CATEGORIZATION OF SOUNDS FOR THE DESIGN OF CLINICAL AUDITORY ALARMS

Joana Vieira[1,2,4] **Rui Marques[1]** **Frederico Pereira[1]**
Vasco Ferreira[3] **Jorge Almeida Santos[2,3]** **Paulo Noriega[4,5]**

[1] CCG: Centre for Computer Graphics, Guimarães, Portugal
[2] Ergonomics & Human Factors Group, ALGORITMI Research Centre, University of Minho, Guimarães, Portugal
[3] School of Psychology, University of Minho, Braga, Portugal
[4] CIAUD Research Centre for Architecture Urbanism and Design, Lisbon School of Architecture, Universidade de Lisboa, Portugal
[5] ergoUX Ergonomics and UX Lab, Lisbon School of Architecture, Universidade de Lisboa, Portugal
Joana.Vieira@ccg.pt

ABSTRACT

An online sound categorization study was carried out to assess the association of everyday sounds with regard to medical equipment audio alarms. There were seven clinical alarm risk categories as described in IEC 6060-1-1-8 as well as an additional alarm category for 'Blood Pressure'. After a headphone screening test, participants categorized all sounds into one of the categories, a methodology adapted from a classic usability methodology called Card Sorting [1]. They were asked to indicate the quality of fit of their evaluation (Poor, Fair, Perfect) after sorting each sound. From the eight categories tested, a consensus was achieved for the categories of Blood Pressure, Cardiac, Power Down and Ventilation. For the other categories of Drug Administration, Temperature, Perfusion and Oxygen, lower consistency was observed, highlighting the difficulty of sound design for specific contexts. Having evidence of agreement among participants is an important output for the general goal of designing a library of informative sounds for medical devices.

1. INTRODUCTION

Clinical auditory alarms are a tool in healthcare facilities to alert and inform caregivers of patient or medical equipment state changes [2]. Nevertheless, both healthcare professionals and patients face auditory alarms as a necessary hindrance which affects not only their mental health and professional capabilities [3], [4], but which also, in the case of patients, affects their effective recovery [5], [6] simply by being exposed to the stressful auditory soundscapes generated by medical devices. Research on how to better design auditory interfaces for healthcare is decades old [7]–[10], but the implementation of changes in a complex sociotechnical context such as healthcare is famously slow [11]. There are recommendations regarding the fundamental requirements that clinical auditory signals used in medical equipment should follow, included in the current global medical device standard IEC 60601 Parts 1–8 [12], revised in September 2020. This standard's previous version proposed simple melodic alarm sounds to distinguish eight alarm sources, each with a high and a medium-priority version. The melodies intended to be mnemonics of what they represent, with the purpose of helping clinicians discriminate the source of the alarms [13]. However, the standard had its own set of problems experimentally demonstrated, namely a lack of diversity between the sounds which contributed to problems with learning and recognizing the alarms [9], [14], [15]. Although the melodies for these short, tonal alarms are different for each alarm source, the timbre/pitch, key, duration, rhythm and tempo are fixed [2], making them very hard to distinguish from one another.

This standard's updated version includes new design requirements, a new library of proposed alarms, and it allows manufacturers to develop their own set of alarms, as long as they abide to its performance metrics.

The current research is part of a project which follows a set of human centered design methodologies [16] applied to the design of clinical auditory alarms. Our final goal is to propose a library of alarm sounds which can safely be implemented in medical devices, improving detection, ease of localization, learning and comprehension of the alarms after psychophysical, neurological and usability tests with end-users.

The current paper describes a methodology used after interviews with healthcare professionals, and the characterization of surgery and recovery rooms in hospitals and precedes the sound design stage. This order of studies allows to proceed with the auditory alarm synthesis grounded on design guidelines which are human, and context centered. One step towards that goal consists in refining and categorizing sounds according to seven pre-established categories. These categories are in IEC60601-1-8 and are Cardiovascular, Drug delivery, Power down, Oxygenation, Perfusion, Temperature, and Ventilation. After interviews with professionals, they have suggested an additional vital sign parameter which would help them if it had an unequivocal sound, which is the Blood Pressure parameter. For that reason, we have considered Blood Pressure a category *au-par* with the other seven.

The main aim of this study is to identify one everyday sound which is strongly associated with each one of the seven IEC60601-1-8 categories (plus Blood Pressure) by a sample of healthcare and non-healthcare participants. Having a final pool of candidate environmental sounds obtained through categorization exercises will allow a more focused sound design phase.

