Universidade de Lisboa

Faculdade de Motricidade Humana

HUMAN-CENTRED DESIGN OF CLINICAL AUDITORY ALARMS

Joana Catarina Fernandes Vieira

Advisor: Prof. Doutor Paulo Ignácio Noriega Pinto Machado

Co-advisor: Prof. Doutor Jorge Manuel Ferreira de Almeida Santos

Dissertation submitted in order to obtain a Doctor of Philosophy degree in the field of Human Kinetics, specializing in Ergonomics

2022

Universidade de Lisboa

Faculdade de Motricidade Humana

HUMAN-CENTRED DESIGN OF CLINICAL AUDITORY ALARMS

Joana Catarina Fernandes Vieira

Advisor: Prof. Doutor Paulo Ignácio Noriega Pinto Machado

Co-advisor: Prof. Doutor Jorge Manuel Ferreira de Almeida Santos

Dissertation submitted in order to obtain a Doctor of Philosophy degree in the field of Human Kinetics, specializing in Ergonomics

Jury

President: Doctor Francisco José Bessone Ferreira Alves

Members of the Jury

- Doctor Elif Ozcan Vieira
- Doctor Maria Emília Capucho Duarte
- Doctor Doutor Francisco dos Santos Rebelo
- Doctor Rui Miguel Bettencourt de Melo
- Doctor Teresa Margarida Crato Patrone de Abreu Cotrim
- Doctor Paulo Ignácio Noriega Pinto Machado

2022

This thesis is dedicated to the memory of Professor Jorge Almeida Santos. Its work was partially supported by grant no. PTDC/PSIGER/31943/2017 - Prototyping medical acoustic environments for the design of new auditory warning signals, co-financed by COMPETE2020 under the PT2020 program.

ACKNOWLEDGEMENTS

This work would not have happened without a challenge made by Professor **Jorge Almeida Santos**. To him, whom I met sullen in a bright classroom, and then curious and cheerful in a dark laboratory, I send a thank you.

To my advisor, Professor **Paulo Noriega**, I would like to thank his confidence, his words of comfort, his comments and suggestions for improvement, and his enthusiasm for a topic he embraced with interest.

I thank **Centro Clínico Académico de Braga**, in particular **Rosana Magalhães**, **Raquel Marques**. Your mediation and care made possible a study that had everything not to happen in the midst of a global pandemic. Likewise, I would like to thank nurse **Florbela Costa** for enabling the acoustic characterizations of the various surgeries, and for doing so with pure interest in the results. To all the **teams of professionals from the Operating Room** that I came across: thank you for the corner in the waiting room and for the curious questions. I thank nurse **Ana Araújo**, to whom I have addressed several doubts about *beeps*. I am grateful above all to my **participants**.

A lonely journey tends to be more reckless and fearless, and I think that is how I ran into the **Alarms Club**. It was the equivalent of meeting my science rock stars, in such a welcoming and fun way. The number of times I quote **Judy Edworthy** must explain how special it was to meet her while eating a steak. The same for **Elif Özcan Vieira** who is a beautiful and passionate person and scientist, **Rosana Sanz** who became a companion, **Joseph Schlesinger** who hid the most comical easter-eggs in our presentations, **Michael Schutz**, and **Carryl Baldwin** who turned this academic meeting into a meeting of friends who like pizza and science.

The words from the bottom of my heart start by going to **João Lamas** and **João Ferreira.** With them I felt less adrift. Not having a musical background made me run into so many difficulties, but all of them were patiently explained to me. The ease with which they decipher code has also helped me countless times. I thank **Emanuel Sousa** for exactly the same, with the addition of the difficult questions and the confidence. I would not have been able to finish this work without **Rui Marques** and **Frederico Pereira**. If the first was always tireless in creating sounds and analyses, the latter was always over his shoulder double checking. I trust a lot of what I present in this thesis because it went through his

thorough analysis. The **Paterson Team**, especially **Tom Barker**, **Vasco Ferreira** and **Guilherme Campos**. It was always comforting to talk to you about alarming subjects.

My former colleagues from CCG. **Sandra Mouta** who does not know it, but is my leadership and scientific reference; **Mariana Silva**, who takes care of everything and everyone and with whom I discussed sound; **Carlos Silva**, with whom in equal times I got mad with, and had fascinating conversations about what goes wrong in the world of science and reality shows; to **Rosane Sampaio** who teaches me with her analytical eyes and surgical observations; **Liliana Fontes**, who always went out of her way to help in whatever she could, and **Inês Fortes**, who I turned to to talk about data, and who helped me by putting into conversation what Excel presented to me in cells.

To **Alexandra Asanovna Elbakyan**, because knowledge must be accessible to everyone.

My best friends of many years. **Mackay**, who has the gift of making everything stop and feels like a breath of fresh air every time. **Rita**, who is so far away but works very hard for that to be a detail. When we talk, I have all her attention and I am old enough to know how valuable that is. **Joui** is my other half. She is a super woman and a super friend. Her words, questions and mindreading helped me a lot. Her recent efforts to make me a part of Olívia's life are extraordinary. I am lucky to have her.

To **Fernando** and **Mabília**. Your unlimited kindness, car rides, bed & breakfast made it possible to attend classes in a city four hours away.

To my first home and family, **Mãe, Pai, Nuno**. It is so comforting to know you are there, no matter what.

Luís, for never asking those questions. For always being by my side challenging me. You keep showing there is more in me than I know of. For being my love, and my peace. *Consta nos astros, nos signos, nos búzios, eu li num anúncio, eu vi no espelho, tá lá no evangelho, garantem os orixás: Serás o meu amor, serás a minha paz. Consta nos autos, nas bulas, nos dogmas, eu fiz uma tese, eu li num tratado, está computado nos dados oficiais: Serás o meu amor, serás a minha paz*[1].

[1] Mas se a ciência provar o contrário, e se o calendário nos contrariar, mas se o destino insistir em nos separar: Danem-se os astros, os autos, os signos, os dogmas, os búzios, as bulas, anúncios, tratados, ciganas, projetos, profetas, sinopses, espelhos, conselhos, se dane o evangelho e todos os orixás: Serás o meu amor, serás, amor, a minha paz. – *in* Dueto, Chico Buarque.

ABSTRACT

Auditory alarms are commonly badly designed, providing little to no information or guidance. In the healthcare context, the poor acoustics of alarms is one contributor for the noise problem. The goal of this thesis is to propose a human-centred methodology for the design of clinical auditory alarms, by making them less disruptive and more informative, thus improving the healthcare soundscape. It implements this methodology from concept to evaluation and validation, combining psychoacoustics with usability and user experience methods. Another aim of this research consisted in understanding the limitations and possibilities offered by online tools for scientific studies. Thus, different processes and methodologies were implemented, and corresponding results were discussed.

To understand the acoustic healthcare environment, field visits, interviews, and surveys were performed with healthcare professionals. Additionally, sound pressure levels and frequency analysis of several surgeries in different hospitals provided specific sound design requirements, which were added to an existent body of knowledge on clinical alarm design. A second stage consisted in prototyping very simple sounds to comprehend which temporal and spectral parameters of sound could be manipulated to communicate clinical information. Parameters such as frequency, speed, onset, and rhythm were studied, and relations between subjective perception and physical parameters were established. In parallel, and heavily influenced by the new *IEC 60601-1-8 - General requirements, tests and guidance for alarm systems in medical electrical equipment and medical electrical systems*, a design strategy with auditory icons was created. This strategy intended to provide as much information as possible in an auditory alarm. To do so, it involved two main components: a priority pointer indicating the priority of the alarm; an auditory icon indicating the cause of the alarm. A third component indicating increasing or decreasing tendency of the vital sign was designed, but not validated with users. After online validation of the priority pointer and auditory icon for eight categories (cardiac, drug administration, ventilation, blood pressure, perfusion, oxygen, temperature, and power down), a new library of clinical auditory alarms is proposed.

Keywords: clinical auditory alarms, psychoacoustics, human-centred design, mixed-methods, auditory icons, healthcare

RESUMO

Os alarmes auditivos são habitualmente mal concebidos, dando poucas informações ou orientações perante a situação que despoletou o aviso. No contexto da saúde, a má acústica dos alarmes é um dos contribuidores para o problema do ruído. O objetivo desta tese é o de melhorar a paisagem sonora em ambientes clínicos, propondo uma metodologia centrada no Humano para o design de alarmes auditivos clínicos, tornando-os menos disruptivos e mais informativos. Essa metodologia é implementada desde o conceito até a avaliação e validação, combinando métodos da psicoacústica com métodos de usabilidade e experiência do utilizador. Outro objetivo desta investigação é o de compreender as limitações e possibilidades oferecidas pelas ferramentas *online* para estudos científicos. Assim, diversos processos e metodologias foram implementados, e os respetivos resultados são discutidos.

Para compreender o ambiente acústico clínico, foram realizadas visitas de campo, entrevistas e inquéritos com profissionais de saúde. Além disso, avaliou-se o nível de pressão sonora e frequências de várias cirurgias em diferentes hospitais. Esta atividade forneceu requisitos específicos de design de som que foram adicionados a um corpo existente de conhecimento sobre design de alarmes clínicos. Uma segunda etapa consistiu na prototipagem de sons simples para compreender que parâmetros temporais e espectrais do som poderiam ser manipulados para comunicar informações clínicas. Parâmetros como frequência, velocidade, envelope e ritmo foram estudados, e as relações entre a perceção subjetiva e os parâmetros físicos foram estabelecidas. Paralelamente, e fortemente influenciado pela nova norma *IEC 60601-1-8 - Requisitos gerais, testes e orientações para sistemas de alarme em equipamentos médicos elétricos e sistemas médicos elétricos*, foi criada uma estratégia de design com ícones auditivos. Essa estratégia pretendia incorporar o máximo de informações num alarme auditivo. Para isso, envolveu dois componentes principais: um ponteiro de prioridade que indica a prioridade do alarme; e um ícone auditivo que indica a causa do alarme. Um terceiro componente de tendência (aumento ou diminuição do valor do sinal vital) foi criado, mas não validado com utilizadores. Após a validação do ponteiro de prioridade e ícone auditivo para oito categorias (cardíaco, administração de medicamentos, ventilação, pressão arterial, perfusão, oxigénio, temperatura e falha de equipamento), propõe-se uma nova biblioteca de alarmes auditivos clínicos.

Palavras-chave: alarmes auditivos clínicos, psicoacústica, design centrado no Humano, métodos mistos, ícones auditivos, saúde

TABLE OF CONTENTS

1 General Introduction 21

1.1 The problem of noise in hospitals 23

1.2 The meaning of a sound 25

1.3 Research objectives 27

1.4 Dissertation Overview 27

2 "Alarms are still a problem!" 29

2.1 Introduction 31

2.2 Historical Context of Clinical Auditory Alarms 32

2.3 Conclusion 45

3 Sirens, church bells and alarms 46

3.1 What is an alarm? 48

3.2 Auditory icons 53

3.3 Masking and Localizability 54

3.4 Workload and Learnability 57

3.5 Conclusion 60

4 Toward Alarm Design 61

4.1 Introduction 63

4.2 Alarm Design Methodologies 64

4.3 Methodology of the present work 71

4.4 Conclusion 72

Step 1. Understand and specify the context of use 73

5 Need finding: Visit to an Operating Room at a public hospital 74

5.1 Introduction 76

5.2 Sound in the Operating Ward 76

5.3 The Operating Room (OR) 77

5.4 The Recovery Room (RR) 77

5.5 Conclusion 81

6 Learning with Healthcare professionals about hospital soundscapes 82

6.1 Introduction 83

6.2 Aim 83

6.3 Material and Methods 84

6.4 Results 87

6.5 Discussion 107

6.6 Conclusion 111
7 Frequency Analysis in Operating and Recovery Rooms 113
7.1 Introduction 116
7.2 Operating Rooms 118
7.3 Aim 121
7.4 Materials and Methods 121
7.5 Results 128
7.6 Clinical Alarm Design Implications 132
7.7 Discussion 134
7.8 Conclusion 136
Step 2. Specify user requirements 138
8 Design Guidelines for Clinical Auditory Alarms 139
8.1 Introduction 140
8.2 Spectral aspects of clinical auditory alarms 140
8.3 Temporal aspects of clinical auditory alarms 143
8.4 Spatial aspects of clinical alarms 144
8.5 Conclusion 145
Step 3. Produce design solutions 149
9 Subjective Perception of Auditory Warning signals (Lab vs. Web) 150
9.1 Introduction 152
9.2 Methodological considerations 153
9.3 Results 159
9.4 Discussion 171
9.5 Conclusion 173
10 Categorization of sounds 174
10.1 Introduction 176
10.2 Method 179
10.3 Procedure 181
10.4 Analysis 183
10.5 Results 186
10.6 Discussion 188
10.7 Conclusion 190
11 Design Strategy of a set of Clinical Auditory Alarms 191
11.1 Introduction 193
11.2 Our Alarm Sound Design Strategy 193

11.3 Our Alarm Sound Design Structure 199
11.4 Test Alarm Set 200
11.5 Conclusion 201
Step 4. Evaluate the design 202
12 Validation of clinical auditory alarms with users: Perception of urgency 203
12.1 Introduction 205
12.2 Aim 209
12.3 Method 210
12.4 Discussion 221
12.5 Conclusion 223
13 Validation of clinical auditory alarms with users: Comprehensibility and referent association 224
13.1 Introduction 226
13.2 Method 227
13.3 Analysis 233
13.4 Results 236
13.5 Discussion 243
13.6 Conclusion 244
14 General Discussion and Conclusions 246
14.1 Methodological considerations 253
14.2 What are the limitations? 256
14.3 Future research perspectives 258
14.4 Conclusion 260
15 References 262
16 Appendices 284
Appendix 1 - Glossary 285
Appendix 2 –Interview with healthcare professionals: Script 288
Appendix 3 – Online questionnaire for healthcare professionals 290
Appendix 4 - Stimuli used in Subjective perception of auditory warning signals 293
Appendix 5 – Stimuli used in Categorization of Sounds 305
Appendix 6 – Stimuli used in Perception of Urgency 307
Appendix 7 – Final library of Clinical Auditory Alarms 309
Appendix 8 – General organization of studies, timeframe and no. of participants 316

LIST OF TABLES

Table 2-1. Six tunes selected to represent each of the six alarm systems or categories (Block, 1992) 36

Table 2-2. Proposed standard-compatible, device-specific alarm sounds. Source: Block et al. (2000) 37

Table 2-3. ECRI Institute's Top 10 Technology Hazards mentioning Alarms between 2012 and 2019. 41

Table 2-4. Description of the auditory icons in the 2020's amendment of the IEC 60601-1-8 standard 43

Table 3-1. The progressive definition of an alarm (Us et al., 2011) 49

Table 3-2. Alarm priorities for clinical alarms. Examples from IEC 60601-1-8. 51

Table 6-1: Participant's background information 85

Table 6-2: Demographics of respondents of the online survey (n=46) 86

Table 6-3: Coding system used to categorize the interviews divided in four main dimensions: Physical space; Sound and Alarms; Alarm-triggered behaviour; and Suggestions for improvements. 88

Table 7-1. Recommended sound pressure levels according to IEC 61672: 2013-2017..... 117

Table 7-2. Sound levels according to surgical specialty (all Leq levels in dB(A)). Results from different studies..... 120

Table 7-3. Average Leq dB(A) for surgeries according to surgical speciality 129

Table 7-4. Average Lmax dB(A) for surgeries according to surgical speciality 131

Table 8-1: Summary of guidelines described in this paper..... 147

Table 9-1. Pairs of words used for the evaluation 154

Table 9-2. Levels of each acoustic parameter 155

Table 9-3. Descriptors per pair of words. A sheet with this information was always near the participant 158

Table 9-4. Values of Kendall's W in Lab and Web conditions 161

Table 9-5. Values of ICC in Lab and Web conditions 161

Table 9-6. Results of linear regression by levels of Frequency (500 Hz, 1500 Hz and 2500 Hz). N = 54. 95% Confidence Interval (only ***R2***> 0.5 are depicted)..... 164

Table 9-7. Mean values and R^2 comparison between linear models of Lab and Web Conditions for the levels of Frequency (500 Hz, 1500 Hz and 2500 Hz)..... 166

Table 9-8. Results of linear regression by levels of Speed (x1, x2, x4). N = 54, 95% Confidence Interval (only ***R2*** > 0.5 are depicted)..... 166

Table 9-9. Mean values and R^2 comparison between linear models of Lab and Web Conditions for the levels of Speed (x1, x2 and x4)..... 168

Table 9-10. Results of linear regression by levels of Rhythm (Regular, Syncopated). N = 54, 95% Confidence Interval (only ***R2*** > 0.5 are depicted)..... 169

Table 9-11. Mean values and R^2 comparison between linear models of Lab and Web Conditions for the levels of Rhythm (Regular and Syncopated)..... 170

Table 10-1. Demographic data of participants 180

Table 10-2: Main differences and similarities 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. 185

Table 12-1. Power function exponents for five warning parameters according to Hellier and Edworthy (1999) 208

Table 12-2. Independent Variables of the current study 209

Table 12-3. k values used as reference for the creation of the two levels priorities.....212

Table 12-4. Characterization of the participants 215

Table 12-5. Kendall's W values for the MT evaluations, non-MT Evaluations, and all evaluations together.....215

Table 12-6. Results of a hierarchical multiple regression 220

Table 13-1. Demographic description of participants 228

Table 13-2. Descriptions of the sounds used in the comprehension and referent association experiment.....229

Table 13-3. Table of evaluation for the Comprehension test (SAE J2830) 235

Table 13-4. Criteria for the definition of auditory icon approval levels 236

Table 13-5. Final selection of auditory icons for each category.....242

Table 13-6. Comparison between learnability performance metrics form IEC 60601-1-8 (2020) and current study's results 242

Table 14-1. Comparison between operating rooms' LAeq between other countries and our study..... 249

LIST OF FIGURES

Figure 1-1. Mary Jameson reading Anthony Trollope's The Warden on an optophone, ca. 1921. Source: Blind Veterans UK. 26

Figure 2-1. Cassette from 1981 with most of the auditory alarms heard nowadays in patient monitoring systems (Sen, 2020) 31

Figure 2-2. 1982-2020 timeline of the main events of clinical auditory alarms 32

Figure 2-3. Auditory warning sound components. Original illustration taken from Patterson et al. (1986) 35

Figure 3-1. Alarm paradigms: Stimulus and Response-based adapted from Stanton, (1994) 48

Figure 3-2. Important concepts from IEC 60601-1-8 to understand the meaning of Alarm in healthcare settings (AAMI, 2013) 50

Figure 3-3. Systems model of alarms. Adapted from Stanton (1994). In the original "SYSTEM" was "PLANT" 51

Figure 4-1. Interdependence of human-centred design activities (adapted from International Organization for Standardization, 2010) 64

Figure 4-2: Adaptation of Edworthy and Stanton's (1995) representation of auditory warning signals design methodology 67

Figure 4-3: Adaptation of Watson & Sanderson (2007) process for the design of auditory displays with Ecological Interface Design (EID) 69

Figure 4-4: Representation of ISO 9241-210, Edworthy and Stanton (1995), Watson and Sanderson (2007), and Sanz-Segura, (2021) models for design of auditory alarms. Boxes with blue filling were steps and activities considered and implemented on the course of the present research work. 71

Figure 5-1. Example of a patients monitoring station (Source: medicalexpo.com/, Digicare Biomedical Technology) – no visual or sound records were made during the needfinding visit. 79

Figure 6-1. Summary of False/True answers for the Physical Space dimension of the survey (N=46) 99

Figure 6-2. Summary of Frequency answers for the Physical Space dimension of the survey (N=46) 100

Figure 6-3. Most relevant parameters for their tasks according to online survey respondents. 101

Figure 6-4. Ranking of the noisiest surgeries according to survey respondents. 102

Figure 6-5. Summary of True/False answers for the Sound and Alarms dimension of the survey. 103

Figure 6-6. Summary of frequency answers for the Sound and Alarms dimension of the survey. 103

Figure 6-7. Words used to characterize alarm sounds according to survey respondents organized by order of frequency. 104

Figure 6-8. Summary of True/False answers for the Alarm-triggered behaviour dimension of the survey. 105

Figure 6-9. Summary of frequency answers for the Alarm-triggered behaviour dimension of the survey. 106

Figure 6-10. Summary of True/False answers for the Suggestions for improvements dimension of the survey. 106

Figure 6-11. Which vital sign parameter should be redesigned according to survey respondents, organized by order of frequency. 107

Figure 7-1. Evaluation of typical sound pressure levels according to the north American Occupational Safety and Health Administration (Osha.gov) 117

Figure 7-2. Brüel & Kjaer's Handheld Analyzer Type 2270 with a type 4189 microphone. Equipment used for the Frequency Analysis during surgeries. 123

Figure 7-3. The three operating rooms and recovery room in Hospital A where measurements were performed. The red marker indicates the location of the sound level meter. 124

Figure 7-4. Measurement in Vascular and Urology Surgeries in Hospital A 124

Figure 7-5. Measurement position inside the Recovery Room with one patient (not pictured) in Hospital A 125

Figure 7-6. Panoramic view of an operating room in Hospital A 125

Figure 7-7. Surfaces of the operating room of Hospital A 126

Figure 7-8. Outline of the surgical rooms where measurements were taken in Hospital B. The red marker signals the location of the SLM during measurements. Each room indicates which surgery was measured inside. 126

Figure 7-9. Measurement position in Hospital B's Room 1, where the Vascular and Urology surgeries took place. In rooms 2 and 5 the positioning was similar. 127

Figure 7-10. Measurement positioning during the Neurosurgery in Hospital B.....127

Figure 7-11. Measurement position during Orthopaedic surgery in Hospital B..... 128

Figure 7-12. View from an empty operating room inside Hospital B..... 128

Figure 7-13. Average of all surgeries' level measurements in third octave bands. Results for each hospital..... 129

Figure 7-14. Average levels in third-octave bands for each type of surgery in both hospitals. The red line marks the 55 dB(A) conversation level (larger).....130

Figure 7-15. Lmax dB(A) distribution for each surgery 131

Figure 7-16. Average fraction of time, or occurrence rate (%), in given interval of SPL for each type of surgery (larger). 132

Figure 7-17. Third-octave band spectra of the High Priority Cardiac Alarm (red) over the background noise of an Orthopaedic surgery (blue). Overlap of frequencies is represented by the darkest tone. 133

Figure 7-18. Third octave band spectra of the High Priority Oxygenation Alarm (red) over the background noise of an Orthopaedic surgery (blue). 133

Figure 8-1. Summary of $\boldsymbol{f0}$ and frequency range recommendations..... 141

Figure 9-1. Image of the evaluation interface with the semantic differential scale under evaluation 154

