TOWARDS A BETTER UNDERSTANDING OF MENTAL MODELS IMPLIED IN SONIC ICON DESIGN AND PERCEPTION

Tifanie Bouchara
Cnam-ENJMIN / CEDRIC-ILJ
2, rue Conté
Paris, 75003, France
tifanie.bouchara@lecnam.net

Mathieu Bouchet
Cnam-ENJMIN / CEDRIC-ILJ
2, rue Conté
Paris, 75003, France
mathieu.bouchet.auditeur@lecnam.net

Nicolas Misdariis
Ircam-STMS Lab / PDS
1, place I. Stravinsky
Paris, 75004, France
nicolas.misdariis@ircam.fr

ABSTRACT

This paper presents our ongoing efforts to determine if there are shared references, i.e. mental models, across designers and potential users of sonic icons in mobile applications. First, 13 sound designer students had to conceive sonic icons regarding 11 common mobile functionalities. Their conceptual models were analyzed through lexical analyses of their self-report on their design and manual annotations of their renderings. Second, the 143 obtained icons were evaluated by 52 naïve listeners through a free categorization task. While deeper analyses are still required, results already indicate that some function/sonic icon links make strong consensus on how they should be realized, considering both the designers' productions and the listeners' clustering. This forms one key result of the study and could lead to useful guidelines in sonic icon design.

1. INTRODUCTION

The present study addresses the global issue of Human-Machine Interface (HMI), especially focused on how to diminish the gap between designers' (conception) and users' (reception) mental models? The study stands into the sonification paradigm by exploring the role of sound in the frame of communication and/or interaction between a human being (a user) and a machine (a computer, or equivalent). It relies on a standard approach of *sonic iconography* that supplies non-visual representations of objects that are manipulated, or actions that are executed, in such situations. But, and besides, this study is also incorporated within a broader, and former, reflection concerning multicultural sonification. This reflection targets new models of sonification addressing issues faced by populations with limited access to new Information and Communication Technologies (ICT). This globally aims at opening the door to the digital world and its associated services (administrative forms, banking, social networks, etc.) to more users and communities, possibly including non-western cultures, or even including persons with special needs such as illiteracy or computer-illiteracy.

It is well known – especially in the ICAD community – that the concept of sonic iconography is initially, and historically, based on two seminal foundations : « *earcons* » [1] and « *auditory icons* » [2] – even if alternatives like « *spearcons* » [3], and recent developments like « *morphocons* » [4] exist in different research or applied forms. This being, in that broad reflection on multicultural sonification – and as a starting point of the present study –, we seek to overcome these traditional sonification paradigms or, at least, the way they are currently implemented, i.e. mostly on the basis of western-culture referential elements (music, technology, environment, etc.). For that, we work on the following hypotheses: existence of a shared s*ound syllabus* as a possible answer to the question of cultural invariants in sound perception, and consequently, possible development of a sonic iconography adapted to socio-cultural specificities and consistent with graphic and visual elements of the interfaces. In summary, the long-term target of this global reflection could be a sort of universal, or at least, culturally-informed way of designing *sonic icons*, in order to better fulfill either ergonomics or aesthetical requirements in new sonification paradigms and processes – involving, for instance, collaborative design protocols [5]. Kind of approaches that are rather similarly and progressively adopted for visual icons [6].

In fact, the present study addresses some preliminary, but fundamental, issues about this topic, that can be questioned as follows: what are the mental mechanisms that allow the understanding of a sonic icon? How does the link between a sonic icon and its associated function operate in the user's mind? How do visual components play a role in the congruent integration of the information? To undertake these investigations, we implemented a methodological and experimental framework that globally relies on two key elements: i/ a specific creation of a set of sonic icons, assumed in a pedagogical way within sound design cursus by nearly professional sound designers, that allows us to understand the design mechanisms underlying the sound production in that case; ii/ a large-scale experimental protocol with a corpus of naive listeners that allows us to understand the mental representation underlying the reception of such sound artefacts and their matching with given functionalities. This 2-fold implementation gave access to data that starts to highlight invariants in the production and reception processes. They are presented and discussed so as to open the way to further and extended investigations within the global reflection topic.

2. RELATED WORKS

2.1. Sonic Icons: taxonomy and use

Since the introduction of the auditory icon concept by Gaver (1986), lots of research have been done by ICAD community on auditory displays and signs, as they allow a quick non-visual transmission of information (see [6], [[8] for reviews).

Several types of non-speech auditory cues were defined, like *auditory icons* (nonmusical sounds that have some resemblance to the objects, functions, and events they are representing) [2], *earcons* (short synthesized musical sound phrases that can be parameterized) [1], *spearcons* (speeding-up speech) [3], *spindex* (speeded pronunciation of the first letter of each menu item) [9]. *Auditory emoticons* [10] are affective icons used to convey emotional content and therefore are very concerned with the question of aesthetics. *Morphocons* [4] also addressed the question of aesthetics by allowing some customization of earcons. Mustonen [11] proposed in 2008 to regroup all these types under the label *auditory sign*. Indeed, he noted that it could be more useful for the community to understand the complex nature of auditory signs in a more pragmatic way than to reinforce the categorization between them. However, the categorization, particularly between auditory icons and earcons is strongly anchored in the ICAD community. The semiotics behind the taxonomy of signs have been extensively studied since [12], and the relation between signal and referent (i.e. iconographic vs symbolic relation) in auditory display was addressed in several research papers ([13], [14]) but is beyond the scope of our study. Therefore, in the rest of this paper we use the term sonic icons to refer to all these non-speech auditory cues.

