



HAL
open science

Two-step detection of water sound events for the diagnostic and monitoring of dementia

Patrice Guyot, Julien Pinquier, Xavier Valero Gonzalez, Francesc Alias

► **To cite this version:**

Patrice Guyot, Julien Pinquier, Xavier Valero Gonzalez, Francesc Alias. Two-step detection of water sound events for the diagnostic and monitoring of dementia. IEEE International Conference on Multimedia & Expo - ICME, Jul 2013, San Jose, United States. pp. 1-6. <hal-01148156>

HAL Id: hal-01148156

<https://hal.science/hal-01148156v1>

Submitted on 4 May 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization



Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>
Eprints ID : 12441

To link to this article : DOI :10.1109/ICME.2013.6607558
URL : <http://dx.doi.org/10.1109/ICME.2013.6607558>

To cite this version : Guyot, Patrice and Pinquier, Julien and Valero Gonzalez, Xavier and Alias, Francesc *[Two-step detection of water sound events for the diagnostic and monitoring of dementia](#)*. (2013) In: IEEE International Conference on Multimedia & Expo - ICME, 15 July 2013 - 19 July 2013 (San Jose, United States).

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

TWO-STEP DETECTION OF WATER SOUND EVENTS FOR THE DIAGNOSTIC AND MONITORING OF DEMENTIA

Patrice Guyot, Julien Piquier

SAMoVA team - IRIT
University of Toulouse - France
{guyot,piquier}@irit.fr

Xavier Valero, Francesc Alías

Grup de Recerca en Tecnologies Mèdia
La Salle - Ramon Lull University - Spain
{xvalero,falias}@salle.url.edu

ABSTRACT

A significant aging of world population is foreseen for the next decades. Thus, developing technologies to empower the independency and assist the elderly are becoming of great interest. In this framework, the IMMED project investigates tele-monitoring technologies to support doctors in the diagnostic and follow-up of dementia illnesses such as Alzheimer. Specifically, water sounds are very useful to track and identify abnormal behaviors from everyday activities (e.g. hygiene, household, cooking, etc.). In this work, we propose a double-stage system to detect this type of sound events. In the first stage, the audio stream is segmented with a simple but effective algorithm based on the Spectral Cover feature. The second stage improves the system precision by classifying the segmented streams into water/non-water sound events using Gammatone Cepstral Coefficients and Support Vector Machines. Experimental results reveal the potential of the combined system, yielding a F-measure higher than 80%.

Index Terms— Healthcare, Activity of Daily Living, Acoustic Event Detection, Water Flow, Computational Auditory Scene Analysis

1. INTRODUCTION

With rising life expectancy, nowadays there is a real challenge in helping the elderly population to keep their autonomy as long as possible. In case of dementia, such as the Alzheimer disease, an early diagnostic allows to give the patient and their relatives all the information and assistance to better deal with the disease. It may help in patient rehabilitation by offering strategies to maintain their security and autonomy at home. From the doctor's point of view, the usual practice in this kind of disease is to obtain behavioural information through observation, self-reporting or conversation with the relatives. Thus, the analysis of possible lack of autonomy in the Activities of Daily Living (ADL) is one of the key points to establish an early diagnostic.

Therefore, new science and technology challenges have recently appeared with the objective of helping doctors to monitor the ADL of patients. The automatized recognition of

water flow events may enable the indexing of several daily-life human activities, such as: activities of hygiene (e.g., *hand-washing, teeth brushing*), diet activities (e.g., *cooking, making coffee*), housework (e.g., *doing the dishes, moping, cleaning*), and leisure (e.g., *gardening*). Consequently, different studies dealing with water flow recognition have been presented in the last years. Some of them support aging in place by monitoring an elder's activities of daily living [1, 2]. Other studies aim at specifically identifying water sounds in bathrooms to obtain behavioral information [3] on activities of personal hygiene with respect to privacy. Finally, an other application directly provides reminder prompts for Alzheimer's patients in the specific context of a hand-washing task [4].

This work is framed on the IMMED project [5], which proposes a video monitoring system to help the doctors in the diagnosis of dementia illnesses. A wearable camera, positioned on the patient's shoulder, allows recording the ADL at his/her personal place of residence. For doctors, the method prevents the bias inherent from carrying out the activities observation in a medical office and also ensures a more ecological approach to monitor the autonomy of the patient in his/her own living place. For the patient, this method is far more comfortable, since he/she does not need to go to the hospital. A psychologist visits the patient to manage the camera and guide him/her with interviews and a set of pre-defined activities. The resulting videos are added to the traditional interviews and questionnaires to assist the doctors to measure autonomy decline.