1.1 Categorization of Environmental Sounds

Categorizing is not defined by a precise set of rules. In what regards sound, it may differ according to listening habits and abilities, and it may differ according to the methodology used to study it. This is a cognitive process strongly guided by the similarity between different entities [17] and is part of a "cognitive economy" where the environment is simplified and information is retrieved with the least cognitive effort [18]. Several studies have intended to understand exactly which perceptual strategy and cognitive processes are used by human listeners to interpret and form categories of sounds, and most use environmental sounds as stimuli. Environmental sounds are "*all naturally occurring sounds other than speech and music*" [19].

There are several everyday sounds taxonomies [18], based on an ongoing effort of understanding how these sounds are interpreted. These include sorting exercises of ambient urban noise [20], human activity sounds [21], or road traffic noise [22]. Category formation generally follows the strategy of similarity between sound sources, places and actions [18]. In tasks where participants had to describe sounds, they usually described the action, the object of the action or the context where the action occurred [23].This behavior matches the statement from Giordano et al. [24] that environmental sounds have meaning in function of their connection with the events and objects that generated the sounds.

Nevertheless, other lower-level strategies have been reported by previous literature, revealing how category formation is based on different cues in different contexts [18]. In 1979 Vanderveer [23] asked participants to sort the items based on the similarity of the sounds and the conclusion was that the basis of the sorting was twofold: they either grouped the sounds because they represented similar events (e.g., drop a can, drop wood) or because they had acoustic similarities. Another study provided additional information, stating that acoustic properties were used to group environmental sounds from non-living objects, and semantic properties were used for environmental sounds from living things [24].

Importantly, Ballas reported categorization is influenced by acoustic variables, ecological frequency (the frequency with which a listener encounters a specific sound event in everyday life), causal uncertainty (as the amount of reported alternative causes for a sound) and sound typicality, but acoustic variables account for about half of the variance in identification time and accuracy. Lemaitre et al. [25] have observed that sounds with high causal uncertainty are grouped together according to acoustic similarities more often than sounds with an identifiable source.

In what might be in the basis of categorization, some authors state that the similarity of two sounds depends on which aspect the participant chooses to focus, and individual differences such as listening habits, abilities and skills might influence to which aspect will one pays more attention [25]. In that experimental study, the authors identified three properties used by participants to describe a sound: (1) acoustic properties, (2) causal properties, and (3) semantic properties [25] and when comparing expert and non-expert listeners, the authors observed that both used different similarities to categorize sounds. Non-expert listeners tended to focus more on the causal properties of the sound event, and expert participants grouped together sounds based on their acoustical similarities. It was concluded that judging sounds based on their acoustical properties requires training, and that non expert participants resorted more often to semantic similarities to sort sounds than expert listeners [25].

This leads us to important methodological intricacies and how instructions and techniques can affect the comprehension of the categorization strategies. To understand which attributes are used to form categories, some include the semantic differential method where participants score concepts on different rating scales [18]. One of the disadvantages of this method includes a potential lack of ecological validity for the comprehension of categorization, as the evaluated concepts or attributes are defined a priori by the researcher [18]. Other methods include sorting or grouping tasks (icons representing each sound simultaneously are presented on a screen and requires the listener to group the sounds by similarity) and forced choice or pairwise comparisons (a pair of sounds is presented sequentially and requires a similarity rating from the listener). Sorting tasks allow participants to use their own criteria (and sometimes their own descriptors), providing greater ecological validity [18], [25]. Most studies do not present participants with predefined categories.

A call for attention has been made Aldrich et al. [26], referring that the used methodology affects the categorization effort. These authors compared the categorization outcomes of pairwise comparisons and grouping tasks. Results have shown that for similar acoustic sounds, the paired comparison methodology resulted in a categorization strategy primarily based on acoustic information. The grouping task encouraged participants to make more use of categorical over acoustic information, although some groups of acoustic features were also found. There was no observed effect of methodology with unfamiliar sounds.

1.1.1 Clinical Auditory Alarms

In the context of alarm design for healthcare settings, to our knowledge, only Bennet et al. [27] have performed similar studies, although strictly focused on the auditory features of alarms. The authors compared the IEC alarm set with an experimental set. Whereas IEC alarms differed on two auditory dimensions (melody and tempo), the experimental set had additional dimensions where it varied. The authors intended to identify the acoustical correlates of urgency perception. The strongest observed correlates for the perception of urgency were, among others: standard deviation and mean of the rhythmic attack slope (i.e., variation in the "transientness" over time), variation of the tonal centroid, and the mean spectral roughness, indicating that fluctuating spectral content is a key to determining perceived urgency. With a different perspective, a recent study used the same technique of Card Sorting to understand if the rationale of categories found on IEC60601-1-8 is adequate [28]. Grouping rationale included Urgency, Monitoring or Device events, Calls, Exits and Responder, resulting in a proposed alarm organizational structure.