Figure 9-2. Depictions of a) Stimuli with 180-ms onset, regular rhythm, speed level 4 (1500 Hz); b) Stimuli with 20-ms onset, 180-ms offset, syncopated rhythm, speed level 1 (500 Hz) 156

Figure 9-3. First screen of the LimeSurvey (LimeSurvey GmbH, 2018) listening test. It included instructions, description of the scales' adjectives, a test sound to adjust the volume and a trial question to test the 0-100 mm slider. 160

Figure 9-4. Correlations between the four acoustic parameters and the six pairs of words – Lab condition. Thick borders include the manipulated parameters..... 162

Figure 9-5. Correlations between the four acoustic parameters and the six pairs of words – Web condition. Thick borders include the manipulated parameters.....163

Figure 9-6. Significant regressions for Frequency as predictor. Relationship between participant's evaluation of a sound as Irritable (***R2***= .54) or Pleasant (***R2*** = .80) and three levels of increasing frequency (Lab). 164

Figure 9-7. Significant regressions for Frequency as predictor. Relationship between participant's evaluation of a sound as Irritable (***R2***= .22) or Pleasant (***R2*** = .33) and three levels of increasing frequency (Web)..... 165

Figure 9-8. Significant regressions for Speed as predictor. Relationship between participant's evaluation of a sound as Fast (***R2***=.71), Positive (***R2***=.60), Preoccupying (***R2***=.73), or Urgent (***R2***=.75) and three levels of increasing speed (Lab) 167

Figure 9-9. Significant regressions for Speed as predictor. Relationship between participant's evaluation of a sound as Fast (***R2***= .56), Irritant (***R2*** = .31), Positive (***R2*** = .39), Preoccupying (***R2*** = .41) or Urgent (***R2*** = .56) and three levels of increasing Speed (Web). 168

Figure 9-10. Significant regressions for Rhythm as predictor. Relationship between participant's evaluation of a sound as Fast (***R2***=.17) and Urgent (***R2***=.14) and two levels of Rhythm (regular or syncopated). 169

Figure 9-11. Significant regressions for Rhythm as predictor. Relationship between participant's evaluation of a sound as Fast (***R2***= .13), Irritant (***R2*** = .15), Pleasant (***R2*** = .09), Preoccupying (***R2*** = .08) or Urgent (***R2*** = .11) and two levels of increasing Rhythm (Web)..... 171

Figure 10-1. Categorization Interface on LimeSurvey 182

Figure 10-2. Percentage of sounds placed in each Category according to participants' profession 184

Figure 10-3. Weighted similarity matrix (item proximity and Quality of Fit) of the final pool of 22 sounds 186

Figure 10-4. Categories in which 22 final sounds (x-axis) were placed..... 187

Figure 10-5. Multidimensional scaling with k=7..... 188

Figure 11-1. Adopted Clinical Alarm Design Strategy..... 194

Figure 11-2. Perceptual correlates of the detection of an alarm. 194

Figure 11-3. Functionality – Perception – Physical relation. With the objective of Detectability, certain percepts need to be triggered and for that, physical components of the sound should be manipulated. 195

Figure 11-4. Perceptual correlates for the comprehension of the Priority of an alarm. ... 196

Figure 11-5. Functionality – Perception – Physical relation. With the objective of understanding Priority, certain percepts need to be triggered and for that, physical components of the sound should be manipulated. 196

Figure 11-6. Perceptual correlates of the identification of the source of an alarm. 197

Figure 11-7. Functionality – Perception – Physical relation. With the objective of Identification, certain percepts need to be triggered and for that, physical components of the sound should be manipulated. 197

Figure 11-8. Perceptual correlates of the understanding of the Tendency of an alarm.... 198

Figure 11-9. Functionality – Perception – Physical relation. With the objective of identifying Tendency, certain percepts need to be triggered and for that, physical components of the sound should be manipulated. 198

Figure 11-10. Anatomy of an Alarm Signal 199

Figure 11-11. Modular structure of an alarm 200

Figure 12-1. Example of a pulse with a) seven harmonics and b) seven harmonics and five inharmonics 210

Figure 12-2. Answering scales of the six listening tests (T1 to T6) (Parizet et al., 2005) ... 211

Figure 12-3. Interface for the Evaluation of Urgency of sounds 214

Figure 12-4. Histogram representing the range of the 100mm scale used by participants. Each bin corresponds to 10 points of the scale. 216

Figure 12-5. Differences in Perception of Urgency between the two levels of the variable Fundamental Frequency. 217

Figure 12-6. Differences in Perception of Urgency between the two levels of the variable Rhythm. 217

Figure 12-7. Differences in Perception of Urgency between the two levels of the variable Inharmonicity. 218

Figure 12-8. Differences in Perception of Urgency between the two levels of the variable Pulse Length. 218

Figure 12-9. Differences in Perception of Urgency between the two levels of the variable Speed (No. Pulses)..... 219

Figure 12-10. Relation between the variables Pulse Length and Frequency in the Perception of Urgency 221

Figure 13-1. The three stages of evaluation of the designed auditory icons with reference to the source of each method..... 227

Figure 13-2. Example page before the comprehension test for 15 auditory icons.....231

Figure 13-3. Page of the referent association test for 15 auditory icons. When participants hovered over each category, they were presented with examples of equipment associated with that category. 232

Figure 13-4. Feedback page presented after each referent-association evaluation. The participant should state how adequate the sound was for that category on a 7-point Likert Scale from “Not adequate at all” to “Totally adequate” 233

Figure 13-5. The three methods of analysis of each evaluation stage with reference to the source of each evaluation method..... 234

Figure 14-1. (Top) Cardiac auditory icon (Bottom) Ventilation auditory icon represented over the background noise of an Orthopaedic surgery. 252

Figure 14-2. Methodological steps followed on this work. The blue squares were implemented, the red squares were not implemented..... 255

Figure 14-3. Guide to optimise the implementation of alarms improvement elements in healthcare (Bach et al., 2018) 258

List of Abbreviations and Acronyms

A/C	Air Conditioned
AAMI	Association for the Advancement of Medical Instrumentation
BSI	British Standards Institute
CO_2	Capnography curve/ carbon dioxide monitoring
dB	Decibel
ECG	Electrocardiogram
ECMO	Extracorporeal membrane oxygenation
ECRI	Emergency Care Research Institute
EID	Ecological Interface Design
fo	Fundamental Frequency
HCD	Human-centred Design
HEPA	High Efficiency Particulate Arrestance
HVAC	Heating, ventilation, and air conditioning
Hz	Hertz
ICC	Intra-class correlation
ICU	Intensive Care Unit
IEC	International Electrotechnical Commission
IOI	Inter-onset interval
IPI	Interpulse Interval
IQR	Interquartile range
ISO	International Organization for Standardization
kHz	Kilohertz
LAF	Sound Level with 'A' frequency and Fast Time weighting
Leq	Equivalent continuous sound pressure level
MAUDE	Manufacturer and User Facility Device Experience
MDS	Multidimensional scaling

NA	Not Applicable
OECD	Organisation for Economic Co-operation and Development
OR	Operating Room
OR(N)	Occurrence rate (dBs)
OSHA	Occupational Safety and Health Administration
PTSD	Post-Traumatic Stress Disorder
PVC	Premature Ventricular Contractions
RR	Recovery Room
SAE	Society of Automotive Engineers
SD	Standard Deviation
SLM	Sound level meter
SNR	Signal to noise ratio
SpO_2	Pulse Oximetry
SRK	Skill-Rule-Knowledge
WHO	World Health Organization

LIST OF APPENDICES

Appendix 1 - Glossary.....285

Appendix 2 –Interview with healthcare professionals: Script.....288

Appendix 3 – Online questionnaire for healthcare professionals.....290

Appendix 4 - Stimuli used in Subjective perception of auditory warning signals 293

Appendix 5 – Stimuli used in Categorization of Sounds.....305

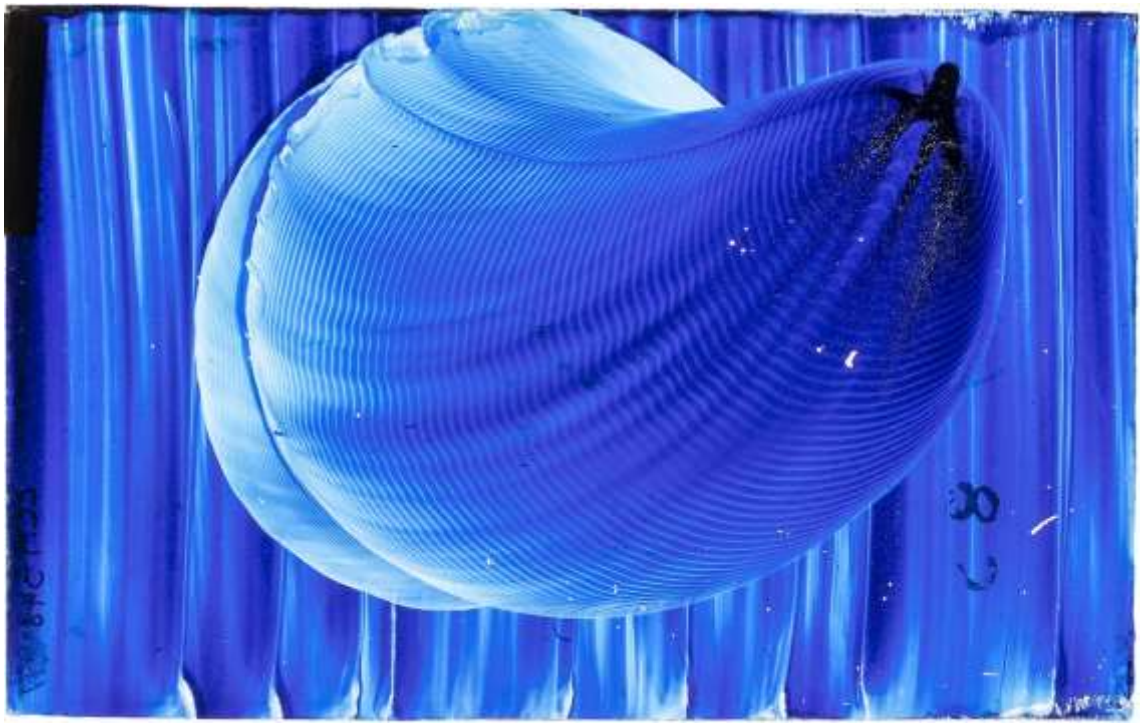
Appendix 6 – Stimuli used in Perception of Urgency 307

Appendix 7 – Final library of Clinical Auditory Alarms 309

Appendix 8 – General organization of studies, timeframe and no. of participants 316

1 General Introduction

The problem of noise in healthcare facilities is intensified by the poor acoustic and informative quality of clinical auditory alarms. The general goal of this thesis is to contribute to an improvement in the healthcare soundscape by proposing a human-centred methodology for the design of clinical auditory alarms and making the alarms less disruptive and more informative. In specific, it will aim at: a) Implementing this human-centred methodology to design clinical auditory alarms from concept to evaluation and validation, combining psychoacoustics methods with usability and user experience methods; b) Understanding which temporal and spectral parameters of sound can be manipulated to communicate clinical information; c) Proposing a library of new clinical auditory alarms.



Margaret Watts Hughes (date unknown). Single pitch Impression Figure, pigment on glass.

1.1 The problem of noise in hospitals

"It (noise) robs the industrious man of his time; it annoys the musical man by its intolerable badness; it irritates the invalid; deprives the patient, who at great inconvenience has visited London for the best medical advice, (…), and it destroys the time and the energies of all the intellectual classes of society by its continual interruptions of their pursuits." (Public Domain Review, 2020). These are the vociferous words of Charles Babbage, co-inventor of the first mechanical computer, written in 1864. The polymath had witnessed the birth of noise with the advent of machinery and the industrial revolution, after millennia of silence (Schafer, 1977).

This aversion to noise gathered a rather large community of silence advocates, of which Florence Nightingale is probably the most famous. The British nurse and social reformer wrote that *"Unnecessary noise is the most cruel absence of care that can be inflicted on sick or well. (…) Every careless clatter or banal bit of banter can be a source of alarm, distress, and loss of sleep for recovering patients"* (Nightingale, 1859).

The increase in technology and machinery has raised the levels of noise in almost all human settings, and hospitals and healthcare facilities have not been spared, often to the patients detriment (Busch-Vishniac, 2019a; Busch-Vishniac et al., 2005). Intrinsic to the background sounds of hospitals are the sounds emitted by medical devices such as ventilators, infusion pumps or pulse oximeters. A wide variety of these medical devices are used throughout Intensive Care Units (ICUs) and recovery and operating rooms, where they monitor the physiological condition of patients. These devices are intended to measure and visually and/or aurally communicate the patient's condition to nearby healthcare professionals. Some devices emit sound as a continuous monitoring tool for staff, while others use sounds as alarms.

Noise in healthcare facilities is a major source of complaints, from patients, professionals, and visitors (Hravnak et al., 2018; Lorenz et al., 2017). Noise from equipment, alarms and voices disrupts patients' sleep and raises their level of stress (Hsu et al., 2012). Successful recovery requires restful sleep, and the lack of it increases patient vulnerability (Similowski et al., 2019).

Alarms are a systemic problem by themselves, in large part because there are too many of them (Block, 2008; Thangavelu et al., 2014). They affect not only patients, but also healthcare professionals. In the clinical environment, this problem is intensified due to

the use of equipment from different manufacturers who use different sounds, which places increased identification demands on the clinical staff (Casey et al., 2018).

A single patient in a surgical floor can trigger hundreds of alarms, and between 80%-99% of them will be unactionable (Gross et al., 2011; Srinivasa et al., 2017). A clinically unactionable alarm refers to when its implied priority is higher than the expected action by the operator. This amount of false alarms has two serious consequences: 1) the constant beeping can contribute to patient delirium (Johansson et al., 2012; Schlesinger et al., 2017); and 2) it also leads to staff burnout, lack of response and apathy, a safety issue referred to as alarm fatigue (Edworthy, 2013b; Kristensen, Edworthy, & Özcan, 2016; Wilken et al., 2017). Direct effects of desensitization to alarms by caregivers can include missed and misinterpreted information, louder voice conversations, louder telephone rings, televisions, in a reaction akin to the Lombard effect, whereby to overcome communication noise, a noisier environment is generated (Bottalico et al., 2020). The cycle leads to stress, frustration, and fatigue, contributing to an overall inefficiency of alarms with potentially serious consequences. Ultimately, alarm fatigue can also lead to the death of a patient because an alarm was ignored, or a device was silenced.

Noise and badly designed alarms are complex problems, and any proposed solution may seem as not enough. If noise and alarms are identified as problems, the most ambitious solution one might come up with would include a completely new and silent soundscape for hospitals. The truth is that current research is heading in two directions: towards disruptive systemic innovation (designing better soundscapes), but also towards incremental improvements in alarm design.

The sound design of medical equipment alarms has been mostly overlooked. Sound is often an afterthought for device manufacturers, who follow a "better safe than sorry" strategy to avoid legal consequences in case a machine does not sound an alarm. There is also a technical legacy where these devices have commonly used very simple piezoelectric speakers, which although robust, small, and cheap, offer limited scope for sound design. For this reason, electronic musician, composer, and author Yoko Sen as a spokesperson from Sen Sound (informal presentation, March 2021) has proposed a three-staged roadmap for the improvement of hospitals' soundscapes:

1) Incremental Change: Redesign legacy sounds from medical devices (for some manufacturers, this will be the first change after roughly 40 years).
2) Radical Change: Design new monitor sounds co-created by healthcare professionals and patients.

3) Transformative Innovation: Design healing environments with soothing soundscapes.

The current research situation places sound designers and researchers of clinical alarms between the first and the second stages of Sen Sound's proposed roadmap. As an example, professor Judy Edworthy – affectionately called "the godmother of alarms" (Rueb, 2019) – has led a revolutionary amendment in IEC60601-1-8 standard, proposing completely new design requirements for clinical auditory alarms.

On the Transformative Innovation side, Sen Sound suggest a change towards a healing environment should be at the heart of a paradigm shift in the sound design of equipment alarms. For example, the Critical Alarms Lab at the University of Delft led by professor Elif Özcan Vieira has developed the CareTunes project (Bogers, 2018). The aim of this project is to create a melodic representation of the patient's physiological signals. A device transcribes a patient's vital signs into songs that sound like electronic dance music. Drums represent the heartbeat; guitars represent oxygen saturation and piano is for blood pressure. When a patient is stable, the tune is harmonious, but it becomes dissonant when a patient's status changes for the worse, thus calling for a caregiver's attention.

1.2 The meaning of a sound

The use of sound for communicating important information to healthcare workers has many advantages and is therefore a widely employed form of machine-human communication. Benefits include allowing listeners to move without the need to keep the device in their visual field, more compliance to auditory warnings over visual warnings (Edworthy & Hards, 1999), attract attention regardless of the listener's current activity (Edworthy & Stanton, 1995), rapid communication and localizable emitting source in the room space.

Even though research has shown that the current alarms have several limitations and possible negative implications, the question of how should alarms be better designed remains unanswered: What sound should be associated with a given information? Should we use abstract or natural sounds? Should we, instead, take advantage of learned associations? The alarm sounds currently used in healthcare sound as an alert. But they can be more informative than that.

In Taiwan, Beethoven's "Fur Elise" means it is time to put the garbage out. In the USA, the same tune means the ice cream truck is coming (99percentinvisible.org, 2016). There are certain tunes that we associate with certain events. This is immediate, effortless, and this

is the power of sounds. Yet, most machines in our daily lives emit simple beeps as status information.

The idea of designing informative sounds is not brand new, although its motivations started on different grounds. In 1912, Edmund Fournier d'Albe, a British physicist, built a device that he called the optophone, which converted light into tones (Mills, 2015). The optophone, or the musical print machine as it became later known, scanned prints and the pattern of light reflected from a given character, triggered a corresponding set of tones in a telephone receiver (Figure 1-1).

Figure 1-1. Mary Jameson reading Anthony Trollope's The Warden on an optophone, ca. 1921. Source: Blind Veterans UK.

It converted typewriting into patterns of audible tones (Chan et al., 2018). The machine sounded the auditory shape of the letter. For instance, the letter "o" would start being scanned from left to right, starting with two close mid tones close to the middle, and as the circle opened, the two tones separated, one going high and the other one going low, and then the tones got together again[2]. The device went through several iterations, but it was adopted by very few people (Mills, 2015).

The direct translation between the visual and the auditory modalities may have determined the lack of success of the technology. We should not design sounds having

[2] see https://vimeo.com/279339111 for an example.

visual stimuli as inspiration. The hearing system has many capabilities, and sounds should be designed for those capabilities. Human's listening systems can identify sound sources, spoken words, and melodies, as well as recognizing patterns. They are mostly able to do so in noisy conditions. For this reason, it is expected for sound in human-computer interfaces be better designed, more informative and more adapted to our actions (Hermann et al., 2011).

This expected relation between information and sound leads to the term *sonification*. Sonification is the data-related generation of sound, exploiting "*the auditory perceptual abilities of human beings such that the data relationships are comprehensible*" (Walker & Nees, 2011, pp.9). Considering sound as a medium for information display, we can systematize auditory displays according to their functions. They can be 1) alarms, alerts and warnings; 2) status, process, and monitoring messages; 3) data exploration and 4) art, entertainment, sports, and exercise (Walker & Nees, 2011).

1.3 Research objectives

The problem of noise in healthcare facilities is intensified by the poor acoustic and informative quality of clinical auditory alarms. The general goal of this thesis is to contribute to an improvement in the healthcare soundscape, based on Judy Edworthy's recent work, by proposing a human-centered methodology for the design of clinical auditory alarms, and making the alarms less disruptive and more informative. In specific, it will aim at:

1. Implementing this human-centered methodology to design clinical auditory alarms from concept to evaluation and validation, combining psychoacoustics methods with usability and user experience methods.
2. Understanding which temporal and spectral parameters of sound can be manipulated to communicate clinical information.
3. Proposing a library of new clinical auditory alarms.

1.4 Dissertation Overview

The current work sets itself in the field of human-centered sound design, particularly focused on the design of clinical alarms of medical devices inside operating and recovery rooms. Although alarm design has been a pressing topic in Human Factors research for decades, the existent sound design methodologies are often theoretical propositions with no visible outcomes or feedback to validate the proposed method.

Chapter 2 presents a historical preview of the issue of clinical auditory alarms. Chapter 3 specifically dwells on the concept of an alarm, priorities of an alarm, and on learnability, workload, masking, and localizability issues associated with auditory alarms. Chapter 4 presents this dissertation's alarm design methodology while comparing it to existent methods. Chapters 5 to 13 are organized according to International Standard ISO 9241-210 Ergonomics of human-system interaction: Human-centred design for interactive systems (International Organization for Standardization, 2010), in a four staged process:

Step 1. Understand and specify the context of use.

Chapter 5 describes an observational visit to an operating ward of a public Portuguese hospital. Chapter 6 presents the results of interviews and survey with healthcare professionals about hospital soundscapes. The stage of gathering evidence was concluded with field analysis inside several surgeries. Chapter 7 presents sound level and frequency analysis of surgeries performed inside the operating rooms of two hospitals.

Step 2. Specify user requirements.

Chapter 8 presents a compendium of design guidelines and requirements for clinical auditory alarms.

Step 3. Produce design solutions.

Chapter 9 makes the first attempt at extracting variables and temporal properties of simple abstract sounds by having participants make subjective evaluations of sounds. This experiment allowed to establish the first mapping between subjective perception and spectral and temporal sound parameters. To test more semantic, information-sound selecting auditory icons to represent categories of alarms. Chapter 11 presents the adopted clinical alarm design strategy which resulted in a prototype set of auditory alarms.

Step 4. Evaluate the design.

Chapters 12 and 13 present the evaluation of the prototype set of auditory alarms with users. We performed a study to validate the perception of urgency of our alarms, and another study to evaluate the comprehensibility and referent association of the alarm set. To conclude, Chapter 14 presents the general discussion, conclusions, and implications of the current work.

The final library of alarm sounds is described in Appendix 7, including links to media files. Appendix 8 presents a systematization of the dissertation's studies, their timeframes and number of participants.

2 "ALARMS ARE STILL A PROBLEM!"

In 2013 Judy Edworthy intitled an editorial for Anaesthesia magazine with a dramatic "*Alarms are still a problem!*" (Edworthy, 2013a). The author has been an important presence throughout the decades of research dedicated to the issue of auditory alarms. This chapter intends to provide a historical preview of the issue of clinical auditory alarms from a Human Factors perspective.

Bain News Service, St. Luke's - A Ward [between ca. 1915 and ca. 1920]. The Library of Congress. No known restrictions on publication.