Former studies employed sonic icons in a lot of contexts, but mainly as a sensory substitution way to compensate for the lack of vision, e.g. giving access to blind users [15] or users situationally impaired because they are driving or already fully engaged in another visually-demanding task [16]. Sonic signs are also used as auditory warnings [17] or to compensate for the small screen space in a mobile situation through auditory menus or auditory scrollbars [18]. Therefore, in general, sonic icons are not supposed to be associated with and perceived simultaneously of the visual icons they complemented or replaced. While in the concept of video games sound designers actually map both sonic and graphical UI to the universe of the game – but with a very empirical approach, few studies explored the relationship between visual and sonic icons in the research community [19].

2.2. Mental representation in sonic icons: conceptual *vs.* mental models

In UI design, either traditional or auditory-based, it is now commonly accepted that a better design is user-centered and takes the mental representation of the users into account. This internal representation of the system is called the mental model and is formed by users' own learning and previous experience [20]. Cognitive models are therefore more and more developed in the ICAD community ([21], [22]).

The current study aims at determining whether there are now functionalities that are sonified often enough in current apps, so shared mental models are emerging in a common and robust way from one user to another. The presence of commonalities in mental models of sonic icons would guide their future design by establishing possible standards.

Finding commonalities in sonic design is a similar approach as the one from studies about cross-culturalism in sonic design, as they are also looking for shared interpretation of sounds, or auditory parameters, between users coming from culturally distant countries. For example, [23] compared the understanding of auditory alarms by participants coming from 6 different countries from 3 different continents and proved that the effect of perceived urgency frequencies is common in all these countries. Similarly, [10] proved that there are some commonalities in the perception of affective sounds, as people coming from USA and Korea mostly (over 80%) prefer the same sound for a certain affective state.

However, our approach differs as we are also interested by the designer's mental model, also called conceptual model. Indeed, one of the biggest dilemmas in human-computer interaction design is the common gap between designers' conceptual models and users' mental models. Mental models are only representations of what users know, or think they know, about a system. They rely on beliefs about the system at hand, not the reality of the system. At the opposite, designers know much more about the system, so they form brilliant mental models of their own creations, leading them to believe that each feature is easy to understand. Users' mental models of the UI are less informed than the conceptual models so are more likely to be more deficient, to lead users to make mistakes and find the design much more difficult to use.

To reduce the gap between mental models of users and conceptual models of designers, some studies proposed to use participatory design methodology (co-design) [24], i.e. to let the final users directly participate during conception steps and inform the conceptual model of the designers. That is not easy to develop for auditory display application due to inherent difficulties in sonic design and a lack of proper tools [25].

We suggest another approach by also observing more deeply the conceptual models from the designers. Interestingly very few auditory display studies have compared the conception methodology and the sonic icons obtained from several sound designers. For example, while [23] compared the perception of auditory alarm by users coming from different countries, they use sounds designed by only one designer. The unique study – to our knowledge – that is comparing the perception of designer and user concerned the listening of an in-vehicle auditory interface recorded in a ecological situation, listened by several naïve listeners and one designer [26]. In that study, the sound design was actually made by somebody else and no information about the conceptual models of the actual designer was analyzed. This study did not compare the sounds designed by several designers to represent the same functionalities.

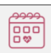
Finally our approach is similar to the one explored in [10] for auditory emoticons: asking several students in sound design programs to individually and concurrently design the sonic icons to represent the same set of functions. However, [10] was more interested in the user interpretation and therefore intentionally removed all similar sounds produced by different designers. On the contrary, the present study aims at first analyzing how much similar sonic designs will be obtained from the designers for the set of functionalities, and for what functionalities, before analyzing the rate of shared interpretation by more naïve listeners.

3. CONCEPTION OF A SET OF SONIC ICONS

3.1. Selection of visual icons and their associated functions

Due to the nature of the global reflection frame, and especially its consideration of low literacy or computer-illiteracy (illectronism), the search for icons and functions to sonify focused on website or applications concerning public services or equivalent. Four web-based applications for mobile devices were analyzed in terms of visual iconic contents: the French Health Insurance (ameli.fr), the French Governmental Tax Service (impots.gouv.fr), the MAAF Mutual Insurance Group (maaf.fr), and the French Employment Agency (pole-emploi.fr). For each of these services, all visual icons for each accessible views/pages were listed. Then a functionality label was associated to each visual icon: for some icons, that were accompanied with a legend, the text of the legend served as the associated label; for other icons, that were only used as graphical widgets, a text label was manually produced. This labeling work was done for each of the 4 application websites.

Then, on the basis of the whole set of icons/function pairs, we selected 12 functionalities commonly used by at least three fourths of the analyzed applications. Six functions corresponded with very general web- items or actions and were represented rather similarly among the websites: 'help', 'scan', 'disconnection', 'wrap/unwrap', 'backward', 'close'. Six other functions corresponded with quite specific web-items or actions: 'home', 'contact', 'my profile', 'my documents', 'calendar', 'professional directory'. Moreover, they appeared with some subtle variations among the websites, either in the function (e.g. 'calendar' was associated with a generic almanac or with a specific personal diary), or the icon itself (e.g., 'my documents' was represented as an open folder or as the iconic PDF file). Table 1 shows the 12-function list and their different visual representations.

SPECIFIC	
'home'	
'contact'	
'my profil'	
'my documents'	
'calendar'	
'professional directory'	

GENERAL	
'help'	
'scan'	
'disconnection'	
'wrap / unwrap'	
'backward'	
'close'	

Table 1: 12 functionalities selected from 4 public services web-app. Each function is linked to a label (translated from French) and, for a better understanding here, associated with one of the 4 token of each visual icon. 2 groups of functions are made out depending on their specific / generic natures.