2. METHOD

In information architecture research a methodology called Card Sorting is commonly applied in visual interface design. It consists in the sorting or grouping of concepts or objects as previously described. This sorting process provides information about terminology, relationships and categories [1]. An important variant in this methodology regards the "open" or "closed" sorting approach. The examples provided thus far consist in open sorting, where participants make up and name their own categories. Closed sorting has predefined categories and is used when trying to establish changes required to an existing structure. Because IEC 60601-1-8 new auditory alarms will need to fit inside the same categories, we have chosen a closed sorting for this study.
The exercise consisted in an online, closed sound sorting activity with evaluation of Quality of Fit. After a screening test for headphones, participants sorted 51 sounds into nine predefined categories: Cardiovascular, Blood Pressure, Drug delivery, Power down, Oxygenation, Perfusion, Temperature, Ventilation and "I don't know / Doesn't fit anywhere". After placing a sound inside a category, the participant self-assessed the quality of fit - Poor, Fair, Perfect – of each categorization. One sound could only fit in one category and no extra categories could be generated.

2.1 Participants

Sixty-seven respondents completed the online categorization questionnaire. Four participants were removed for failing the screening test or for systematically categorizing all 51 sounds into the "I don't know" category. Sixty-three respondents were considered eligible for the final sample. Thirty-three referred not having previous experience with medical devices and thirty referred they have interacted with medical devices before. This experience came from their professional and student experience in hospitals or emergency teams, but also from experience as caregivers of hospitalized family members, own rehabilitation, and monitoring and finally, as medical devices designers and advisors. Twenty-nine were healthcare professionals (registered nurses, physicians, or medical students). Most respondents were Portuguese. 63% were female and 38% were male. Further demographic data is displayed Table 1.

2.2 Stimuli

Fifty-one stimuli were presented. Previous literature [29] had made available a list of tested sounds, mostly auditory icons to be used in the same medical categories. Although the materials were merely descriptive, we searched for sounds matching the descriptions on online databases (freesound.org). When a sound was not available, we recorded it ourselves. The sounds consisted in everyday sounds (some environmental, some familiar from sci-fi imagery, etc.).

Experience with medical devices		**Profession**	
Yes	30	*Physician*	14
No	33	*Nurse*	9
Age		*Medical Student*	3
18-29	35	*Student*	10
30-39	21	*Researcher*	5
40-49	4	*Designer*	3
50-59	1	*Psychologist*	1
60-69	2	*Computer scientist*	2
Musical training		*Consultant*	2
Yes	20	*Industrial Engineer*	2
No	43	*Acoustic Engineer*	2
Location		*Occupational Therapist*	1
Portugal	23	*Marketing Director*	1
Netherlands	8	*Medical Device design*	1
USA	8	*Medical Lecturer*	1
UK	3	*Logistics Management/HFE*	1
Germany	2	*Filmmaker*	1
Ireland	2	*Finance/Economics*	1
Australia	1	*NA*	3
Spain	1		

Table 1. Demographic data of participants

Each category had, on average, six different sounds. These categories were: Oxygenation, Ventilation, Cardiovascular, Artificial perfusion, Temperature, Drug administration, and Equipment or Power failure and Blood Pressure. As part of the pool of sounds, the alarms of the 2020 update to the standard were included without the priority dimension. All sounds were trimmed to between 3 and 4 seconds in duration. They were stored as 24-bit wav format, with a sample rate of 44.1 kHz. All files were normalized to the same Leq(A) value. Harmonicity Noise Ratio, Spectral Centroid, Loudness, Roughness and Sharpness' measurements for the 51 sounds are available in Table 2 of the Supplemental Digital Content (http://tiny.cc/FA2020JVieira)[1].

2.3 Procedure

The study was conducted via the online survey platform LimeSurvey [30]. Initially, participants were asked demography questions including whether they were healthcare professionals or had musical training. Before the categorization task, participants performed a listening

[1] Supplemental digital content is available for this article. Direct URL: http://tiny.cc/FA2020JVieira

screening test. To help to ensure that all participants performed the listening test while using headphones, we implemented the intensity-discrimination task designed by Woods, Siegel, Traer and McDermott [31]. This task involves tones that sometimes have a phase difference of 180° between stereo channels. The antiphase tones are heavily attenuated when played through loudspeakers but are not attenuated over headphones, thus making it possible to differentiate between listening setups. The task consists of a three-alternative forced-choice (3-AFC) which asks participants '*Which tone is quietest?*'. Regarding the volume of the sounds, participants were asked to find a comfortable level and not to adjust it until the end of the experiment.
The main task consisted in grouping the 51sounds into nine categories. Figure 1 represents the categorization interface.

