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 : 12440

To link to this article : DOI :10.1109/ICASSP.2013.6637757
URL : http://dx.doi.org/10.1109/ICASSP.2013.6637757

To cite this version : Guyot, Patrice and Pinquier, Julien and André-Obrecht, Régine *Water sound recognition based on physical models*. (2013) In: IEEE International Conference on Acoustics, Speech, and Signal Processing - ICASSP 2013, 26 May 2013 - 31 May 2013 (Vancouver, Canada).

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

WATER SOUND RECOGNITION BASED ON PHYSICAL MODELS

Patrice Guyot, Julien Pinquier, Régine André-Obrecht

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

ABSTRACT

This article describes an audio signal processing algorithm to detect water sounds, built in the context of a larger system aiming to monitor daily activities of elderly people. While previous proposals for water sound recognition relied on classical machine learning and generic audio features to characterize water sounds as a flow texture, we describe here a recognition system based on a physical model of air bubble acoustics. This system is able to recognize a wide variety of water sounds and does not require training. It is validated on a home environmental sound corpus with a classification task, in which all water sounds are correctly detected. In a free detection task on a real life recording, it outperformed the classical systems and obtained 70% of F-measure.

Index Terms— Water, drop, bubble, activity of daily living, acoustic event detection, computational auditory scene analysis

1. INTRODUCTION

While the water sounds of, e.g., rain, rivers or sea are prevalent in nature, and an emblematic part of human's sonic history [1], they are also indicative of many of our everyday home activities, such as cooking or washing. Unsurprisingly, the ability to detect such sounds automatically has therefore become a crucial part of a recent trend of systems monitoring Activities of Daily Living (ADL) in the context of elderly assistance and healthcare.

More precisely, recognizing ADL that involve water is used to assist elderly people in staying autonomous at home [2, 3, 4], to provide reminding prompts for Alzheimer's disease patients [5], or help doctors diagnose age-related dementia [6]. The IMMED project [7], which is the context of our work, aims to provide doctors with annotated video recordings of their patients, shot in their own place of residence, to help them measure and diagnose patterns of autonomy loss that would otherwise be difficult to observe. The main objective of the project is to develop algorithms to automatically segment the video recordings into key patient ADL, to make their navigating easier for doctors. As several of the ADL of interest for doctors use water, e.g. *washing hands*, *brushing teeth*, *doing the dishes*, *cooking* or *gardening*, the automatic water recognition task, merged with other sources of information [8], is expected to contribute significantly to the usability of the system. Beyond such healthcare applications, water activity detection is also used to monitor water usage [9, 10, 11] and to detect water leakage and water waste [12, 13, 14].

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) [15]. This challenge has motivated several recent attempts at adapting the classical audio pattern recognition methodology to the specificity of the task.

A typical difficulty of real-life AED [16] comes from the possible overlapping of many sound events. Target sounds may occur simultaneously with a background of e.g. speech or traffic noises, which will make their recognition less accurate. Some studies propose therefore to attached microphones directly on water pipes. A conjunction of four microphones were used to recognize eight daily living activities in a house, with the two features of Zero-Crossing Rate (ZCR) and root mean square [2]. In 2008, Irbaz et al. [9] proposed to use a sensor feeding a classification algorithm that relied on Mel-Frequency Cepstral Coefficients (MFCC) and a k-Nearest Neighbor (KNN) classifier. In another study, Chen et al. [3] proposed to detect activities from a single microphone positioned near the washing basin, also using MFCCs but a more sophisticated Hidden Markov Models (HMM) classifier. In one of the most sophisticated set-up to date, Taati et al. [5] 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; all achieved comparable performances regardless of what classifier was tested.

Nonetheless, another difficulty of real-life AED lies in the heterogeneity of data collected from different places. Two recordings of e.g. water sounds can be acoustically very different for a variety of incidental reasons, such as the presence of a different mix of other sounds, the acoustics of the room, microphone position or quality, the physical properties of the appliances (e.g. sink and faucet), and the activity which is carried out. Both issues combine to make a very hard case for the classical statistical pattern recognition approach. We therefore choose to depart from the machine learning approach that has been prevalent to date. In a previous work, with our own system based on a wearable device, we introduced an original feature, spectral cover, and found it to be more robust to speech and environmental background noises than state-of-the-art methods [17].

