



## Identification of categories of liquid sounds

Patrice Guyot, Olivier Houix, Nicolas Misdariis, Patrick Susini, Julien Pinquier, Régine André-Obrecht

### ► To cite this version:

Patrice Guyot, Olivier Houix, Nicolas Misdariis, Patrick Susini, Julien Pinquier, et al.. Identification of categories of liquid sounds. *Journal of the Acoustical Society of America*, 2017, 142 (2), pp.878-889. 10.1121/1.4996124 . hal-01578629

**HAL Id: hal-01578629**

**<https://hal.science/hal-01578629>**

Submitted on 2 Dec 2021

**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.



## Open Archive Toulouse Archive Ouverte

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's version published in:

<http://oatao.univ-toulouse.fr/19147>

**Official URL:** <https://asa.scitation.org/doi/10.1121/1.4996124>

DOI : <https://doi.org/10.1121/1.4996124>

**To cite this version:** Guyot, Patrice and Houix, Olivier and Misdariis, Nicolas and Susini, Patrick and Piquier, Julien and André-Obrecht, Régine *Identification of categories of liquid sounds*. (2017) The Journal of the Acoustical Society of America, 142 (2). 878-889. ISSN 0001-4966

Any correspondence concerning this service should be sent to the repository administrator: [tech-oatao@listes-diff.inp-toulouse.fr](mailto:tech-oatao@listes-diff.inp-toulouse.fr)

# Identification of categories of liquid sounds

Patrice Guyot, Olivier Houix, Nicolas Misdariis, and Patrick Susini

*Institut de Recherche et Coordination Acoustique/Musique - Centre National de la Recherche Scientifique - Université Pierre et Marie Curie, 1 Place Igor Stravinsky, 75004 Paris, France*

Julien Pinquier and Régine André-Obrecht

*Institut de Recherche en Informatique de Toulouse, Université Paul Sabatier, 118 Route de Narbonne F-31062 Toulouse Cedex 9, France*

Sounds involving liquid sources are part of everyday life. They form a category of sounds easily identified by human listeners in different experimental studies. Unlike acoustic models that focus on bubble vibrations, real life instances of liquid sounds, such as sounds produced by liquids with or without other materials, are very diverse and include water drop sounds, noisy flows, and even solid vibrations. The process that allows listeners to group these different sounds in the same category remains unclear. This article presents a perceptual experiment based on a sorting task of liquid sounds from a household environment. It seeks to reveal the cognitive subcategories of this set of sounds. The clarification of this perceptive process led to the observation of similarities between the perception of liquid sounds and other categories of environmental sounds. Furthermore, the results provide a taxonomy of liquid sounds on which an acoustic analysis was performed that highlights the acoustical properties of the categories, including different rates of air bubble vibration.

[<http://dx.doi.org/10.1121/1.4996124>]

## I. INTRODUCTION

Water plays an important role in human activities. Settlements and dwellings have always been dependent on its presence. People use water to drink and cook, for sanitation, and domestic chores. These three kinds of activities produce sounds that can be heard at home. In addition, water sound has sometimes been used to mask other sounds, for example, road annoyance.<sup>1</sup> A further example is the courtrooms of Middle Eastern palaces that often included a fountain whose noise could be useful to mask confidential conversations.

These sounds produced by liquid, or by an element in interaction with liquid, are henceforth referred to as liquid sounds. From the listener's viewpoint, they are usually considered as environmental sounds which can be defined as "all naturally occurring sounds other than speech and music" (Gygi *et al.*,<sup>2</sup> p. 1). Experimental research has been conducted to investigate the way human beings interpret these environmental sounds. It appears that they continuously monitor their environment by identifying and interpreting useful sounds.<sup>3</sup>

Numerous recent academic works have reported sounds involving liquid. Water sounds are considered in the field of experimental psychology with the analysis of environmental sound perception.<sup>2-5</sup> The presence of liquid sound in the context of daily living has led to applications capable of activity recognition and waste detection in the framework of automatic sound analysis.<sup>6-11</sup>

Up to now there have been no specific studies on the perception of liquid sounds. Indeed, liquid sounds have been

presented among other environmental sounds. Thus, intra-category differences are squeezed by the presence of other categories, which is known as context effect.<sup>12</sup> From the viewpoint of sound detection, signal processing approaches are generally based on generic methods such as machine learning techniques. In this framework, liquid sounds are mostly detected like any other sound without a specific approach. Moreover, the sound events targeted are not clearly defined due to the lack of an established classification.

In addition to these recent works, it appears that the physical production of water sounds has long been subject to analysis in the acoustic domain. Based on this knowledge, new models for sound synthesis have been proposed.<sup>13-15</sup> Nevertheless, these different studies in the fields of perception, sound detection and acoustics have not been considered jointly.

The aim of the present work is to fill-in different gaps in liquid sound studies. It mainly focuses on the category of liquid sounds. Stimuli are restricted to water and liquid events occurring only in the home, a place that provides an impressive variety of sounds. To explore this diversity, a sorting task during a psychophysical experiment involving listeners is conducted to highlight liquid sound categories. In the end, liquid sounds are analyzed from their physical production to their perception and then considered through acoustic analysis. Our results will be discussed in relation to the acoustic modeling reported by previous studies.

This paper is organized as follows. Section II merges and reviews different related works on water sound production and perception. The experimental protocol is described in Sec. III and the results are discussed in Sec. IV. The final section provides an acoustic analysis of different kinds of audio features.

## II. RELATED WORKS

### A. Acoustic of liquid sounds

#### 1. Air-bubble

Acoustically, liquid sounds can differ considerably depending on the physical cause that produces them. For instance, the iconic sound of water drops is close to a short sine sound, while a shower produces a continuous flow closer to a colored noise. Other kinds of sound events are also observable, like squeezing a mop or boiling water.

From the viewpoint of physics, it has long been known that water by itself hardly makes any sound.<sup>16</sup> Water sounds mostly come from the entrainment of air bubbles trapped within water. This is observable in nature, for instance, when a wave, which moves silently on the water's surface, breaks. By breaking, the wave traps air and thus produces a sound.

Much physical research has been devoted to this air-bubble phenomenon; see Ref. 17 for an extensive review. The oscillations of air bubbles can be modeled by a spring-mass system. The bubble forms a cavitation that can be compared to a spring overlapped by a certain amount of water. According to the Minnaert formula,<sup>18</sup> this model produces a unique frequency that depends on the size of the bubble and which can be expressed as

$$p(t) = A_0 \sin(\omega_0 t) e^{-\delta t}, \quad (1)$$

where  $A_0$  is the initial pressure amplitude,  $\omega_0$  is the resonance frequency,  $t$  is the time, and  $\delta$  is the damping factor (in the reciprocal of the time unit).

As the bubble rises, the mass of the water overhead decreases. Consequently, the vibration frequency of the bubble increases. This produces the famous sound of a drop falling into water, which can be modeled as a pure sine wave with an upward chirp. The acoustical models developed have led to different methods for synthesizing water sounds.<sup>13–15</sup> The addition of simple air-bubble sounds creates various water sound events (see Ref. 19 for a video demonstration of Moss's method), that can be produced by diverse physical causes.

#### 2. Other vibrations

Air bubble vibration is not the only phenomenon involved in what we call water sounds. Small shock waves are produced when water drops fall and break on a surface, making the contact region move at supersonic speed.<sup>20</sup> In addition, solid object vibrations may be caused by water. For instance, the use of a faucet generally makes the faucet vibrate as does filling a glass, which makes a glass noise. These solid vibration sounds can be heard simultaneously with bubble sounds.

The entrainment of air bubbles is generally mentioned as the main source of water sounds.<sup>21</sup> As pointed out by a more recent study, air bubbles can appear when liquid drops impinge on a solid surface.<sup>22</sup> Finally, this microscopic phenomenon seems relevant for characterizing liquid sounds as it occurs in each of them.

### B. Perception of liquid sounds

#### 1. From acoustics to perception

The acoustic variability of water sounds is considerable due to the different numbers and sizes of bubbles, possible solid interactions, and possible specific temporal patterns. Events like *water drop*, *shower*, *boiling water*, *flushing*, and *filling* produce distinctly different sounds. Thus, although the air-bubble phenomenon is relevant from the physical viewpoint, it is less obvious from the standpoint of human perception.