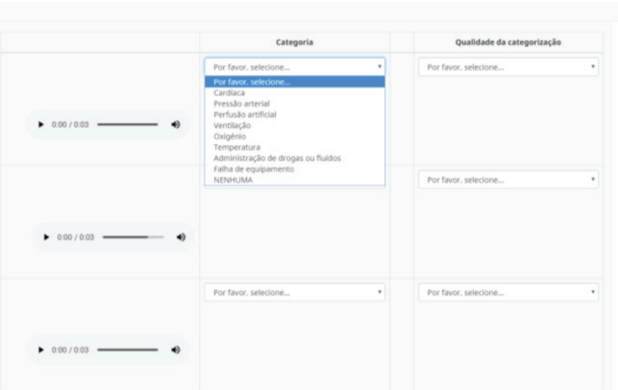

Figure 1. Categorization Interface on LimeSurvey

The sounds were presented in a random order for each participant. The instructions mentioned that participants should be wearing their headphones and that they could hear the sounds more than once.

They were also provided a short description of what could be included in each category.
Additionally, participants were requested to indicate how good they thought their categorization was, by stating how confident they were that that sound could represent the selected cause of the alarm. The quality of fit scale was Perfect, Fair, Poor or NONE (if not categorized). No time limit was imposed.

3. ANALYSIS

The two groups (H - Healthcare and non-H – non-Healthcare professionals) were analyzed separately. Both populations classified a large number of sounds as not applicable to any category (NAs), with healthcare professionals classifying 41.5% of sounds as not belonging to any category and non-healthcare professionals categorizing 27.5% as such. The most attributed categories for both populations were Ventilation, Oxygen and Cardiac. A chi-square test of independence was performed to examine the relation between profession and categorization. The relation between these variables was not significant, $x^2(8, N=63) = 8.98$, $p > 0.5$, meaning that there was not a significant difference between both groups regarding categorization.
Analyzing the categorization of each group, the similarities regarding categorization are noteworthy. Table 2 includes the main similarities (green) and differences (red) between of the categorization of each group in percentage. These percentual values correspond to the number of times each stimulus was placed in a category.

Category	Stimuli	H (%)	Non-H (%)
Blood Pressure	cuff inflate	48	42
Cardiac	sound_korotkoff	68	67
	norm cardio	37	32
	metal sheet	42	24
Power down	Robot power down	39	38
	Engine sci-fi	32	30
Ventilation	darth vader	42	34
	mask	56	32
	bellows	50	25
	AC	8	31
Oxygen	bellows	2	46
	constant bubbles	19	8
	bubbling	14	8
	breathe out	10	44
	decompression	13	38
Drug administration	water inside glass	23	22
	spread pills	14	27
	take pills	11	26

Table 2. Main differences (red) and similarities (green) between participants with Healthcare (H) and non-Healthcare (non-H) professions. The Perfusion and Temperature categories had a residual number of classifications and are not represented.

Both groups strongly associated the sound "cuff inflate" with the category Blood Pressure (H – 48% ; non-H- 42%). The "sound_koroktoff" (H – 64% ; non-H-69%), "norm cardio" (H – 34% ; non-H-33%) sounds were categorized as "Cardiac" sounds by both groups. Similarly to the "Power down" category with the sounds "Robot power down" (H – 38%; non-H-39%) and "Engine sci-fi" (H – 33%; non-H-31%). The sounds "darth vader" (H – 40% ; non-H-35%) and "mask" (H – 55% ; non-H-30%) were classified in the Ventilation category by healthcare and non-healthcare professionals. The differences between both were observed especially in the Ventilation and Oxygen categories. Healthcare professionals strongly classified the sound "bellows" to Ventilation (47%) while the non-healthcare professionals strongly classified the "AC" sound in that same category (32%).Curiously, the "bellows" sound was classified by non-healthcare professionals as an Oxygen sound (48%). Regarding the Oxygen category, healthcare professionals preferred "constant bubbles" (19%), and "bubbling" (14%) sounds, while non-healthcare professionals selected the "breathe out" (43%) and "decompression" (40%) sounds. Additionally, healthcare professionals selected "metal sheet" sound for the Cardiac category (40%) more than non-healthcare professionals (25%). Finally, for the Drug administration category both groups selected the "water inside glass" sound (23%) but non-healthcare professionals preferred the "spread pills" (28%) and "take pills" (27%) sounds.

To understand whether both groups' data could be pooled, a similarity analysis was performed using SynCaps V3 [32]. A similarity matrix displays the percentage of times each possible pair of sounds appeared together in the same category across the sample of participants. A similar score of 0.46 would indicate that 46% of the times, two specific sounds (e.g. bubbling and constant bubbles) were placed in the same category.