An automatic method for ADL indexing is developed to facilitate the watching of the videos by the doctors. By this way, the doctors watch directly activities of special relevance to evaluate the patient autonomy. Thus, the automatic water sound recognition task, merged with other audio and video informations [6], is expected to contribute significantly to the automatic segmentation of the video in ADL.

However, in this project, unlike previous studies on water sound recognition, patients in multiple homes were recorded. Hence, the data considered is very heterogeneous. A classical pattern recognition approach may find difficulties to match the data due to the huge range of situations that can be found

in a home environment. To overcome this problem, in this paper we propose a two-step method for water sound event detection. Section 2 presents the state of the art in the field of acoustic event detection in real life and, more specifically, on water sound event recognition. Section 3 describes the proposed two-step recognition system, which is experimentally evaluated in Section 4. Finally, Section 5 draws up the conclusions and perspectives for future work.

2. RELATED WORK

From the technical point of view, the task of water sound detection is a special case of the more general problem of Acoustic Event Detection (AED). AED task offers a large set of applications, such as: jingle detection on broadcast streams [7], life-logging in real life [8], surveillance systems [9], and healthcare [4].

In this domain, the major part of the studies follow a typical recognition scheme that computes acoustics descriptors on short frames of the signal, thus composing patterns from the different events to recognize. Descriptors are usually extracted from the temporal domain (energy, zero-crossing rate, etc.) or from the spectral domain (spectral centroid, spectral flux, spectral roll-off, etc.) [10]. Among them, the Mel-Frequency Cepstral Coefficients (MFCC), which come from speech recognition systems, have become very popular, however their efficiency in noisy environments is discussed in comparison to Gammatone Cepstral Coefficients [11].

Then, a learning set manually annotated is employed to train certain supervised pattern recognition system that performs the detection of new sound events [12]. Among the classifiers, some of the most popular are: k-Nearest Neighbor (kNN), Gaussian Mixture Models (GMM), Support Vectors Machine (SVM) and Hidden Markov Models (HMM), the latter modelling the temporal evolution of the signal.

This kind of approach has demonstrated very competitive results in studio production context such as radio, music or television [7]. In that case, the data is homogeneous enough, and a train set can accurately represent all the data variety.

However, the implementation of an AED system in a real life context entails a considerably higher difficulty due to the high likeliness of finding overlapping sound events, among others. This problem, which is the main object of study in the closely related field of Computational Auditory Scene Analysis is also observed in the case of water flow recognition. Speech or traffic noises from nearby roads are typically overlapping sounds that considerably increases the difficulty of the recognition problem. Another difficulty in AED lies on the low homogeneity of the target data. In the case of water flow, the sound depends on the physical properties of the appliances (e.g. sink and faucet), and the activity which is carried out. The combination of both issues make real-life AED, and in extension water flow detection, a very hard case for the classical pattern recognition approaches. In [13], Zhang

and al. used an interesting approach a two blocks system for ball hit detection in racket sports games. The two blocks are respectively composed of a temporal energy peak detection and a spectral MFCC-based refinement. With this appropriate features, good results of ball hit detection are achieved with a low computational complexity and an easy training process.

Previous studies on water sound recognition carried out experiments employing train and test sets recorded in the same house in order to avoid too much heterogeneity on the data [1, 2, 3, 4]. Two of them used sound data obtained from microphones directly attached on water pipes. In [1] a conjunction of four microphones were used to recognize eight daily living activities in a house, as *dishwasher*, *shower*, *toilet* or *clothes washer*. The acoustic model was based on the original zero-crossing and root mean square features, with a SVM classifier. In 2008, Irbaz et al. [2] proposed a system that relied on MFCC and a KNN classifier. This system aims to detect water usage activity patterns in a kitchen environment. In another study, Chen et al. [3] proposed to detect bathroom activities with a single microphone positioned near the washing basin, also using MFCC but in this case a HMM as classifier. In one of the most sophisticated set-up to date, Taati et al. [4] focused on the task of detecting the activity of hand-washing from a combination of audio and video features, collected by a camera fixed above the sink. The audio features used in the system were the signal to noise ratio, ZCR, spectral centroid, spectral roll-off, spectral flux and MFCC. Different classifiers were tested and achieved comparable performances.