Despite all the variety, all of these approaches rely on similar assumptions: that water sounds are continuous flow sounds that can be modeled as generic audio textures. However, in reality, most of the ADL involving water can occur without a distinct water flow. We contend here that what makes liquid sounds so specific, and perceptively salient, is rather their temporal granularity, which results from the acoustics of individual air bubbles.

In this paper we present a original method for water sound detection which is based on a physical models. This method does not require training and can be applied to a wide variety of water sounds. Section 2 describes the water sound modelling. The details of our system are explained in section 3. Sections 4 and 5 present respectively its development and test experiments.

This work is supported by a grant from Agence Nationale de la Recherche with reference ANR-09-BLAN-0165-02, within the IMMED project.

2. WATER SOUND MODELLING

Water sounds have been studied for a long time. It has been known for almost one century that water by itself hardly makes any sound at all [18]. It is mostly when air is trapped by water that sounds can be heard. Perhaps surprisingly, liquid sounds come mainly from harmonic vibrations resulting from the entrainment of air bubbles. These bubbles can be created for example by drops falling into water. Their time-frequency localization makes some water sounds perceptively easy to recognize [19].

Leighton gave large revue of the acoustic bubble vibration phenomenon [20]. Recent studies in sound synthesis also use models based on air bubble vibration. Doel proposed a model of synthesis based on water drops falling into water to create complex liquid sounds [21]. Some recent improvements have been suggested, which take into account the sound radiation [22], or different form of bubble oscillation [23]. We describe in this part a simple model of bubble radiation, based on the studies of [20] and [21].

The impulse response of a radially oscillating air bubble is given by:

$$\iota(t) = asin(2\pi ft)e^{-dt} \tag{1}$$

where f is the resonance frequency, d the damping factor, a the amplitude, and t the time index. According to the Minnearts formula [24], the resonance frequency depend of the radius r of the bubble:

$$f = 3/r \tag{2}$$

The damping factor d can also be expressed in function of the radius:

$$d = 0.13/r + 0.0072r^{-3/2} \tag{3}$$

A perceptual clue that we are hearing a liquid sound comes from the rising pitch of a bubble formed close enough the surface. The rising bubble is modelled by setting the frequency as a function of time according to:

$$f(t) = f_0(1 + \sigma t) \tag{4}$$

where σ is related to the vertical velocity of the bubble.

We can see on Figure 1 the chirp form of a water drop sound. In addition to the rising pitch, we notice a short large-band noise at $t = 0$. We suppose that this noise is caused by the impact of the water drop. As water sounds are mainly cause by drops in this study, we will include impact sounds in our model as attack transients.

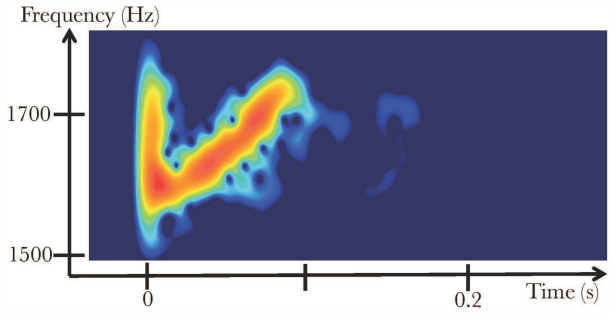

Fig. 1. Spectrogram of a drop falling into water.

3. WATER DROP DETECTION SYSTEM

These acoustic studies gave us material to create a water sound recognition system based on water drop sound detection. The detection in the spectral domain is based on the acoustic model previously described. The bubble sounds will be detected in a spectrogram as localized time-frequency phenomenons, preceded by an attack. Their time duration depends on their pitch.

Three steps make up our system (Fig. 2), which are selection on a filter bank, decision on a time-frequency zone and a final post-treatment .

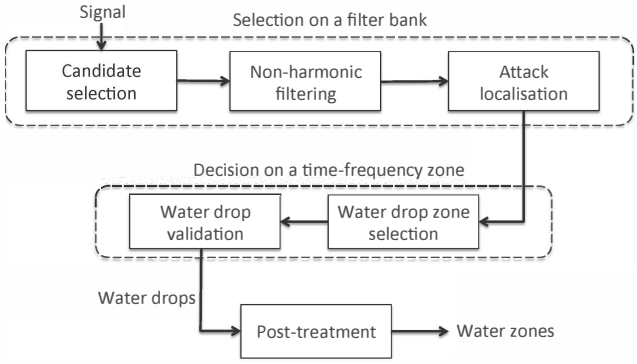

Fig. 2. Flow diagram of our water recognition system.