An average maximum similarity score was computed by averaging the similarity scores across all the sound pairs. In the healthcare professionals' group, the average maximum similarity score was 0.63 (SD=0.13), indicating moderate similarity. In the non-Healthcare professionals' group, the average maximum similarity score was also moderate, 0.52 (SD=0.09). Joining both groups resulted in an average maximum similarity score of 0.66 (SD=0.95), indicating that joining both samples slightly improved the similarity between them. Additionally, a hierarchical cluster analysis suggested one single cluster, further reflecting similarity in the categorizations. Considering the lack of significant differences between categorization from both groups, and the results from the similarity analysis, the following analysis were performed with an aggregated sample of healthcare and non-healthcare populations.

A further refinement of the pool of sounds for the following analysis was obtained with a weighted similarity matrix (Figure 1 in Supplemental Digital Content http://tiny.cc/FA2020JVieira) in which the weight corresponded to the average quality of fit (from 1 – Poor to 3 - Perfect) of each pairing. The sounds with the best self-assessment from participants regarding their categorization and with the smaller number of NA categorizations were selected for further analysis, resulting in 22 sounds. Temperature category was removed from analysis due to the reduced number of categorizations.

4. RESULTS

Categorization results are in Figure 2. To confirm the number of dimensions of the categorization, we used the weighted similarity matrix to perform multidimensional scaling (MDS), a multivariate analysis to visualize the similarity between samples through two dimensional plots, representing the distances between pairs of sounds.

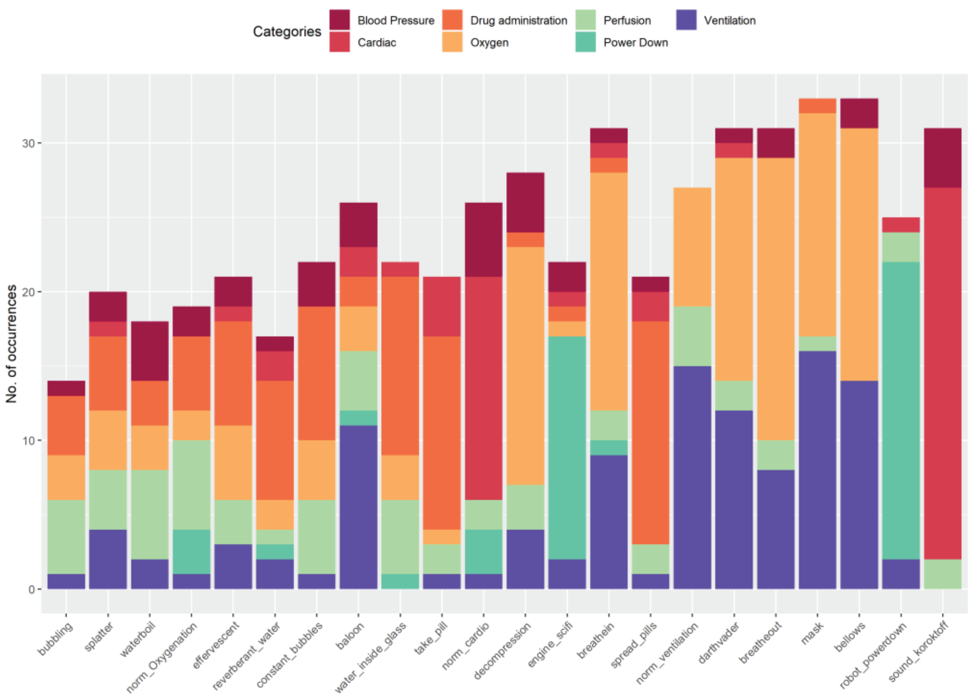

Figure 2. Categories in which 22 final sounds (x-axis) were placed

A preliminary analysis (cluster dendrogram) to understand the number of clusters that should be inputted for the MDS proposed K=7 (Figure 3). MDS confirmed a relation between "robot powerdown" and "engine_scifi", which were frequently grouped under the Power Down category, and are depicted in dark yellow in Figure 3. "Sound_koroktoff" and "norm_cardio" (IEC60601-1-8 sound for Cardiac alarm) are here represented also in the same group, in blue. Both were often grouped either under the Cardio or the Blood Pressure categories. The rest of the sounds are distributed among five groups, with air related sounds representing two major groups on the left, in red and green in Figure 3. All these sounds were categorized under the Ventilation or Oxygenation categories, and the "balloon" sound is more detached from the rest of the sounds. This might have influenced the departure of "norm_ventilation" (IEC60601-1-8 sound for ventilation) and "breathe_in" from the red group. Water related sounds constitute the other three clusters, here in pink, purple and water-green on the right side of the plot. One cluster consists of sounds of "water boiling", "constant bubbles", "effervescent" or "splatter". The pink cluster had more sorts of liquid sounds with "reverberant water", "norm Oxygenation" (IEC60601-1-8 sound for Oxygenation) and the "take pill" sounds.