Finally, listeners do not appear to be aware of the air-bubble phenomenon. In “everyday listening” mode, as defined by Gaver,<sup>23</sup> listeners try to infer the sound source and its physical properties. In the case of water sounds, they usually seek to combine sounds with a higher level of physical cause than water bubbles, such as the action of *filling a glass*, which involves liquid and solid interactions.

#### 2. Specific studies on liquid sounds

In addition to physics and sound synthesis, water sounds have also been considered in psychophysical experiments. A study by Cabe focused on the action of filling a cylindrical vessel, for instance, a glass of water.<sup>24</sup> In this kind of action the container's fundamental resonant frequency increases as the container is filled with water. This auditory information can support the perception of filling. An experiment showed that monitored subjects correctly filled receptacles using only auditory information. However, other experiments have shown that the best performance is obtained with multimodal information.

In another study, Maria Geffen identified scale-invariance in the time-frequency structure of running water sounds.<sup>5</sup> Results of an experiment showed that the perception of these sounds did not change with the modification of playback speed. Moreover, according to the authors, a scale-invariant sound could be perceived as a water sound.

Finally, the perception of running water sounds seems to depend on their global structure. As pointed out by Bregman,<sup>25</sup> the perception of global properties generally relies on details, which in our case are individual air bubble vibrations.

#### 3. Liquid sounds among environmental sounds

Beyond specific studies on liquid sounds, the perception of environmental sounds has been addressed through classification experiments. As a result, liquid sounds form a subclass apart.

Vanderveer's PhD thesis can be considered as a milestone in environmental sound research.<sup>26</sup> In an experiment, she asked participants to write what they heard when listening to recorded sounds. The results showed that the participants described the action, source, and area where the action took place more than the sound itself in terms of auditory features.

Different studies on the categorization of environmental sounds have since been conducted. Marcell *et al.* obtained 27 categories corresponding to the sound sources or more

abstract concepts.<sup>27</sup> A specific category linked to water/liquid sounds was cited and included six sounds labeled as *pouring water*, *water bubbling*, *river*, *water dripping*, *water draining*, *frying food*. In addition, water sounds can be found in other categories such as a bathroom or kitchen that are more related to the context of listening.

In Ref. 2, Gygi *et al.* proposed a free labeling experiment using 50 sounds and found 13 categories, including the category *water*. In another study, Giordano *et al.* focused on the cognitive differences between hearing living and non-living stimuli as well as hearing action and non-action sounds.<sup>28</sup> In Ref. 4, Lemaître *et al.* investigated the influence of listeners' expertise and causal uncertainty<sup>29</sup> on the categorization of environmental sounds. Participants were asked to indicate the cause (object and action) of each sound (101 recordings of activities usually occurring in a kitchen). Remarkably, "water" was the noun most cited in the verbalization of the participants. This underlines the relevance of a specific category related to liquid sounds among environmental sounds.

In a more recent article,<sup>3</sup> Houix *et al.* used as a starting point the phenomenological taxonomy proposed by Gaver.<sup>23</sup> This taxonomy of simple physical interactions producing sounds provided an organization of environmental sounds classified by physical interactions. The liquid sounds based on Gaver's taxonomy are presented in Fig. 1.

Houix *et al.* proposed an experimental framework to reveal the perceptual structure underlying a large set of everyday sounds in terms of sound event related to several physical states and actions proposed by Gaver. The results demonstrated the relevance of the main categories (liquids, solids, gases) identified by Gaver. Nevertheless, the hybrid categories involving different kinds of interaction, such as between liquid and solid, do not appear in the participants' classification. The sounds that implied water were considered as liquid sounds, even if interaction with a solid was audible.

The second part of the study revealed perceptual categories that somewhat differed from Gaver's assumptions. Indeed, the categories recovered were based on the physical actions that produce sounds and revealed a distinction between discrete interactions (e.g., impacts) and continuous interactions

(e.g., tearing). In addition, the subcategories of liquid sounds proposed by Gaver (drip, pour, splash, ripple), seemed arbitrary and hardly matched with water sound events such as water flow, shower, or sponge squeezing. Therefore, they are not representative of all kinds of liquid sounds.

#### 4. Classification of liquid sounds

Liquid interactions may produce diversified sound events (*water drop*, *shower*, *boiling water*, *filling*). Several studies indicated that these sounds are well-identified and grouped together as a category.<sup>3,4,27</sup> Although the liquid category seems to form a subclass of environmental sounds, to the best of our knowledge no perceptual study has ever focused on this specific category.

In order to investigate the cognitive organization of this set of sounds, we conducted a new psychophysical experiment. We focused on a particular set of sounds related to healthcare applications: liquid sounds from home activities. In a sorting experiment, three kinds of similarities have been observed: acoustical (timbre), causal (source or action), and semantic (place, occasion or concept).<sup>4</sup> The present study is in line with previous works that proposed a taxonomy of sounds based on physical interactions.<sup>3,23</sup> Moreover, it aims at providing a causal classification of liquid sounds for activity recognition.<sup>6-8,10</sup> Therefore, we asked the participants to focus on causal similarities rather than on acoustic and semantic ones.

Much effort was made to compile a corpus containing various causes related to liquid sounds. In the following parts, we describe the participants, the procedure used to form the corpus, and the apparatus used for the listening tests. The sorting task and its results will be presented and discussed afterwards.

### III. EXPERIMENT

#### A. Participants

The sorting experiment was performed by 30 (13 women, 17 men) participants, that were paid for their participation. The average age of the participants was 25.5 (standard deviation = 4.5). All reported normal hearing.

#### B. Stimuli

The following parts describe the process that leads to obtain 64 various liquid sounds that can be heard at home. All the sounds were formatted in 16 bits WAV files at a sampling rate of 44.1 kHz.

##### 1. Inventory of the set of sounds

Establishing a corpus is a critical element of a sorting task. In particular, the corpus must reflect the variability of liquid sounds occurring in a real life context in terms of causes related to different home activities. Thus a large set of liquid sounds as well as hybrids (liquid/solid) were considered in the present study.

To create the set of sounds, we looked for lexemes identifying all the sounds generated by a liquid interaction that

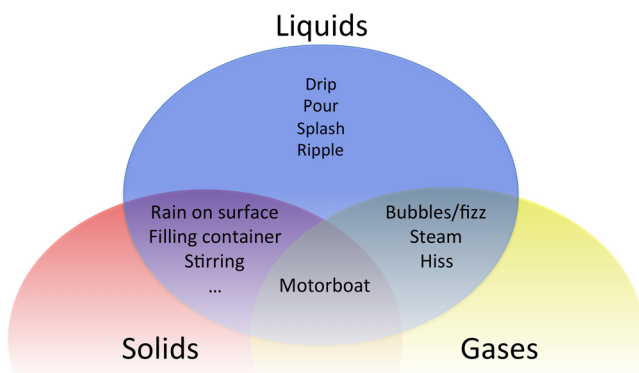


FIG. 1. (Color online) Example of liquid sounds using Gaver's taxonomy (Ref. 23). The liquid, solid, and gas physical sources overlap on hybrid regions.



may occur in the home. To determine these terms, we used the following approaches.

- **Gaver's taxonomy:** A basic list was created from the liquid and hybrid categories of Gaver's taxonomy. This study provided us with several types of actions and sounds (for example, *splash* and *drip* from the liquid sources and *filling a container* from the hybrid categories).
- **Questionnaires:** This first list of words was supplemented by a questionnaire addressed by email to thirty colleagues. It contained the following instruction: *make a list of all sound events involving the production of a liquid element (more or less viscous material) that one might hear inside a house*. Following the return of the questionnaires, different tokens were added to the list (for example the verb *to spray*).
- **Wordnet project:** We used the Wordnet project<sup>30</sup> which defined a set of semantic relationships between words. We used the words already collected as entries and considered their neighbors in the network as words that could be added. These relations allowed us to extend our lexemes to other words involving liquid (for example, the verb *to slosh*).