3.2. Production of sonic icons attached to visual icons

3.2.1. Sound designers pool

A pool of 13 sound designers came from two sound design majors within two different high schools (Master level). All the students from the majors participated in this sonic icon production without any subsequent selection. Hence, seven students were studying in the School of Art and Design in Le Mans – ESAD TALM Le Mans – (M1). All of them were male, aged from 21 to 30 years old and came from rather different educational backgrounds: acoustics, design, architecture, fine arts, audio technologies or music studies. Six other students were studying in the National School of Video Games and Interactive Digital Media in Angoulême – ENJMIN Angoulême – (M1). All of them were also male[1], aged from 20 to 22 and, however, came from rather similar educational backgrounds: sound design or sound integration in video games. All the sound designers received course credits for their participation and their renderings were evaluated by their teachers independently of this study.

3.2.2. Sound design instructions

The sound designers carried out two different tasks during two subsequent experimental steps.

The first step aimed at assessing the identifiability of the selected visual icons (see Sec. 3.1) through an online questionnaire. The 28 visual icons gathered from the 4 applications with regards to 12 functionalities (see Table 1) were sequentially displayed through a web page in a random order; the same random order was used for all the participants. The sound designers were asked to freely label their interpretation of each icon by typing below it a maximum of two words (verbs or nouns).

The second step aimed at producing the set of sonic icons. The students were then informed of the 12 functionalities to sonify and were provided with the 4 associated visual icons such as presented in Table 1, without any other information about visual icons or functionalities. They were given instructions at the beginning of the exercise, specifying the context of the study, the design brief and the format of the deliverable. They were asked to render 12 sound files together with a 2-page report presenting "the statement of intent", i.e. their global inspirations, working methodology (type of tools / synthesis paradigm), conception and realization principles. No constraint about duration or type of sound was specified.

3.2.3. Available time for sound design

The time accorded for production was rather different for the 2 groups of sound designers. For the sound designers from ESAD Le Mans, the production of the sonic icons formed a part of the final exam of a course module dedicated to sound perception and design (psychoacoustic, auditory perception / cognition) and sound computing (basic physics, digital sound synthesis). They had 3 weeks to realize the whole set of sounds (homework). On the other hand, the sound designers from ENJMIN Angoulême realized the sonic icons in a 3-hour session (classroom work) with then more limited means and tools to realize the same exercise, but with the same requirement in terms of sound quality and explanation material. All students were asked to work alone, or at least to make an individual rendering – despite the fact that they may have shared ideas or methods during the process.

4. EXPERIMENTAL INVESTIGATION OF SOUND – FUNCTION RELATIONSHIP

4.1. Participants

52 participants (22 women) were recruited by the RISC – information network on cognitive sciences (risc.cnrs.fr). They were aged between 18 and 45 years old and they reported to have normal or corrected-to-normal vision and normal hearing. They gave consent prior to the experiment and were paid for their contribution. At the beginning of the test, they were given written instructions, the global context of the study was explained, and the synopsis of the test was detailed. Moreover, they had the support of the experimenter at any moment of the test to answer additional questions or deal with any technical / digital problem.

4.2. Stimuli

143 stimuli were used for the experiment; They correspond to 11 sonic icons – associated to 11 functions above 12 – produced by the 13-sound designer pool (see Sec. 3.1.1). Note that, after a first qualitative analysis of the whole sound production (see Sec. 5.1), the function 'wrap/unwrap' was finally decided to be removed from the test corpus as it has led

[1] All students were male as no female student attended those classes at that time. This highlights the necessity to train more female in sound design, at least in France, but this discussion is beyond the paper topic.

to singularities among the sound designers: this peculiar function seemed to be not well and commonly understood, some of them chose to specify it with 1 composite sound whereas some others chose to design 2 separate sounds. This situation may have led to difficulty and confusion for the implementation and the conduction of the listening test. The sound files taxonomy then observed the following rules: a letter from A to M to indicate the sound designer (the 6 ones from ENJMIN Angoulême corresponding to letters from A to F) and a number from 1 to 11 to indicate the sound/function in a given order: 1.–'home', 2.–'help', 3.–'professional directory', 4.–'calendar', 5.–'contact', 6._'disconnection', 7.–'my documents', 8.–'close', 9.–'my profile', 10.–' backward', 11.–'scan'. Sound samples can be accessed at this address: recherche.ircam.fr/equipes/design/ICAD20media/index.html

This being for sake of experimental feasibility, subsets of this 143-item corpus were formed by grouping the 11 sonic icons produced by 4 different sound designers. This partition mainly prevented from leading to participants' fatigue and demotivation, and more practically from saturating the experimental user interface. But this forced us to thoroughly fix these partitions among the participants (in a pseudo-random process), so as to be sure that each sound of the corpus was listened to the same number of times.

Apart from that, the technical characteristics of the corpus were as followed: non-compressed audio encoding format (wav), duration from 0.52 to 1.63 sec., 44.1 kHz sampling frequency and 16-bit resolution. The sounds of the corpus were all normalized at 70 dB RMS (+/– 2 dB).

4.3. Apparatus

The experiment was conducted in a professional audiobooth (IAC) equipped, for this occasion, with the following audio chain components: a Mac Pro (3,7 GHz Quad-Core Intel Xeon E5 / OS X v.10.11.5), a RME Fireface 800 audio interface for the D/A conversion and for the amplification (headphones output), and professional headphones (Sennheiser HD650).