2.1 Introduction

For the past 40 years clinical technology has developed dramatically. Medical devices are intelligent, often connected to each other into a system, with usable and accessible graphical user interfaces.

One feature has remained mostly unchanged, and that feature is the set of auditory alarms associated with each equipment. Figure 2-1 represents a cassette from March 1981 with seven recordings of auditory alarms. According to informal reports from manufacturers’ design leads, most of the sounds from medical devices in use nowadays come from this 40-year-old tape. And although these patient monitoring sounds are functional, that is not enough anymore (Sen, 2020).

Figure 2-1. Cassette from 1981 with most of the auditory alarms heard nowadays in patient monitoring systems (Sen, 2020)

The history of auditory alarms, from its start to the current paradigm shift in sound design, is more eventful than one might think. In fact, in 2008 Frank Block has added a wonderful chapter by publicly apologizing to the medical community for having contributed to the definition of not so adequate auditory alarms, in his editorial intitled "*For if the trumpet give an uncertain sound, who shall prepare himself to the battle?" (I Corinthians 14:8, KJV)*”.

Very recently, these clinical auditory alarms have suffered profound changes with the update of the standard IEC 60601-1-8 - 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 Equipment and Medical Electric (2020). The standard's update was released in September 2020 together with .wav files of new clinical auditory alarms, aiming at a new soundscape of healthcare environments with not only functional, but informative auditory alarms.

2.2 Historical Context of Clinical Auditory Alarms

The following sections will provide a historical and general context to the design and development of alarms in clinical context. Figure 2-2 shows an overview of the timeline that will be described in the following sections.

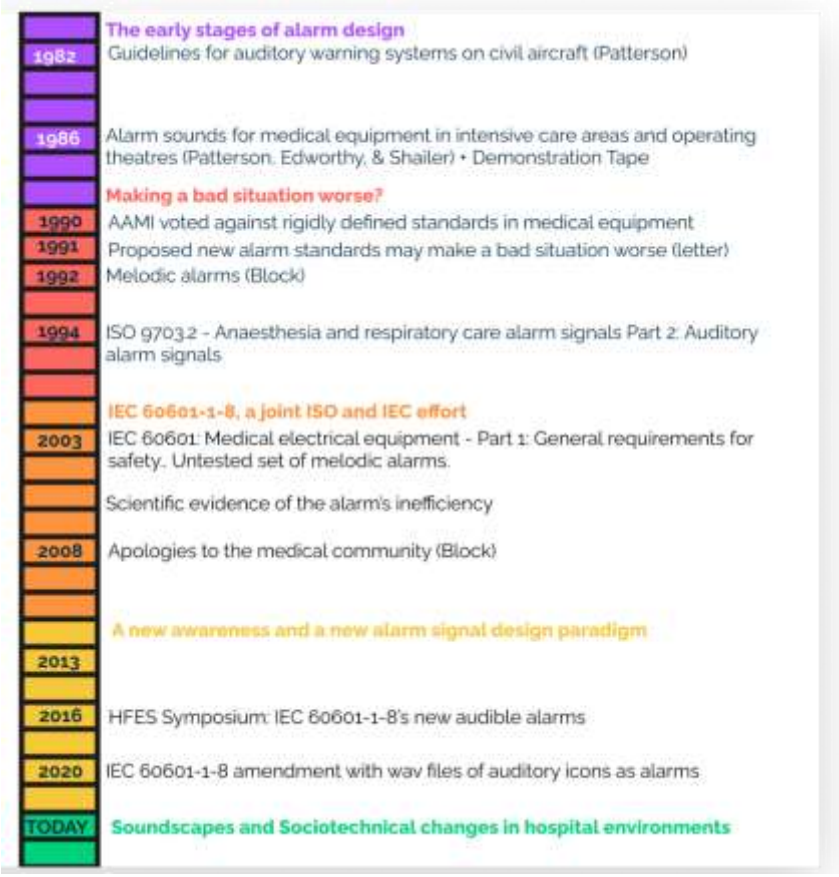


Figure 2-2. 1982-2020 timeline of the main events of clinical auditory alarms

2.2.1 The early stages of alarm design

The modern history of alarm research and design has started with Roy D. Patterson. For the first time, psychoacoustic knowledge and principles were put into practice for the design of alarms in applied contexts. In 1982 he wrote "Guidelines for auditory warning

systems on civil aircraft", a report prepared for the Airworthiness Division of the Civil Aviation Authority focusing on the systems used inside flight-decks of commercial aircraft.

Besides alarm design, questions regarding the number of alarms (some aircraft had 16 alarms), and the role of voice alarms were also considered. Some of the documented problems at the time included (1) the level of warnings, which was too loud, disrupting thought and team communication; (2) the abrupt onsets and offsets of the warnings, evoking startle reactions; (3) the lack of distinctive temporal variation in the alarms, with some being continuous, and very hard to distinguish from one another. In what concerns voice warnings, Patterson (1982) affirmed that they should be used for support and not replace warning sounds, as there was already a lot of speech on the flight-deck and they would take a relatively long time to present their information. This care with sound design, innovative at the time, was requested out of ergonomic reasons. Because the warning sounds were so rude, as the author puts it, the crew turned them off.

Patterson gave great importance to temporal distinctiveness allied to unique spectral components. An example of a description for the "undercarriage unsafe" warning sound went like this: "*A burst of six pulses defines the warning sound. The basic grouping of four clustered pulses followed by two, irregularly-spaced pulses provided the rhythm of the sound which, combined with the spectral characteristics stored in the waveform, gives the sound its distinctiveness.*" He also provided a method of predicting the masked threshold of warnings.

For the first time, the anatomy of an alarm was detailed, and each minor component was isolated and described. Terms such as sound pulse, rise and fall times, shape of the onset, and offset, and pulse spacing, were explained, and related to perceived urgency. Regarding spectral characteristics, Patterson referred frequency limits, pitch, harmonics, harmonic spacing, among many others. The alarms resulting from these design guidelines were validated with real listeners.

Shortly after, Patterson was joined by Judy Edworthy to perform another report, now in the healthcare context. "Alarm sounds for medical equipment in intensive care areas and operating theatres" (Patterson et al., 1986), was a seminal work for clinical auditory alarms. Its goal was to support standards work performed by the Alarms Working Group of the British Standards Institute (BSI), Health Care Committee 16 (HCC 16). The introduction text is an impressive characterization of the ergonomic problems at the time. It is particularly powerful because in 2021 a similar report could start in exactly the same way:

> *"Hospitals are employing an ever-increasing amount of medical equipment with auditory warnings to signal potentially dangerous conditions or equipment malfunction. As a result, there has been a rapid proliferation of auditory warnings particularly in intensive care areas and operating theatres. Generally, the warning sounds are too loud, many are high-pitch tones that are difficult to localise, and most are confusing because they have been introduced without consideration of other sounds in the environment.*
>
> *Currently, auditory signals provided on medical equipment may both resemble other signals with entirely different meanings yet differ from signals with a similar meaning. In practice, staff have to learn a fresh set of warning sounds each time they move to a new environment and confusion can be caused when a new piece of equipment is added to the existing stock. There is, in short, a need to rationalise warning signals for use in hospitals."*

The report was an output of a grant to produce a demonstration warning set to support not only the British but also international standards' work. This demonstration set included seven pairs of emergency and cautionary warnings for the following categories: 1) general; 2) oxygenation; 3) ventilation; 4) cardiovascular; 5) artificial perfusion; 6) drug administration; 7) temperature.

Figure 2-3 illustrates the course of a warning sound with its several components. The upper row shows the basic pulse, which is an acoustic waveform, with rounded onset and offset. The middle row shows the burst, which is a set of pulses in a pattern. The bottom row shows the complete time course of the warning. The waveform in the pulse carries the spectral information which never changes. In the illustrated warning, four clustered pulses (I-VI roman numerals) followed by two irregularly spaced pulses give the warning its rhythm. The inter pulse interval is varied within the burst to vary the perceived urgency (Patterson & Mayfield, 1990). The height of the "houses" indicates the relative intensity of the bursts. The designed alarms were based on the previous guidelines established by Patterson for civil aircraft, and the tape can be heard here[3].

The efforts for clinical alarm standardization continued, and the so-called "Patterson-Edworthy sounds" were implemented in a Fisher-Pakel humidifier. However, their use in anaesthesia was controversial, and American and international standards organizations did not reach a consensus regarding alarm tones for medical equipment.

[3] http://th.id.au/alarms/sounds/Patterson%20Alarm%20Tape%20Mono.mp3

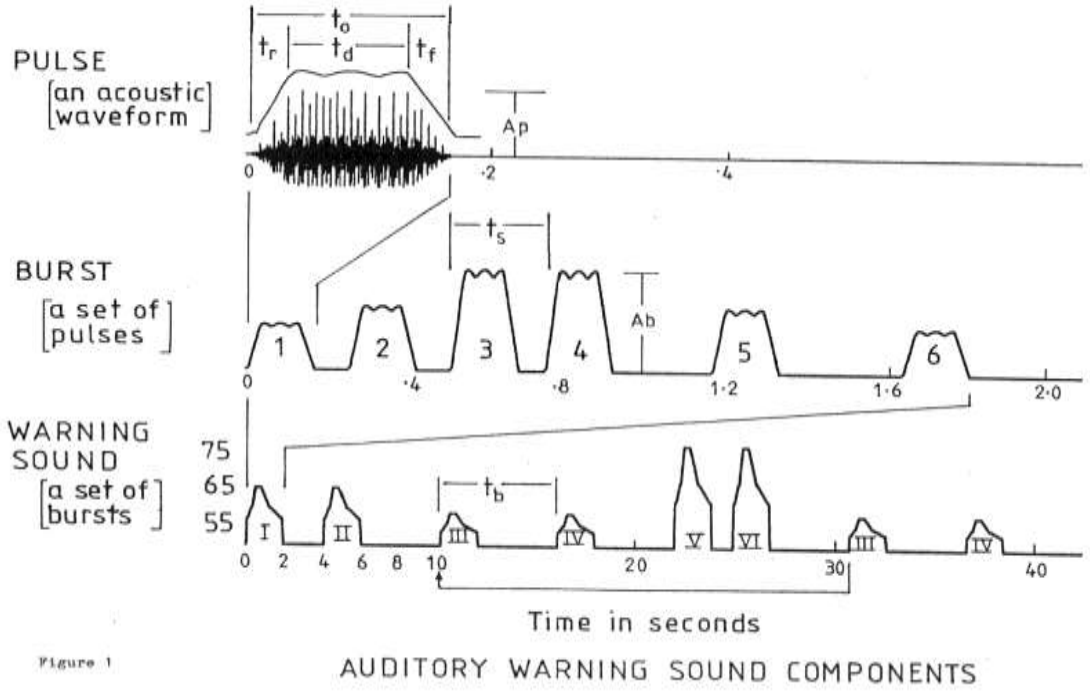


Figure 2-3. Auditory warning sound components. Original illustration taken from Patterson et al. (1986)

ISO TC1 SC3 WG1 ended up not endorsing these alarms.

2.2.2 Making a bad situation worse?

Then *"Proposed new alarm standards may make a bad situation worse (letter)"* happened. A letter signed by Matthew Weigner, M.D. sent to the *Anesthesiology* publication stated that because these sounds and the studies who used them were more than ten years old, they might no longer be applicable (Wiegner, 1991). The argument was that advances in monitoring technology might not be compatible with the alarm requirements, and their use in non-integrated devices would only worsen noise and stress in operating rooms (OR) environments. Importantly, the author of the letter claimed the need to test the alarms not in individual devices but considering the entire OR environment. In October 1990, the Association for the Advancement of Medical Instrumentation (AAMI) voted against rigidly defined standards in medical equipment until significant scientific studies or clinician's input was obtained. Frank Block (1992) added that the tones have never been tested inside an operating room and that "*some observers have suggested that these tones would be difficult to learn and would provoke emotions of amusement, or possibly offense, in a clinical environment.*" (Block, 1992, pp. 286).

In fact, Frank Block starred the next chapter of clinical alarm design and standardization. While researching an alternative for the Patterson-Edworthy sounds, Block reviewed more than 2000 classical and popular tunes. Six songs were selected to represent each of the six alarm systems' structure (Table 2-1). All were no.1 hits in USA, except for "I Left my Heart in San Francisco". No theoretical rationale was provided to back the shift towards melodic alarms (Block, 1992).

Table 2-1. Six tunes selected to represent each of the six alarm systems or categories (Block, 1992)

CATEGORY	SONG TITLE	ARTIST	YEAR
Oxygenation	Love is Blue	Paul Mariat and His Orchestra	1968
Ventilation	The Wayward Wind	Gogi Grant	1956
Cardiovascular	I Left my Heart in San Francisco	Various Artists	1954
Temperature	Chariots of Fire	Vangelis	1982
Artificial Perfusion	The Twist	Chubby Checker	1960
Drug Administration	Raindrops Keep Fallin' On My Head	B.J. Thomas	1970

All songs were converted to sine waves to match the pure tones found on monitors. These melodies were tested in a panel discussion with 79 participants, and results suggested that familiarity with pop songs could help produce a set of familiar alarms. Participants chose the correct category better than chance. There were differences due to nationality, and further studies were said to be needed to better comprehend the efficacy of these sounds, as well as determining how to change them since there would have been copyright issues.

In 1994, ISO TC1 SC3 WG1 released "ISO 9703.2 - Anaesthesia and respiratory care alarm signals Part 2: Auditory alarm signals". This edition did not make use of any of Patterson's psychoacoustic cues except for the simplified version of Patterson's 'general' sound. The sounds were based on simple beeps, and no category-specific sounds were included. Alarm signals were prioritized according to high (five-beat rhythm played twice at 10 second intervals), medium (three-beat rhythm played once at 30 second intervals), or low priority. The standard specified a range of fundamental pitches and required at least four harmonics to improve detection and direction location of the alarm signal. One of the problems pointed to this standard was that manufacturers opted to use a fixed pitch. This resulted in similar devices making different sounds and different devices making similar sounds.

2.2.3 IEC 60601-1-8, a joint ISO and IEC effort

In the early 2000's a Joint Working Group from the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) started working on a collateral standard to IEC 60601: Medical electrical equipment - Part 1: General requirements for safety. This collateral standard became what we know today as IEC 60601-1-8, `General requirements and guidelines for the application of alarms in medical electrical equipment.'

IEC 60601-1-8 was an improvement compared to ISO 9703.2 because it provided guidance on how to implement psychoacoustic cues, and it defined standard melodies for the six categories, including a new category: Power Down. Every sound was assigned a mnemonic and a changing pitch suggesting the nature of the device (for instance, rise-and-fall for the Ventilation category resembling mechanical ventilation). As examples for the alarms' mnemonics, the Oxygen alarm signal sounded like the word "ox-y-gen" and the cardiovascular alarm sounded like the word "car-di-o-vas-cu-lar". The alarm signals proposed by Block et al. (2000) are in Table 2-2. The second and third columns indicate the alarm pitch's respective musical key using the ABC musical notation. As an example, C4 corresponds to the middle C note.

Table 2-2. Proposed standard-compatible, device-specific alarm sounds. Source: Block et al. (2000)

AUDITORY ALARM CATEGORY	MEDIUM PRIORITY ALARM (MNEMONIC)	HIGH PRIORITY ALARM (MNEMONIC AND OTHER INFORMATION)	TYPICAL EQUIPMENT
General	C4 C4 C4	C4 C4 C4 - C4 C4 (Fixed pitch)	Other equipment which does not readily fall into one of the following categories, including but not limited to electrical or non-oxygen gas supply systems, EEG monitors, intracranial pressure monitors, laparoscopic gas insufflation systems, calfcompressor systems, etc. Optionally this sound may be used by any medical equipment
Oxygen	C5 B4 A4 (OX-Y-GEN)	C5 B4 A4 - G4 F4 (OX-Y-GENA-LARM; slowly falling pitches; top of a major scale; falling pitch of an oximeter)	Pulse oximeters, transcutaneous/ tissue oxygen monitors, oxygen analyzers, oxygen concentrators, oxygen gas supply lines

(CONT.) AUDITORY ALARM CATEGORY	MEDIUM PRIORITY ALARM (MNEMONIC)	HIGH PRIORITY ALARM (MNEMONIC AND OTHER INFORMATION)	TYPICAL EQUIPMENT
Ventilation	C4 A4 F4 (VEN-TI-LATE; RISE AND FALL)	(VEN-TI-LA-TI-ON; VEN-TI- LATE A-LARM; RISE AND FALL AND FALL Old `NBC Chime'; inverted major chord; rise and fall of the lungs)	Workstations which included artificial ventilators (but which do not include cardiac monitors); artificial ventilators, spirometers, CO_2 monitors, ventilator disconnect (airway pressure) monitors, etc.
Cardiovascular	C4 E4 G4 (CAR-DI-AC)	C4 E4 G4 - G4 C5 (CAR-DI-AC A-LARM Trumpet call; call to arms; major chord)	Workstations which include cardiac monitors, multi-parameter monitors which include cardiac monitors, heart rate monitors, invasive or non-invasive blood pressure monitors, cardiac output monitors, peripheral perfusion monitors (plethysmographs), transesophageal echo, fetal heart rate monitors
Temperature or delivery of energy	C4 D4 E4 (TEM-P'RA-TURE)	C4 D4 E4 - F4 G4 (TEM-P'RA-TURE A-LARM; slowly rising pitches; bottom of a major scale; related to slow increase in energy or (usually) temperature)	Temperature monitors, heated air humidifiers, infant radiant warmers, neonatal incubators, patient heating or cooling systems, blood or fluid warmers; electrocautery, ultrasound, laser, X-ray or MRI systems, nerve stimulators
Drug delivery	C5 D4 G4 (IN-FUS-ION)	C5 D4 G4 - C5 D4 (IN-FUS-IONA-LARM; Jazz chord (inverted 9th); drops of an infusion falling and `splashing' back up)	Volumetric infusion pumps, syringe drivers, anesthetic agent delivery systems or analyzers
Artifcial perfusion	C4 F#4 C4 (PER-FU-SION)	C4 F#4 C4 - C4 F#4 (PER-FU-SION A-LARM Artificial sound; Tri-tone; similar to `Yo-ee-oh' of the Munchkins in `The Wizard of Oz')	Cardio-pulmonary perfusion pumps (`heart-lung machines') and associated equipment, intra-aortic balloon pumps, ventricular assist devices; artificial hearts; renal dialysis systems
Power failure	C5 C4 C4 (POW-ER FAIL) (GO-INGDOWN)	C5 C4 C4 C5 C4 (POW-ER GO- ING DOWN; falling pitch as when the power has run down on an old Victrola)	Any device when it experiences loss of power or other major failure of the device
Low priority alarm	E4 C4 (low priority) (IN-FO; MESS-AGE; ding-dong; doorbell or hostess call)	--	

The authors of this proposal claimed formal testing was needed for this set of alarm signals, to avoid the problems from the previous standards which did not go through user testing. Unfortunately, and similarly to the previous standards, Table 2-2 represents very closely what was proposed in IEC 60601-1-8 's Annex F (ANSI/AAMI/ IEC 60601- 1-8:2006 & A1:2012, 2013). The proposed alarms were selected for the IEC update without undergoing formal testing.

One of the first scientific studies using IEC's set of alarm signals reported their urgency was inappropriate (Mondor & Finley, 2003). In this study, the repetition rate did not influence urgency ratings. The most dominant variable in determining urgency was pitch. Authors suggested that using the number of pulses as a vehicle for urgency was a poor design choice as listeners needed to wait until the end of the signal to interpret its urgency.

Regarding discrimination of overlapping alarms, another study has shown that the melodic alarms could not be discriminated (Lacherez et al., 2007). Out of 14 nurses, only 2 could identify all alarms by the end of the experiment, and their comments were that the IEC 60601-1-8 alarm sounds were too similar. To simulate the environment in which nurses are likely to hear the alarms, the authors have varied not only the overlap (after the first note or after the third note), but also the workload while learning the alarms. Participants performed an arithmetic task, and it did not worsen accuracy at identifying alarms (Lacherez et al., 2007). A similar study intended to evaluate the effect of the alarm's mnemonics on clinicians' performance in identifying the IEC 60601-1-8 melodic alarms (Wee & Sanderson, 2008). No significant effect of mnemonics was observed on speed, accuracy, or retention, undermining the mnemonics' main goal of helping clinicians to map melodies to systems. The authors state that mnemonics work better when people generate their own. Regarding retention, after two learning sessions only 1 of the 22 nurses obtained 100% accuracy at identifying the alarms in two successive tests. As for the interpretation of priorities, performance was systematically faster and more accurate for medium priority alarms than for high priority alarms – even though the perception of urgency was correct.

Following evidence that the alarms were not easily learned by professionals, Frank Block wrote a public apology to Roy Patterson and the community in general in a 2008 editorial of *Anesthesia and Analgesia* (Block, 2008). Block clarified Patterson's intention with the Patterson-Edworthy sounds, which were never clear to the evaluating committee. The sounds were never designed to coexist in one environment. Instead, a general sound should capture the healthcare professional's attention to a central alarm display in an

integrated workstation to identify the problem. Only on rare occasions would the organ/category system sounds be used – the cardiovascular alarm signal in a cardiac care unit. Additionally, the alarm signals were designed to be discriminated, and they varied in timbre, tempo, rhythm, and number of pulses. Unbeknownst to the standards groups and critics, all alarm signals had a mnemonic in English, French and German. As an example, the Oxygen alarm signal sounded like the word "ox-y-gen" and the cardiovascular alarm sounded like the word "car-di-o-vas-cu-lar." It became clear that IEC 60601-1-8 set of alarms were not consistently related and resulted in two different sounds for each organ system or category (3-note and 5-note patterns for medium and high priority, respectively). While Patterson designed 8 alarms, the new set of alarms consisted of 17 sounds, which was much more than the literature reported people could learn. As a recommendation, Block suggested designing a new set of sounds using the Patterson-Edworthy sounds as basis for modification and improvement. As an apology, Block wrote:

> *"This author would like to take this opportunity to apologize to the medical community for his role in derailing the Patterson alarm sounds nearly two decades ago. I now believe that the Patterson sounds were genius, and that they should have been adopted 10 or 20 years ago."*
>
> – Frank E. Block (2008)

2.2.4 A new awareness and a new alarm signal design paradigm

The issue of clinical alarms was slowly brought to the spotlight not by researchers and some clinicians as up until then, but by hospitals and healthcare institutions. In 2011 a summit focused on Medical Device Alarms happened on Herndon, Virginia. Its desired outcome was a list of priority issues that the healthcare community could commit to address related to clinical technology alarms. As with all complex scenarios, it was concluded that the problem and the solution were multi-dimensional – but achievable over the next 10-20 years.