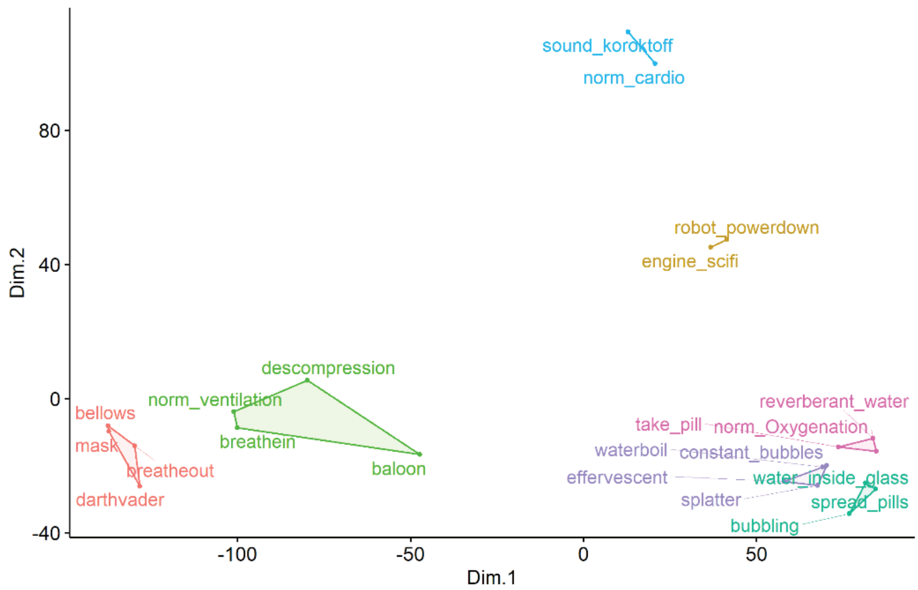

Figure 3. Multidimensional scaling with k=7

Finally, the water green cluster was constituted by "water inside glass", "spread pills" and "bubbling". All the sounds in the Water-related sounds group were distributed among the Perfusion, Oxygen and Drug administration categories.

5. DISCUSSION

The categorization efforts of all sixty-three participants revealed consistency between the healthcare and non-healthcare populations. The main difference resided in the highest number of NA categorizations by the healthcare participants. The main similarities were observed in the Blood Pressure category, with the "cuff_inflate" sound as the one most often selected for that category. The Cardiac, Power Down and the Ventilation categories were also concordant, but the main differences were found in the Oxygen, Perfusion and Drug administration categories. The two groups did not show a statistical relationship regarding categorization and profession, and both were analyzed together after a similarity analysis indicated a similarity of 66%. A weighted similarity matrix considering the Quality of Fit assessed by the participants after each categorization allowed the selection of the 22 best evaluated sound categorizations. The resulting two-dimensional depiction after MDS clearly represents the logic behind the sounds' categorization. The sounds for the Power Down (robot power down; engine scifi) were put together as a single category, and so were the Cardiac sounds (norm cardio; sound korokotoff). Along Dimension 1 we can find air-related sounds to the left and water related sounds to the right. On this high-level division, air-related sounds were generally placed in the Ventilation category, whereas water-related sounds were categorized among Oxygen, Drug administration and Perfusion categories.

Looking at the categorization before clustering, there is a misunderstanding between Ventilation and Oxygenation categories, with a lot of the sounds categorized as Ventilation being also categorized as Oxygenation. To non-Healthcare participants one may find it legitimate to associate Oxygen sounds to other air-related sounds. But in fact, Oxygen relates to a molecular analysis of the blood (pulse oximeters, transcutaneous or tissue oxygen monitors). This association was not observed among Healthcare professionals who categorized as Oxygen the water-related sounds. After proximity analysis and clustering, the sound designed for the Oxygenation category by the norm IEC60601-1-8 (norm Oxygenation) was placed together with other water-related sounds. The group of water-related sounds was the most divided, even though its three clusters are close to each other. Perfusion, Drug administration and Oxygen categories seem to associate with liquid sounds, but the idiosyncrasies of each category are not evident enough to lead to sound design guidelines. One might argue that Perfusion sounds can be associated with movement of a liquid in a closed space while Drug administration is more associated with water pouring inside a receptacle, or with the act of handling actual pills, but this division needs further exploration of the acoustic properties associated with each.