The experimental user interface was developed and run on the TCL-LabX environment (v.3.13) [27]. This software was chosen for this open-access nature and, more deeply, for being adequate to the experimental protocol (see Sec. 4.4) we chose to implement this listening test. In fact, its graphical interface assumed the representation of the sound corpus items (Figure 1), the management of all participant's actions (playback / move / select items, enter text– see Sec. 4.4 for further details), and the data backup in a conventional textual format. The user interface was rendered in a 21.5" monitor (AMD FirePro D300) with a resolution of 1920x1080 pixels.

4.4. Procedure and tasks

Following a similar approach as in [28], the experiment was composed of two consecutive steps.

The first step consisted in a free categorization approach where a participant was presented with a subset of 44 sounds (i.e. 11 sonic icons from 4 designers, see Sec. 4.2) that she/he had to group into categories. The sounds within the subset were initially randomly labeled with a number from 1 to 44, and were then displayed aligned in the TCL-LabX Graphical User Interface (Figure 1). Information about the nature of the categories were given to the participants in the instructions page: they had to represent different functionalities (of the use of a smartphone, a computer, etc.) for which some generic examples were given (call, send a message, etc.). In the GUI, participants could listen to each sound as often as needed during that step, by simply double clicking on the graphical box associated with the sound. Participants had then to categorize the sounds by dragging each box in the space of the screen so as to form the desired categories (Figure 1). The number of obtained category and the number of sonic items per category were free. This step ended once the participant was satisfied with her/his categorization and validated the structural and spatial arrangement of the interface.

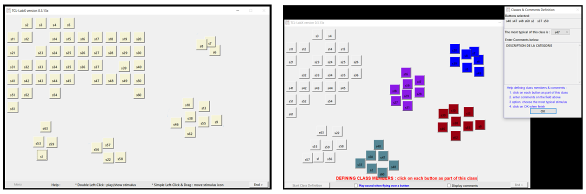


Figure 1: Depiction of the TCL-LabX GUI. (left) Free categorization (1st step): the sonic icons of the corpus appear as boxes in the interface that the participant has to move in the screen space to form the perceptual categories. (right) free verbalization (2nd step): each group of sonic icons is selected and labelled and a prototype sound per category is chosen.

The second step consisted in a free verbalization approach where the participants were asked to enter information about each category they made in the previous step. This information mainly consisted in a brief comment (one word or expression) for each class and the selection, within it, of the sample that seemed be the most representative of the group – namely the *prototype* (Figure 1). This step ended once the participants informed all the categories they made and validated their labels and prototypes in the current interface.

For sake of equity in the experimental design (see explanations in Sec. 4.2), the experiment was divided into two independent blocks (each block being composed with both categorization and labeling steps). In each block, the participant was presented with 11 sounds from 4 random sound designers among the total of 13. The 4 designers were different from one block to another, meaning that each participant was presented with 8 designers' sound productions in total. The random selection of the 4 designers for each block was counterbalanced across the participants so as, at the end of the experiment, each designer appeared 32 times.

A short break of 10 minutes was allowed between the two blocks to avoid auditory fatigue. During that break the participant had to fill out a short questionnaire to collect basic personal data (age, gender, native country, musical practice). The entire experiment took 60 to 90 minutes per participant.

5. RESULTS

5.1. Qualitative analysis of the sound design strategies

5.1.1. Analysis of the produced sounds

The 12-sound corpus initially produced by the 13 designers (156 items) was considered regarding two points of view.

Firstly, we put this production in the frame of the formalized approach of auditory display by trying to sort these sonic icons in two main and referenced classes: « *earcons* » and « *auditory icons* ». The results of this binary classification showed interesting trends in the functions—designers representation space. Regarding functions, the general category ('help', 'scan', 'disconnection', 'wrap/unwrap', 'backward,' 'close') was mainly treated with earcons (70% among the 78 items), whereas the specific category ('home', 'contact', 'my profile', 'my documents', 'calendar', 'professional directory') was similarly treated either with earcons or auditory icons (roughly 45% each). In that frame, it is worth noticing that some functions were rather unanimously treated with auditory icons by the whole pool of sound designers ('contact', 'my documents', 'scan') whereas some other functions were rather unanimously treated with earcons ('home', 'my profile', 'close', 'backward', 'disconnection', 'help'). Moreover, it is also interesting to notice that one sound designer adopted a kind of hybrid approach by mixing the two fundamental concepts in the same sonic icons [29]. Finally, despite obvious specificities in the cursus, and the current conditions of the design sessions, no significant differences were observed with regards to the scholar origins of the sound designers – i. e. the previous proportions were rather the same in the two sub-groups of the pool (ESAD Le Mans or ENJMIN Angoulême).

Secondly, we tried to have a phenomenological description of these sound productions by applying to it a reduced listening and description process, respectively on the basis of Schaeffer's approach [30] and Carron et al. recent works [31]. By means of a – work-in-progress – sound lexicon able to give standard definition and illustrations of reduced sound properties, each sonic icon was annotated with words related to either general (e.g. dynamics), morphological (e.g. ascending/descending) or timbral (e.g. warm, dull) dimensions. This approach resulted in an expert acoustic description of the sonic icons attached to the 12 studied functions. It potentially led to extracting invariants such as: presence of roughness for the function 'my 'documents', or roundness for the function 'help' – even if these analysis elements was made in a rather explorative way and still stands globally difficult to interpret (see Sec. 6 for discussions).