In 2013, AAMI - Association for the Advancement of Medical Instrumentation made a webinar intitled "How to Identify the Most Important Alarm Signals to Manage". After analysis of the Manufacturer and User Facility Device Experience (MAUDE) Database, they were able to determine the most common types of alarm problems. Between January 2010 and December 2012, 393 occurrences were reported, 383 of which were Death Reports. 306 included "Device Alarm System Issue" and 63 referred to "Not Audible Alarm". Most were related to the Patient Monitor, followed by the Hospital Ventilator. Alarm Issues

have been reported since 2012 (until 2019) as one of the Top Ten Health Hazards according to ECRI Institute (Table 2-3). This list highlights technology safety topics that warrant attention.

Table 2-3. ECRI Institute's Top 10 Technology Hazards mentioning Alarms between 2012 and 2019.

YEAR	PLACE	NAME	DESCRIPTION
2012	1st	Alarm Hazards (ECRI Institute, 2012)	Alarm fatigue, urgency mapping, restored settings, etc.
2013	1st	Alarm Hazards (ECRI Institute, 2013)	The sheer number of alarms has become problematic. Caregivers can become overwhelmed trying to respond to the alarms, or they can become desensitized, which can lead to missed alarms or delayed response, placing patients at risk.
2014	1st	Alarm Hazards (ECRI Institute, 2014)	Excessive numbers of alarms—particularly alarms for conditions that aren't clinically significant or that could be prevented from occurring in the first place—can lead to alarm fatigue, and ultimately patient harm.
2015	1st	Alarm Hazards (ECRI Institute, 2015)	Inadequate Alarm Configuration Policies and Practices
2016	2nd	Missed Alarms (ECRI Institute, 2016)	Missed Alarms Can Have Fatal Consequences
2017	3rd	Missed Alarms (ECRI Institute, 2017)	Missed Ventilator Alarms Can Lead to Patient Harm
2018	4th	Missed Alarms (ECRI Institute, 2018)	Missed Alarms May Result from Inappropriately Configured Secondary Notification Devices and Systems
2019	4th	Improperly Set Ventilator Alarms (ECRI Institute, 2019)	Improperly Set Ventilator Alarms Put Patients at Risk for Hypoxic Brain Injury or Death

On an ambitious note, AAMI's President has issued in 2011 the challenge that "*by 2017, no individuals will be harmed by adverse alarm events*" (Logan, 2011). Additionally, alarm management has been established as a National Patient Safety Goal in 2013 by the Joint Commission, USA's largest hospital accreditation organization. They issued a sentinel event alert warning of the dangers of alarm desensitization and urging hospitals to audit up alarm management protocols and alarm setting guidelines (McKinney, 2014). Strategies of the healthcare facilities included identifying the most important alarm signals to manage and making alarm audits analysing alarm load, alarm notification and alarm content (ECRI, 2014). Hospitals mobilized towards that goal and simple and effective approaches resulted in fewer alarms (ECRI, 2013).

At the same time, in the research arena, Judy Edworthy (2013b) writes a review about clinical auditory alarms where an important reflection is made about the type of alarms proposed in the standards. To the best of our knowledge, this is the first moment where auditory icons are recognized as a possibility as clinical alarms (Ulfvengren (2003) had already used them in aviation). According to the author, it seemed intriguing that alarms had so many recognition problems when our daily lives are populated with hundreds of sounds which we know and understand. These sounds are usually produced by the objects or events that make those sounds, and there is a direct mapping between sound and event (a tyre skidding means a car might be about to crash). These are referred to as auditory icons. Importantly, the strength of the relationship between sound and object it represents determines the ease with which it can be learnt (Edworthy, 2013b).

This idea was put to practice shortly after with a study comparing three sets of alarms in terms of their learnability (Edworthy et al., 2014). The authors used the tonal alarms specified in IEC 60601-1-8, a set of abstract alarms randomly selected from a database and a set of auditory icons, alarms designed as indirect metaphors of the eight standard functions. Results indicated a significant difference in learnability. Auditory icons were learned significantly better, and the randomly selected abstract alarms were learned significantly better than the alarms specified in the standard. The authors hypothesized that the first had more acoustic variability than the latter. The main conclusion of this study was that auditory icons could be serious candidates for alarms in the clinical setting. The idea to start revisiting the design of alarms in the IEC standard starts blossoming after studies showing that the early Patterson-Edworthy sounds were easier to learn than IEC's set of alarms (Atyeo & Sanderson, 2015; Edworthy, 2015). Its intent is officially announced at a Human Factors and Ergonomics Society symposium (Edworthy & Baldwin, 2016) where several members discussed the idea of developing new audible alarms for IEC 60601-1-8 (Edworthy et al., 2016), as well as a framework for evaluation (design, usability, testing sounds in realistic work environment, and masking).

The goal goes beyond alarm signals' design, and the categories of risk are also discussed (Edworthy et al., 2017). IEC 60601-1-8 specifies eight risk categories, based on the six proposed by Kerr (1985) (Hypoxia, Ventilation problems, Cardiovascular problems, Interruption to artificial perfusion, Drug administration error and Thermal risk) which were based on the ways tissue damage could occur. This results in the same medical device producing more than one alarm. Other proposals include equipment-based alarms, but this would probably result in a proliferation of alarms *au-par* with a proliferation of devices. Conversely, the causes for tissue damage are unlikely to change, and this makes

Kerr's system future-proof, as well as with a reduced number of categories, which seems advantageous from a Human Factors perspective (Edworthy et al., 2017). From an ICU's clinician's perspective, an anaesthesiologist in particular, categorization can make sense in terms of immediate harm or in terms of who needs to hear the alarm. In the OR, Kerr's categorization may not be as useful. In this situation, increasing the number of cardiac categories or subdividing it may be beneficial (blood pressure, heart rate, heart rhythm, cardiac output). This categorization choice could increase the amount of auditory information conveyed during periods when attention is directed elsewhere (Edworthy et al., 2017).

In 2020, IEC 60601-1-8 released its major update (2020), specifically in what regards its alarm signal specifications. Following numerous research demonstrating the increased learnability, localizability, resistance to masking and recognizability of auditory icons (Bennett et al., 2019; Bolton et al., 2018; Edworthy et al., 2017, 2018; McNeer et al., 2018), they were included in the new standard as a proposal for alarm signals for alarm systems (Table 2-4).

Table 2-4. Description of the auditory icons in the 2020's amendment of the IEC 60601-1-8 standard

FUNCTION OF ALARM	AUDITORY ICON METAPHOR
Equipment or supply failure	Starting up a motor that shuts down suddenly
Cardiovascular	"Lup-dup"; heartbeat sound
Perfusion	Liquid disturbance, water churning, bubbles
Drug administration	Shaking pill bottle
Oxygen	Irregular, stylized dripping/saturation
Ventilation	A single inhale followed by an exhale
Temperature	Whistling kettle

The categories have remained the same, and the structure of each alarm is as follows: Priority Pointer + Auditory Icon

There are three priority pointers, for low, medium, and high priority alarms. The auditory icon is only associated with medium and high priority pointers. The "General" category has no auditory icon associated, only a pointer (high priority) is used. There are seven auditory icons, each for a risk category. The pointer and the icon overlap so that the maximum amount of information is provided in the least amount of time.

The first description of these auditory icons can be found in Edworthy et al. (2014), and the process that led to their choice consisted in searching in sound libraries for indirect

metaphors for the alarm functions. Team meetings with psychoacoustic researchers were used for selection until agreement was reached over how adequate were the sounds. Criteria included: 1) each could be readily, but indirectly, associated with IEC functions; 2) that there was acoustic variability across the set; 3) and that the sounds had the potential to be acoustically appropriate and audible in the work environment.

The current state of the art in clinical alarms is as stated in this standard. While auditory icons are suggested as alarm signals for medical devices, sonic branding (equipment brands having their own sonic identity) is considered as a real and current trend. For this reason, manufacturers are encouraged to develop and test their own set of alarms, as long as they abide by the standard's performance metrics.

2.2.5 Soundscapes and Sociotechnical changes in hospital environments

The reality is that alarm signals are but a small part of a hospital acoustic environment. The soundscape, or sonic environment, can deeply affect the patient experience. Patients are more and more considered as consumers or guests. As such, their stay is a curated journey with the goal of providing them a good experience (Spence & Keller, 2019). Research shows that hospitals with a better patient experience do better financially (Trzeciak et al., 2016). A classic study has already provided evidence how of little environment changes like being inside a room with a window looking on a natural scene shortened postoperative hospital stays and gave fewer negative evaluations to nurses (Ulrich, 1984).

In recent years there has been an effort to investigate how healthcare facilities' soundscape affects patients' recovery time, dignity and, overall well-being of patients and professionals alike (Ecophon, 2018). Inside a hospital environment, hearing becomes more important and relevant to a patient because the visual environment is dull and boring (Spence & Keller, 2019). Electronic musician, composer and author Yoko Sen from Sen Sound has been a prominent voice in drawing attention to this systemic problem, pointing out that healthcare facilities' soundscapes are closer to noise than to melodic sounds, provoking annoyance, and negatively affecting patient experience (Sen & Sen, 2020). Past investigations have reported on the longer recovery times and health risks on patients' exposure to unsuited noise levels (Buxton et al., 2012; Science Daily, 2012). Sleep deprivation affects the immune system, directly affecting mortality rates in hospitals. This means that uncontrolled hospital noise can be directly considered as dangerous.

Some hospitals are putting great efforts into reducing their noise levels or even designing healing soundscapes (Rueb, 2019; Ruiter, 2015). The New Children’s Hospital in Helsinki partnered with Aalto University, and created a series of ambient soundscapes (International Sound Awards, 2019). Each floor represents a theme such as ocean, jungle or forest, and all spaces are designed to make children comfortable. Also Brian Eno is known to have designed a quiet room in the Montefiore hospital (Brown, 2013) that soothed and calmed patients with what he called functional music. Similarly, the project Sonic Environments for Healing used generative music systems to improve the atmosphere of hospital waiting rooms (Lazarevic, 2015). MacDonald and Schlesinger (2018) created a device that would control music in the operating room by integrating it with the vital signs from the anaesthesia monitor. This device would automatically reduce the volume of the music whenever it detected critical events such as a slowing heart rate, reduced oxygen saturation or blood pressure, to allow the team to better communicate and concentrate on the patient. More recently, the Critical Alarms Lab at the University of Delft has developed the CareTunes project (Bogers, 2018; Özcan et al., 2021), aiming at a melodic representation of the patient’s physiological signals.

2.3 Conclusion

The design of alarms and of healing soundscapes is becoming more and more a topic of interest, and manufacturers and hospitals are increasingly following human and patient centred design processes. In a very interesting exercise of discussing the right approaches in the current state of things (Ozcan et al., 2018), the authors consider the technological advances and solutions, and the ways society engages with them by means of continuous tracking of heart rate, sleep, or behaviour. According to their predictions, very soon the number of stakeholders interested in and influencing the nature of alarms will increase. Besides constant monitorization, existent technologies such as noise cancellation could be applied for patient settings, as well as the use of spatialized alarms so that healthcare professionals easily detect the source of the alarms in the midst of a populated recovery room. Multimodality is also being studied with haptic alarms replacing auditory alarms. Independently of the way technology and research lead their efforts, a sound, robust, holistic, and collaborative approach to design is necessary.

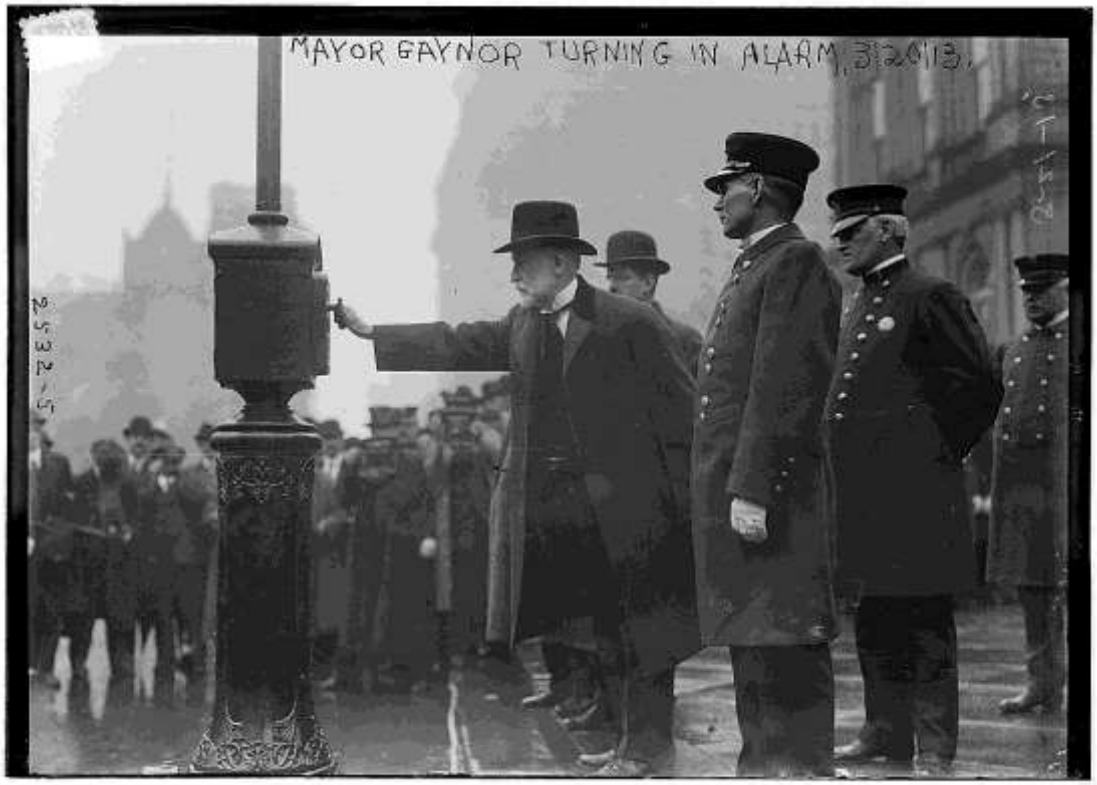
3 SIRENS, CHURCH BELLS AND ALARMS

"In Greek mythology the Sirens were nymphs who destroyed those who passed their island by means of their singing, at once piercing yet dulcet as honey. Circe warned Odysseus of the Sirens and so enables him to elude their fatal song by plugging the ears of his men with wax and having himself bound to the mast of his ship. The Sirens signify mortal danger to man and this danger is broadcast by means of their singing. (...).

Sirens and church bells belong to the same class of sounds: they are community signals. As such, they must be loud enough to emerge clearly out of the ambient noise of the community. But while the church bell sets a protective spell on the community, the siren speaks of disharmony from within."

- Schafer (1977)

This chapter includes important concepts for clinical auditory alarms' design.



Bain News Service (1913), *Mayor Gaynor turning in alarm*. The Library of Congress. No known restrictions on publication.

3.1 What is an alarm?

The term "alarm" has two common uses: it is used to refer both to a stimulus and to a response. What could be a problem of definition, reflects itself in different paradigms (Figure 3-1). A *stimulus-based* model states that the existence of an alarm in the environment influences the individual. The *response-based* model proposes that the stimulus causes an alarm state in the individual, leaving space for interpretation of the stimulus, depending on the individual exposed to it. This differs from the stimulus-based model in the sense that the latter considers all alarms to be homogenous and equally identified as alarming by all individuals. According to Stanton (1994), the stimulus-based model is favored by the engineering approach, while the psychological approach believes in the response-based model.

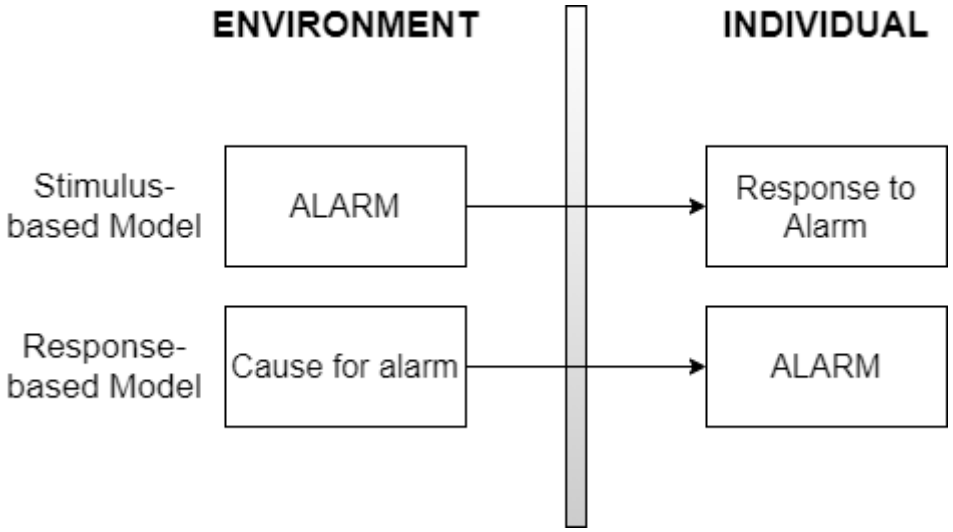


Figure 3-1. Alarm paradigms: Stimulus and Response-based adapted from Stanton, (1994)

This difference in interpretation has heavily influenced our auditory alarms. Usually beeps or buzzers, they often startle and irritate, and are consequently deactivated. For this reason, very early on the use of technology and products with auditory alarms, multidisciplinary Human Factors teams have focused on the design of auditory alarms.

Efforts to combine both perspectives include the progressive definition proposed by Us et al. (2011), represented in Table 3-1. The definition stating that "*an alarm is a signal which signifies to the operator that a response requiring state threatens or prevents the accomplishment of a goal*" is generic enough but is missing an important concept when we transpose it to the healthcare context. The concept of thresholds and limits is of the utmost importance in clinical alarms. For this reason, we searched for a healthcare application of the alarm concept.

Table 3-1. The progressive definition of an alarm (Us et al., 2011)

DEFINITION	EXPLANATION
An alarm is a signal signifying to an operator that an abnormal state has occurred.	The signal: 1) alerts the operator, arises attention, warns and gives notice → perceptual level 2) indicates danger, malfunction, error condition, process deviation, and unexpected event → cognitive level (operators' expectations, experience (i.e. norms) or knowledge). 'Abnormal state' comprises process deviation, error condition, malfunction, and neared or overrun limits.
Alarm is a signal which signifies to the operator that an abnormal state needs a response.	'Response' means that a reaction is required. An abnormal state is a state which threatens or prevents the accomplishment of a goal.
Generic definition of alarm: "An alarm is a signal which signifies to the operator that a response requiring state threatens or prevents the accomplishment of a goal."	

Importantly, clinical literature does not refer to an alarm, but to alarm systems and alarm signals. The term alarm system is used to make it clear that the alarm signal is not independent from the parts of medical equipment that detect alarm conditions and generate alarm signals (AAMI - Association for the Advancement of Medical Instrumentation, 2009). Alarm systems indicate the presence of (generally) temporary, clinically relevant, or potentially dangerous conditions in real time, and alarm signals are the signals generated by alarm systems. IEC 60601-1-8 standard provides important concepts to understand the main definition of Alarm Condition, referring to the state of the alarm system when it has determined that a potential or actual hazardous situation exists for which operator notification is required (Figure 3-2).

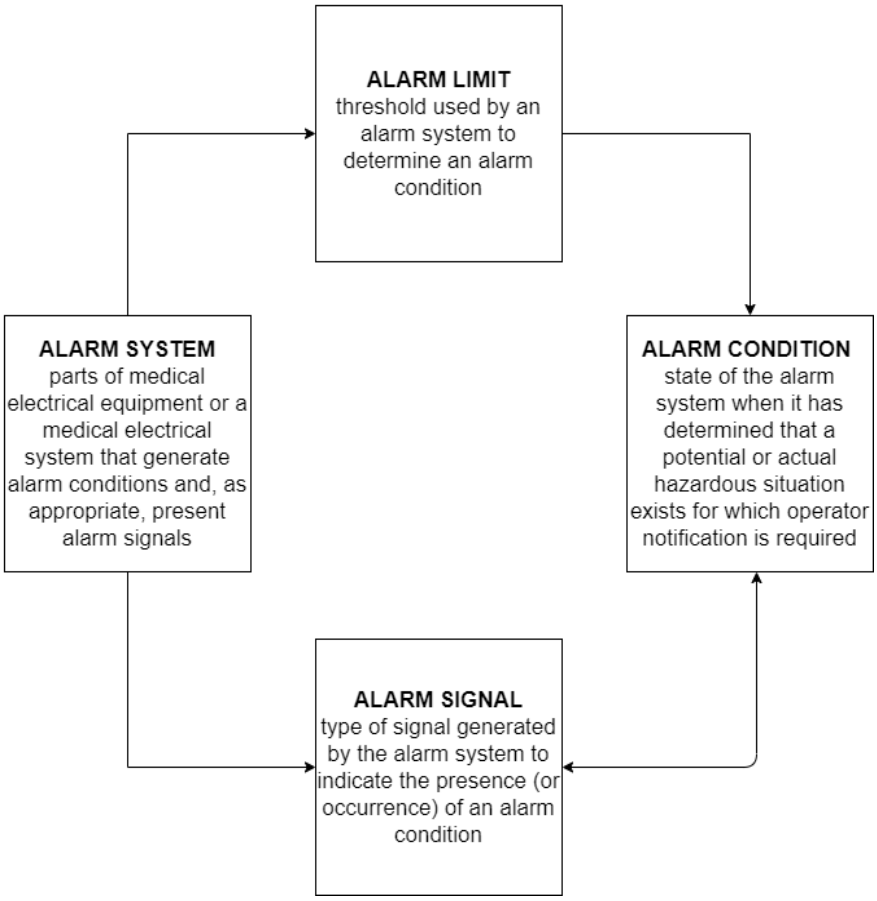


Figure 3-2. Important concepts from IEC 60601-1-8 to understand the meaning of Alarm in healthcare settings (AAMI, 2013)

For the purpose of the current work, we will consider the following definition of an alarm:

An alarm is a signal which signifies to the operator that a response requiring state is occurring. This signal is activated by an alarm system whenever a crossing of an alarm limit is detected and thus generates an alarm condition.

Stanton (1994) has characterized the flow of information between the system and the user considering the role of the alarm of warning the operator of impending danger, in different degrees of severity. The systems model is represented in Figure 3-3 and details the entire process of activating and responding to an alarm. According to this model, the individual's attention must be attracted to the area where the alarm lies, and the alarm must communicate information about the occurring event. The individual then should acknowledge the alarm, and decide what action to take next, according to the given information. The input given by the individual may be the correct one, or the incorrect one - thus leading to errors in the system.