3.1. Selection on a filter bank

3.1.1. Candidate selection

From the equation (1), we assume that a water drop sound, at a given time, have all its energy in a single frequency band. The first step of our system consists in identify possible candidates, as frequency bins of high energy in a filter bank. From a spectrogram, we compute frequency bands of 200 Hertz width with an overlapping of 100 Hertz. We normalize each frequency bin by the total energy of the time frame. We select a bin as a candidate if its energy is superior to a threshold $T_{Candidate}$ according to:

$$\frac{E_{bin}}{E_{frame}} > T_{Candidate} \tag{5}$$

Air bubbles are most of the time limited by size. Larger bubbles may appear, for instance when a large stone is thrown into the water. As this activity is quite unusual in a household environment, we decided to remove candidates which correspond to large bubbles. Equation (2) implies a link between the radius of the air bubble and the frequency of the sound. We remove low frequency candidates under $T_{lowf} = 800$ Hz.

3.1.2. Non-harmonic filtering

Equation (1) implies that air bubbles do not have harmonic partials. We suppress in this step candidates that have been detected on harmonic time frames to be robust to speech and various harmonic sounds. The problematic is to remove sounds with harmonic partials like speech but keep water drops, which are, as pure sine waves, absolutely harmonic.

We suppose that harmonic sounds like speech are most of the time characterized by a low fundamental frequency, which is under

T_{lowf}. We use the Yin algorithm [25] to estimate the fundamental frequency f_0 of a signal at a given time.

However, the Yin algorithm produces errors in a noisy context. We use its aperiodicity measure $ap(t)$ to validate the f_0 found. The most noisy parts of the signal are maintained, because in this cases the Yin is not reliable enough to determine f_0. We fix an aperiodicity threshold T_{ap} to suppress candidate in frames t where:

$$f_0(t) < T_{lowf} \quad \& \quad ap(t) < T_{ap} \tag{6}$$

3.1.3. Attack localization

This step aims to determinate the attack time for each water drop candidate. We suppose this time is localized before the candidate. The attack time is allocated to the local minimum energy bin between the candidate and 100 ms before the candidate. We assume this time to be the beginning of the water drop event. If the retrieved time for the minimum is the same that the candidate time, we consider there is no attack and remove the candidate.

3.2. Decision on a time frequency zone

In the first step, we selected candidates on a filter bank. Now, we consider the full spectrogram, and we are going to validate the candidates in larger time-frequency zones.

3.2.1. Water drop zone selection

The damping factor d depends only on the frequency of the water drop, as we see in eq. (2) and eq. (3). To determine a time zone, we look for a time t where:

$$|asin(2\pi ft)e^{-dt}| < \epsilon \tag{7}$$

which is true if: $t > \frac{ln(\epsilon/a)}{d}$.
By fixing a threshold ϵ expressed in decibel, and supposing $a = 1$, we are able to determine a time zone in function of the frequency of the candidate.

Moreover, for each candidate, we fix the frequency range at 500 Hz around the candidate in order to be robust to pitch variations (eq. 4). Thus, we consider a rectangular zone around the candidate found.

3.2.2. Water drop validation

We define a zone before (pre-zone) and a zone after (post-zone) the selected water drop zone. By considering the discrete aspect of water drops, there should be less energy in the pre-zone (E_{pre}) and the post-zone (E_{post}) than in the water drop zone (E_{drop}). The pre-zone and post-zone have the same size than the water drop zone.

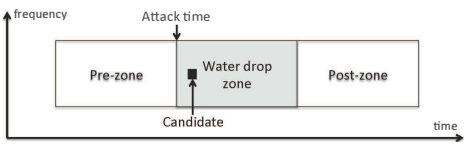

Fig. 3. Time frequency water drop zone.

We validate the water drop candidate if:

$$E_{pre} < 0.5 * E_{drop} \quad \& \quad E_{post} < 0.8 * E_{drop} \tag{8}$$

As the water drop length is an estimation and depends on several factors like the recording conditions, we use a weaker condition for the Post-zone. Finally, we removed all other candidates found in a validate water drop zone.