5.1.2. Analysis of the reported intentions

Apart from the sound material themselves produced during the first step of this study (see Sec. 3), there have been a lot of semantical and methodological information that were reported by the sound designers about their intent, inspirations, framework, tools, etc. Even if no exhaustive lexical analysis has yet been made on this corpus, some elements of understanding were preliminary extracted from the 2/3-page reports quite uniformly written by each sound designers of the pool. Thus, we learnt from this observation – to be continued (see Sec. 6 for discussions) – that, for instance, a large panel of synthesis techniques were used (wavetable/sampling, additive/subtractive, frequency modulation, physical modeling), the search of identity or homogeneity among the sonic icons was one of the main constraints self-imposed by the sound designers themselves (even if no specifications had been made on that point in the design brief) and that the shared references that inspired some of the creative ideas were taken from the sonic ecosystem of current social networks (Facebook, Instagram, etc.) or famous communication applications (Skype, Messenger, etc.). Moreover, focusing on the iconography itself, some icons/functions seemed to have made consensus on the rendering approach: for instance, 'my documents' was rather systematically based upon the paper metaphor, using either the swishing of book leaves, the friction on a sheet of paper, or the noise of a pen writing. On another hand, the sonification scheme of the 'close' function seemed to have been built with short and dull sounds or, more abstractly, with descending melodies – both of them trying to explicitly suggest the closing of a door.

5.2. Quantitative analysis of the perceptual categorization

5.2.1. Data formatting

Thanks to the usability and versatility of the TCL-LabX software, the raw data output from each participant's experiment took the form of 4 files:

i/ a list-file remaining the sound files used in the test corpus and their correspondences with the graphical boxes used in the user interface (see Figure 1 above);

ii/ a log-file listing some system information (version, paths, etc.), plus all the actions the participant made on each of these graphical boxes during the test (so that one can rebuild the organizational scheme that led to the final categorical representation);

iii/ a class-file describing the constitution of the sound categories (experimental step 1);

iv/ a text-file containing, for each class, the participant's comments and prototype selection (experimental step 2).

At the current step of analysis, we just have looked into the categorization data corresponding to the class-file. This being the specificity of this protocol was included in its experimental design that can be considered as incomplete. In fact, as already mentioned (Sec. 4.4) each participant listened to a sub-group of the whole corpus made of the 11-sound production from 2x 4 sound designers (block 1 and 2). This led to 2 partial categorization sub-matrices (44 x 44) coding the fact that 2 sounds (i and j) were put in the same class by a participant – boolean value in the {i,j} cell of the matrix. Each of these individual sub-matrices had then to be replaced in a global dissimilarity matrix (143 x 143) by layering all the sub-matrices (2x52= 104 matrices) which overlapped insofar as each matrix cell {i,j} (corresponding to a sound) had the same number of layers, actually 32 – i.e. the number of times each sound was listened to during the whole experiment, along the 52 participants. In other words (and for sake of clarity), if M participants would have put into the same class (or cluster) the "function k of designer i" and "function l of designer j", then the value of the cell (11*i+k, 11*j+l) is M. Furthermore, the matrix is symmetric.

The next and last step of formatting consisted in the process of a global co-occurrence matrix from the global dissimilarity matrix by averaging it in its third dimension – i.e. within the layers corresponding to participants and participants' blocks. This being (and after some verification procedures), the initial data were correctly formatted to enter into a type of analysis that aimed at pointing out the distance relationships between all the items of the matrix, i.e. the different sonic icons produced for each function by several sound designers.

5.2.2. Data Analysis

Two kinds of conventional data analysis were then undertaken, on the basis of the formatted data (Sec. 5.1.1): i/ a cluster analysis with the aim of having an overview of the categorial structure underlying the data, and furthermore, the mental representation associated to this corpus of sonic icons; ii/ a multidimensional scaling (MDS) analysis with the aim of having a more precise view on some local grouping – despite a rather high dimensionality of the data (see details below).

The cluster analysis is performed by means of hierarchical cluster classification algorithm, using the average method, and its result is depicted in a conventional dendrogram representation (Figure 2) that depicts the level of categorization (vertical position in the tree) between each of the corpus items (the lower a node – and the shorter the path – is between two items, the stronger these items belong to the same category). For a better reading of this graphical output, a demarcation threshold is chosen (by expertise) at 0.2, that corresponds to a compromise between a reasonable number of categories and a rather high level of information. We can first observe that, this idiosyncratic threshold makes emerging the same number of categories as the number of studied functions, but with very variable numbers within each of them (min = 3, max = 42). A partial zoom in the dendrogram (Figure 3) gives several deeper insights. Some categories gather items associated to same functions (e.g. 6, 7 or 11) produced by different designers either from ESAD Le Mans or ENJMIN Angouleme (respectively, {E, L, K}, all except {F, G, J}, and {A, I, H, K D, G}). Some other categories strongly mix different functions and designers. There are also some pairs of items that are placed at the lowest level of the tree meaning that they were always associated together, be it that they correspond to the same functions (e.g. 9, 10 or 7 and 11 - 2 times each) or they differ in functionality ({1,9}, {3,6}, {8,5}, {6,4}, {2,10},{9,10}, {9,10}, {3,4}, {7,4}).

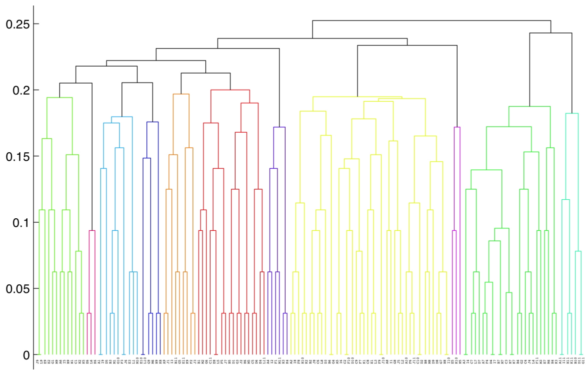


Figure 2: Dendrogram given by the cluster analysis of the formatted data resulting from the free categorization experiment (Sec. 5.2.1). The demarcation threshold is set to 0.2 and lets appear 11 distinct categories.