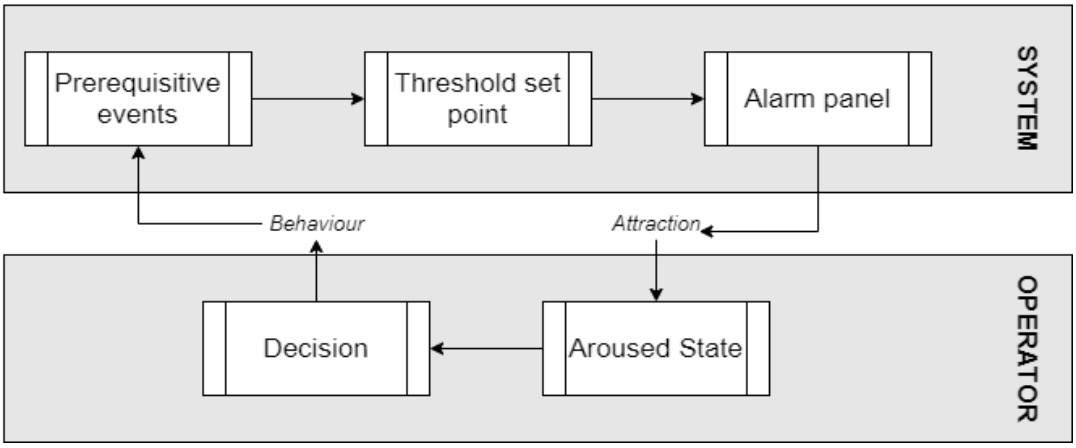


Figure 3-3. Systems model of alarms. Adapted from Stanton (1994). In the original "SYSTEM" was "PLANT".

The concepts of severity and priority are also extremely important when describing alarms. In the case of clinical alarms, their priorities are established by manufacturers. These are assigned after risk analysis considering severity and rapidity of onset of harm if the alarm condition is not attended to. The priority of the alarm should suggest to the operators the speed of their response to the alarm condition. The speed of response could be Immediate, Prompt, or Delayed. Examples from events in each category are presented in Table 3-2.

Table 3-2. Alarm priorities for clinical alarms. Examples from IEC 60601-1-8.

SPEED OF OPERATOR RESPONSE	EXAMPLES	IEC 60601-1-8 PRIORITY OF ALARM SIGNALS
Immediate problems which are likely to cause patient injury or death within seconds to several minutes if uncorrected.	Asystole; Ventricular fibrillation; Failure of a cardiac support device (intra-aortic balloon pump, cardiopulmonary bypass machine); Sustained high airway pressure; Extreme hypoxemia; Sustained high-energy radiation beam	HIGH PRIORITY ALARM SIGNAL
Prompt problems which do not cause patient injury or death until at least several to many minutes have elapsed.	Many cardiac arrhythmias; High or low blood pressure; Apnea (unless prolonged or associated with extreme hypoxia); Mild hypoxemia; High or low pCO_2	MEDIUM PRIORITY ALARM SIGNAL
Delayed problems which cause patient injury only after many minutes to hours have passed.	Failure of an infusion pump for maintenance of intravenous fluids; Failure of an enteral feeding pump; Failure of a patient weighing system	LOW PRIORITY ALARM SIGNAL

The last column represents the priority of the alarm signals associated with each event. High priority alarms signals should be reserved for alarm conditions that require immediate response, medium priority signals to alarm conditions requiring prompt responses and low priority signals are used for alarm conditions that need operator awareness. Some medical equipment does not require high priority alarm signals (AAMI, 2013).

One final concept that needs to be separated from the concept of alarm is the information signal. These are merely informative and do not have a priority level associated. Examples include the ECG waveform, SpO2 tone and the Fluoroscopy beam-on indication.

3.1.1 What is a good alarm?

Begault (1993) states that the design of audio interfaces should require the expertise of acousticians, musicians or psychoacousticians to avoid the following mistakes: Sounds which are repetitive, loud, and simple; Sounds used as alarms instead of carriers of information; Sounds with poor quality. On a similar note, Schafer (1977) highlights that an acoustic designer should understand the environment to be tackled, have training in acoustics, psychology, sociology, and music. Smith et al. (2004) have established general criteria that auditory warnings and alarms should aim for, including:

- auditory warnings should serve to alert the operators to the fact that there is a situation, or emergency that requires their attention;
- auditory warnings should provide the operators with information about the nature or identity of the situation or emergency that they are required to attend to;
- auditory warnings should guide the user toward the appropriate course of action to deal with the situation or emergency. This third criterion is often neglected, with complementary information being given by a visual display.

Edworthy and Hellier (2005) list how could clinical alarms be better:

- warning sounds are standardized.
- the acoustic properties of alarms are given proper consideration.
- the learnability of alarms is given proper consideration.
- prioritisation of alarms is possible within the systems.
- the urgency of alarm sounds matches their criticality.
- intelligent alarm systems are used.

Sanderson, (2009) has established important design criteria which should contribute to an increased effectiveness of alarms:

- Design and evaluation teams should have experts from the application domain, such as anaesthesiologists, critical care nurses, etc.; and psychoacousticians; behavioural scientists or human factors experts; experts in the standards development process.
- When evaluating clinical alarms, researchers should consider individual differences, learning criteria, retention criteria, localisation of the alarms, urgency mapping, performance under representative work situations, tolerance to the alarms and reliability of the alarms.

It seems clear that designing a good alarm is far from trivial, for it requires technical competencies in acoustics, experimental methodologies, Human Factors and standards, besides knowledge in the application domain. The alarm should also include as much information about the nature of the alarm condition, and how to proceed to mitigate the alarm condition.

3.2 Auditory icons

The most common auditory warnings are pure-tone beeps or very simple melodies. This is probably what comes to mind when we think of the word “alarm”. However, auditory interfaces can take several forms, and warnings can be speech-based, or earcons (abstract, constructed sounds, often very short musical fragments (Lemmens et al., 2001)), or they can be auditory icons. These are sounds that create a strong semantic link to the category they represent (IEC; 2020).

Auditory icons originated in informatics, and they are short, one-shot sound messages, consisting of short snippets of sounds that are present in everyday life (Cabral & Remijn, 2019). Gaver, the first researcher to propose the use of auditory icons, has defined them as caricatures of naturally occurring sounds (Gaver, 1987). This makes them complex sounds. The first auditory icons were used in computer interfaces to enhance the visual desktop metaphor with an auditory dimension. Examples include selecting objects (a ‘hitting’ sound), opening files (a ‘whooshing’ sound), copying files (a ‘pouring’ sound) or putting files into the trashcan (a ‘metal trashcan’ sound or a ‘paper-crunching’ sound) (Cabral & Remijn, 2019).

Shortly after their first use, Mynatt, (1994) proposed a set of factors influencing the usability of auditory icons. Since the early 90s, their use has increased in all sorts of fields, including healthcare.

In automotive studies, it has been found that drivers responded faster to auditory icons than to tonal or speech based warning signals (e.g. Graham, 1999; Ho & Spence, 2005; Lee et al., 2002). Human beings have the capacity of extracting information from the environment by identifying naturally occurring sounds. The effectiveness to respond to auditory warnings can be expanded exploring this ability, also because humans have good memory capacity for environmental sounds and they are easier to associate with particular target events (Smith et al., 2004).

Smith et al. (2004) made two experiments to understand the effectiveness of auditory icons when compared to speech and abstract sounds as to ease of learning and retention in different workload conditions over a one-week period. The results revealed that the speech auditory warnings were learned more rapidly (2.1 trials), followed by the auditory icons (3.4 trials) and then the abstract sounds (8 trials). The retention task made three and eight days later revealed that the learned associations for speech and icons were reproduced almost in perfection. The same did not happen with abstract sounds (Smith et al., 2004). These results indicate the potential that icons have as auditory warnings. Icons are easy to learn and require minimum training, but their main drawback is that many objects do not have a familiar or obvious representation in iconic form (Graham, 1999).

On more fundamental terms, it has been observed that acoustical heterogeneity in auditory alarm sets makes them easier to learn (Edworthy et al., 2011). In a study comparing auditory icons and abstract sounds, it has been observed that auditory icons performed well and were easy to learn, but they did not necessarily outperform some of the abstract sounds, which were also heterogeneous. It could be that auditory icons are easier to learn due to a combination of two factors: their acoustic heterogeneity and their object-sound association.

3.3 Masking and Localizability

In psychoacoustics, a sound that renders other inaudible to a human listener is called masker, with the inaudible one the maskee (Fallis, 2009). In detectability tasks, where it is asked to a participant to listen for and report on the audibility of a target signal in noise, usually the maskee is the target signal, with the masker taken as the noise signal. Noise can have much more power than a target signal, but if the spectral regions of noise and

target are sufficiently distant, the target will be audible. On the other hand, if both regions coincide, noise needs only to have residual more power than the target to mask it. This is the fundamental of spectral masking of two simultaneous sounds. Spectral masking does not entirely depend on the total power of masker and maskee, but rather on their relative power distribution along frequency (Zwicker & Fastl, 2013). One objective of good alarm sound design is so it can be reliably detected at relatively low volume levels, preventing buildup of overall ambient sound levels in the work environment, which have been shown to have adverse effects on healthcare workers, and patients (Ecophon, 2018; Edworthy et al., 2018; Patterson, 1982; Graneto & Damm, 2013).

In healthcare settings, up until the early 1990's the technology implemented in medical equipment limited the range of alarm types, with most having only capability for simple beeps and buzzes. Today´s equipment with micro-electronics are capable of complex sounds reproduction, enabling a valuable contribution to the issue of masking through careful alarm sound design (Edworthy, 1994).

Alarms should be at a level that guarantees detectability, but below levels that result in disruption of performance and annoyance. On a published study investigating warning systems on civil aircraft, the problem of alarm masking by background noise inside the flight-deck was described, and a *power spectrum model* method for its prediction was presented (Patterson, 1982). This method implemented a model of the auditory filters based on a mathematical function that correlated well with experimentally measurements (Patterson, 1976). Through a power spectrum analysis of background noise and alarm signals, it would be possible to determine the appropriate sound levels for the alarms so to guarantee its audibility in statistical terms. This method predicted the *masking thresholds* of steady non-tonal noise in function of frequency. The threshold is the predicted level for a target to have a detectability with 75% correct answers (Green & Swets, 1966). Patterson´s alarm design criteria set alarm levels to be +15 dB above this threshold. This value minimizes undesirable near-threshold effects such as interference, poor frequency discrimination and poor localizability (Scharf, 1964). The criteria also established an upper limit for the alarm warnings at +25 dB above masking, noting that, from a certain level, the improvement in detectability with increasing level reaches a ceiling, with the aggravating factor that high levels lead to issues of annoyance and communication breakdown, thus increasing potentially hazardous situations.

To date, various methods for masking prediction have been developed, with several based on psychoacoustic models. Psychoacoustic models relate human perception to physical

quantities of sound, using mathematical functions that are based on a large number of experimental data. Models that mimic the human excitation patterns on the basilar membrane have proven successful on the prediction of masking (Hasanain et al., 2015).

In the healthcare context , various researchers have considered the issue of alarm masking (Judy Edworthy, 1994; Konkani et al., 2012; Toor et al., 2008). In a recently published study (Bolton et al., 2019), the prediction of spectral masking using a previously established method (Hasanain et al., 2017) was experimentally validated with application to IEC60601-1-8 conformant clinical alarms. The experiments evaluated the simultaneous masking of two concurrent (tonal) alarms, with masking prediction based on the alarms' primary harmonics interaction. Results confirmed that participants had more difficulty on detecting signals the method predicted as masked, with misses close to 50%.

Another important requirement in alarm design is the ability to localize where in space did the alarm sound from. In interviews with nurses (cf. Chapter 6), some have shared having difficulty in determining from which equipment is the alarm coming from, especially in rooms with several patients. Localization studies show that a narrow bandwidth of a sound reduces its localization acuity. On the other hand, broad band noise (sound energy is distributed over a wide section of the audible range) improves the localization efforts, better than a combination of tones as is usually found in alarms. A broadband noise also is more useful in reverberant spaces for it helps resolve ambiguities derived from reflections from walls, ceiling and other objects (Catchpole et al., 2004). The compromise between having noise as alarm and a simple tone is complexifying the auditory signal. For this reason one of Paterson's (1982) first design requirements for alarms included having at least four harmonics. This would make alarms more resistant to masking (if one frequency component is masked, others are available) and aid localizability.

In practical terms, the need of having localizable alarms can be explained with 1952's Hick-Hyman law. It states that reaction times increase logarithmically with the average stimulus information or uncertainty. Wallace and Fisher (1998) replicated this effect with auditory stimuli. In their experiment, participants had six speakers around them in the horizontal plane. A tone was played, and the participant had to indicate from which speaker the tone came from. The probabilities of a sound coming from a determined speaker varied (Wallace & Fisher, 1998). The results demonstrated that the amount of possible audio locations increased the amount of time participants took to determine the location of the sound. If we envision a recovery room with twenty patients (as is common), each

connected to at least one medical equipment, we can start to understand the importance of being able to determine as accurately as possible the physical location of the alarm. Another important auditory phenomenon was mentioned in this study, the front to rear confusion. When two speakers are located and activated symmetrically, there is more room for confusion as to the sound location than when they are asymmetrical. Importantly, small head movements can provide the information necessary to resolve the ambiguity[4].

The 2012 amendment of IEC 60601-1-8 included specification for the harmonic structure of the alarms, and these values were low-frequency, which improved localizability and reduced aversiveness (Edworthy et al., 2017). The localizability of these alarms was compared with the new 2020 amendment alarms in a localization study (Edworthy et al., 2017). The 2020 alarms consist of auditory icons which are harmonically more complex. Participants had to hear a sound, and indicate which speaker, from a range of eight speakers placed around the participant, emitted the sound. Results showed that localization accuracy was at a high level for responses to auditory icons, and it was significantly lower for the old IEC alarms, which have five harmonics. These results demonstrate that sounds with more harmonics are easier to localize than those with fewer.

3.4 Workload and Learnability

Localizability is affected by the operator's workload at the time of the alarm. In six varying workload conditions, Edworthy et al., (2018) asked participants to indicate the location of alarms from one of eight possible speaker locations. Distracting tasks included reading, arithmetic tasks, be exposed to ICU noise or doing the reading or arithmetic tasks in the ICU noise. Performance was best with no secondary task (one alarm missed in ten occurrences). In the heaviest workload conditions, one alarm in every four was missed.

It is also important to understand the effect an alarm has on performance in a high workload condition. Lacherez et al. (2016) tested whether identifying or ignoring alarms

[4] Sound localization implies an egocentric reference in terms of spatial perception. Every measurement and orientation have on account the listener's position. The location of the sounds is commonly made in three dimensions: azimuth, from left to right; elevation, from bottom-up and the distance of the listener to the sound source (Goldstein, 2007). To localize sounds with more accuracy, the auditory system uses cues, like the Interaural time difference and Interaural level difference. They are based on the principle that sounds originating from a side will reach one ear before and louder than the other ear.

influenced a digits-forward task[5]. Alarms interfered with serial recall (recall a list of items in a specific order) when the task consisted in identifying them, but not when the task was to ignore them. Also, there was no difference between familiar or unfamiliar alarms, both were equally ignorable. A result which is a relevant design consideration, was that spoken alarms were harder to ignore, making them an inappropriate choice for high workload environments. In practical terms, a worker engaged in a high priority task will be able to prioritize their work over the identification of the alarm, and familiarity with the alarm does not lead to involuntary processing of the alarm.

A relevant cognitive phenomenon related with the detection of auditory events is inattentional deafness. In music, inattentional deafness is defined as the inability to perceive an unexpected musical stimulus when attention is focused on a certain part of the piece. Koreimann et al. (2014) asked participants to listen to 1 minute and 50 seconds of Richard Strauss' *Thus Spoke Zarathustra* and count the number of timpani beats. The unexpected event was a 20-second electronic guitar solo. Results were equivalent for musicians and non-musicians, and while 81% of the control group noticed the e-guitar solo, only 43% of the experimental group noticed the unexpected event while counting the number of beats.

Another variation of the phenomenon was tested with moving events (Dalton & Fraenkel, 2012). The authors created a binaural scene with two simultaneous conversations, one between two women and another between two men. Halfway, a male character walked through the room repeating the phrase "*I'm a gorilla*" for 19 seconds. Half of the participants had to attend to the female conversation and the other half attended the male conversation. Results have shown that the absence of attention can leave people "deaf" to a sustained and dynamic auditory event that would be noticeable on normal listening conditions. 90% of the participants who were attending the male conversation noticed the gorilla. 70% of the participants attending to the female conversation did not notice the gorilla. This shows the importance of where the focus of attention is pointed at in detecting unusual events.

Also visual load can affect the detection of auditory events (Macdonald & Lavie, 2011). While performing a high or low-load visual task (discriminating line length or line colour, respectively), participants had to detect a pure tone that was presented at the same time

[5] Participants are read a sequence of numbers and asked to repeat the same sequence back to the examiner in order.

of the visual task's display. Seventy-nine percent of participants in the high-load visual discrimination task failed to notice the presence of this tone. This phenomenon was later tested in applied settings with alarms (Dehais et al., 2013).

These cognitive scenarios have an effect when, for instance, learning alarms while performing other tasks (Edworthy et al., 2013). A study simulated the cognitive load of an ICU ward where alarms co-occur more often with drug calculation, preparation and administration tasks, patient observation and talking. Whenever participants had to learn alarms while performing secondary tasks, there were several confusions between the alarms, especially when they shared acoustic and verbal labelling similarities.

Several strategies for learning and retaining the meaning of alarms have been studied. Edworthy and Hards (1999) asked participants to learn twelve warning sounds. There were four learning methods under evaluation: one group learned the sounds according to verbal labels given to them by the experimenter; a second group generated their own verbal labels; a third group learned the sounds using graphic images (the waveform of the sound) given to them by the experimenter; and the fourth group generated their own graphic images. Also, the type of warning sound was varied, between environmental sounds, semi-abstract monitoring sounds and abstract sounds used at the time as hospital warnings. Regarding learning strategy, the best results were found when learning occurred with verbal labels when labels were provided by the experimenter. With participants generating their own cues, performance was good both with verbal and with graphic labels. Environmental sounds were easier to learn than the others, but only in the conditions were the experimenter provided the labels.

Also using clinical alarms, the learnability of different sets of alarms was evaluated (Edworthy et al., 2014). One set included IEC 60601-1-8 tonal alarms, the other included random abstract alarms and a third set was made of auditory icons. After ten blocks of eight trials, the auditory icons were learned significantly better than the other two sets. These results are in line with those regarding the easier learnability of heterogeneous sounds with built associations to events. The better learnability of auditory icons as clinical alarms has since then been evidenced in several other studies (Judy Edworthy et al., 2018; Judy Edworthy et al., 2017). Another advantage of auditory icons is the larger number of alarms a person can memorize, which can be as many as 30. However, also with this type of sounds, confusion arises when sounds have similar characteristics (Frimalm et al., 2014).

3.5 Conclusion

The complexities of defining an alarm are only outnumbered by the rules for designing a good alarm. To name a few, alarms should be easily detected without startling, and they should be easily learned. Several studies on the matter of learnability have started focusing on the use of auditory icons as warning signals, because results were consistently demonstrating how they were not only easier to learn and to retain, but also more difficult to mask by background noise. The issue of masking and of facilitating the localizability is of extreme importance in noisy healthcare contexts. Using complex sounds with broad-band components – more easily implemented in auditory icons – has proved beneficial. Finally, the learnability and the detection of sounds are affected by the workload of the listener, and hospitals are a known high-workload environment. A cognitive phenomenon known for visual stimuli, inattentional blindness (Simons & Chabris, 1999), was also found for auditory events, which was called inattentional deafness (Koreimann et al., 2014). It is important to be aware of these cognitive limitations when designing alarms, processes, and services in general.

4 Toward Alarm Design

This chapter, named after Schafer's call to revolution "*Toward Acoustic Design*" (Schafer, 1977) sets the tone to all ensuing experimental work. The numerous processes of designing tangible products within a people-centric perspective are well known and documented. Transferring and adapting these processes to intangible products such as sounds, more specifically auditory alarms, is less common but equally important.

The four-step process proposed by ISO 9241-210 Human-centred design for interactive systems (ISO, 2010) is the thread that ties the research narrative together. These steps are: 1) To understand and specify the context of use; 2) To specify the user requirements; 3) To produce design solutions to meet these requirements; 4) To evaluate the designs against requirements. The possible activities within each step are described, and a clinical alarm design process is proposed in the scope of this research work.

Vivian Leroy Crisler (1885-1953), an acoustics researcher at the National Bureau of Standards, Washington, D.C., looking over a loudspeaker horn used in sound penetration tests [between 1923 and 1929] The Library of Congress. No known restrictions on publication.

4.1 Introduction

The main premise of this work is that it is developed around the needs of the people who work with and are exposed to clinical auditory alarms. It is Human centred.

Out of the dozens of design processes, ranging from the UK Design Council's Double Diamond (Design Council, 2007), to the Stanford d.school Design Thinking Process (Patnaik, 2009), the main pillar of the following activities is the International Standard ISO 9241-210 Ergonomics of human–system interaction — Part 210: Human-centred design for interactive systems (International Organization for Standardization, 2010).

This standard is a general, application-agnostic approach "*that aims to make systems usable and useful by focusing on the users, their needs and requirements, and by applying human factors/ergonomics, and usability knowledge and techniques. This approach enhances effectiveness and efficiency, improves human well-being, user satisfaction, accessibility and sustainability; and counteracts possible adverse effects of use on human health, safety and performance*" (pp. vi). A human-centred approach guides itself according to the following principles:

- the design is based upon an explicit understanding of users, tasks, and environments.
- users are involved throughout design and development.
- the design is driven and refined by user-centred evaluation.
- the process is iterative.
- the design addresses the whole user experience.
- the design team includes multidisciplinary skills and perspectives.

These principles in their turn, are consubstantiated in interdependent activities, as depicted in Figure 4-1.