3.3. Post-treatment

We apply a smoothing at the end of the process to suppress isolate water drops. We consider a sliding window of duration s_t. We removed water drops when their number in the window is always less than s_n.

4. DEVELOPMENT

We made a development corpus by selecting two excerpts of 5 seconds each from the IMMED corpus [7]. One of them is composed of water drop sounds resulting of a *do the dishes* activity. The second one is an excerpt of speech. We added 20 sounds download from the Freesound Project [26]. 15 of them are water sounds as *dripping water*, *splash in a swimming pool*, *tap water*, *river*, or *boiling water*. 5 others are various home environmental sounds as *alarm clock*, *doorbell*, *door-opens-and-shuts*.

All this files are converting in Wave format at 16 bits and 16 kHz sampling rate. Spectrograms are computed with a Fast Fourier Transform (FFT) algorithm, with a 512 points hamming window and a sliding factor set at 2. We fixed our thresholds on this development corpus and we obtained:
$T_{candidate} = 0.15, T_{ap} = 0.6, \epsilon = -15dB, s_t = 2.5s$ and $s_n = 5$.

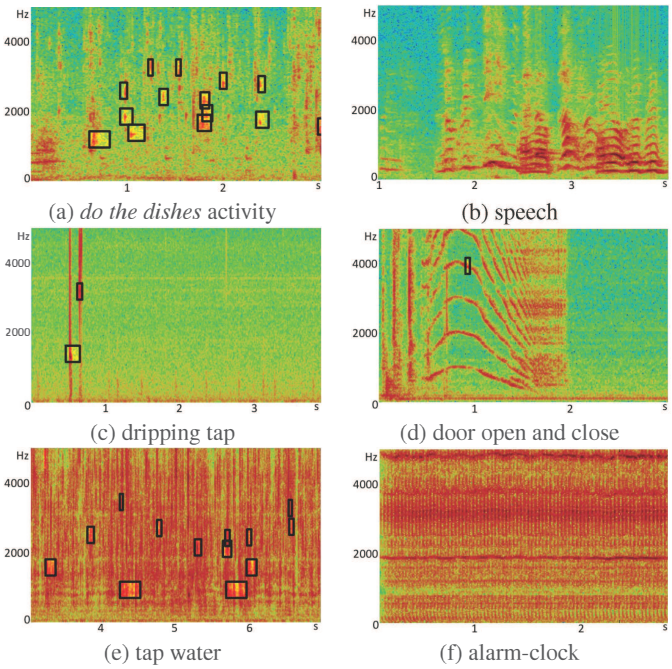

Fig. 4. Recognition on various excerpts.

We can see on Fig. 4 results on our system before the post-processing step. Figure 4-a and 4-b show three seconds excerpts on our IMMED corpus. We can see that water drops are well detected. Through to the non-harmonic filtering, the speech excerpt does not show candidates. The 4 others files come from the Freesound

Project. An isolated false alarm, due to an high pitched harmonic sound is visible on Fig. 4-d. The post-processing step is then sufficient to remove this error, however it also remove the two water drops in Fig. 4-c. Figure 4-e is very interesting: we can see in presence of a very noisy water flow sound that some water drops sounds stand out and are detected by our system.

5. EXPERIMENTS

5.1. Classification on a household corpus

We used the BBC-Sound-Effect-Library{#3: *Household*} to create a test corpus. This corpus contains 47 files, for a total duration of one hour. It is composed by home environmental sounds: 9 corresponds to water sounds and 38 to other sounds. It contains sounds like *wash-hand basin*, *bath*, *toilet flushed*, *dripping tap*, and also *front door open and close*, *telephone*, *domestic chiming clock*, *electric jigsaw*.

The objective is to classify this corpus into two classes: *water* and *other*. The *water* class is chosen by our system if it remains at least one water drop after the post-processing step. Table 1 shows results of our system in term of number of files detected as water. All water sounds had been recognized by our water drop system. The F-measure is 66 % in terms of number of file detected. One false alarm is due to the *frying an egg* sound, which imply liquid, even if there is not presence of water. Other false alarms are specific sounds, for instance *sawing a piece of plywood* or *match being struck*. We assume that those textured sounds are composed of numerous pitched impacts that are detected by our system.