## 2. Collecting audio files

Using different sound databases, we selected all sounds whose label corresponded to a previously identified lexeme. We collected a set of more than 700 audio files from the following sound databases: Auditory Lab,<sup>31</sup> BBC Sound Effects Library—Original Series,<sup>32</sup> Blue Box Audio Wav,<sup>33</sup> Hollywood Edge Premiere Edition I, II, and III,<sup>34</sup> Sound Ideas General Series 6000,<sup>35</sup> and SoundScan v2 Vol. 61.<sup>36</sup>

Then, we eliminated redundancies, defined as acoustically similar sounds produced by the same cause. In this task, we focused on maintaining the diversity of sound events produced by one or several liquid elements that can possibly interact with other materials. Many sounds, such as similar water flows, were deleted. At the end of the process, we obtained 81 different sounds. We describe below how this corpus was formatted for the experiment.

## 3. Editing

We edited the sounds to keep the unit of event and not complex sequences. For example, the sequence “*open a*

*beer—pour into glass—fizz*” can be split into three events. Moreover, the sounds were edited in order to avoid subgroups of different lengths. We obtained a mean duration of 3.4 s with a standard deviation of 1.2 s (Table III shows duration of each class obtained).

## 4. Ecological sound pressure level adjustment

As the sounds were taken from different databases, the recording conditions differed between the stimuli, especially the sound pressure level. A recent study by Susini *et al.*<sup>37</sup> showed that when loudness is a factor in comparing sounds, it will typically dominate participants' judgments, especially in sorting tasks. On the contrary, when sounds are normalized in loudness, they may not reflect their ecological level related to specific distances in a real context; for example, water drops produced at 90 dB SPL will probably not be perceived as water drops at one-meter distance. This assumption was confirmed in Ref. 38; it appears a significant effect of the level change on recognition score. In order to avoid any unrealistic perception of everyday sounds in a sorting task comparison, in Refs. 3 and 4, equalization was specific to each sound source, and designed to match the typical level of the source when heard at a distance of one meter in a real context.

For equalization, a tap water flow was calibrated in order to obtain the same amplitude as for a listener positioned in front of the source (see Fig. 2). This sound was used as a standard sound for the ecological sound pressure level adjustment procedure, because it is similar to a quasi-stationary broadband noise. The participants listened to pairs consisting of the reference sound and another sound from the corpus. They had to adjust the sound level of the stimulus to obtain a realistic loudness, as the sound sources were at a distance of 1 m. This task is rather subjective. However, the stimuli of the corpus are usual sounds that are daily heard at home.

Nine people, different from the other tasks, participated to this loudness calibration. At the end of the experiment, we changed the sound pressure level of the stimuli according to the median level chosen by the participants. In this calibration step, we noticed that some sounds presented a higher inter-subject variability than others, for instance, sounds including fast boiling such as *Frying-An-Egg* or *Water on hot plate*. The following step allowed removing some of these sounds from the corpus.



FIG. 2. (Color online) Calibration of a recording sound to obtain the same level (in dBA peak) in the listening cab as under real conditions in a bathroom.

## 5. Final selection

Results of a previous experiment showed that the causal uncertainty of the sounds<sup>29</sup> strongly influences the strategy used to sort them.<sup>4</sup> Sounds with a high uncertainty value—which means sounds difficult to identify—were clustered together based on their acoustical similarities.

Therefore, we removed badly identified sounds to emphasize causal similarity confidence ratings. These ratings were assumed to be strongly correlated to the real capacity to identify the sound. The participants had to listen to each stimulus (two possible plays) and answer the question “*Are you able to determine the cause of the sound?*” by choosing one of the following answers:

- *I have no idea,*
- *I am not really sure,*
- *I hesitate between several types of causes,*
- *I am almost certain,*
- *I can perfectly determine the cause of the sound.*

The sounds were rated by 13 participants, different from the other tasks. Finally, we removed 17 stimuli that were given low identifiability ratings, for example, *Pouring ketchup from bottle, heavy glugs.*

## C. Apparatus

The sounds were played by a Macintosh Mac Pro (Mac OS X v10.6) workstation with an RME Fireface 400 sound card. The stimuli were amplified dichotically over a pair of Yamaha MSP5 loudspeakers. The participants were seated in a double-walled sound isolation booth (IAC) located at the *Institut de Recherche en Informatique de Toulouse* and received written instructions (in French) explaining the tasks to be performed.

## D. Procedure

The sorting experiment was conducted on a corpus of 64 sounds. The procedure had two steps. The first one was a sorting task where the 30 participants had to group the sounds according to causal similarities. The following is an excerpt of the instruction for this task:

*“Create classes of sounds depending on the physical event that caused the sound. For example, for the sounds ‘firecracker,’ ‘Popped balloon,’ and ‘tire bursts,’ the physical event that caused the sounds could be ‘explosion.’”*

In the second step, following the sorting task, the participants were asked to write the properties shared by the sounds in each class. We used the software `TCL-LABX` (Ref. 39) to perform these two tasks.

The first step lasted 26 min on average. Different classes of sounds were obtained; the number of classes per participant ranged from 2 to 18 with an average of 10.

## E. Data analysis

Each individual partition was coded as a  $64 \times 64$  distance matrix, where the rows and the columns are sounds. A value of 0 between a row and a column indicates that the participant grouped the sounds together, whereas a value of 1 indicates that the sounds were not placed in the same group.

The first step of the analysis aims at calculating agreement between participants, in order to identify eventual outliers. As in Ref. 40, we used the  $Rv$  coefficient to calculate the agreement between each pair of participants. The  $Rv$  coefficient between two individual matrices  $S_n$  and  $S_m$  is defined as

$$Rv(S_n, S_m) = \frac{\text{trace}\{S_n^T S_m\}}{\sqrt{\text{trace}\{S_n^T S_n\} \text{trace}\{S_m^T S_m\}}}. \quad (2)$$

It reflects the agreement between two participants as a function of the number of sounds put together by both participants. It was computed on each pair of participants (see Ref. 40 for more details). This gave a  $30 \times 30$  matrix of  $Rv$  coefficients. A principal component analysis on this matrix was used to analyze the agreement.

One participant was specifically isolated from the group. This participant established three classes labeled as *kitchen*, *toilet*, and *bathroom*. Thus, he used contextual similarity to perform the classification, contrary to the instructions given which asked them to perform a classification based on the physical event that caused the sound. Therefore, the data obtained from this participant were excluded from the analyses.

The individual partitions were averaged to form a co-occurrence matrix. This matrix reflects the number of times that two sounds were grouped together by the participants. We performed on it an agglomerative hierarchical cluster analysis. In this algorithm, we chose the distance between *the furthest*, *the shortest*, *the median*, *the Weighted average distance (WPGMA)*, and the *unweighted average distance (UPGMA)*, using `MATLAB`.<sup>41</sup> To determine the best distance, we examined their resulting cophenetic correlation coefficients.<sup>42</sup> We chose the UPGMA method, which gave the best cophenetic correlation coefficient (0.86), and computed a dendrogram from this distance.

The classical methods used to identify classes consist in using a threshold on the distance of the dendrogram. We preferred to use the more relevant consistency coefficient.<sup>43</sup> This coefficient is computed as a ratio between the height of a link in a cluster hierarchy and the average height of links below it. This method provides compact cluster cleavages at different distances.

We chose two thresholds on the consistency coefficient in order to obtain different granularities: a general level of 3 classes (A, B, C) and a specific level of 8 classes (A, B1–B6, C). Figure 3 illustrates these two levels of clustering and a list of sounds is given in Table I. The next step was the identification of semantic labels for these classes.

As in Ref. 3, we performed a lexical analysis to interpret the clusters. The `TXM` (Ref. 44) software was used to analyze the terms employed by the participants in their descriptions.