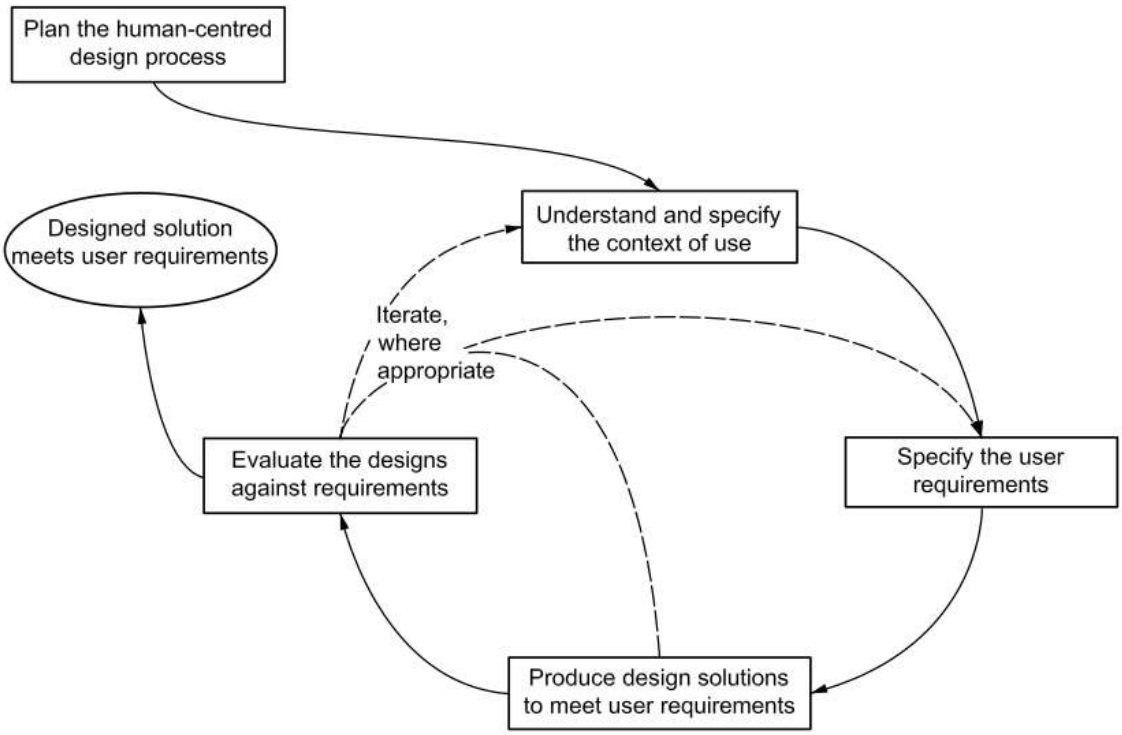


Figure 4-1. Interdependence of human-centred design activities (adapted from International Organization for Standardization, 2010)

For design purposes, the main four activities include:

1. Understanding and specifying the context of use;
2. Specifying the user requirements;
3. Producing design solutions;
4. Evaluating the design;

These four activities will guide the main sections of the current work. The research methods included in each activity were inspired by previous research and literature. To obtain a final design framework that will sustain the current research, we will enrich ISO's 9241-210 four main pillars in a stepwise fashion, including alarm design perspectives in a chronological order. Finally, we will present the motivations and constraints that led us to the adopted alarm design path.

4.2 Alarm Design Methodologies

In 1995 Judy Edworthy and Neville Stanton published "A user-centred approach to the design and evaluation of auditory warning signals: 1. Methodology". It was one of the first efforts to determine specific steps aimed at auditory warning signals. The authors adapted the standardized procedure for evaluating public information symbols (ISO, 2014a) onto the auditory modality. The pertinence of this choice resides in the similarities between

visual and auditory symbol design. An adapted representation of this method is presented in Figure 4-2, which is built on ISO's 9241-210 four main activities. It includes eleven steps which will be briefly detailed.

1. **Establish the need for warning(s) for given referent functions.** The expected output of this phase is a set of functions for which sounds will be required. The authors refer to these functions as *referents*.
2. **Existing and modified sounds.** When there are referents with associated warnings, they can be deemed as inappropriate and need to be redesigned. One advantage of redesigning existents warnings is that their target audience already has learned to associate them with their referents. One possible path in this situation is to gather all existing warning sounds and make the improvement modifications.
3. **Generate trial sounds.** This step of the process provides the creative space to generate new sounds, even for situations where warning sounds already exist. This stage should include both the designer and a potential user of the warnings. Some methods proposed by the authors include free-association or structured interviews. The difficulty of these methods resides in the need for music or acoustics skills to translate ideas into sounds.
4. **Appropriateness ranking/rating test.** The ranking test is the first evaluation stage. It consists in having respondents listening to the generated sounds and evaluating them according to their appropriateness to a given referent. The output of this step should be a set of two to three sounds for each referent (a final set, along with some reserve warnings).
5. **Design trial warning set.** At this stage, the designer should make modifications to the set of sounds that resulted from step 4. By the end of this stage there should be one warning for each referent, plus some reserve sounds.
6. **Learning/Confusion test.** The aim in this step is to observe whether two or more referents are associated with referents with similar meanings. Attention should be given not only to the number of confusions but also to the nature of confusions. Confusing sounds can be modified or replaced.
7. **Urgency mapping test.** The respondents for the first stage of this test should be people with good working knowledge of the environment in which the warnings will be heard. The urgency levels need to have been

defined in the first stage "Establish the need for warning(s) for given referent functions". The urgency of the referents is rated on a 1-5 scale. The respondents for the second stage of the test should be a different group of participants without any knowledge of the referents. This means their urgency evaluation on the same 1-5 scale will be exclusively based on psychoacoustic characteristics.

8. **Design prototype warning set.** By this time, the designer has a prototype warning set which should be comprehensive and compatible.
9. **Recognition/Matching test.** The Recognition/Matching test consists in presenting a warning sound, one by one, along with all the referents. The participant will have to match the warning to the most appropriate referent. The referents are all present in each trial. The warning set and the list of referents should be played once or twice before starting the test. The level of 66% of correct matching of symbol to referent is proposed as a criterion[6].
10. **Generate standardized verbal descriptions** (excluded from representation). Two versions of standardised verbal descriptors are generated: a general and colloquial description of the warning and another with acoustic descriptors. The former is likely to be the names given to the sounds by the people working in that environment. The latter should be descriptive enough to allow replication of the sound (spectral, pitch and temporal characteristics).
11. **Operational test.** The operational test should recreate the working environment as much as possible via simulation in the laboratory. One important phase includes the auditory test to assure warnings are heard without being too loud, considering an accurate background noise simulation. Other relevant tests include evaluating confusability, degeneration of the recognisability of the warning while performing other tasks, and trial scenarios.

A few years after Edworthy and Stanton's study, Watson and Sanderson (2007) wrote "Designing for attention with sound: challenges and extensions to ecological interface design". The paper intended to verify whether Ecological Interface Design (EID) principles

[6] The 66% cut-off point is also the one proposed by SAE J2830 standard, applied in Chapter 13's study.

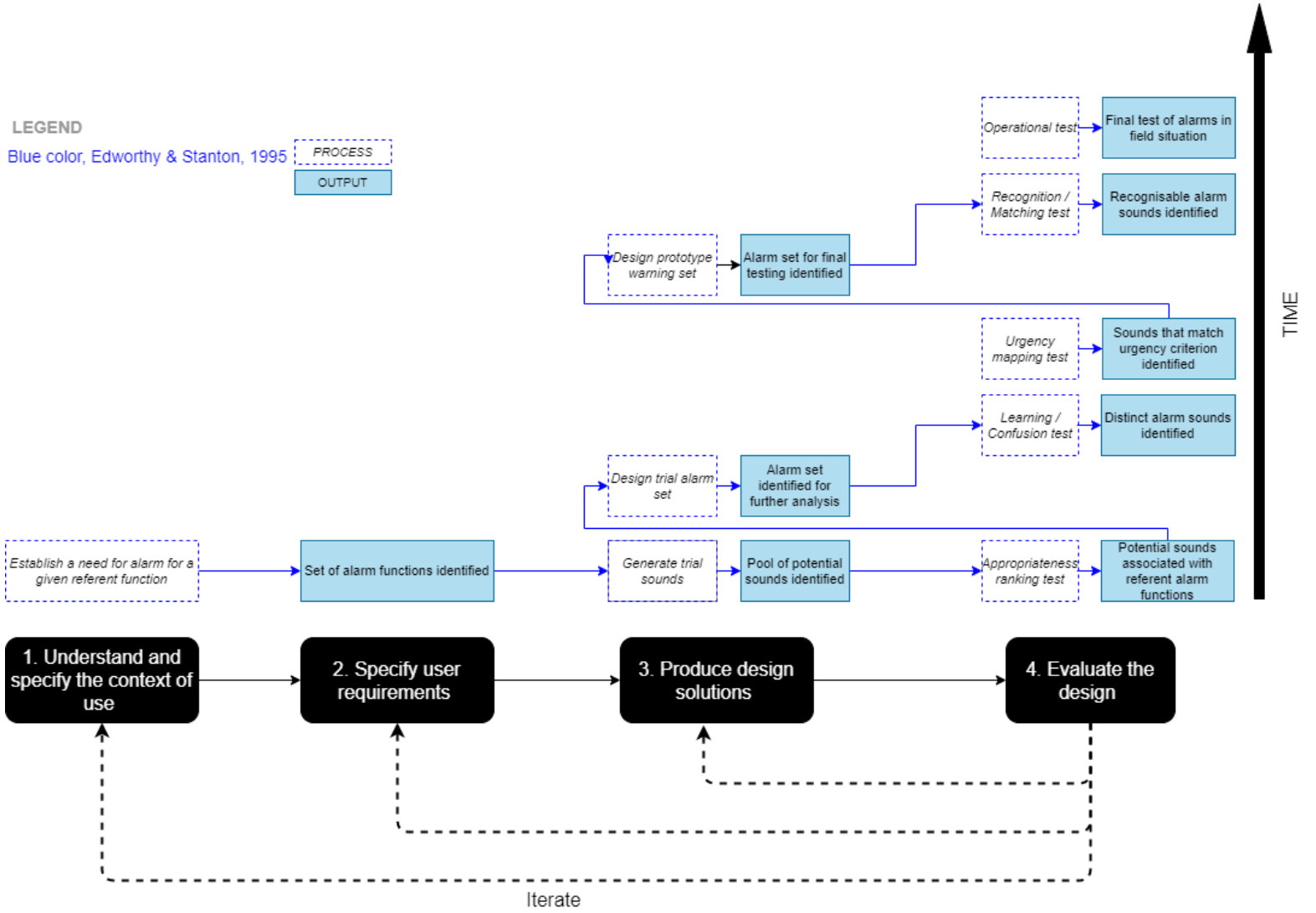


Figure 4-2: Adaptation of Edworthy and Stanton's (1995) representation of auditory warning signals design methodology

(Vicente, 2002) could be applied to the design of auditory displays. EID is a theoretical framework that supports the design of human-computer interfaces in complex sociotechnical systems. Its goal is to stimulate the use of *skill* (S) and *rule*-based (R) behaviour (to use little cognitive resources), but also to support more effortful and error-prone *knowledge*-based (K) behaviour. This SRK taxonomy comprises three design principles: 1) *Skill*-based behaviour: Workers should be able to act directly on the interface; 2) *Rule*-based behaviour: There should be a consistent one-to-one mapping between the work domain constraints and the perceptual information in the interface; 3) Knowledge-based behaviour: The interface should represent the work domain in the form of an abstraction hierarchy to serve as an externalized mental model for problem solving.

Because up to 2007, date of the publication, EID had never been implemented with auditory interfaces, a case study with anaesthesia monitoring set the stage to try it. The authors have adapted its stages, and its representation is in in parallel with ISO's 9241-210's four main activities in Figure 4-3. This process consists in nine steps, all following a similar line of thought as ISO 9241-210 and Edworthy and Stanton.

1. Problem identification
    - **Is there a problem to solve?** Gather evidence from incident reports, operator reports, field observations, etc.
2. Need Analysis
    - **What work is to be supported?** Determine functional structure of work with domain analysis.
    - **What information is needed?** Extract variables, constraints, and temporal properties.
    - **Who needs the information?** Determine whose work is affected by state changes and when.
    - **What level of cognitive control is required?** Determine right levels of cognitive control over control tasks in work domain using SRK.
3. Design Synthesis
    - **What modalities are indicated?** Identify best modalities to achieve required levels of cognitive control.
    - **Semantic mapping: meaning preserved?** Decide kind of auditory display, sound mappings, number of streams.
    - **Attentional mapping: attention preserved?** Preserve appropriate conditions for individual attention and team attention.

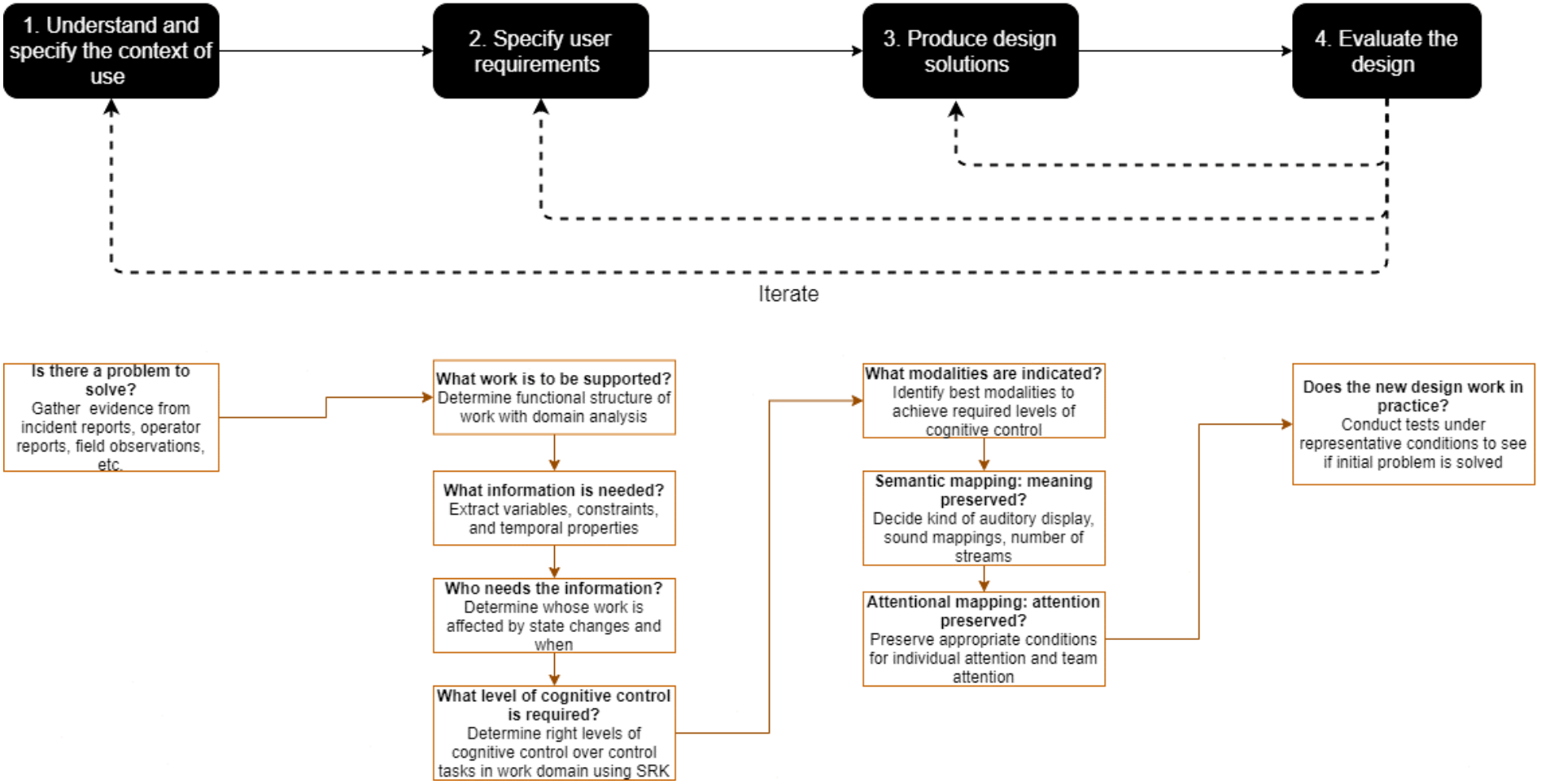


Figure 4-3: Adaptation of Watson & Sanderson (2007) process for the design of auditory displays with Ecological Interface Design (EID)

4. Evaluation
    - **Does the new design work in practice?** Conduct tests under representative conditions to see if initial problem is solved.

More recently, Sanz-Segura, (2021) have proposed a framework for the design of critical alarms. This comprehensive framework encapsulates a total of six phases that range from equipment considerations to user response compliance. We will focus on the alarm design process from the designer's perspective.

System Events (events which occur during monitoring with sensors)

- **Data**. Systems are in environments with devices that monitor physiological variables. These events provide real time information about the system's status and the equipment's condition.
- **Events**. Data is analysed and prioritized withing predefined ranges which lead to certain events. A change or a history of changes in vital signs of a patient states a recovery trend or a life-threatening condition.

1. Information Quality (technical requirements of the equipment and the acoustic properties of the sensory channel)
    - **Information equipment.** Technological systems which manage data and transmit this information through wearables, central alarm systems or others.
    - **Physical parameters.** The physical attributes of an alarm signal are measurable parameters and should consider psychophysical responses (i.e., threshold of hearing) and psychoacoustical parameters (i.e., pitch, duration, loudness, timbre) or localizability.
2. Designated Action (action of the operator in response to alarms)
    - **Alarm users.** The designated operators are the ones responsible for responding to the alarms.
    - **User response.** Tasks differ by the time that operators have to respond to alarms and the number of operators assigned to a task.

In sum, Edworthy and Stanton's 1995 method is heavily inspired in the ISO methodology for graphical symbols. Watson and Sanderson in 2007 proposed to incorporate skill-rule-knowledge taxonomy onto the design of auditory interfaces. More recently Sanz-Segura made a comprehensive framework considering several phases, but on the case of design, the author considered system events, information quality and designated actions. All three

methods are grounded on the human-centred design ISO standard defending a four staged process of comprehending the context of use, specifying user requirements, producing design solutions, and evaluating the design.

4.3 Methodology of the present work

The purpose of the present work is to design a library of clinical alarms adequate for medical equipment inside operating and recovery rooms.

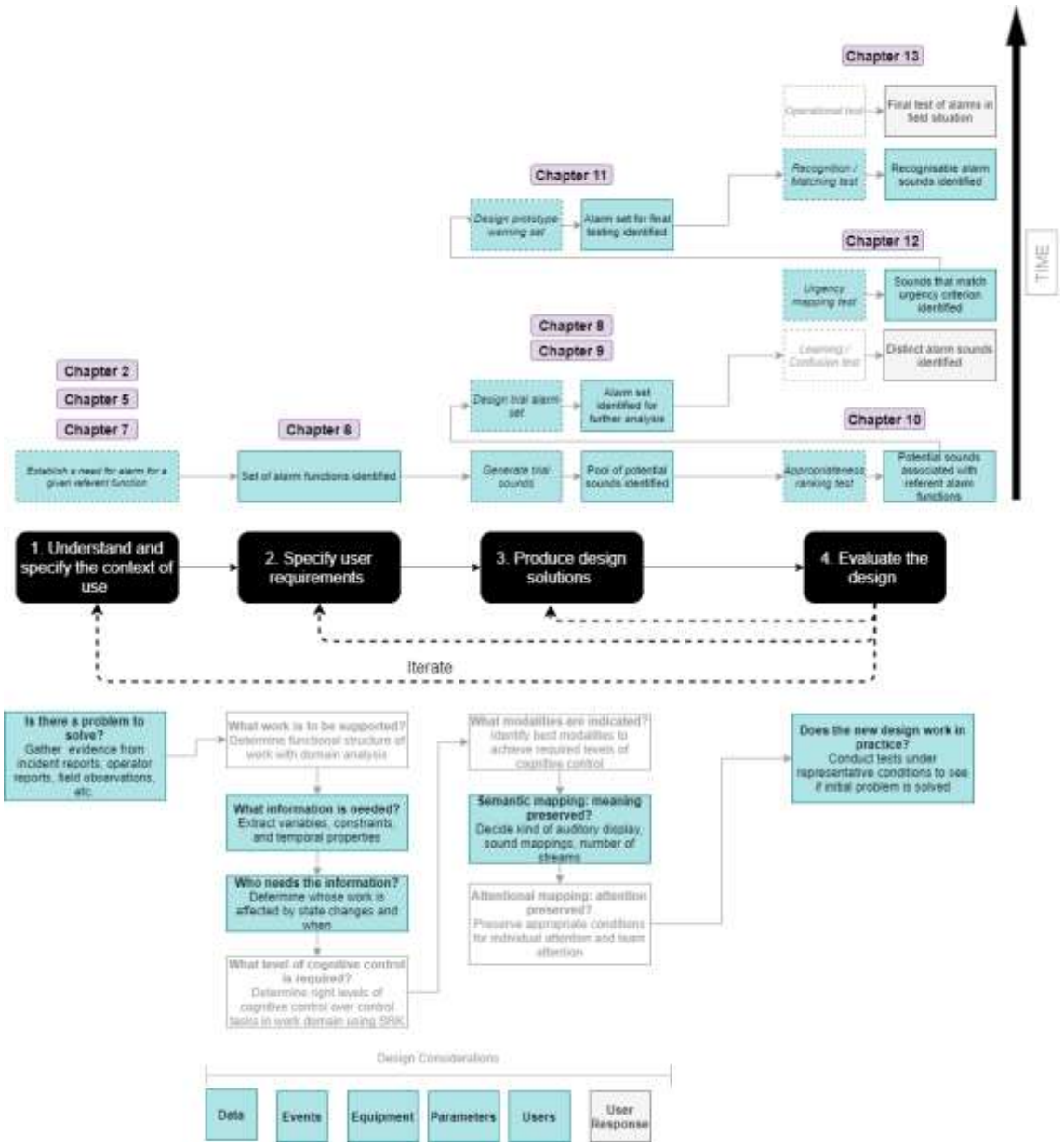


Figure 4-4: Representation of ISO 9241-210, Edworthy and Stanton (1995), Watson and Sanderson (2007), and Sanz-Segura, (2021) models for design of auditory alarms. Boxes with blue filling were steps and activities considered and implemented on the course of the present research work.

The previous methodologies have established a great number of best practices which were followed considering all available resources and constraints. The following chapters will describe activities which fit in the process illustrated in Figure 4-4 (implemented actions have a blue background) and grounded in all the methodologies discussed thus far.

A final "operational test" in field situation as proposed by Edworthy and Stanton proved impossible to accomplish due to the difficult access to field situations. Additionally, the COVID-19 pandemic and its safety measures difficulted the opportunity to use laboratorial 3D sound immersion as an alternative. We believe the stepwise and iterative fashion of the implemented design process make the resulting alarm sounds a robust output, capable of answering all main user and context requirements.

4.4 Conclusion

An auditory warning design methodology was proposed by merging existent design methods. The main pillar is the four staged Human-Centred Design (HCD) process, followed by 1995's Edworthy and Stanton adaptation from the visual symbol standard for the auditory modality. Important strategies were collected from Watson and Sanderson's (2007) application of Ecological Interface Design to the design of auditory warnings. The current IEC 60601-1-8 2020 amendment has followed a very robust methodology that was not yet reported but has close resemblance with the design steps suggested in this chapter.