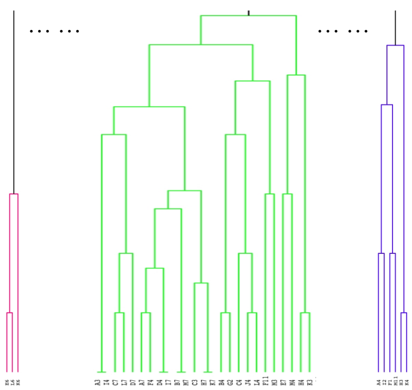


Figure 3: Zoom on the general dendrogram representation depicted in Figure 2. Each part (left, center, right) represents one category and illustrates either grouped (same icons from several designers) or mixed (several icons from several designers) sub-structures.

The MDS analysis is performed by means of a non-metric multidimensional scaling algorithm and uses a conventional normalized stress criterion to define the dimensionality of an acceptable solution. In fact, the stress measures the difference between observed and estimated data, according to the number of dimensions used in the model. Here, different dimensions were tested (from 2 to 5) until the stress factor reaches an acceptable value under 20%, (actually 18%) that starts to be considered as « fair » [32]. The output of this analysis is then a 5-dimension space were the items (sonic icons) are optimally placed with regards to their distance in the global dissimilarity matrix.

Moreover, as this kind of result turns out to be rather difficult to visualize and analyze (10 projections of the combined pairs within the 5 dimensions), and with regards to the global aim of the study (search for invariants in the sonic icons constitution and interpretation), we decided to perform an additional process based on the variability of the function/icon sub-groups of the corpus. A standard deviation was computed for each of the 11 sub-groups on each of the 5 dimensions of the resulting space and was sorted in increasing order. Then, a counting of the two first dimensions that most often minimize this deviation led to two combinations of 2-d spaces ((1,2) and (2,5)) that we consider, at first order, as the most informational – and easy to observe – results with regards to the context and the complexity of the study.

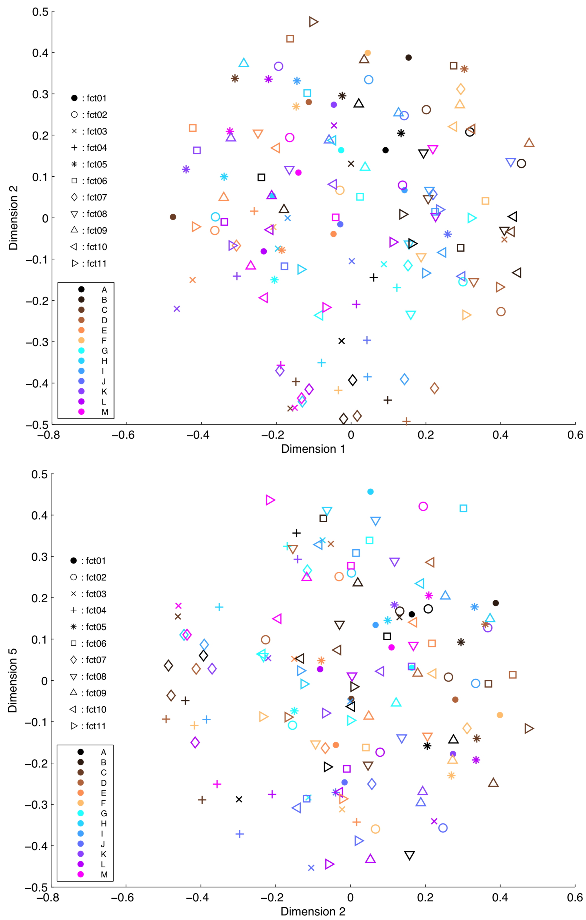


Figure 4: Perceptual spaces given by the MDS analysis of the formatted data resulting from the free categorization experiment (Sec. 5.2.1). Two sub-spaces are selected: (1,2) top, and (2,5) bottom. Icons/functions are depicted with symbols and sound designers with colors (see legend).

These (partial) graphical representations (Figure 4) of the relationships between the icons of the corpus led us to other kinds of insights. First, we mostly encounter again some of the previous results, especially with regards to the strong grouping of icons/functions No. 6, 7 and 11. In a certain extent, we can also consider that icon/function No. 1 is also rather concentrated on sub-region of the 2-5 space (center-right). Second, this colored depiction – and furthermore, the others 2-D configurations not included here for sake of space and readability – allowed to assume that no strong tendencies concerning sound designer's creative inspiration or identity come out, as mostly along each function, the colors are widely span over the whole range of the space.

6. DISCUSSION

The present study provides interesting insights about the general topic of sonic icons, in two distinct directions: the conceptual approach adopted to realize the production of such sonic artefacts, and the mental model used when decoding or receiving this information. In a larger extent – and this is one of a long-term claim –, it may lead to setting up rules or guidelines for the design of sonic icons attached to user interface functionalities and able to be understood by a large span of people (from different ages, social backgrounds, countries, cultures, etc. …).