These results point the way to a high-level division between the sounds and the predefined categories, but the analysis is not without its limitations. First, the analysis with a weighted similarity matrix searches for pairs of sounds usually grouped together, which resulted in the exclusion of sounds which might be often associated by themselves to one category. This happened to the Blood Pressure category, with one unequivocal sound (cuff inflate) categorized as such.

Further analysis should focus on the acoustic properties which these sounds have in common. Based on this studies' results it is not possible to define whether categorization was based on acoustic properties, causal properties or semantic properties as proposed by Lemaître [25]. From our results, it seems the followed strategy was based on similarity between sound sources (air, water, percussion) as proposed by Bones [18]. Similar to Aldrich [26], it seems the grouping task encouraged the use of categorical information instead of acoustic information.

6. CONCLUSION

An online closed sound sorting exercise was executed with the purpose of selecting sounds for the seven categories of clinical alarms suggested in IEC 60601-1-8, plus an additional category for Blood Pressure. A multidimensional scaling proposed seven clusters which we could identify as Power Down and Cardiac sounds. Two additional clusters grouped air-related sounds, associated to the Ventilation category. The last three clusters consisted of water-related sounds associated with the Perfusion, Oxygen and Drug administration. This categorization suggests that some categories are more easily characterized with sound than others. Due to the analysis with a similarity matrix which privileged frequency of pairings to determine categories, the Blood Pressure category was not included in the multidimensional scaling analysis. The Temperature category was excluded for having the least attributed sounds from participants.

Nevertheless, the current study allows to suggest types of sounds to the following categories:

- Cardiac – Sound Korotkoff
- Blood Pressure – Cuff Inflate
- Power down – Robot Power Down
- Ventilation – Mask/Darth Vader
- Drug administration – Spread pills/Take Pill/ Water inside glass

The sounds placed on the Oxygen and Perfusion categories are systematically grouped together, demonstrating the need for further refinement and distinction between these two categories and respective sounds.

Future studies will include sound design of clinical alarms based on the current results, and the analysis of the categorization based on the sounds' acoustic properties. Applications might include the enrichment of machine learning algorithms for the design of auditory alarms.

7. ACKNOWLEDGEMENTS

This work was supported by grant no. PTDC/PSI-GER/31943/2017, cofinanced by COMPETE2020 under the PT2020 program, and supported by FEDER.

8. REFERENCES

[1] W. Hudson, "Card Sorting," in *The Encyclopedia of Human-Computer Interaction*, 2nd ed., R. F. Soegaard, Mads and Dam, Ed. Aarhus, Denmark.: The Interaction Design Foundation., 2014.

[2] R. R. McNeer, D. B. Horn, C. L. Bennett, J. R. Edworthy, and R. Dudaryk, "Auditory Icon Alarms Are More Accurately and Quickly Identified than Current Standard Melodic Alarms in a Simulated Clinical Setting," *Anesthesiology*, pp. 1–9, 2018.

[3] E. Institute, "Ecri Institute Honors The Johns Hopkins Hospital For Innovations In Alarm Management," *TechNation*, no. March 2013, 2013.

[4] I. Busch-Vishniac, "Hospital Soundscapes: Characterization, Impacts, and Interventions," *Acoustics Today*, vol. 15, no. 3, p. 11, 2019.

[5] Ecophon, "Noise causes sleep disruption for critical care patients," *Ecophon Group*, 2018.

[6] D. Birdja and E. Özcan, "Better Sleep Experience for the Critically Ill : A Comprehensive Strategy for Designing Hospital Soundscapes," 2019.

[7] R. Patterson, J. Edworthy, and M. Lower, "Alarm sounds for medical equipment in intensive care areas and operating theatres," London, 1986.

[8] E. J. Hellier, J. Edworthy, and I. Dennis, "Improving auditory warning design: quantifying and predicting the effects of different warning parameters on perceived urgency.," *Human factors*, vol. 35, no. 4, pp. 693–706, 1993.

[9] J. Edworthy *et al.*, "The Recognizability and Localizability of Auditory Alarms: Setting Global Medical Device Standards," *Human Factors*, vol. 59, no. 7, pp. 1108–1127, 2017.

[10] J. R. Edworthy *et al.*, "Getting Better Hospital Alarm Sounds Into a Global Standard," *Ergonomics in Design: The Quarterly of Human Factors Applications*, vol. 26, no. 4, pp. 4–13, 2018.

[11] D. A. Norman and P. J. Stappers, "DesignX:

Complex Sociotechnical Systems," *She Ji: The Journal of Design, Economics, and Innovation*, vol. 1, no. 2, pp. 83–106, 2015.