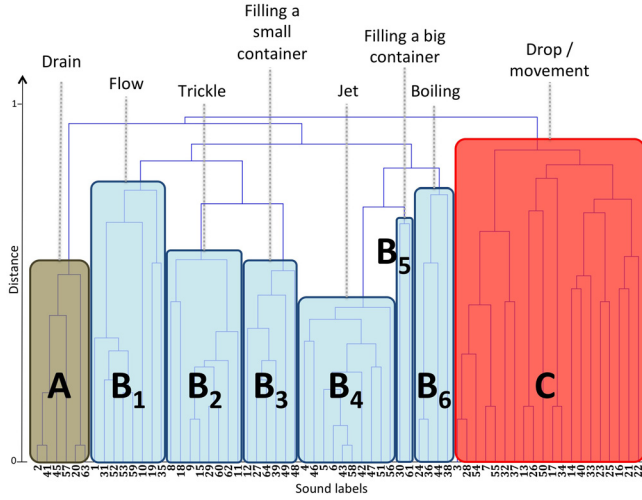


FIG. 3. (Color online) Dendrogram of liquid sounds computed from the unweighted average distance. The classes were identified from the consistency coefficient with two levels of clustering (level 1: A, B, C; level 2: A, B1-B6, C). The class labels stem from the lexical analysis. The sound labels refer to the list of sounds in Table I.

TABLE I. The following table presents the sounds used in the experiments. Each sound is identified by a number  $n$  that was used in the dendrograms (for example in Fig. 3). A cross in the column *Ident* means that the sound was misidentified and removed from the corpus before the classification task (see Sec. III B 5). The last column presents the original index *Ind* of the sounds.

$n$	Label	Duration	<i>Ident</i>	<i>Ind</i>
1	Filling up kitchen sink	4.99		# 1
2	Water gurgling	2.67		# 2
3	Water dripping into kitchen sink	1.78		# 3
4	Moving head-held shower head	4.14		# 4
5	Water running	4.1		# 5
6	Pressure tap	2.25		# 6
7	Single drip in empty Styrofoam	1.32		# 7
8	Watering Can to filled Styrofoam cup	3.71		# 8
9	Pouring corn starch solution in goo	4.54		# 9
10	Pouring into large anti-echo container	2.9		#10
11	Pouring from watering can into empty Tupperware	3.56		#11
12	Watering can into empty vase	3.72		#12
13	Slosh in a two liter plastic bottle	2.02		#13
14	Slosh in a glass vase	2.23		#14
	Slosh milk jug	1.01	X	#15
	Slosh wood balls in milk jug	3.05	X	#16
	Slosh in Styrofoam cup	2.03	X	#17
	Splash cardboard	0.39	X	#18
	Fingers in goo	1.45	X	#19
	Splash metal	1.65		#20
15	Splash metal	2.93		#21
16	Fingers in water	0.63		#22
	Splash rigid plastic	0.82	X	#23
17	Squeezing sponge over full Tupperware	3.96		#24
	Squeezing sponge	1.6	X	#25
18	Teapot being filled with boiling	5.04		#26
19	Frying-an-egg	2.51		#27
20	Toilet flushed	4.49		#28
21	Someone having a bath	2.64		#29

TABLE I. (Continued)

$n$	Label	Duration	<i>Ident</i>	<i>Ind</i>
22	Someone having a bath	1.19		#30
23	Bubble pop	0.99		#31
24	Various bubbles	4.08		#32
25	Bubbles	2.55		#33
26	Slosh bottle	3.55		#34
	Glue splat	1.38	X	#35
	Sloshy hit	1.26	X	#36
27	Filling bottle	3.85		#37
28	Dripping	5.43		#38
29	Dripping	3.21		#39
30	Filling	5.25		#40
31	Water tap flowing	3.09		#41
	Drips, splatty, fall onto leaves	2.53	X	#42
32	Drip, single, falls into liquid	1.2		#43
33	Water plops in metal bucket	4.86		#44
34	Drips, water, from wringing out wet cloth	1.69		#45
35	Rain, medium, splatty, fall on asphalt from roof of house	1.96		#46
36	Bubbles, cauldron of liquid	4.56		#47
37	Ice dropping into glass of water	2.55		#48
38	Soda pouring into empty glass with fizz	2.87		#49
39	Pouring liquid	2.21		#50
40	Multiple water cooler glugs	2.81		#51
41	Water going down drain	5.01		#52
42	Shower running	5.12		#53
	Faucet, run water into sink	5.01	X	#54
	French fries dropped into hot grease and sizzle	1.85	X	#55
43	Hamburgers frying on grill	3		#56
44	Water boiling in a pot	4.26		#57
45	Kitchen sink draining with lots of gurgling	4.73		#58
46	Water on running through shower head...			
	...with a constant gurgling drain	4.13		#59
47	Large sink filling with water	3.65		#60
	Pouring ketchup from bottle, heavy glugs	3.53	X	#61
	Trigger spray bottle	1.38	X	#62
48	Spray into glass	1.48		#63
49	Beer poured into glass and fizzing	4.69		#64
50	Pour liquid into glass with ice	2.66		#65
51	Fill bathtub	3.4		#66
52	Water running slow, tub partly filled	4.3		#67
53	Water running fast, tub partly filled	2.12		#68
54	Water dripping into full tub	3.71		#69
55	Water dripping into metal sink	4.89		#70
56	Water running in porcelain sink	3.68		#71
57	Drain porcelain sink	4.84		#72
58	Water spray in full tub	4.2		#73
59	Water run, large tub partly filled	4.51		#74
60	Sink flow	2.69		#75
61	Filling pouring	4.45		#76
62	Water flow	2.93		#77
63	Water flush	3.87		#78
	Water on hot plate	1.34	X	#79
64	Bottle of wine	4.95		#80
	Fizzing bottle opening	1.08	X	#81



For each sound, we merged the descriptions made by each participant to describe the group containing the sound and obtained a set of descriptions for each sound. They were aggregated according to the two levels of clustering. For each class, we identified the terms with the most occurrences and calculated their specificity<sup>45</sup> with respect to the total set of terms. Specificity is high if the word occurrences are abundant within a specific class, in comparison with other classes. The three most specific terms used can be seen in Table II. In this table, we omitted the words shared by all classes (water, noise, sound, liquid). For each class, the table presents the translated terms and the original terms in French, as well as their number of occurrences and their specificity. From the set of the most specific terms in each class, we proposed the generic labels shown in Fig. 3: *drain, flow, trickle on a surface, filling a small container, jet, filling a big container, boiling, and drop/movement*.

## IV. DISCUSSION

### A. Comparison with sound labels

The classes obtained and their labels were considered in terms of relevance. Cluster A is the most compact class of the dendrogram. Water in pipe sounds seems well identified and forms a category apart from all the other sounds. Cluster B shows more variety with six sub-classes. Among them, sub-class  $B_5$  is made up solely of two sounds merged at a long distance. Therefore, we do not consider  $B_5$  as a relevant category in what follows. Cluster C is the largest with 20 sounds. It is also the least compact due to long distance merging. Despite this, this cluster does not contain obvious sub-classes and therefore exhibits more variety and less consensus.

We made a comparison between the original labels of the stimuli and the descriptions of the classes. It shows a strong correlation between the terms. However, the  $B_4$  subclass most

likely contains a misidentified stimulus: *Hamburgers frying on grill*. Listening revealed a close acoustic resemblance with the other stimuli of the class that correspond to a *jet*. Both are close to a colored noise. This highlights a failure in the process of removing misidentified sounds.

Overall, the clusters obtained were quite compact and associated with well-defined labels. Therefore, our experiments provided a classification of liquid sounds with the following clusters: *drain, flow, trickle on a surface, filling a small container, filling a big container, jet, boiling, and drop/movement* (see Fig. 3).

### B. Interpretation

Data from the sorting task experiment were analyzed to reveal several categories of liquid sounds. Based on a quantitative lexical analysis of participants' descriptions, a label for each category was obtained. As expected, the labels emphasized the physical characteristics related to each cluster. These characteristics are discussed in the following.

#### 1. Temporality

Listening analysis revealed different temporal properties for the main clusters. The A and B classes were composed of sounds slightly longer on average than those of C class (3.2 versus 2.6 s). Moreover, the related sound event of the A and B classes appeared to be extendable over time. For example, the duration of the boiling sounds was determined by the editing step, but the related sound events were generally longer than a few seconds. Finally, the sounds from these two classes presented a fairly stable envelope (for instance, in the *flow* sub-class).

In turn, the duration of the sound events from the C class were generally short, although they could be repeated. This class was mainly composed of isolated water drops, single air bubbles, short actions (squeezing a sponge), or repetitive

TABLE II. The most specific verbalization for each class (A, B, C) and sub-class ( $B_1$ – $B_6$ ). For each class the original term is added to the English translation. “Occ” presents its number of occurrences in the class and “Spec” its specificity.