In one of our previous works, we tested such traditional approaches, and we observed that they were not suitable in the context of IMMED project which use a wearable camera in different places [14]. In the latter study, we introduced the Spectral Cover audio descriptor which, in conjunction with a simple thresholding decision, became quite robust for water sound event detection in noisy environments. The results were promising, but still let much room for improvement (average F-measure of 66%). In this work, we aim at taking advantage of such findings and combine them with other state of the art techniques in order to increase the accuracy in the identification of water sound events.

3. FULL SYSTEM DESCRIPTION

The following method use the audio part of the video recorded by the camera. The water event detection system is carried out by a system composed of two differentiated blocks (see Figure 1). The first block (segmentation) deals with the heterogeneity of the data and detects the salient events within the continuous sound signal. Without a specific training, this block segments the sound stream extracting a set of key sound events which are potential candidates to belong to the water category. The second block (classification) takes the isolated key events as input data and classifies them into two classes: water/no water. In the next sections, both segmentation and

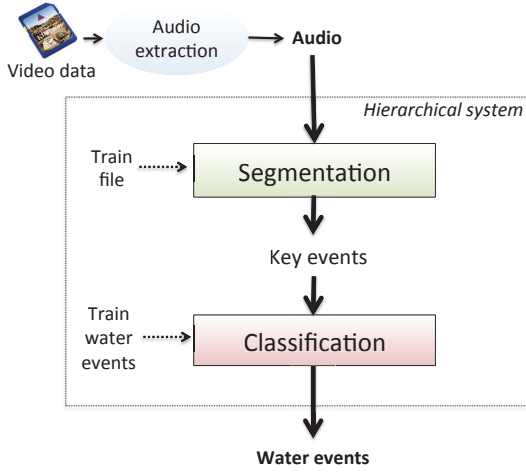


Fig. 1. Diagram of the hierarchical system.

classification blocks are described in detail.

3.1. Segmentation

This part is based on the Spectral Cover (SC), which had been found to be more robust to speech and environmental background noises than classical low level descriptors or MFCC in a water detection task [14]. SC reacts to large spectral bands and boosts the high frequencies.

The Spectral Cover is given by:

$$SC = \frac{\sum_i (ampl(w_i) * w_i)^2}{[\sum_i ampl(w_i)]^\gamma}, 1 \leq \gamma \leq 2 \quad (1)$$

where w_i represent the frequencies, $ampl(w_i)$ their amplitudes, from the Fourier transform, and $\gamma = 1.5$.

This feature is used in simple but effective system, only based on a threshold, which allows the recognition of water flow sounds longer than two seconds:

- The SC is computed on each frame of a Fourier transform (Hamming window of 80 ms and overlap of 50%). The minimum of the spectral cover values is extracted on 2s overlapped zones. Since the water events are characterized for yielding high SC values, a zone is considered as a candidate if the value exceeds a certain threshold T_1 (experimentally set). With concatenating the juxtaposed values, we obtain a first selection.
- However, there are other sound events which present higher values of SC than the water sounds, for instance sounds from vacuum cleaner. Thus, a second threshold is needed to reject this kind of sound events. If 85% of the frames included in the candidate segment have

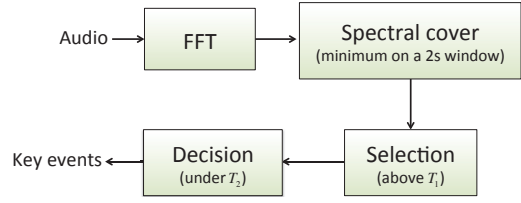


Fig. 2. Diagram of the segmentation block.

a spectral cover value superior to a second threshold T_2 , the segment is rejected as water flow; if not, it is accepted.

At this stage, the precision of the system is rather low (see line 1 of the Table 2). Therefore, a second step (classification) is needed in order to reject the key events which are not composed of water.

3.2. Classification

The classification block aims at refining the recognition accuracy of the system by classifying the isolated key events extracted by the segmentation step. Bearing in mind this goal, a detailed spectral analysis of the signal is performed. Rather than using the popular MFCC, the Gammatone Cepstral Coefficients (GTCC) are employed to perform this task, given the results reported in [11], where they outperformed MFCC in non-speech audio classification. The GTCC are bio-inspired signal features that analyse the sound signal in the spectral domain by mimicking the human auditory system. More precisely, they model the spectral response of the cochlea. The GTCC follow a similar computation scheme than MFCC. However, instead of Mel filters they employ Gammatone filters, which follow the equation 2 [15]:

$$g(t) = K t^{(n-1)} e^{-2\pi B t} \cos(2\pi f_c t + \phi) \quad t > 0 \quad (2)$$

where K is the amplitude factor; n is the filter order; f_c is the central frequency in Hertz; ϕ is the phase shift; and B represents the duration of the impulse response. The latter is related to the Equivalent Rectangular Bandwidth, which is a psychoacoustic measure of the auditory filter width at each point along the cochlea [15]. Following the model proposed by Glasberg and Moore [16], the ERB is calculated as:

$$ERB = \frac{f_c}{9.26} + 24.7 \quad (3)$$

The GTCC are computed in 30 ms frames with an overlap of 15 ms, as in [11]. The filter bank is adapted to the sampling frequency of the audio recordings: 40 Gammatone filters cover the frequency range from 20Hz (minimum audible) to the Nyquist frequency. After applying the Discrete

Cosine Transform, 13 GTCC are obtained. The mean and standard deviation of the GTCC coefficients along the frames within a water candidate segment are calculated, yielding a 26-dimensional feature vector. Subsequently, the principal components of the resulting feature vector are computed, obtaining compacted sound patterns. The classification of such patterns is performed by a supervised machine learning technique. In this work, three very different learning paradigms are considered: KNN, GMM and SVM.

KNN is an instance-based machine learning technique where a new sample is classified based on the majority class of its nearest neighbors. A commonly used distance measure is the Euclidean distance (4) [17].

$$d(x, y) = \sqrt{\sum_{k=1}^K (x_k - y_k)^2} \quad (4)$$

where x and y represent two different samples in the K -dimensional space.

The GMM is a probabilistic learning algorithm that takes a weighted sum of N simple Gaussian functions (represented by the mean and covariance of the data) to model the probability density function of each class to be recognized. With those models, conditional probabilities are computed for each class (5).

$$p(x/X_k) = \sum_{j=1}^{m_k} p(x/j)P(j) \quad (5)$$

where X_k are the samples for each class, m_k is the number of Gaussian components, $P(j)$ is the prior probability that x was generated by Gaussian component j , and $p(x/j)$ is the mixture component density. The Expectation-Maximization algorithm is employed to optimize the classification by identifying the parameters yielding maximum likelihood parameters of each class [12].

The SVM is a kernel-based machine learning technique that maps the input samples into a high dimensional space and finds the hyperplane that optimally separates the two classes. The optimal separating hyperplane is chosen following the criteria of maximizing the distance to the closest training instances (support vectors). Non-linear kernel functions such as polynomial (6) and Gaussian Radial Basis Function (7) are typically employed map the input feature space X to another high-dimensional feature space F . This process is essential to tackle non-linearly separable problems.

$$K(x, y) = (\langle x, y \rangle + 1)^d \quad (6)$$

$$K(x, y) = \exp\left(\frac{-(x - y)^2}{2\sigma^2}\right)^2 \quad (7)$$

where d is the polynomial degree and σ^2 is the variance of the Gaussian function.

4. EXPERIMENTS

4.1. Corpus

The audio corpus is provided by the IMMED project: 20 videos of patients making activities of daily living in their own personal residence. The total length is 7 hours 30 minutes, with records from 6 to 44 minutes duration. The recordings were carried out with a camera (HD Go-Pro Fisheye), and the audio part (with sampling rate of 12 KHz) was extracted from the video signal. The camera is positioned on the shoulder of the patients who execute ADL in the presence of a psychologist who suggest activities. The listening of those audios reveals the presence of multiple environment noises, such as opening and closing of doors, footsteps, kitchen utensils, and rubbing on the recording device. Some particular sounds like water flow or ring phone give obvious clues on the patient activity.

As the patient and the psychologist are in constant discussion, and due to the position of the microphone, speech is very present and can be loud in comparison to the other sounds. This strong presence of speech increases the difficulty of the water flow detection task for three reasons. First, it masks some low-level water events, which could be missed by the automatic recognition on the audio part. Secondly, due to the important acoustical variability of speech and overlapping between speech and water sounds, the data are very heterogeneous and the classical pattern recognition approach is difficult to use. Thirdly, some speech excerpts, like fricative consonant, show audio features close to water flow sounds and can be detected as false alarms.

All the recordings were manually annotated into two classes: water/no water. As we disposed of both the audio and the video, the ground truth obtained is reliable. The sound events labelled as "water" contained at least one of the two sound events: the noisy sound created by the faucet, and the splash sound created by water impacts and bubbles. Overall, there were 85 water flow events lasting 25 minutes in total.

