.119	1.958	6.853	0.140	0.699	**74.266**	2.937
13	1.923	15.385	3.846	21.154	0	0	0	0	26.923	0	0	0	**30.769**
Overall Good Recognition Percentage: **73.042%**													

Table 4. Confusion matrix for Leave-one-out method, GMM, LFCC, EMonitor'age only sounds

Rec. class / Class	01	03	04	05	08	10	11	12
01	**36.364**	9.091	0	4.545	27.273	0	9.091	13.636
03	0	**17.391**	0	8.696	47.826	0	21.739	4.348
04	2.174	4.348	**47.826**	11.957	18.478	2.174	10.870	2.174
05	2.000	4.000	26.000	**38.000**	20.000	6.000	2.000	2.000
08	0	0	0	8.333	**50.000**	0	38.889	2.778
10	0	0	5.848	14.620	1.170	**78.363**	0	0
11	1.515	0	0	0	22.727	0	**72.727**	3.030
12	1.678	4.895	2.517	2.238	8.811	0.979	4.615	**74.266**
Overall Good Recognition Percentage: **68.596%**								

Table 5. Confusion matrix for TrainWorld on living lab sounds, TrainTarget on 80% Emonitor'age sounds, Tests on 20% EMonitor'age sounds, GMM, MFCC

Rec. class / Class	01	03	04	05	08	10	11	12
01	**9.091**	4.545	63.636	0	4.545	0	0	18.182
03	0	**8.696**	82.609	0	0	4.348	0	4.348
04	0	0	**93.478**	0	0	1.087	1.087	4.348
05	0	0	86.000	**0.000**	2.000	8.000	2.000	2.000
08	0	2.778	22.222	0	**58.333**	0	11.111	5.556
10	0.585	0.585	25.146	0	0.585	**71.930**	0.585	0.585
11	0	0	39.394	0	12.121	4.545	**34.848**	9.091
12	0.420	0.420	12.028	0	0.280	0.979	0	**85.874**
Overall Good Recognition Percentage: **74.128%**								

Table 6. Confusion matrix for TrainWorld on living lab sounds, TrainTarget on 80% Emonitor'age sounds, Tests on 20% EMonitor'age sounds, GMM, LFCC

In case of the "Leave-one-out" method, one sound is used for testing and all the others are used in the training set. Sounds in the training set are used to generate the GMM models for the classes they belong to. The extracted sound is then tested for membership against every class. The process is repeated for each and every sound available.

In case of the method used for obtaining results in tables 5 and 6, TrainWorld and TrainTarget are used for modelling data and for producing a GMM. TrainWorld is a generic application capable to obtain a GMM World Model using the Expectation-maximization (EM) algorithm. TrainTarget is used for modelling data using the MAP (maximum a posteriori) and EM methods. Both TrainWorld and TrainTarget are part of the LIA-RAL software from the University of Avignon and are described in [6].

5. Conclusions

The sound environment is very rich in information, but the noise presence and the distant recording create difficulties for the analysis.

The best results so far have been obtained by creating the models from laboratory sound data base, by adapting all models using 80% of the sounds recorded in the nursing home (EMonitor'age project), all with LFCC coefficients.

The recognition rate is comparable with the one obtained in [5] using only sounds recorded in controlled conditions.

The work is in progress in order to obtain a more reliable system.

6. Acknowledgement

The authors wish to thank BPI France which has funded the EMonitor'age project and all consortium members.

R E F E R E N C E S

[1] *A. Dufaux,* (2001), Detection and recognition of Impulsive Sounds Signals, PhD thesis, Faculté des Sciences de l'Université de Neuchtel, Switzerland.

[2] *A. R. Abu-El-Quran, R. A. Goubran, and A. D. C. Chan*, Security Monitoring Using Microphone Arrays and Audio Classification, IEEE Trans. Instrum. Meas., vol. 55, no. 4, pp. 1025-1032, Aug.2006.

[3] *S.-H. Shin, T. Hashimoto, and S. Hatano,* Automatic detection system for cough sounds as a symptom of abnormal health condition, IEEE Trans. Inf. Technol. Biomed., vol. 13, no. 4, pp. 486-493, Jul. 2009.

[4] *M. Sehili, B. Lecouteux, M. Vacher, F. Portet, D. Istrate, B. Dorizzi, J. Boudy,* Sound Environnment Analysis in Smart Home, LNCS 7683 - Ambient Intelligence, ISBN 978-3-642-34898-3, DOI 10.1007 - 978-3-642-34898-3 14, Springer Berlin-Heidelberg 2012, pp. 208-223

[5] *M. A. Sehili, D. Istrate, B. Dorizzi, J. Boudy,* Daily Sound Recognition Using a Combination of GMM and SVM for Home Automation, EUSIPCO 2012, 27-31 aot 2012, Bucharest - Romania, pp. 1673-1677

[6] *J.-F. Bonastre, F. Wils, and S. Meignier,* Alize, a free toolkit for speaker recognition. In ICASSP'05, IEEE, Philadelphia, PA (USA), March, 22 2005