6.1. Discussion of outcomes

Qualitative outcomes (Sec. 5.1) firstly force us to notice that « *auditory icons* » and *« earcons »* still stand as seminal foundations for sonic iconography. This is coherent with their historical status and must, of course, be due to an educational trace (most of the lectures on sonification or sound in HMIs naturally develop these concepts), especially relevant in the present case as the sound designers were students at the time of the study – and then surely attended to that kind of lectures. Nevertheless, it could be worth forcing the conception towards other sonification approaches (see Sec. 2), thus coping with a kind of fixation process [33] in that domain. What we could learn also on the basis of these qualitative transcriptions is that some function/sonic icon links make a strong consensus on how it should be realized. This is especially the case for 'scan' or 'my documents' that were rather unanimously considered in the metaphoric point of view, with a choice of metaphors leading to a strong perceptual pairing that clearly appeared in the clustering analyses. With these examples, we start to reach a kind of shared – not to mention universal – mental representation of these precise information that would actually be interesting to test in different cultures, or social context. Anyhow, the transcription of sound designers verbatim has not been fully accomplished and might need a more grounded textual or lexical analysis to provide richer and deeper information on the sound design process. Apart from that, the semantic annotation approach adopted to describe the corpus of sonic icons appeared to be promising but also suffered from a lack of methodology – and then gave some data rather difficult to tap. As in Carron's doctoral work [34], we should have implemented a more structured annotation session, ensuring first that the semantic dimensions were commonly understood, and then operating annotations with a wider expert panel.

Quantitative outcomes (Sec. 5.2) draw an average mental model of representation of the sonic icons, among a pool of more than 50 participants. They bring some information, especially with regards to perceptual clustering of some types of icons – that somehow confirm the sound designers' inspirations, as previously mentioned. But, because of their complex and noisy/fuzzy nature, they can also lead to the several following assumptions: the experimental paradigm was difficult to achieve, the participants adopted different strategies (focus on the function specificity, the designer's identity or both), or a large part of the sonic icons didn't convey enough or right information in terms of functionality. This being, these outcomes might also be due to a basic methodological issue: the fact the sound corpus to be tested was too wide to be able to consider it in full, in a complete experimental design (amplifying the noise of the data), and consequently, the fact that the data analysis were highly multi-dimensional that prevented from observing the results in a simple and efficient way. These elements might eventually lead to do over the experimental paradigm or deal with data analysis and visualization methods more in depth, beginning with finding a solution to reduce the dimensions of analyses.

Finally, the sounds were designed to occur while activating a function or selecting an icon, to strengthen an intuitive awareness or enrich the function or confirm, whereas the listening experiments were carried out with the sounds out from their context to compare them and group them freely.
Furthermore, while the sounds were globally normalized to allow this comparison, the sound level should have been thinking as an informative parameter by the sound designers, e.g. to distinguish between different urgency or intensity of functions. While this normalization and decontextualization were useful to let sound dimensions emerge, they may have impaired the interpretation of the sounds. Therefore, there is still a necessity to verify the effectiveness, the audiographical consistency, and the global interpretation of the proposed sound designs in a more ecological task, for instance by integrating them in a fake interface.

6.2. Future research

Besides the short-term improvement of the protocol or the further analyses suggested in the precedent section (Sec. 6.1), other long-term research questions are still to be addressed.

The first issue concerns the role of cultural origin, from a geographic angle (country) or from a socio-demographic angle (age, field of work, literacy or computer-literacy abilities). This problematic has already been studied in auditory displays [10], [23] but only on the listener/user point of view. Influence of culture on the designer conceptual model is still to be considered.

There is also a general lack of knowledge concerning the role of vision on auditory mental representations, from either the designer or user perspective. More research should be done to better understand how the presence of visual icons may guide the designers during the design phase of or may influence the interpretation of sonic icons by listeners.

7. ACKNOWLEDGMENT

The authors are grateful with the contribution of the students from ESAD Le Mans (supervised by one of the co-author) and from ENJMIN Angoulême (supervised by Prof. Cécile Le Prado-Natkin).

8. REFERENCES

[1] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg, “Earcons and icons: their structure and common design principles”. *Hum.-Comp. Int.*, vol. 4, 1989.

[2] W. W. Gaver “Auditory icons: using sound in computer interfaces”. *Hum.-Comp. Int.*, vol. 2, 1986.

[3] B. Walker, J. Lindsay, A. Nance, Y. Nakano, D. Palladino, T. Dingler, and M. Jeon, “Spearcons (speech-based earcons) improve navigation performance in advanced auditory menus”. *Human Factors*, vol 55, no. 1, p. 157-182, 2012.

[4] G. Parseihian, and B. F.G. Katz, “Morphocons: a new sonification concept based on morphological earcons”. *J. Aud. Eng. Soc.* Vol. 60, no. 6, 2012.

[5] T. Kvan “Collaborative design: what is it? ” *Automation in construction*, *9*(4), 409-415, 2000.

[6] A. R. Sarkar, G. Sanyal, and S. Majumder, “Participatory design for selection of icons to represent daily activities of living for a vision-based rehabilitationcum-assistance system for locked-in patients”, *Disabil. Rehabil.: Assist. Tech.*, vol. 50, p.282-292, 2020.

[7] Á. Csapó et G. Wersényi, “Overview of auditory representations in human-machine interfaces”, *ACM Comput. Surv.,* vol. 46, pp. 1-23, 11 2013.

[8] T. Hermann, A. Hunt, and J. Neuhoff, *The sonification handbook*. Logos Publishing House, Berlin, 2011.

[9] M. Jeon, B. Walker, “Spindex (speech index) improves auditory menu acceptance and navigation performance”. *ACM Trans. Access. Comput.*, *3*(3). 2011.