The four strategies used as inspiration contribute to a careful design methodology which has in mind both the environment and the listener of the alarm.

STEP 1. UNDERSTAND AND SPECIFY THE CONTEXT OF USE

5 Need finding: Visit to an Operating Room at a public hospital

This chapter describes a tour of the Operating Ward of a public Portuguese hospital. The ward in question was polyvalent, and it performed all surgeries. It communicated with the Emergency Room. The purpose of this visit was twofold: 1) to know and characterize the physical space to be aware of the equipment disposition and use, and 2) to informally talk to the healthcare staff and understand their opinions about the soundscape they work in every day.

Bain News Service (between ca. 1915 and ca. 1920). St. Luke's -- Men's Ward. The Library of Congress. No known restrictions on publication.

"This revolution will consist of a unification of those disciplines concerned with the science of sound and those concerned with the art of sound. Acoustic ecology is the study of sounds in relationship to life and society. This cannot be accomplished by remaining in the laboratory."

(Schafer, 1977, pp. 205)

5.1 Introduction

Robert McKim, former head of Stanford University's product design program, coined the term Needfinding for the qualitative research approach which studies people to identify their unmet needs (Patnaik & Becker, 1999). The method intended to bring designers and end users closer. Some principles of Needfinding include focussing on the needs and not the solutions; going to the end user's environment to make the need more specific; collecting different types of data such as interviews and observations. Once scrutinized, these needs will suggest areas for innovations, research, or improvement (Patnaik & Becker, 1999).

The following sections describe a visit to the operating ward of a hospital, which allowed to observe the physical space, listen to the sonic space and to contact with the end-users of this study's output.

5.2 Sound in the Operating Ward

The operating ward at this public Portuguese hospital has 120 professionals working 24/7, having approximately 40 people per shift. The central bloc has six operating rooms and one recovery room. All operating rooms have hexagonal shape (except for one which was rectangular) to avoid dust or particles accumulating on the corners. All ceilings were closed with HEPA filters (High Efficiency Particulate Arrestance).

It became clear from walking on the corridors that auditory alarms are not circumscribed to the rooms. On the corridors, there are gas alarms which sound whenever there are oxygen breakdowns. Additionally, there are machines in some walls which are used for the transportation of small items such as medication and ampules between floors, using air propulsion. According to the workers, whenever a new item arrives, a beep sounds and a light turns on, and due to their constant movement, workers mentioned the light is more relevant to them.

A varied number of brands of equipment was observed in each room. Although there is the understanding that maintaining the same brand of equipment is an advantage, the

selection criteria of equipment are strongly economical. After one machine is damaged and its repair turns out to be more expensive than buying a new one, the hospital simply looks for machines with the same characteristics with the best price in the market. This results in older equipment still in good shape being used with newer equipment, both from different brands.

5.3 The Operating Room (OR)

During surgeries, doctors either listen to music or talk, and it is very rare to listen to noises coming from outside the operating room. Usually, six people are inside: three nurses, two surgeons, and one anaesthesiologist. It happened that during the tour a vascular surgery was happening, and two sounds became immediately evident: the ventilation system and the pulse oximetry with its rhythmic beep. According to the nurse, another important sound source during a surgery is the electric scalpel, which produces a constant noise with two types of sound: one for cutting and another for coagulation.

Another relevant source of noise while preparing for surgeries is the noise of the plastics being removed from the sterilized instruments, along with their disposal in the metal tray.

5.4 The Recovery Room (RR)

The initial focus when starting this visit was the OR, however, after conversations with the professionals in the Recovery Room (RR), it became clear that they had two completely different soundscapes, and the latter had more design and co-existence of equipment challenges.

This RR was rectangular, with eight beds. On average, there are three patients per nurse in the RR. However, if there is a very complicated patient, there will be one exclusive nurse for this patient. These three patients are placed as close as possible to one another for the nurse's convenience. The space description seems rather clean, but the professionals inside mentioned:

> *"Now when you have all patients, three nurses, anaesthesiologists and someone is passing by and listens… this turns into a major mess. Listen to the noise of the AC once again, it's something no one imagines!"*

The first time we visited the RR, there were a nurse and a patient. We analysed the vital signs monitoring machine (Dräger A S/3) and were first introduced to the pulse oximetry monitoring sound. Pulse Oximetry or Saturation measures the amount of oxygen-saturated haemoglobin relative to total haemoglobin (unsaturated + saturated) in the

blood. Normal arterial blood oxygen saturation levels in humans should range between 95–100 percent. If the level is below 90 percent, it is considered low and called hypoxemia. Whenever this parameter gets higher or lower, its pitch changes. This means professionals can infer if the oxygen saturation of the blood is higher or lower just by the pitch change. It is a sonification of a monitoring sound. If it gets to the alarm's threshold, which varies between 100 and 90, and if it gets lower, the pitch gets higher and more shrilly. Once it hits 89 it starts beeping repetitively and will not stop until someone turns it off. Importantly, no medical equipment ever turns itself off. Pulse Oximetry is the only parameter which has this sort of pitch modulation and is considered a monitoring sound and not an alarm, as it is constantly providing auditory feedback.

In all other auditory alarms, only the interpulse interval (speed) varies between beeps. In the case of heart rate and ECG, if there are variations, a long beep will sound. It serves to call the nurses' attention to the monitor, and a visual alert appears with more specific information.

- *"Do you resort a lot to this monitor?*
- *Yeah, we need to. Because a lot of noises are the same. They are just an audio signal and I need to come here and check to what corresponds that signal.*
- *Imagine you have a lot of patients…*
- *This makes a BEEP and we need to come here and see what's happening, that's why we usually have three patients."*

In some equipment it is possible to manage the level of the alarm, but in others that is not possible.

In the main desk of the Recovery Room, there is a central monitor which shows the basic data trends of all eight vital signs monitors. This monitor has the sum of all monitors, and professionals can look and have a sense of the trends of the patients' vital signs. The monitor presents eight little windows with the basic data from all patients (Figure 5-1).

Figure 5-1. Example of a patients monitoring station (Source: medicalexpo.com/, Digicare Biomedical Technology) – no visual or sound records were made during the needfinding visit.

According to the professionals inside the RR, this is something very important to them. On a more intuitive level, it became common the reference to a sixth sense, referring to the experience they gather which makes them predict what will happen with which patient.

- *Is there any time when the room is full, each nurse has three patients and then a machine beeps and they get confused as from which machine does the sound comes from?*
- *Yeah. They need to ask or look. And make decisions. That's the human link of the process, but all monitors interpret information, and that helps us already.*

On the second visit to the RR, there were four patients, with two nurses and one anaesthetist. They have mentioned the beeps do not annoy them, as they know what they all mean. For them, the strange thing is not to listen to them. However, all agreed that the beep sound serves to alert them, but it does not inform them. They all need to go to the monitor and see what the problem is. After talking about the Pulse Oximetry monitoring sound, they have admitted that it would be useful to have more pitch modulation in other parameters. At this point of the visit, we heard a strong sound which triggered a conversation between everybody:

- *There, it was that one (points to a patient).*

- *This means he just had his blood pressure measured.*
- *So you never confuse them?*
- *Nop! We can distinguish the beeps. You can distinguish when the patient is desaturating, you can know whether the frequency is rising or decreasing…*
- *And that's it…!*
- *Oh, I have to look!! But, but, if I have the thresholds set in the machines, for the beeps…*
- *Imagine it's possible to introduce an alarm, for example, for the size of the standard deviation of heart rate.*
- *Ah!*
- *In a coronary unit that can be very important!*
- *Yes of course!*
- *You have a lot of patients, and there is no sound for that sort of thing.*
- *In fact, no, there is nothing to differentiate it.*
- *The sound indicates us some things, but we need to look to the monitor.*
- *The idea would be to have all the alarms like the oximeter.*
- *Yes, to have that in all parameters would be very interesting. To say if the parameter is more to the left or to the right, that would be important. Even if it would be just one more parameter, it would be very important already. The ECG for example.*

The noise sources add up in such complex environments, and the professionals mentioned other equipment, such as the infusion pumps inside the rooms which are very noisy, and the equipment used to keep the beds warm. When it overheats it turns itself off with a long beep. Also, the nasal oxygen administrator makes a constant noise. There is a constant sound that seems to particularly disturb some workers, and that is the sound of the air conditioning and the ventilation systems, present inside all rooms.

Most complaints related with sound refer to the noise made by mobile phones, parallel conversations, the intercom, and air conditioning. The resident staff did not complain about the noise made by the equipment. Nevertheless, by the end of the day, there is a tiredness that they are only aware of in a silent environment. One nurse mentioned the noise levels during the day can be similar to the ones of an airport, stating that:

"It bothers. It bothers, you reach the end of the day very tired."

Towards the end of the visit, the wishful thinking exercise led the nurses to talk about smart alarms, and how the equipment should detect the type of patient and adjust the alarm's thresholds accordingly.

5.5 Conclusion

This chapter intends to set the starting context of this work, which was followed by more interviews with healthcare professionals. It was extremely important to begin this research by seeing the space and talking to the people who work there daily. This need finding activity allowed to bring the recovery room as a research scenario, presenting localization challenges, as well as coexistence and masking concerns from multiple sounds originating in different equipment. Finally, we heard for the first-time healthcare professionals saying that having more informative alarms would be an advantage to their practice. The need was found.

6 Learning with Healthcare professionals about hospital soundscapes

This study envisions the redesign of clinical auditory alarms using a human-centred methodology, encompassing two data collection techniques. Individual interviews were performed with six Portuguese healthcare professionals to understand their daily practices inside the operating and recovery rooms regarding clinical auditory alarms. Additionally, an online survey collected further information from forty-six respondents currently working in operating rooms. Regarding equipment, professionals interact with equipment from different manufacturers, resulting in a diversity of auditory alarms and a difficulty in associating a sound to an event. About Alarm-triggered behaviour, most stated that it would be advantageous to their work to have unambiguous sounds for specific parameters. Having a thoughtful and human-centred redesign of clinical auditory alarms will create a more positive soundscape and environment inside the operating and recovery rooms.

6.1 Introduction

Hospitals are known to be noisy environments, affecting the well-being of both healthcare professionals and patients. The consequences of a noisy environment may be a slower recovery of the patients as elevated sound pressure levels result in impaired sleep, increased stress, delayed post-illness rehabilitation, escalate restlessness, increase respiratory rates and increase heart rates (Ecophon, 2018).

Scientific studies and projects (Birdja & Özcan, 2019; Ozcan et al., 2018; Özcan et al., 2018) have been defending and proposing positive soundscapes for these environments, demonstrating that sounds from nature or harmonic music can have positive effects on health (Rueb, 2019). These studies generally follow human-centred design principles as proposed by ISO 9241-210 - Human-centred design for interactive systems. Among others, the initial stages of development should include an explicit understanding of users, tasks and environments, and users should be involved throughout design and development of the product.

One of the elements contributing to this environment are clinical auditory alarms (Madaras & Apps, 2012). Ever since 2012, auditory alarms have been listed as one of the ten main technology hazards inside hospitals by ECRI Institute[7] (ECRI Institute, 2019). These hazards range from potential errors like modifying alarms without restoring them to their original settings, to false alarms and alarm fatigue.

IEC 60601-1-8 includes general requirements for basic safety and essential performance, and the alarms used with that equipment (Edworthy et al., 2013). Although the updated standard proposes dramatic modifications, such as the use of auditory icons instead of melodies, it will also allow stakeholders to design their own alarm signals if they comply with the standard's performance metrics.

To do so, it is necessary to understand how clinical staff interacts with their daily sonic environment, which tasks do they perform, and how these are affected by sounds.

6.2 Aim

The current study aims at characterizing operating and recovery rooms as experienced by healthcare professionals in terms of sounds and noise, and the tasks and common

[7] ECRI Institute is a non-profit organization that establishes best practices to improve patient care using applied scientific research in healthcare.

behaviours healthcare professionals perform inside them. Individual interviews were performed to gather clinical-staff-centred requirements about auditory alarms, and to understand which and how could current auditory alarms be improved. Additionally, an online survey collected further information from respondents currently working inside operating rooms.

6.3 Material and Methods

6.3.1 Study Design

To contribute to a comprehensive and complete understanding of the research problem, this study used a multimethod approach. We followed a triangulation design model (Creswell et al., 2004) by gathering both qualitative and quantitative data. Firstly, qualitative interviews were conducted with six healthcare professionals to identify and describe key questions and problems. This data was later used to develop an online survey which intended to gather responses from a larger number of operating room professionals regarding auditory warning signals. Similar emphasis was given to both forms of data.

6.3.2 Qualitative methodology: Interviews

6.3.2.1 Interview guide and administration

Semi-structured interviews were carried out to explore the participants opinions about the presence of sound inside the Operating Room (OR) and Recovery Room (RR), particularly about auditory alarms coming from the OR and RR equipment. The interview consisted of a semi-structured protocol exploring the following areas in a fluid and flexible sequence: (1) Demographics and professional path; (2) Daily activities inside the OR and RR; (3) General opinion about sound in the OR and RR; (4) Equipment and corresponding auditory alarms; (5) Improvements in the auditory alarms. The interview scripts can be found in Appendix 2. Some interviews were face-to-face while others were made via Skype. Interviews lasted between 30 to 50 minutes, and all were audio recorded with the participant's permission.

6.3.2.2 Participants

Six Portuguese healthcare professionals were invited and informed of the aims of the study. All worked in the operating rooms of different hospitals. At the time of the interviews, participants' ages ranged from 29 to 58 years. Three were nurses, two were surgeons and one was an anaesthetist. All interviewees were female. Table 6-1 shows

participants' background information. Participants are identified throughout the text according to their profession (S = Surgeon, N = Nurse, A = Anaesthetist) and order of the interview (e.g.: S1 = Surgeon interview 1). Requirements to participate included working in the (OR) and having normal hearing.

Table 6-1: Participant's background information[8]

	S1	S2	N1	N2	N3	A1
Age	29	31	54	31	58	31
Gender	F	F	F	F	F	F
Specialty	Surgeon General surgery	Surgeon Obstetrics	Nurse	Nurse	Nurse	Anaesthetist
Years on the OR	3	5	25	1	20	5

Notes: S = Surgeon, N = Nurse, A = Anaesthetist. S1 = Surgeon interview 1

6.3.3 Quantitative methodology: Online Survey

6.3.3.1 Survey Design and administration

Based on the data from the interviews, the survey was divided into 7 main sections. The first section requested (1) Demographic information from the respondent (e.g., job title, years of experience in the OR). The following sections were about: (2) Physical space and equipment in the OR; (3) Reactions to auditory alarms; (4) Vital signs parameters; (5) Perceptions of the patients' opinions on auditory alarms; (6) Recovery Room and (7) Alarm descriptions. After the interviews, it became evident that a missing category relating exclusively to the Recovery Room was missing, hence, this extra category was included.

Some questions provided general statements collected during the interviews with which respondents had to state were True or False according to their experience or opinion. There was always available the "I don't know" option. Other statements were answered via Likert Scales relating to frequency ("Always", "Very Frequently", "Occasionally", "Rarely", "Very Rarely", "Never"). Finally, other types of questions consisted in multiple choice. All questions from sections 2-7 were required and there were no open questions for ease of filling. All questions had a comment section which respondents could use to add more information. The questions from the online survey can be found in Appendix 3.

The survey was implemented online via Limesurvey (LimeSurvey GmbH, 2018) from May to August 2019 and was shared using personal contacts of the authors, healthcare

[8] Information gathered at the time of data collection (2017)

professionals private Facebook groups, Twitter, and LinkedIn (with the hashtags #sounddesign #nomorebeeps #alarms #healthcare) and the SurveySwap platform. The survey took on average 7 minutes to fill. The requirements for filling the online survey were working on an operating and/or recovery room and having normal hearing.

6.3.3.2 Participants

Forty-six respondents answered the online questionnaire and Table 6-2 presents the detailed characterization with percentage on the left and total number on the right for when respondents could choose more than one option. Of all the respondents, 74% were female. The average age was 42 years (SD=12), and all currently worked in the operating room. The average time of experience in this ward was of 15 years (SD=11), however, the majority has had experience in other hospital areas such as the ICU, emergency room, psychiatrics, paediatrics, among others. Thirty-four respondents were from Portugal, and twelve were from other countries. Thirty-four of the respondents were nurses, six were assistants, three were anaesthesiologists, and there was one respondent for each of the job titles of surgical technician, supply techs and certified registered nurse anaesthetist.

Table 6-2: Demographics of respondents of the online survey (n=46)

GENDER	%
Female	74
Male	22
N/A	4
JOB TITLE	**%**
Nurse	74
Assistant	13
Anaesthesiologist	7
Surgical technician	2
Supply techs	2
Certified registered nurse anaesthetist	2

MAIN TASKS IN OPERATING OR RECOVERY ROOM	N
Instrumentalist Nurse	39
Circulating nurse	35
Anaesthesiologist Nurse	20
Anaesthesia Care	9
Assisting surgeon	7
Polyvalent	4
Supervising/teaching	2
Watch	2
Recovery Room	2
Electronic record	2

YEARS OF EXPERIENCE IN OPERATING ROOM	%
> 30	9
> 20	20
> 10	26
≤ 10	23
≤ 5	22
COUNTRY OF RESIDENCE	**%**
Portugal	74
United States of America	7
Canada	5
Malaysia	2
Costa Rica	2
UK	2
Netherlands	2
Australia	2
Argentina	2
Uzbekistan	2

IN WHICH SURGERIES DO YOU WORK MORE OFTEN?	N
General Surgery	27
Urology	20
Otolaryngology	18
Colon and Rectal Surgery	17
Neurological Surgery	17
Thoracic Surgery	12
Gynaecologic Oncology	12
Plastic and Maxillofacial Surgery	12
Paediatric Surgery	11
Orthopaedic Surgery	11
Obstetrics and Gynaecology	10
Vascular Surgery	9
Oral and Maxillofacial Surgery	6
Ophthalmic Surgery	5
Dental	1

The surgeries where respondents worked more often were General Surgery, Urology, Otolaryngology, Neurological Surgery, Colon and Rectal Surgery. Other surgeries where respondents worked were Thoracic Surgery, Obstetrics and Gynaecology, Gynaecologic Oncology, Ophthalmic Surgery, Oral and Maxillofacial Surgery, Orthopaedic Surgery, Plastic and Maxillofacial Surgery, Paediatric Surgery, Vascular Surgery and Dental.

6.4 Results

6.4.1 Qualitative data from interviews

6.4.1.1 Code system

After transcription of the interviews, all material was coded using computer software MAXQDA10 (VERBI Software, 2010). The code system used to categorize the interviews is described in Table 6-3 and was divided in four main dimensions: Physical space; Sound and Alarms; Alarm-triggered behaviour; and Suggestions for improvements, matching the structure of the interview.

Table 6-3: Coding system used to categorize the interviews divided in four main dimensions: Physical space; Sound and Alarms; Alarm-triggered behaviour; and Suggestions for improvements.

PHYSICAL SPACE

Code name	No. of occurrences
Physical space. Information regarding the physical spaces inside the OR	10
Equipment. Information related to the equipment used	40
Positioning. Information about position of the equipment and position of the professionals	16

SOUND AND ALARMS

Code name	No. of occurrences
Sound Description. How sound is described, adjectives and vocalizations	51
Parameters. Referred monitoring parameters	41
General Opinion on Sound. How interviewees feel about sounds inside the OR	25
Types of surgery. Descriptions about different types of surgeries	22
Importance of sound. Personal opinions about the audio alarms and feedback	18
Perceptions of Opinion of the patients. Information regarding what patients think and say about the auditory alarms	15
Anecdotal Stories. Anecdotes involving auditory alarms	14
Importance of silence. Personal comments about the importance of silence	5
False alarms. Comments about false alarms inside the OR	4

ALARM-TRIGGERED BEHAVIOUR

Code name	No. of occurrences
Reactions to sound. Descriptions of reactions to audio alarms	34
Common practices. Common tasks and practices after hearing an alarm	23
Listen without looking. Comments about being able to listen and understand without looking to a monitor	9

SUGGESTIONS FOR IMPROVEMENTS

Code name	No. of occurrences
Parameters which could be improved. Opinion about which auditory alarms parameters could be improved	13

Results were clustered into four main dimensions: (1) *Physical space, (2) Sound and Alarms, (3) Alarm-triggered behaviour* and *(4) Suggestions for improvements*. *Physical space* describes the equipment and positioning of healthcare professionals in rooms or in relation to specific equipment; *Sound and Alarms* is the descriptive content with which interviewees described the operating and recovery rooms soundscapes; *Alarm-triggered behaviour* includes good or common practices professionals perform when listening to an

auditory alarm and finally, *Suggestions for improvements* includes answers to the specific question of which vital sign parameter interviewees considered in need of improvement.

Data is presented according to the dimensions of the code system, supported by examples from the participants' discourse.

6.4.1.2 Physical space

This dimension describes the equipment and positioning of healthcare professionals in rooms or in relation to specific equipment.

Positioning

According to the interviewees, in the operating room during surgery, the monitoring tower is usually near the head of the patient. In relation to the nurse, the tower is on the front-side because they seat next to the patient. Although this is usually the position of the nurses in the operating theatre, it might differ according to the room and the surgical specialty.

Participant A1 referred she always tried to position herself in such a way where she can see both the computer screen – where she makes the surgeries' records- and the monitor screen. Nevertheless, sometimes that is not possible, as the organization depends on the hospital which always tries to maintain the same equipment distribution.

Regarding the recovery room, patients enter and leave as surgeries occur throughout the day. Each patient has a monitor, and all monitors are on the right side of the patient. The room is organized in such a way that it is possible to see all beds from a central counter. Depending on the care each patient needs, they are placed in different rooms. If the patient needs a personalized and frequent vigilance, he or she goes to the intermediate care unit which can be a smaller room. Nevertheless, there are larger recovery rooms with 14 beds which are regularly full. In this situation, these 14 beds would be the responsibility of three nurses, according to N2. A1 acknowledged there is a problem in this situation when it comes to identifying which machine emits the audio alarm.

> *"It has happened before, the alarm starts ringing in the middle of all those patients and I go: Well, where does this come from? In this sense different alarms would be an advantage" (A1).*

Similarly, N2 works on a recovery room with 12 beds and sometimes there are several alarms beeping from several monitors. This nurse reported having difficulty in identifying from which bed are the alarms coming from "*You need to check one by one*" (N2). N3

reported never having this difficulty, since the monitor emitting the alarm also has a red light.