Table 1. Classification results on the BBC Sound effect library.

	Water sounds	Other sounds
Corpus	9	38
Water drop system	9	9

5.2. Detection on a real life recording

The second experiment is made on the IMMED project [7]. For privacy reasons, this corpus is only available to physicians and people involved in this project. We selected a 21 minutes file consisting of an elderly person doing different activities at home. As it was recording in a real life context, a lot of environmental sound events occur, as the opening of noisy doors, the singing of birds, some noises of cooking utensils, etc. It contains 82 seconds of the *do the dishes* activity. However, the water flow sound can difficultly be heard, that makes this activity very hard to recognize by classical methods. The results of this experiment are presented in table 2 in terms of duration. We can see that almost all the activity has been detected by our system. The F-measure of this experiment is 70 %. False alarms are mainly due to manipulation of objects such as plastic bags, or shocks on the recording device. For comparison purpose, we added in table 2 results of a previous system which focused on water flow detection [17] with a F-measure of 45 %. A classical state-of-the-art approach (GMM/MFCC) obtained only 36 % of F-measure as highlighted in [17].

6. CONCLUSION

We presented on this paper an original water sound recognition method which is based on physical models. Two experiments have been proposed. The first is a classification task made on a referenced database composed of home environmental sounds. The second is a detection task on a file recorded at the home of an elderly person in real life conditions. Both results are very encouraging. Some improvements may be considered.

Table 2. Detection results on a real-life recording.

	Water sounds	Other sounds
Water Sound in the database	82 s	1171 s
Water drop system Output	63 s	35 s
Water drop system results	**F-measure 70%**	
Spectral Cover system [17]	F-measure 45%	
GMM / MFCC [17]	F-measure 36%	

Even if some air bubble sounds stand out from a noisy flow (Fig. 4-e), we suppose that water sounds are sometimes too noisy for the drops to be detected. We could therefore consider the fusion of our system with a water flow detection system. This fusion could make possible to recognize both the noisy sound of a water flow and the perceptively salient sounds of air bubbles. Furthermore, the presence of flow or drops let us plan to use these two systems for an activity identification task, for example between the activities *do the dishes* and *filling a container*.

7. RELATION TO PREVIOUS WORKS

This study presents a new method for water flow detection, with applications for elderly assistance. Several papers have been presented in this field this last few years [2, 3, 5, 9]. All this papers used state-of-the-art technics for sound recognition, features like energy, ZCR, MFCC, and classifiers like GMM, KNN or SVM. Two drawbacks can be seen on the proposal technics.

Firstly, this study all use fixed microphones placed on the water pipes or above the sink. We suppose this place allows them to avoid lot of environmental noises, in comparison of the wearable device used in the IMMED project [7]. Furthermore, the main drawback we can underline is that the train and test data have been recorded in the same place. It seems clear that the methods closely depend on the homogeneity of the data, and are not designed to be used in different places or different context. In [17], we proposed a method without training based on a specific feature for water flow recognition in a noisy environment.

Secondly, all this previous works rely on similar assumptions: that water sounds are continuous flow sounds. The micro-temporal evolution of the water sounds was not analyzed. We contend here that water flow is not the only instance of water sounds. Old acoustic studies [18] as well as recent works in sound synthesis [21] and perception [19] told us that water sounds are mainly composed of air bubble vibrations. Our system is based on the physical phenomenon at the origin of the water sound: the air bubbles. A spectral approach allows us to recognize groups of air bubble sounds in various excerpts. The system is slightly independent of the recordings conditions, including the appliances (e.g. sink and faucet). This method can be fitted on a wide variety of context, from the swimming pool to the bathroom, and can help us to recognize activities of daily living, including *water boiling*, *do the dishes* or *washing hands*.

8. REFERENCES

[1] R.M. Schafer and R. Murray, *The tuning of the world*, Knopf New York, 1977.

[2] 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.

[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] E. Thomaz, V. Bettadapura, G. Reyes, M. Sandesh, G. Schindler, T. Plötz, G.D. Abowd, and I. Essa, "Recognizing water-based activities in the home through infrastructure-mediated sensing," in *Proceedings of the 2012 International Conference on Ubiquitous Computing, UbiComp '12*. ACM, 2012, pp. 85–94.