4.2. Segmentation block setup

For the segmentation block, thresholds are experimentally fixed for all files from a development file of the corpus [14]. This system obtained a F-measure of 66%. The next classification block can not recognize water sounds which have not been previously detected as key events in this first block. However, we suppose it is precise enough to avoid false alarms. Therefore, we use a best recall configuration in this first segmentation block to avoid missed water events. In this study, T_2 has been lowered to create the best recall configuration. We used $T_2=25000$. With this configuration, we obtained a recall of 93% (see line 2 of the Table 2).

4.3. Classifier block setup

The keys events outputted by the segmentation block are cut in segments of one second. These segments are grouped by videos and annotated according to the ground truth in two classes (water/no water). The classification follows a 10-fold cross validation scheme, using 18 videos to train and 2 videos to test the system, repeating this process 10 times with different train-test sets distribution. Train data are never recorded in the same house or with the same patient than the test data. This cross validation scheme is used in all the following experiments.

Moreover, an experiment was built to empirically compare different classifiers. Table 1 shows accuracy results with different machine learning techniques. KNN uses 3 neighbours and Euclidean distance; GMM employs 10 Gaussians and SVM uses a Radial Basis Function kernel. SVM performed a slightly better classification of such sound patterns. This classifier is therefore selected among other machine learning techniques for the next experiments.

Table 1. Classifiers results on the output of the segmentation block.

	Precision	Recall	F-measure
SVM	89%	88%	88%
GMM	90%	85%	87%
KNN	86%	88%	87%

For another comparison, we made an experiment with the chosen classification method applied on the full corpus. In this experiment, the classification block is tested directly on the audio files divided into one second blocks. The line 3 on Table 2 shows results of the classification block on the full corpus. The results are quite unsatisfactory. Moreover, there are similar, but slightly better, to state of the art technic (GMM and MFCC) tested in our previous work [14]. This statement reinforces our belief that direct application of machine learning techniques is not a suitable solution.

4.4. Results of the hierarchical system

In this experiment, the full hierarchical system is tested on the 20 videos of our IMMED corpus. As table 2 shows the hierarchical system obtain 82% of F-measure. In this context, a large majority of water sounds are detected. Most of the water sounds events of the ground truth are pointed at least one time. Therefore results are widely acceptable for the needs of our application. With respect to previous works, the results are notably improved. This fact underlines the benefit of incorporating a prior segmentation block that eases the task to the SVM, which only needs to classify the isolated key sound events.

Table 2. Results on the IMMED corpus.

	Precision	Recall	F-measure
Segmentation block (previous work)	54%	83%	66%
Segmentation block (best recall)	28%	93%	44%
Classification block	48%	86%	53%
Hierarchical system	79%	86%	82%

Nevertheless, results can still be improved. Our hierarchical system is limited by a recall of 93% in the segmentation block. This means that some water sounds are not detected by our system. A depth listening of the corpus reveals some water sounds which do not present a distinct water flow. A parallel study aims at analysed liquid sounds that are not composed of flow [18]. Finally, some water sounds are shorter than 2 seconds, the limit of the segmentation block. In this case, missed water sounds have low impact on the final application because these sounds do not precisely refer to a specific activity.

5. CONCLUSION AND PERSPECTIVES

In this paper, we presented a two step-method for water flow recognition in real life conditions. This method is built in the context of a larger system aiming to monitor and assist daily activities for elderly people. Unlike previous studies in this domain, the final application is designed to be used in different places. Our system is based on a two-step method: a robust segmentation with low level features is followed by a high level classification scheme. The segmentation block is based on a feature which shows solid performances in a noisy context. The classification block uses GTCC features and a SVM classifier. Both are combined yielding outstanding results in comparison to previous methods in the same context (82% of F-measure versus 66%, the best of previous works).

This system will be implemented in a healthcare application for activities monitoring. This method will contribute to assist doctors in their diagnostic and help in patient rehabilitation. Beyond such healthcare applications, this water flow recognition method could be used for others application such as monitor water usage [2] or detect water waste [19].

Finally, our event detection system is currently limited to water flow detection. However, the methodology and the approach followed for this method could be extended to other classes of characteristic sounds, as in [13]. A two-step system based on a robust segmentation on low level features and high level classification refinement can be efficient in a real-life context to detect many other sounds, with various applications like security or sport.

6. ACKNOWLEDGMENTS

This work is supported by a grant from Agence Nationale de la Recherche with reference ANR-09-BLAN-0165-02, within the IMMED project. It is also acknowledged the financial support of Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya, in the framework of the Pyrenees Work Community (CTP).