[10] M. Jeon, J-H. Lee, J. Sterkenburg, and C. Plummer, “Cultural differences in preference of auditory emoticons: USA and South Korea”. in *Proc. of the 21st Int. Conf on Aud. Disp. (ICAD 2015)*, July 6-10, Graz, Styria, Austria, 2015.

[11] M-S. Mustonen, “A review-based conceptual analysis of auditory signs and their design”, in *Proc. of the 14th Int. Conf. on Aud. Disp. (ICAD 2008),* Paris, France, 2008.

[12] C. S. Peirce, *Logic as Semiotic: The Theory of Signs* chez Philosophical Writings, C. S. Peirce, Éd., Dover Publications, 1902.

[13] D. Oswald, “Non-speech audio-semiotics: A review and revision of auditory icon and earcon theory”. in *Proc. of the 18th Int. Conf on Aud. Disp. (ICAD 2012),* Atlanta, GA, USA, 2012.

[14] A. Petocz, P. E. Keller et C. J. Stevens, “Auditory warnings, signal-referent relations, and natural indicators: Re-thinking theory and application”, *J. Exp. Psychol. Appl.*, vol. 14, pp. 165-178, 2008.

[15] G. Wersényi, “Auditory representations of a graphical user interface for a better human-computer interaction”, in *Proc. of the 6th Int. Symp. on Comp. Mus. Mod. and Retriev. joinded to the 15th Int. Conf. on Aud. Disp. (CMMR/ICAD-2009)* Copenhagen, Denmark, 2009.

[16] J. Tardieu, N. Misdariis, S. Langlois, P. Gaillard, C. Lemercier. “Sonification of in-vehicle interface reduces gaze movements under dual-task condition”. *Appl. Ergon.*, vol. 50, 2015.

[17] J. Edworthy, S. Reid, S. McDougall, J. Edworthy, S. Hall, D. Bennett, J. Khan and E. Pye, “The Recognizability and Localizability of Auditory Alarms: Setting Global Medical Device Standards” *Human Factors*, vol. 59, 2017.

[18] P. Yalla, and B. Walker, “Advanced Auditory Menus: Design and Evaluation of Auditory Scroll Bars” in *ACM SIGACCESS Conf.,* Halifax, Nova Scotia, Canada*,* 2008.

[19] A. Pirhonen, “Semantics of Sounds and Images - Can They Be Paralleled?” in proc. of the 13th *Int. Conf. on Aud. Display (ICAD2007)*. 2007.

[20] J. M. Carroll, J. R. Olson, and M. Helander *(Ed.)* “Chapter 2 - Mental Models in Human-Computer Interaction : Research Issues about What the User of Software Knows”, *Handbook of Human-Computer Interaction,* National Academy Press*,* 1988.

[21] A. Guillaume, L. Pellieux, V. Chastres, and C. Drake “Judging the Urgency of Nonvocal Auditory Warning Signals: Perceptual and Cognitive Processes” *J. Exp. Psychol. Appl.,* vol. 9, 2003.

[22] S. Roddy et D. Furlong, “Embodied Cognition in Auditory Display” in *Proc. of the 19th Int. Conf. on Aud. Disp. (ICAD-2013)*, Lodz, Poland 2013.

[23] S. Langlois, C. Suied, T. Lagear et A. Charbonneau, “Cross Cultural Study of Auditory Warnings” in *Int. Conf. on Aud. Display (ICAD2008),* Paris, France, 2008.

[24] S. Landry et M. Jeon, “Participatory Design Research Methodologies: A Case Study in Dancer Sonification” in *Proc. of the 23rd Int. Conf. on Aud. Disp (ICAD2017),* 2017.

[25] N. Misdariis, P. Susini, O. Houix, R. Rivas, C. Cerles, E. Lebel, A. Tetienne and A. Duquesne. “Mapping Sound Properties and Oenological Characters by a Collaborative Sound Design Approach – Towards an Augmented Experience”, In Proc. of CMMR – Comp. Mus. Multidisc. Res. conf., Marseille, 2019.

[26] M. Droumeva, and I. McGregor, “A Method for Comparative Evaluation of Listening to Auditory Displays by Designers and Users” in Proc. of the *20th* Int. Conf. on Aud. Display*, New York, USA*, 2014.

[27] http://petra.univ-tlse2.fr/spip.php?rubrique49

[28] Susini et al., “Perceptive study and recommandation for sonification categories” in Proc. of the *9th* Int. Conf. on Aud. Disp. *(ICAD-2003)*, Boston, 2003.

[29] J. Tardieu, N. Misdariis, S. Langlois, P. Gaillard, and C. Lemercier (2015). Sonification of in-vehicle interface reduces gaze movements under dual-task condition. *App. Ergon.*, *50*, 41-49. 2015.

[30] P. Schaeffer, *Treatise on Musical Objects: An Essay Across Disciplines*. Univ. of California Press. 2017.

[31] M. Carron, T. Rotureau, F. Dubois, N. Misdariis, and P. Susini. Speaking about sounds: a tool for communication on sound features. *Journal of Design Research*, *15*(2), 85-109. 2017.

[32] J. F. Jr Hair, R. E. Anderson, R. L. Tatham, and W. C. Black. *Multivariate data analysis*. Fifth Edition. Prentice-Hall:New Jersey. 1998.

[33] N. Crilly, N. Fixation and creativity in concept development: The attitudes and practices of expert designers. *Design Studies*, *38*, 54-91, 2015.

[34] M. Carron. *Méthodes et Outils pour Définir et Véhiculer une Identité Sonore* (Doctoral dissertation). Université Paris 6 (UPMC), 2016.