Class A			Class B			Class C		
Label	Occ	Spec	Label	Occ	Spec	Label	Occ	Spec
flush ( <i>chasse d'eau</i> )	75	57.9	shower ( <i>douche</i> )	273	24.7	drop ( <i>goutte</i> )	145	18.5
suction ( <i>aspirer</i> )	41	37.4	jet ( <i>jet</i> )	108	18.4	object ( <i>objet</i> )	32	16.0
drain away ( <i>evacuer</i> )	35	25.0	strong ( <i>fort</i> )	97	15.3	wring ( <i>essorer</i> )	33	14.1
Class $B_1$			Class $B_2$			Class $B_3$		
Label	Occ	Spec	Label	Occ	Spec	Label	Occ	Spec
bath ( <i>baaignoire</i> )	31	6.2	urinate ( <i>uriner</i> )	24	14.2	glass ( <i>verre</i> )	66	30.8
open ( <i>ouvert</i> )	21	5.3	trickle ( <i>filet</i> )	32	8.3	pour ( <i>verser</i> )	60	20.2
sink ( <i>lavabo</i> )	20	5.3	surface ( <i>surface</i> )	46	8.0	container ( <i>réceptient</i> )	48	16.3
Class $B_4$			Class $B_5$			Class $B_6$		
Label	Occ	Spec	Label	Occ	Spec	Label	Occ	Spec
shower ( <i>doucher</i> )	178	60.5	powerful ( <i>puissant</i> )	4	3.3	boiling ( <i>ébullition</i> )	20	18.4
strong ( <i>fort</i> )	53	16.0	jet ( <i>jet</i> )	10	3.1	bubble ( <i>bulle</i> )	19	10.5
rain ( <i>pluie</i> )	46	12.3	clean ( <i>nettoyer</i> )	2	2.4	boil ( <i>bouillir</i> )	16	7.5

actions (sloshing a bottle). The envelope of these sounds showed considerable variations of amplitude that could be seen as rhythmic patterns.

Previous work on solid sounds showed two main categories that differ according to the type of interaction (discrete or continuous) involved in sound production.<sup>3</sup> This distinction also seems relevant in the case of liquid sounds if we consider the A and B classes on the one hand and the C class on the other hand. Thus the A and B classes contained *continuous* sounds (for example, water flow) while the C class included *discrete* sounds (for example, multiple water drops).

## 2. Other kinds of similarities

**a. Size.** Apart from temporal variation, the size of the objects involved in the interaction may have a significant effect on the classification. Within a given shape and material, a bigger object generally produces a lower pitched sound. The size of the container seems to be relevant in the classification because the two sounds produced by a big container were relatively isolated from the other clusters. In addition, we observed parallelism between the pairs of sub-classes *Trickle/Filling a small container* and *Jet/Filling a big container*. Therefore, size and rate could both affect the global perception of these sounds.

Moreover, the size of objects may be considered through the size of air-bubbles. Hence, throwing a stone into water produces a deeper sound than a water drop because the resulting bubbles are bigger. Unfortunately, this kind of sound does not appear in our corpus limited to sounds related to home activities.

**b. Physical surface.** Many of the sounds of the corpus represent water falling either on a liquid or on a solid surface. Sub-classes  $B_2$  and  $B_4$  contained only sounds of water falling on a solid surface. The term *ground* was frequently used in the description. In sub-class  $B_1$ , by contrast, listening

to stimuli 1, 31, 52, 53, and 59, which were merged at short distance, revealed water flowing into water. Surprisingly, the concept of water falling into water does not appear in the description. This can be explained by a detailed individual analysis of the verbalization. On the one hand, the formulations were various: “*water flowing into water*,” “*in a container already filled with water*,” “*in a liquid*.” On the other hand, the terms used to express this cause (*water*, *liquid*) were not specific to this class.

Moreover, the cause of the sound led to more verbalization when it resulted from an interaction between different materials (e.g., liquid and solid). For instance, we could speak about a “water drop” sound, when a water drop falls into water. In the case of water falling onto a solid surface, the material involved was more verbalized. We could use, for example, the description “filling a plastic cup.” To overcome this lack of description, we have renamed class  $B_1$  as *Flow into water* in the following.

In the contrary, in the class C, different sounds were merged whether they corresponded to water drops falling onto a surface, or bubble sounds. Therefore, we assumed that the discrete/continuous property was predominant on the physical surface involved.

## C. Overview

The Fig. 4 proposes a synthetic taxonomy that combine the obtained results on liquid sounds with a previous study on solid sounds.<sup>3</sup> The branch *Liquids* summarizes our observations. The label of the  $B_1$  sub-class is now *Flow into water*. Moreover, the two sub-classes  $B_3$  and  $B_5$  were merged in one class labeled *Filling*. Overall, the properties that shape this taxonomy by physical cause are the rate of interaction and the material involved. This taxonomy is organized by discrete and continuous interactions, respectively from left to right. In the sub-classes of B, it can be seen that some sound categories contain discrete elements (*Flow into water*,

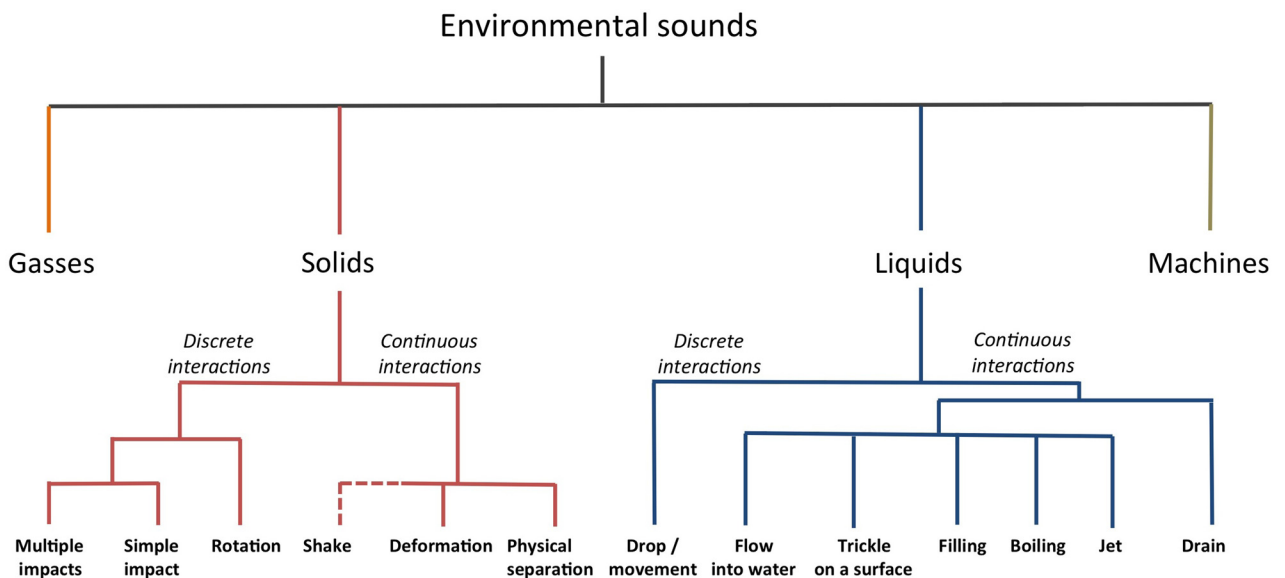


FIG. 4. (Color online) Taxonomy of environmental sounds which incorporates previous results on environmental sounds (Ref. 3), and the present work on liquid sounds.

*Trickle on a surface*) while others are rather continuous (*Jet*). Other sub-classes, such as *Filling* and *Boiling* do not fall into this scale, and possibly include various rates.

Finally, the temporal aspect of environmental sounds appears to be decisive for their categorization, as showed by the global results on liquid and solid sounds. We assume that the envelope of these sounds leads listeners to determine the type of interaction (discrete or continuous). The material involved in the interaction seems to appear at a second level of the categorization. By contrast with the temporal shape, it can be identified by the spectral content of the sounds.