7. REFERENCES

- [1] J. Fogarty, C. Au, and S.E. Hudson, "Sensing from the basement: a feasibility study of unobtrusive and low-cost home activity recognition," in *Proceedings of the 19th annual ACM symposium on User interface software and technology*. ACM, 2006, pp. 91–100.
- [2] A. Ibarz, G. Bauer, R. Casas, A. Marco, and P. Lukowicz, "Design and evaluation of a sound based water flow measurement system," in *Smart Sensing and Context, Proceedings of the Third European Conference, EuroSSC 2008*. 2008, pp. 41–54, Springer.
- [3] J. Chen, A.H. Kam, J. Zhang, N. Liu, and L. Shue, "Bathroom activity monitoring based on sound," *Pervasive Computing*, pp. 65–76, 2005.
- [4] B. Taati, J. Snoek, D. Giesbrecht, and A. Mihailidis, "Water flow detection in a handwashing task," in *Canadian Conference on Computer and Robot Vision (CRV)*. IEEE, 2010, pp. 175–182.
- [5] R. M egret, V. Dovgalecs, H. Wannous, S. Karaman, J. Benois-Pineau, E. El Khoury, J. Pinquier, P. Joly, R. Andr e-Obrecht, and Y. Ga estel, "The immed project: wearable video monitoring of people with age dementia," in *Proceedings of the international conference on Multimedia*. ACM, 2010, pp. 1299–1302.
- [6] J. Pinquier, S. Karaman, L. Letoupin, P. Guyot, R. M egret, J. Benois-Pineau, Y. Gaestel, and J.-F. Dartigues, "Strategies for multiple feature fusion with hierarchical hmm: application to activity recognition from wearable audiovisual sensors," in *Proceedings of the 21st International Conference on Pattern Recognition, ICPR*. 2012, IEEE.
- [7] J. Pinquier and R. Andre-Obrecht, "Jingle detection and identification in audio documents," in *International Conference on Acoustics, Speech, and Signal Processing, ICASSP*. IEEE, 2004.
- [8] M. Al Masum Shaikh, M.K.I. Molla, and K. Hirose, "Automatic life-logging: A novel approach to sense real-world activities by environmental sound cues and common sense," in *Computer and Information Technology. ICCIT*. IEEE, 2008.
- [9] C. Clavel, T. Ehrette, and G. Richard, "Events detection for an audio-based surveillance system," in *International Conference on Multimedia and Expo, ICME*. IEEE, 2005.
- [10] G. Peeters, "A large set of audio features for sound description (similarity and classification) in the cuidado project," 2004.
- [11] X. Valero and F. Al as, "Gammatone cepstral coefficients: Biologically inspired features for non-speech audio classification," pp. 1684–1689, 2012.
- [12] C.M. Bishop et al., *Pattern recognition and machine learning*, vol. 4, springer New York, 2006.
- [13] B. Zhang, W. Dou, and L. Chen, "Ball hit detection in table tennis games based on audio analysis," in *Pattern Recognition, 2006. ICPR*. IEEE, 2006.
- [14] P. Guyot, J. Pinquier, and R. Andre-Obrecht, "Water flow detection from a wearable device with a new feature, the spectral cover," in *Proceedings of the 10th International Workshop on Content-Based Multimedia Indexing, CBMI*. IEEE, 2012.
- [15] R.D. Patterson and J. Holdsworth, "A functional model of neural activity patterns and auditory images," *Advances in speech, hearing and language processing*, vol. 3, no. Part B, pp. 547–563, 1996.
- [16] B.R. Glasberg and B.C.J. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hearing research*, vol. 47, no. 1, pp. 103–138, 1990.
- [17] S. Chu, S. Narayanan, and C.C.J. Kuo, "Environmental sound recognition with time–frequency audio features," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 17, no. 6, pp. 1142–1158, 2009.
- [18] P. Guyot, J. Pinquier, and R. Andre-Obrecht, "Water sound recognition based on physical models," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP. Forthcoming*. IEEE, 2013.
- [19] T.T. Vu, A. Soka, H. Nakajo, K. Fujinami, J. Suutala, P. Siirtola, T. Alasalmi, A. Pitkanen, and J. Roning, "Feature selection and activity recognition to detect water waste from water tap usage," in *Proceedings of the 17th International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA)*. IEEE, 2011, vol. 2, pp. 138–141.