[5] 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.

[6] Y. Gaëstel, C. Onifade-Fagbemi, F. Trophy, S. Karaman, J. Benois-Pineau, R. Mégret, J. Pinquier, R. André-Obrecht, and J.-F. Dartigues, "Autonomy at home and early diagnosis in alzheimer disease: usefulness of video indexing applied to clinical issues. the immed project," *Journal of Alzheimer's and Dementia*, vol. 7, no. 4, pp. 245, 2011.

[7] R. Mégret, V. Dovgalecs, H. Wannous, S. Karaman, J. Benois-Pineau, E. El Khoury, J. Pinquier, P. Joly, R. André-Obrecht, and Y. Gaëstel, "The IMMED project: wearable video monitoring of people with age dementia," in *Proceedings of the international conference on Multimedia*. ACM, 2010, pp. 1299–1302.

[8] J. Pinquier, S. Karaman, L. Letoupin, P. Guyot, R. Mégret, 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.

[9] 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.

[10] J. Froehlich, E. Larson, T. Campbell, C. Haggerty, J. Fogarty, and S.N. Patel, "Hydrosense: infrastructure-mediated single-point sensing of whole-home water activity," in *Proceedings of the 2009 ACM International Conference on Ubiquitous Computing, UbiComp '09*. ACM, 2009, pp. 235–244.

[11] T. Campbell, E. Larson, G. Cohn, J. Froehlich, R. Alcaide, and S.N. Patel, "WATTR: a method for self-powered wireless sensing of water activity in the home," in *Proceedings of the 2010 International conference on Ubiquitous computing, UbiComp '10*. ACM, 2010, pp. 169–172.

[12] Y. Terao and A. Mita, "Robust water leakage detection approach using the sound signals and pattern recognition," in *Proceedings of Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*. Society of Photo-Optical Instrumentation Engineers, 2008.

[13] T.T. Vu, A. Sokan, 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.

[14] K. Kappel and T. Grechenig, "Show-me: water consumption at a glance to promote water conservation in the shower," in *Proceedings of the 4th International Conference on Persuasive Technology*. ACM, 2009, p. 26.

[15] X. Zhuang, X. Zhou, T.S. Huang, and M. Hasegawa-Johnson, "Feature analysis and selection for acoustic event detection," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP*. IEEE, 2008, pp. 17–20.

[16] A. Mesaros, T. Heittola, A. Eronen, and T. Virtanen, "Acoustic event detection in real-life recordings," in *Proceedings of the 18th European Signal Processing Conference, EUSIPCO*. IEEE, 2010.

[17] 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.

[18] Sir W. H. Bragg, *The World of Sound*, Bell, London, 1921.

[19] M.N. Geffen, J. Gervain, J.F. Werker, and M.O. Magnasco, "Auditory perception of self-similarity in water sounds," *Frontiers in Integrative Neuroscience*, vol. 5, 2011.

[20] T. G. Leighton, *The acoustic bubble*, vol. 10, Academic Press, 1997.

[21] K. Van den Doel, "Physically based models for liquid sounds," *ACM Transactions on Applied Perception (TAP)*, vol. 2, no. 4, pp. 534–546, 2005.

[22] C. Zheng and D.L. James, "Harmonic fluids," in *Proceedings of Transactions on Graphics (TOG)*. ACM, 2009, vol. 28, p. 37.

[23] W. Moss, H. Yeh, J.M. Hong, M.C. Lin, and D. Manocha, "Sounding liquids: Automatic sound synthesis from fluid simulation," *ACM Transactions on Graphics (TOG)*, vol. 29, no. 3, pp. 21, 2010.

[24] M. Minnaert Sc. D., "On musical air-bubbles and the sounds of running water," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 16, no. 104, pp. 235–248, 1933.

[25] A. De Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *J. Acoust. Soc. Am*, vol. 111, no. 4, pp. 1917–1930, 2002.

[26] V. Akkermans, F. Font, J. Funollet, B. De Jong, G. Roma, S. Togias, and X. Serra, "Freesound 2: An improved platform for sharing audio clips," in *International Society for Music Information Retrieval Conference, ISMIR, Late-breaking Demo Session*, Miami, Florida, USA, 24/11/2011 2011.