The distinction made by Gaver<sup>23</sup> between liquid and hybrid interactions can be observed to a lesser extent. Sounds containing liquid and solid sources appear in the following classes: *Drop/movement*, *Trickle on a surface*, *Filling*, *Jet*, and *Drain*. Apart from the *Drop/movement* class which contains various elements, all the sounds of these classes are partially produced by solid interactions. Unlike Gaver’s taxonomy, in which the hybrid class is considered at the same level as the liquid and solid classes, the hybrid categories are sub-classes of the larger liquid class in the present taxonomy.

A specific case remains, that of the *Drain* class which is characterized by a considerable solid component with resonating pipes. We assume that the continuous excitation of these pipes, as well as their specific frequencies of resonance as solid objects, made this class very specific. Moreover, we observed that this type of sound was particularly related to home activities.

## V. ACOUSTIC ANALYSIS

An acoustic analysis was performed on the corpus in order to investigate the acoustic properties of each class of the taxonomy. We used the MIR toolbox [MIRtoolbox 1.6.1 (Ref. 46)] to compute different audio descriptors on each stimuli of the corpus: basic features from the temporal and spectral domain (*duration*, *zero crossing rate*, *spectral centroid*, *spread*, *skewness*, *spectral flux*, *mfcc*, *brightness*), and higher level features on dynamics (*low energy*) and rhythm (*event density*). Inter-stimuli and intra-stimuli statistics (*mean*, *standard deviation*, *minimum*, and *maximum*) were also performed on these raw results. Through this process, we aimed at analyzing our stimuli and choosing the best set of audio descriptors that may be used to retrieve the classification made by the participants.

Table III shows several quantitative results through the following features: duration, spectral centroid, brightness,

spectral flux, and event density. It presents the average values and standard deviation for each class and the global results for all the sounds in the last column. The spectral centroid represents the weighted mean of the frequencies of the stimuli. Brightness (in percent) measures the amount of energy above a cut-off frequency (3 kHz).<sup>47</sup> Intra-stimuli variations were observed through the standard variation of the spectral centroid and through the spectral flux computed on successive frames (50 ms and half-overlapping). The spectral flux shows the average distance between the spectrum of successive frames. Finally, the event density estimates the average frequency of audio events from the average number of note onsets per second. The envelope of the signal was first computed by a low pass filter (infinite impulse response filter) with a low time constant value ( $\tau = 1$  ms) that preserves smooth variations of the signal. Then onset positions are chosen as local maxima.

### A. Timbre and material

Liquid produces high-pitched sounds, as indicated by the values of the spectral centroid. Similarly, brightness is very high on average with 57% of the energy above 3 kHz. These values could be influenced by vibrations of solid elements that produce lower pitched sounds. Indeed, the stimuli of our corpus involves a wide and diverse range of materials. For instance, the labels of the class C contain different shapes and materials (*metal bucket*, *rigid plastic*, *large container*). Consequently, the brightness is rather lower in the class A that is characterized by pipe vibrations. Higher brightness values can be observed in classes  $B_1$  and  $B_4$ , characterized by almost no solid vibration.

The intra-stimuli timbre variations over time are generally significant, as shown by the standard deviation of the centroid, which is globally higher than 1 kHz. The spectral flux appears to be correlated with the amount of liquid in movement. Hence, it is particularly high for classes  $B_4$  and  $B_5$ , implying fast and turbulent water movements. It is lower for smaller displacement, in particular for classes  $B_2$ ,  $B_3$ ,  $B_6$ , and C.

### B. Air bubbles

In the framework of acoustics, earlier studies pointed out that liquid sounds mainly come from air-bubble entrainment (see Sec. II A). Therefore, the acoustic characteristics of the sounds may be analyzed through air-bubble vibrations, in particular, according to the number of bubbles. For

TABLE III. Audio descriptors computed on each class of the taxonomy. The table shows the average value and standard deviation of different audio features: duration (s), spectral centroid (kHz), brightness (%), centroid std. (intra-stimuli standard variation of the centroid, kHz), spectral flux, and event density ( $s^{-1}$ ).

	A Drain	B1 Flow	B2 Trickle	B3 Filling (small)	B4 Jet	B5 Filling (big)	B6 Boiling	C Drop/mvt	Global
Duration	$4.3 \pm 0.9$	$3.3 \pm 1.2$	$3.4 \pm 1$	$3.4 \pm 1.4$	$3.8 \pm 0.8$	$4.9 \pm 0.6$	$3.4 \pm 0.7$	$3.9 \pm 1.4$	$3.4 \pm 1.2$
Centroid	$3.8 \pm 1.9$	$7 \pm 1.4$	$6.4 \pm 0.8$	$5.5 \pm 1.9$	$7.7 \pm 1$	$6.2 \pm 2.5$	$4.2 \pm 4$	$4.6 \pm 1.8$	$5.6 \pm 2.1$
Brightness	$38 \pm 23$	$69 \pm 8$	$65 \pm 14$	$56 \pm 19$	$81 \pm 8$	$64 \pm 33$	$39 \pm 37$	$47 \pm 20$	$57 \pm 23$
Centroid std.	$0.8 \pm 0.4$	$0.8 \pm 0.3$	$1.5 \pm 0.4$	$1.5 \pm 0.4$	$0.3 \pm 0.2$	$0.4 \pm 0.2$	$0.4 \pm 0.1$	$1.7 \pm 0.7$	$1.1 \pm 0.7$
Spectral flux	$10.4 \pm 7.5$	$10.3 \pm 5.4$	$4.2 \pm 2.3$	$4 \pm 2.9$	$18 \pm 3.6$	$23.9 \pm 8.7$	$4.3 \pm 2.9$	$3.4 \pm 4.3$	$8 \pm 7.3$
Event density	$29 \pm 9$	$33 \pm 15$	$21 \pm 8$	$17 \pm 16$	$53 \pm 15$	$46 \pm 8$	$64 \pm 23$	$13 \pm 9$	$29 \pm 21$

example, a water flow produces a larger amount of air-bubbles than water drops. Therefore, in the context of sound synthesis, thousands of overlapping bubbles sounds are required to create complex liquid sounds such as waterfalls.<sup>13</sup> This large number of overlapping bubbles sounds can hardly be extracted from an acoustic analysis. However, we suppose that it could be correlated with low-level acoustic features.

In order to analyze this phenomenon, we used the event density. Based on the onset rate, this feature is supposed to estimate the average frequency of acoustic events, such as music notes. The onset detection depends on pick picking on the envelope of each audio files. We suppose that this feature could reflect the number of individual bubble sounds, in cases where the onset detection is sensitive enough for this kind of audio event. In this scope, we tested different values of the time constant  $\tau$  (from 0.002 ms to 1 s) to compute the envelope.

The results are presented here with  $\tau = 1$  ms that maximize the inter-classes separation. The event density of the audio files ranges from a few events to hundred events per second. We suppose these values can be seen as a proxy for the number of vibrating air bubbles. Event density is rather low for the *C* class and reach high values for classes  $B_4$ ,  $B_5$ , and  $B_6$ , which we assume to be characterized by a larger amount of bubbles. Finer differences of bubble rate can be observed in the sub-classes of B. For instance, the difference between sub-classes  $B_2$  and  $B_4$  is related to the rate of the *flow* which becomes a *jet*. The event density increases gradually within classes  $B_3$  (*filing a small container*),  $B_1$  (*flow*), and  $B_4$  (*jet*), which may reflect an increasing number of bubbles. By listening to the stimuli, we noticed that the sounds of subclass  $B_2$  are composed of a continuous noise (water flow) overlapping with discrete and identifiable elements (water drops). On the other hand, in subclass  $B_4$ , the flow masks the discrete events which are hardly audible.

As seen in Sec. II A, the vibration frequency of the bubble depends on its size. Hence, successive vibrations from bubbles of different sizes result in spectral centroid variations. By contrast, according to the law of large numbers, a

very large amount of bubbles implies a small temporal variation in the average size of the bubbles. Consequently, the variations of the spectral centroid are lower in classes  $B_4$ ,  $B_5$ , and  $B_6$ .