[12] AAMI, *ANSI/AAMI/ IEC 60601- 1-8:2006 & A1:2012 MEDICAL ELECTRICAL EQUIPMENT – Part 1-8: General requirements for basic safety and essential performance – Collateral Standard: General requirements, tests and guidance for alarm systems in medical electrical equip*. 2013.

[13] J. Vieira, J. A. Santos, and P. Noriega, "A Review of Design Guidelines for Clinical Auditory Alarms," in *Health and Social Care Systems of the Future: Demographic Changes, Digital Age and Human Factors.*, Lisbon: Springer, 2019, pp. 325–333.

[14] P. Sanderson, A. Wee, E. Seah, and P. Lacherez, "Auditory alarms, medical standards and urgency," in *International Conference on Auditory Displays*, 2006, pp. 24–27.

[15] J. Edworthy, E. Hellier, K. Titchener, A. Naweed, and R. Roels, "Heterogeneity in auditory alarm sets makes them easier to learn," *International Journal of Industrial Ergonomics*, vol. 41, no. 2, pp. 136–146, Mar. 2011.

[16] International Organization for Standardization, "Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems," 2010.

[17] O. Houix, G. Lemaitre, N. Misdariis, P. Susini, and I. Urdapilleta, "A lexical analysis of environmental sound categories," *Journal of Experimental Psychology: Applied*, vol. 18, no. 1, pp. 52–80, 2012.

[18] O. Bones, T. J. Cox, and W. J. Davies, "Sound Categories: Category Formation and Evidence-Based Taxonomies," *Frontiers in Psychology*, vol. 9, no. July, pp. 1–17, 2018.

[19] B. Gygi, G. R. Kidd, and C. S. Watson, "Similarity and categorization of environmental sounds," *Perception and Psychophysics*, vol. 69, no. 6, pp. 839–855, 2007.

[20] C. Guastavino, "Categorization of environmental sounds," *Canadian Journal of Experimental Psychology*, vol. 61, no. 1, pp. 54–63, 2007.

[21] D. Dubois, "Categories as acts of meaning: The case of categories in olfaction and audition," *Cognitive science quarterly*, vol. 1, no. 1, pp. 35–68, 2000.

[22] J. Morel, C. Marquis-Favre, D. Dubois, and M. Pierrette, "Road traffic in urban areas: A perceptual and cognitive typology of pass-by noises," *Acta Acustica united with Acustica*, vol. 98, no. 1, pp. 166–178, Jan. 2012.

[23] N. J. Vanderveer, "Confusion errors in identification of environmental sounds.," *Journal of the Acoustical Society of America*, vol. 65, no. S60, 1979.

[24] B. L. Giordano, J. McDonnell, and S. McAdams, "Hearing living symbols and nonliving icons: Category specificities in the cognitive processing of environmental sounds," *Brain and Cognition*, vol. 73, no. 1, pp. 7–19, Jun. 2010.

[25] G. Lemaitre, O. Houix, N. Misdariis, and P. Susini, "Listener expertise and sound identification influence the categorization of environmental sounds.," *Journal of Experimental Psychology: Applied*, vol. 16, no. 1, pp. 16–32, 2010.

[26] K. M. Aldrich, E. J. Hellier, and J. Edworthy, "What determines auditory similarity? The effect of stimulus group and methodology.," *Quarterly journal of experimental psychology (2006)*, vol. 62, no. 1, pp. 63–83, 2009.

[27] C. Bennett, R. McNeer, and C. Leider, "Urgency Analysis of Audible Alarms in The Operating Room.," in *12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, 2011, pp. 771–775.

[28] M. C. Wright, S. Radcliffe, S. Janzen, J. Edworthy, T. J. Reese, and N. Segall, "Organizing Audible Alarm Sounds in the Hospital : A Card-Sorting Study," *IEEE Transactions on Human-Machine Systems*, 2020.

[29] C. Bennett, R. Dudaryk, N. Crenshaw, J. Edworthy, and R. McNeer, "Recommendation of New Medical Alarms Based on Audibility, Identifiability, and Detectability in a Randomized, Simulation-Based Study," *Critical Care Medicine*, vol. 47, no. 8, p. 1, 2019.

[30] LimeSurvey GmbH, "LimeSurvey: An Open Source survey tool." Hamburg, Germany., 2018.

[31] K. J. P. Woods, M. H. Siegel, J. Traer, and J. H. McDermott, "Headphone screening to facilitate web-based auditory experiments," *Attention, Perception, and Psychophysics*, vol. 79, no. 7, pp. 2064–2072, 2017.

[32] Syntagm, "SynCaps V3." 2018.