Other experiments confirmed the interest of the event density to characterize different classes of water sounds. For instance, with lower values of the time constant ( $\tau < 1$  ms), the event density increases until 700 events per second, but the relation between classes remains steady. However, with bigger values ( $\tau > 1$  ms), the event density takes account only of the main acoustic events, and the order of the classes changes.

### C. Overview

Finally, the classes of the stimuli distinguished themselves especially for the brightness that highlights different surface vibrations and the event density that we suppose to be correlated with air bubble vibrations. Figure 5 shows the distribution of the stimuli across these two features.

In this framework, *boiling* ( $B_6$ ) constitutes a unique class. Listening revealed that some sounds of this class are characterized by low-pitched sounds produced by large bubbles. This is reflected by the values of the centroid. Furthermore, in the light of acoustic studies, we assumed that the boiling class differed from the others by the absence of impact. Therefore, it was characterized by a low spectral flux in spite of a large number of bubbles.

In the end, the three main classes (A, B, C) of liquid sounds are not discriminated in the timbre space. On the contrary the sounds are distributed homogeneously according to the two dimensions (brightness, event density). Indeed, as it was requested, participants should have grouped sounds based on causal similarities (source or action), and not according to acoustical similarities (timbre).

## VI. CONCLUSION

In this paper, we reviewed works on liquid sounds from different scientific fields such as acoustics, perception, and signal processing. These works allowed us to obtain a cross-

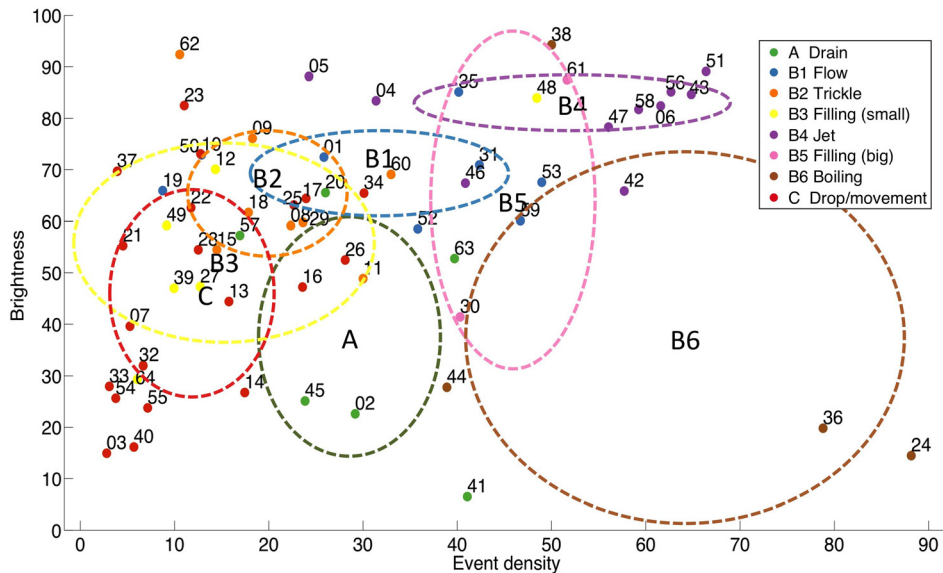


FIG. 5. (Color online) Distribution of stimulus classes according to event density and brightness. Event density reflects the number of air-bubbles involved in the sound. Brightness (percentage of energy above 3 kHz) is generally correlated with the presence of low-pitched solid vibration sounds, except for the *Boiling* class in which larger air-bubbles account for low-pitched sounds.



disciplinary vision of the water sound phenomenon. They also revealed a lack of knowledge in the domain of water sound perception and classification. To address this gap, we asked listeners to classify liquid sounds that can be heard at home during a behavioral experiment. The results led us to consider the following perceptive categories: *drop/movement*, *flow into water*, *trickle on a surface*, *filling*, *boiling*, *jet*, and *drain*. In addition to these classes, our taxonomy can be interpreted in terms of discrete and continuous interaction as was done for solid interaction sounds in Ref. 3.

Therefore, the results led us to observe structural and transformational invariants in the perception of environmental sounds,<sup>23</sup> an observation that was at least validated between solid and liquid sounds. For a given class of sounds such as solid or liquid, temporal aspects appear decisive for differentiating the type of interaction involved. Within the water sound category, these temporal aspects can be observed at a finer level, through the rate of air bubbles vibration. Other elements are also involved in the identification of the physical causes of these sounds, such as materials and air bubble size.

Based on the categories of liquid sounds, we performed an acoustic analysis. This analysis confirms the relevance of the air bubble vibration phenomenon that was analyzed by a rhythm feature, the event density. Other interesting results was obtained through the use of spectral features such as brightness.

Finally, our taxonomy could be used to formulate different sound detection algorithms in the context on liquid sound detection at home with different applications including healthcare<sup>6,7,10,48,49</sup> and waste detection.<sup>8,9,11,50–52</sup> It should allow future improvements in the development of specific approaches for water sound detection.<sup>53–55</sup> These results may also be extended to water sounds in natural environments which would lead to other applications, like sound-sccape annotation<sup>56</sup> and ecology.

## ACKNOWLEDGMENTS

This work was funded by a grant from the Agence Nationale de la Recherche under reference No. ANR-09-BLAN-0165-02, as part of the IMMED project, and by a grant from the Centre National de la Recherche Scientifique, Grant No. GDR 720 ISIS.

<sup>1</sup>S. Skoda, J. Steffens, and J. Becker-Schweitzer, "Road traffic noise annoyance in domestic environments can be reduced by water sounds," in *Proceedings of Forum Acusticum*, European Acoustics Association, Madrid, Spain (2014), pp. 7–12.

<sup>2</sup>B. Gygi, G. R. Kidd, and C. S. Watson, "Similarity and categorization of environmental sounds," *Percept. Psychophys.* **69**, 839–855 (2007).

<sup>3</sup>O. Houix, G. Lemaitre, N. Misdariis, P. Susini, and I. Urdapilleta, "A lexical analysis of environmental sound categories," *J. Exp. Psychol.: Appl.* **18**, 52–80 (2012).

<sup>4</sup>G. Lemaitre, O. Houix, N. Misdariis, and P. Susini, "Listener expertise and sound identification influence the categorization of environmental sounds," *J. Exp. Psychol.: Appl.* **16**, 16–32 (2010).

<sup>5</sup>M. N. Geffen, J. Gervain, J. F. Werker, and M. O. Magnasco, "Auditory perception of self-similarity in water sounds," *Front. Integrat. Neurosci.* **5**, 15–25 (2011).

<sup>6</sup>J. Chen, A. H. Kam, J. Zhang, N. Liu, and L. Shue, "Bathroom activity monitoring based on sound," in *Pervasive Computing* (Springer, New York, 2005), pp. 47–61.

<sup>7</sup>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, New York, 2006), pp. 91–100.

<sup>8</sup>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, Washington, DC, 2008), Vol. 6932, pp. 69322D–69322D-9.

<sup>9</sup>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, New York, 2009), pp. 26:1–26:6.

<sup>10</sup>B. Taati, J. Snoek, D. Giesbrecht, and A. Mihailidis, "Water flow detection in a handwashing task," in *Proceedings of the Canadian Conference on Computer and Robot Vision, CRV* (IEEE, Piscataway, NJ, 2010), pp. 175–182.

<sup>11</sup>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, Piscataway, NJ, 2011), Vol. 2, pp. 138–141.

<sup>12</sup>A. Parducci and D. H. Wedell, "The category effect with rating scales: Number of categories, number of stimuli, and method of presentation," *J. Exp. Psychol.: Human Percept. Perform.* **12**, 496–516 (1986).

<sup>13</sup>K. Van den Doel, "Physically based models for liquid sounds," *ACM Trans. Appl. Percept.* **2**, 534–546 (2005).

<sup>14</sup>C. Zheng and D. L. James, "Harmonic fluids," in *Proceedings of Transactions on Graphics, TOG* (ACM, New York, 2009), Vol. 28, pp. 37–48.

<sup>15</sup>W. Moss, H. Yeh, J. M. Hong, M. C. Lin, and D. Manocha, "Sounding liquids: Automatic sound synthesis from fluid simulation," *ACM Trans. Graphics* **28**, 21–33 (2010).

<sup>16</sup>Sir W. H. Bragg, *The World of Sound* (Bell, London, 1921), pp. 1–216.

<sup>17</sup>T. G. Leighton, *The Acoustic Bubble* (Academic, Cambridge, MA, 1997), Vol. 10, pp. 1–613.

<sup>18</sup>M. Minnaert, "On musical air-bubbles and the sounds of running water," *London Edinburgh Dublin Philos. Mag. J. Sci.* **16**, 235–248 (1933).

<sup>19</sup>"Sounding liquids: Automatic sound synthesis from fluid simulation," <http://youtu.be/MHBVinfmKo> (Last viewed April 5, 2017).

<sup>20</sup>G. J. Franz, "Splashes as sources of sound in liquids," *J. Acoust. Soc. Am.* **31**, 1080–1096 (1959).

<sup>21</sup>M. Strasberg, "Gas bubbles as sources of sound in liquids," *J. Acoust. Soc. Am.* **28**, 20–26 (1956).

<sup>22</sup>J. S. Lee, B. M. Weon, J. H. Je, and K. Fezzaa, "How does an air film evolve into a bubble during drop impact?," *Phys. Rev. Lett.* **109**, 204501 (2012).

<sup>23</sup>W. W. Gaver, "What in the world do we hear?: An ecological approach to auditory event perception," *Ecol. Psychol.* **5**, 1–29 (1993).

<sup>24</sup>P. A. Cabe and J. B. Pittenger, "Human sensitivity to acoustic information from vessel filling," *J. Exp. Psychol.: Human Percept. Perform.* **26**, 313–324 (2000).

<sup>25</sup>A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA, 1994), pp. 1–777.

<sup>26</sup>N. J. VanDerveer, "Ecological acoustics: Human perception of environmental sounds," Ph.D. thesis, Cornell University, Ithaca, NY (1979).

<sup>27</sup>M. M. Marcell, D. Borella, M. Greene, E. Kerr, and S. Rogers, "Confrontation naming of environmental sounds," *J. Clin. Exp. Neuropsychol.* **22**, 830–864 (2000).

<sup>28</sup>B. L. Giordano, J. McDonnell, and S. McAdams, "Hearing living symbols and nonliving icons: Category specificities in the cognitive processing of environmental sounds," *Brain Cogn.* **73**, 7–19 (2010).

<sup>29</sup>James A. Ballas, "Common factors in the identification of an assortment of brief everyday sounds," *J. Exp. Psychol.: Human Percept. Perform.* **19**, 250–267 (1993).

<sup>30</sup>"Wordnet, a large lexical database for english," <http://wordnet.princeton.edu/> (Last viewed April 5, 2017).

<sup>31</sup>"The auditory lab at carnegie mellon university," <http://www.psy.cmu.edu/~auditorylab/website/index/home.html> (Last viewed April 5, 2017).

<sup>32</sup>"Bbc sound effects library—original series," <http://www.sound-ideas.com/Product/152/BBC-Sound-Effects-Library-Original-Series> (Last viewed April 5, 2017).

- <sup>33</sup>“Blue box,” <http://www.bestservic.de>, Best Service GmbH, Munchen, Germany (Last viewed April 5, 2017).
- <sup>34</sup>“Hollywood edge premiere edition 1, 2 and 3,” <http://www.hollywoodedge.com/premiere-edition.html>, The Hollywood Edge, Hollywood, CA (Last viewed April 5, 2017).
- <sup>35</sup>“Sound ideas general series 6000,” <http://www.sound-ideas.com/Product/42/General-Series-6000>, Sound Ideas, Ontario, Canada (Last viewed April 5, 2017).
- <sup>36</sup>“Soundscan v2 vol. 61,” Ultimate Sound Bank, Paris, France.
- <sup>37</sup>P. Susini, O. Houix, and G. Saint Pierre, “The effect of loudness on the perceptual representation of sounds with similar timbre,” *Acta Acust.* **101**, 1174–1184 (2015).
- <sup>38</sup>P. Susini, O. Houix, L. Seropian, and G. Lemaitre, “How loudness affects everyday sounds recognition?,” *J. Acoust. Soc. Am.* **140**, 3267 (2016).
- <sup>39</sup>P. Gaillard, “Laissez-nous trier! tcl-labx et les tâches de catégorisation libre de sons” (“Let us sort it out! tcl-labx and the categorization tasks free of sounds”), in *Le sentir et le dire: Concepts et méthodes en psychologie et linguistique cognitives (Feeling and Saying: Concepts and Methods in Cognitive Psychology and Linguistics)* (Editions L’Harmattan, Paris, France, 2009), pp. 189–210.
- <sup>40</sup>H. Abdi, D. Valentin, S. Chollet, and C. Chrea, “Analyzing assessors and products in sorting tasks: Distatis, theory and applications,” *Food Qual. Pref.* **18**, 627–640 (2007).
- <sup>41</sup>MATLAB, *MATLAB and Statistics Toolbox, version 8.0 (R2012b)* (The MathWorks Inc., Natick, MA, 2012).
- <sup>42</sup>P. Legendre and L. F. J. Legendre, *Numerical Ecology*, 2nd ed. (Elsevier Science, Amsterdam, Netherlands, 1998), pp. 312–335.
- <sup>43</sup>A. Jain and R. Dubes, *Algorithms for Clustering Data* (Prentice Hall, Englewood Cliffs, NJ, 1988), pp. 1–320.
- <sup>44</sup>“Txm,” <http://perso.ens-lyon.fr/serge.heiden/txm/files/software/TXM/0.7.7/> (Last viewed April 5, 2017).
- <sup>45</sup>P. Lafon, *Dépouillements et Statistiques en Lexicométrie (Scores and Statistics in Lexicometry)* (Slatkine, Geneva, Switzerland, 1984), Vol. 24, pp. 1–218.
- <sup>46</sup>O. Lartillot and P. Toivainen, “A matlab toolbox for musical feature extraction from audio,” in *Proceedings of the 10th International Conference on Digital Audio Effects, DAFx* (2007), pp. 237–244.
- <sup>47</sup>P. N. Juslin, “Cue utilization in communication of emotion in music performance: Relating performance to perception,” *J. Exp. Psychol.: Human Percept. Perform.* **26**, 1797–1813 (2000).
- <sup>48</sup>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, New York, 2012) pp. 85–94.
- <sup>49</sup>Y. Gästel, 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,” in *Proceedings of the Alzheimer’s Association International Conference on Alzheimer’s Disease, AAICAD*, Vol. 7, Alzheimer’s & Dementia: The Journal of the Alzheimer’s Association (Elsevier Science, Amsterdam, the Netherlands, 2011), pp. S245–S245.
- <sup>50</sup>A. Ibarz, G. Bauer, R. Casas, A. Marco, and P. Lukowicz, “Design and evaluation of a sound based water flow measurement system,” in *Proceedings of the Third European Conference of Smart Sensing and Context, EuroSSC* (Springer, New York, 2008), pp. 41–54.
- <sup>51</sup>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* (ACM, New York, 2009), pp. 235–244.
- <sup>52</sup>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* (ACM, New York, 2010), pp. 169–172.
- <sup>53</sup>P. Guyot, J. Pinquier, and R. André-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, Piscataway, NJ, 2012), pp. 1–6.
- <sup>54</sup>P. Guyot, X. Valero, J. Pinquier, and F. Alías, “Two-step detection of water sound events for the diagnostic and monitoring of dementia,” in *Proceedings of the International Conference on Multimedia and Expo, ICME* (IEEE, Piscataway, NJ, 2013) pp. 1–6.
- <sup>55</sup>P. Guyot, J. Pinquier, and R. André-Obrecht, “Water sound recognition based on physical models,” in *Proceedings of the 38th International Conference on Acoustics, Speech, and Signal Processing, ICASSP* (IEEE, Piscataway, NJ, 2013) pp. 793–797.
- <sup>56</sup>D. Krijnders, M. E. Niessen, and T. C. Andringa, “Annotating soundscapes,” in *Proceedings of the International Congress on Noise Control Engineering, INTERNOISE* (Institute of Noise Control Engineering, Reston, VA, 2009), pp. 3613–3621.