Equipment

Regarding equipment, the main devices were described. The vital signs monitor was the most referred medical device. These monitors can display between five to six status of the patient's vital parameters, depending on the patient and on the surgery. According to the participants, some might emit alarms whereas others might not. Those which do emit alarms might be different from room to room. A1 said "*When I go to Gastro, they have different monitors. Those monitors turn off automatically".* Not only is the equipment different, the alarms coming from those monitors are also different. *"In the OR, all equipment is from the same manufacturer, but sometimes when we grab transport monitors or another monitor, the sound changes a little"* (A1). According to N1 when the manufacturers are different, the auditory alarms "*can be higher or lower but if something happens, the alarm goes in continuous mode so, they all have common characteristics".* N3 referred what seems to be a common practice in hospitals nowadays, with the hospital adopting the same monitoring equipment "*for economic reasons, mainly for maintenance"* but, she adds *"for our work, that is easier*". An interesting point made by both N1 and N2 was that the difference between auditory alarms was not particularly relevant, because "e*ither way, whenever an event occurs, we have to look at the monitor to see what it is"* (N1). N2 added that they do that, because "*the corresponding bar starts blinking".*

The sound level of the monitors can be adjusted differently for the monitor signals and the alarms. Alarms sound whenever a defined threshold is exceeded. These thresholds should be adjusted according to each patient, even though the machine's system has default values. N1 reported that "*usually the alarm is always in maximum volume so that when it sounds, it is really annoying, it really bothers"*. During surgery, auditory alarms from the anaesthetics machine monitor can be silenced every 2 minutes.

As for the surgeon's perspective, it differs from the nurses' and anaesthesiologists', mainly because their focus on sound is different, and both surgeons stated having no alarm localization problems. According to S2, during surgery the beeps may be similar, but the place the sound comes from makes her know why it beeped, "*the monitoring equipment is at the patient's head while our laparoscopic equipment is by the patient's feet*" (S2). S1 referred that some electrical equipment such as the scalpel emits an electric signal which interferes with the monitoring signal of other parameters. This was confirmed by N3 who

said "*we see the isoelectric line which isn't true, we see the other parameters and we see that everything is alright. Then the alarm becomes a little annoying*".

Other equipment that was mentioned by participants included: the A/C system ("*I have no reason to complain about it"* (A1); "*it doesn't make any noise*" (N1); air systems to warm the patients, described by A1 as "*very annoying*"; aspirator systems, like anaesthesia aspirators, secretions aspirator, liquids aspirator, surgical fumes, also described as "*very annoying*" (A1); ultrasound to block the nerves, which does not emit any sound; the neurostimulator, which sometimes emits a sharp noise, "w*hen it gets closer to the nerve, it has a lower vibration and you can see the oscillations on the screen*" (S1); perfusers, usually silent but with a continuous alarm when something is not right; the blood pressure machine, in which the intervals it takes the blood pressure levels are configured, and which "*gives a visual signal. It fills the cuff and before fixating the number, it blinks and sometimes it makes a sound, but it's kind of low, you need to look at the monitor*" (N3); and finally, the ventilator with which the problem of different manufacturers also happens: *"there are 4 rooms with different brands of ventilators and in those rooms, the alarms are different" (*N2).

6.4.1.3 Sound and Alarms

The Sound and Alarms dimension includes the descriptive content with which interviewees described the operating and recovery rooms' soundscapes.

General Opinion on/the importance of Sound

In the beginning of the interview, participants explained their thoughts about the sound inside the OR, without particularizing yet the case of auditory alarms. A1 said that noise is part of their daily lives, from conversations – sometimes excessively high – to music, but noise is very important for a very basic thing: patient monitoring. N2 referred the OR is a very noisy place, and that it demands some effort to not let it affect their work. For N3, the noises made by people are much louder than the monitoring beeps. S1 agreed, in the sense that surgeons also get used to the noises of the OR. For this participant, the most annoying noises are the conversations and the moment of changing shifts. However, as all interviewees agreed, people get used to working in such an environment, and because of that, people quickly learn to which sounds should they be more sensitive to (N1 used the term "*selective sensitivity*"). N2 added other elements, like noises associated to instrumentation - which depend on the person handling them, the metal instruments themselves, the emergency bells, and the A/C. N3 mentioned that in their day-to-day routines they do not really feel the noise, but *"after a day of work, when I am outside,*

especially at night, I feel so good, like…ah…so peaceful… If we think that, it's because something was bothering before, we just didn't notice!".

For S1, the strident characteristic of the alarms is important because it serves the purpose of alerting towards a change. This participant referred the monitoring machine is very important for the anaesthetist but for a surgeon the most important audio signals are the ones from cut and coagulation.

With a different opinion from all other participants, S2 described the OR as a relatively silent environment where people do not speak very loudly, and the music is generally calm. This participant made the distinction between the operating theatre - where the noise one listens to is mostly equipment-related, and the delivery block - where the noise is associated with the anxiety of the moment, and the fact that in here the patient is seldom alone, but with the partner.

All participants manifested that sound – "noise" as they frequently called it – was important because it served as an indicator of the patient's states, sometimes without having to look at the monitor. A1 exemplified that her work implies making records in the computer, and she cannot look at the computer and monitor at the same time, so the sound is a fundamental information. In her case as an anaesthetist, the pulse oximetry sound is very informative, and is also the most important for her practice. The sound might differ from device to device, but without significant impact in their work.

Parameters and Anecdotal Stories

It was explained that the oximeter signal, which indicated the peripheral saturation of the patient, is the only parameter which uses pitch modelling to communicate a given value. Participants mentioned it provides a "*constant noise which gives the heart rate, and a constant audio pitch according to the peripheral saturation"* (A1). The same participant referred the oximetry audio signal is:

> *"Very intuitive. Extremely intuitive. Especially the beep, what you listen to is the beep, the pitch will indicate how the patient is saturating plus the rate, and it's something that is impossible to teach but which you quickly interiorize. And you can understand if something is right or wrong, very quickly…just intuitively, it's not even rational, it's automatic".*

Being so rich, it ends up being a fundamental signal for anaesthetists. As A1 put it when asked, "*it is an extremely well-designed sound*". A1 told a story about a colleague who was in maternity leave and called another colleague. They were both talking on the phone with

an anesthetized patient in the room. When the other colleague on the phone heard the monitor, she said "*Look, your patient is with 88 heart rate and 97% saturation. That was exactly right! Just with the sound, it's automatic. Something you don't teach. You learn that on your day to day.*"

Participants described sounds reflecting other health parameters, namely: heart rate; blood pressure; capnography curve (CO_2); Premature Ventricular Contractions (PVC) and temperature. The parameters to be monitored differ from patient and from surgery. Parameters like PVC, invasive or non-invasive blood pressure might be required, but the monitor assumes the monitorization once the respective cable is placed and connected.

Sound Description

At first, we planned to register the vocalizations participants made when attempting to describe a sound. There was a category in the coding system aimed at compiling the words used to describe the sounds and associate them with monitored parameters. Most participants used the term "noise" when referring to audio signals. However, it quickly became clear that it was very complicated to just talk about sounds, as interviewees felt descriptive words were lacking, and they felt somewhat embarrassed of vocalizing the sounds.

When talking about the monitor of the anaesthesia machine, A1 mentioned that "*if the blood pressure is lower, a very strong, constant and even irritating warning is activated, very repetitive".* This participant said that what annoys people is not the "*beep*", but the "*intense*" noises.

Other described noises come from vacuum-like equipment. A1 mentioned the device used to warm the bed, and the blood and suction aspirators "*when we turned them of…it is such a relief, it feels really nice to listen to that silence, because that noise is really annoying*". N1 agrees with this opinion about all types of aspirators inside the operating theatre as "*they make an extremely inconvenient noise, we are constantly hearing that "sllurrssp" noise, it's a noise that interferes with our ability to absorb other noises".*

Different descriptions were made talking about the neurostimulator ("*it's lower, when it gets closer it makes a "gggrrr", not so much a beep*", A1) and the blood pressure warning, also a lower sound. N1 characterized the blood pressure warning as "*bitonal, something like "ni-ni-ni"" (sounding like an ambulance*). This participant described the electric scalpel as "*piiiiiiiiiiiiiiiiii pi piiiiiii*" when it is cutting through some tissue. Other equipment, like the oxygen cannula, makes a noise "*like someone is blowing, "fuuuuu*" (N1). Finally, when

electrodes fall and the machine is missing input signal, N3 characterized the noise as "*miiiiiua*".

By far, the most common onomatopoeia was the **beep**: "All those beep beeps" (N1), "I guess they are all beeps! I am not sure if all the devices are beeps, but I would say yes! Some longer than others, some are beeeeeeeeeeep" (S2), **Pi**[9] "the Pi can be higher or lower, something differing in pitch" (N1) or "bi bi's from monitoring".

Some of the adjectives used more often to describe the sounds were: annoying, unpleasant loud, fast, nice, strident, terrible, high pitch, repetitive, irritating. A cross-modal characterization was made by N2 who said the blood pressure signal "*had the sound of the curve*".

Types of surgery

Participants gave their description about the sound environment in different surgeries. It was unanimous that Orthopaedics is "*definitely the noisiest" (N2)* surgical specialty. On the day of the interview A1 had worked on an orthopaedic surgery and said, "*The noise from an orthopaedics room is extremely annoying*". N2 added that on orthopaedics surgeries, with the electric motors, saws, and other equipment, it is difficult to hear the alarms. Similarly, N3 mentioned that "*orthopaedics is a very tiresome specialty. I don't know if due to the noise or other factors, but I always leave more tired from an orthopaedics room than from a general surgery".*

Other addressed specialties were Obstetrics, Urology, Neurosurgery and Plastic Surgery. According to S1, obstetrics is not particularly silent but due to the emotional component of the caesareans, not exactly the equipment. Urology was described as "*noisy",* and Neurosurgery as calmer, N2 added that "*here, the noise bothers".* These are longer surgeries; they might take an entire day "*so we might hear that constant noise the whole day. There the alarms from the monitors are nicer"* (N2). Finally, Plastic Surgery was described as having more delicate equipment and instruments.

Perceptions of the Opinion of the patients

Often the patients during surgery are fully awake and listening, and we wanted to understand if patients have ever manifested any opinion regarding the monitoring or alarm sounds.

[9] All vocalizations are transcribed from the original Portuguese description.

A1 referred some patients complain about the noise when they have regional anaesthesia, or they ask to be put asleep "*you know, if you listen to hammers, nails, saws…sometimes I leave them sedated because of the noise".* As an alternative, N1 mentioned that she usually places some headphones on the orthopaedics patients, so they do not hear the noises from the electric saws. So, most complaints are related not with the monitoring noises but with the intense noise from surgery and conversations.

Nurses worry about having a pleasant environment for everyone so there is often music, and they also communicate a lot with the patient. As mentioned, participants reported having heard no complaints about the monitoring noises, nevertheless, S1 acknowledged they can generate some anxiety. This surgeon has inclusively seen patients getting frightened by the changes in the audio signals "*Sometimes when an alarm sounds, they get scared and ask what is happening, they are aware of the alarms*".

In the recovery rooms, N2 reported the patients with light anaesthesia complain about the monitoring noises and say things like "*When will I leave this room?*".

False alarms

Although not the focus of this research work, it is inevitable to talk about false alarms when discussing audio warning signals. All professionals interviewed characterized well when these would sound, and why. A1 mentioned alarms which might sound because the monitor finds the values observed as meaning something is wrong, but which might be correct for a given patient.

> *"Another thing that happens quite frequently is, I haven't ventilated the patient, he is breathing spontaneously, but I've connected an oxygen port to supplement. Often, if I'm not careful with the ventilator, it will fire an alarm because it doesn't have any input of the exhaled air, it's like the patient is in apnea, and those are alarms we deal with daily, and we already know when they will sound and why" (A1)*

In the case of obstetrics, S2 referred the whole environment as interesting as a case-study because it is noisy and with a lot of false alarms – which generate a lot of anxiety to the couples. N2 talked about the case where electrodes fall during surgery (pressure cuff, oximeter sensor on the finger), or there is signal interference. The nurses manage to identify, through sound, what is disconnected and see all other parameters indicating that all is normal.

6.4.1.4 Alarm-triggered behaviour

Alarms are triggered whenever a value is over or under a defined threshold. Monitors have default values, which can be configured according to the patient's normal values. According to N2, whenever a parameter alarms in the monitor, the auditory signal is the same for all parameters. Other equipment like ventilators, also sound alarms in case of a change in the ventilation of the patient.

Reactions to sound

Regarding reactions to alarms once they sound, as anaesthesiologist A1 said, the noises she will or will not attend to will differ from surgery to surgery. She said the most important signal for her is the pulse oximeter, which is an advantage because it can be easily interpreted. Nevertheless, according to her, it is impossible not to look at the monitor after an alarm sounds. N1 added that, as a nurse, the usual reaction to sound is to look at the monitor because the sound of the alarm is always the same. Because of the monitor, there is never any doubt as to who is the patient emitting the signal, as confirmed by N2 and N3. After, a common action is to silence the alarm by pressing a button. S1 also mentioned that when something sounds different, they tend to look at the monitor to see what is going on. As a surgeon, for S2 the alarm does not work as tool, and when something starts beeping, it creates some anxiety. They then ask the nurse or the anaesthetist what is the situation because it might be a consequence of something they are doing to the patient.

Common practices

In terms of common practices after the alarms sound, it was made clear that no anaesthetist turns the alarms off because that might lead to a big failure after a mere distraction. For them, and A1 spoke considering the professionals she knows, the monitoring sounds need to be there and are a fundamental part of their work. Nurses who work in the OR deal with different surgeons, surgeries, and patients, and during surgery, some surgeons like to listen to music – sometimes very loud. Regarding alarms, nurses have the habit of repeatedly silencing the alarms for 2 minutes (sometimes they change the thresholds of the parameter, others just press the "silence" button). After these two or three minutes the alarm feature resumes.

In the recovery room, nurses say their attention is not exclusively caught by alarms. They know the patients, and they read their symptoms in the way they look and in the monitors. N3 said that alarms are helpful in situations where, for example, a nurse is busy with one patient but if she listens to an alarm, she will immediately turn her head and see what is going on.

Listen without looking

Professionals value the ability to be able to identify what is going "in the background".

> *"Without looking at the monitor I know how much the patient is beating if he is saturating well or not. That sound for us is fundamental. (…) I am looking at the patient and while I am listening that constant noise – beep – I don't worry, when I start listening to an abrupt change I go, hey there this is getting serious" (A1).*

Similarly, N2 said there is a sound from the monitor which you do not need to look at all, which is when the patient is assystolic (possible heart stop). "*In that case the sound is different, pretty annoying, when we hear it, we basically run*" (N2).

N1 summarized the problem of having similar alarms for all parameters. According to this nurse, some monitors have the same signal for all events while other monitors have different signals, which in the end, makes professionals always having to look at the monitor to understand what the problem is.

> *"For example, in our monitors with the same alarms for any parameter, maybe it would be important to have a specific sound, which was unequivocally different from an alarm from another parameter. It makes all the difference, and we could be trained to identify each alarm and understand exactly what it means."*

6.4.1.5 Suggestions for improvements

One of the main questions of the interview consisted in understanding which parameter's alarm could be improved with a re-design. This had similar responses from all participants, which admitted never having thought of that, and were rather unanimous in identifying important parameters. The Blood Pressure alarm was referred as an important one, and one to which all professionals are very sensitive to. However, it is important to notice that the most important parameters differ according to the anaesthesia the patient has. As N3 puts it, *"It depends. If it's a patient with general anaesthesia, co2, oximetry are important parameters. If it's a local-regional patient, because of the vagal attacks he might have, maybe blood pressure and heart rate. Temperature for example, is not so important as heart rate."*

Participant A1 referred the blood pressure parameter as important to redesign: "*Maybe the blood pressure, having a specific alarm. Yeah, that one. Blood pressure does not have a constant noise associated, only when it's out of its regular values. It is a low*

sound. It could be improved – how? I have no idea!". Similarly, N1 suggested blood pressure as well: "*The Blood pressure for example. There is a measurement which is abruptly lower, we need to act immediately, we are very sensitive with blood pressure. All the others depend on it*". Finally, S1 agreed that "*Blood pressure and heart rate are the most important ones for us because these are the ones more related to the hemodynamic state of the patient, which will make him bleed more or less – we are always concerned with that."*

6.4.2 Quantitative data from online survey

Similarly to the previous section, results were clustered into four main dimensions: (1) *Physical space, (2) Sound and Alarms, (3) Alarm-triggered behaviour* and *(4) Suggestions for improvements*. Descriptive statistics are reported as frequency and percentage (%) for categorical data.

6.4.2.1 Physical space

The summary of all answers in the Physical Space dimension is in Figure 6-1 and Figure 6-2.

Positioning

To the question of whether the position of nurses inside the operating room differed according to surgical specialty, 87% of the respondents answered the statement was true, but some added that anaesthetist nurses always stood at the head of the patient. In the case of circulating or instrumentalist nurses, their position could be anywhere, depending on the surgery. As a general statement regarding the distribution of equipment in all their rooms, 70% of the respondents said it was True the Hospitals tended to maintain the same equipment distribution, and 46% did not know if the positioning differed from hospital to hospital.

Recovery Room

As for the Recovery Room (RR) respondents worked in, they all had a very different number of beds. Eighteen respondents worked in a RR with 10 beds, five worked in a RR with 12 and 15 beds, three worked in a RR with 9 beds, and one to two respondents worked in a RR with 1 to 20 beds. 87% reported having visual aids (e.g.: lights) inside the RR that signal when an alarm is triggered (one responded having four different brands for 10 beds, so it depended on the brand), and 67% reported as being False that the auditory alarms

are more important here than inside the operating room. Respondents added in the comment section that they are equally important, but some suggested that in this context the nurse may not be continually at the bedside, and that a higher number of patients require more vigilance performed by a reduced number of professionals.

Equipment

According to 91% of the respondents, they interact daily or weekly with equipment from different manufacturers and 83% reported working with vital signs monitors from different manufacturers.

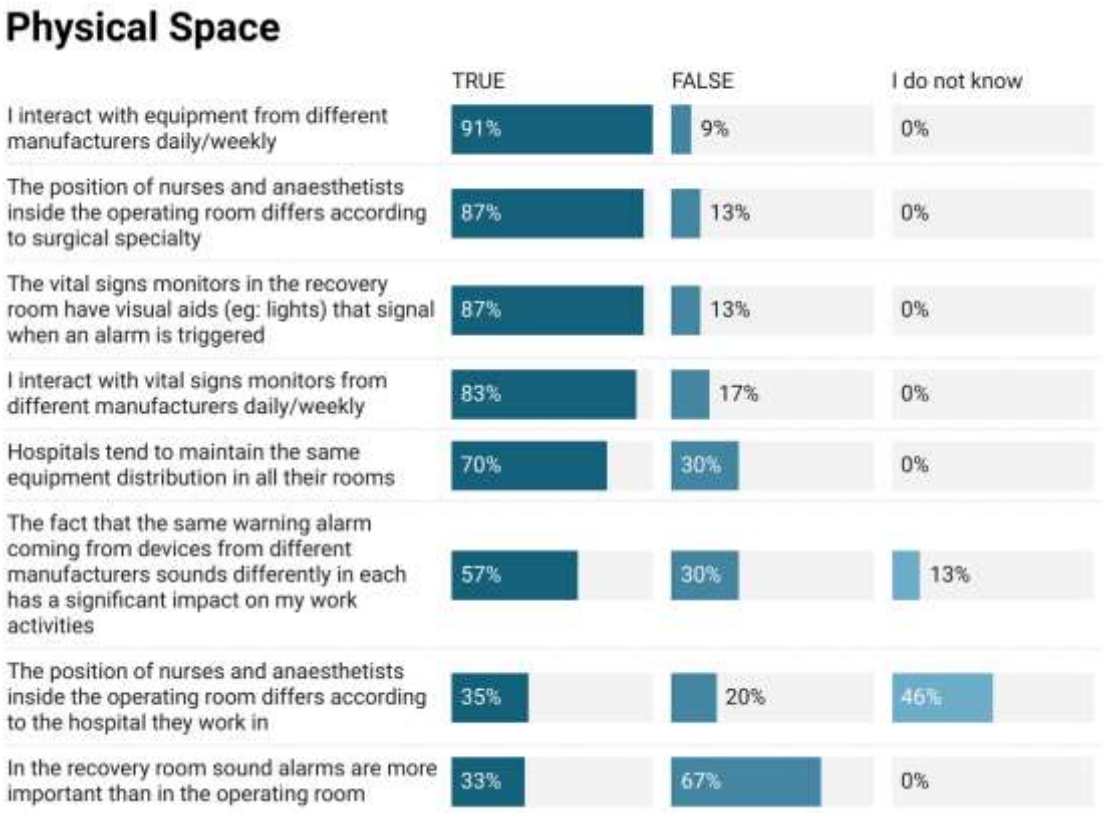


Figure 6-1. Summary of False/True answers for the Physical Space dimension of the survey (N=46).

When respondents interacted with monitors from different manufacturers, 35% reported noticing "Very Frequently" that the sounds were different, even though they communicated the same events or parameters. It might happen that in some practices they have the same manufacturer for each type of device, considering different devices communicate different events. The fact that the alarms sounded differently according to manufacturer had, for 57% of the respondents, an impact on their work.

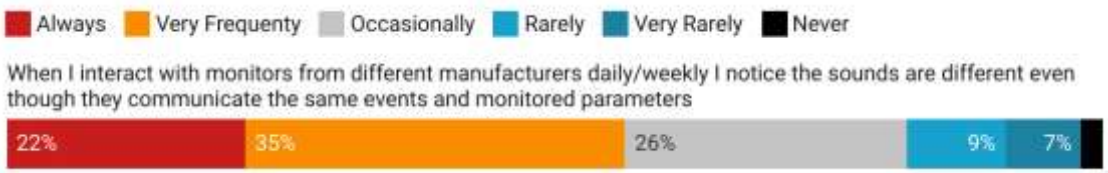


Figure 6-2. Summary of Frequency answers for the Physical Space dimension of the survey (N=46).

6.4.2.2 Sound and Alarms

The summary of all answers in the Sound and Alarms dimension is in Figure 6-5Figure 6-6 and Figure 6-6.

General Opinion on/the importance of Sound

For 91% of the respondents, noise is part of their professional life, and 83% affirmed that auditory alarms had a significant impact on their work activities. Some of the comments added these are important to get the attention of the professionals, but can also become a source of annoyance for being so insistent.

The main sources of noise inside the operating room according to respondents are auditory alarms from medical devices (31%), conversations (27%), Music/Radio/TV's (19%), Personal devices (14%), and Instrumentation (9%). Other noises sources added in the comment section were body heaters and liquid aspirators.

Considering the continuous sound of air conditioning, 41% reported being disturbed by it "Occasionally", and 17% "Very Frequently".

Parameters

For the respondents work, the most important alarms are monitoring alarms (77%), followed by laparoscopic alarms (23%). The most relevant parameters for their tasks are presented in Figure 6-3 (participants could select more than one and, on average, each selected 4 parameters).

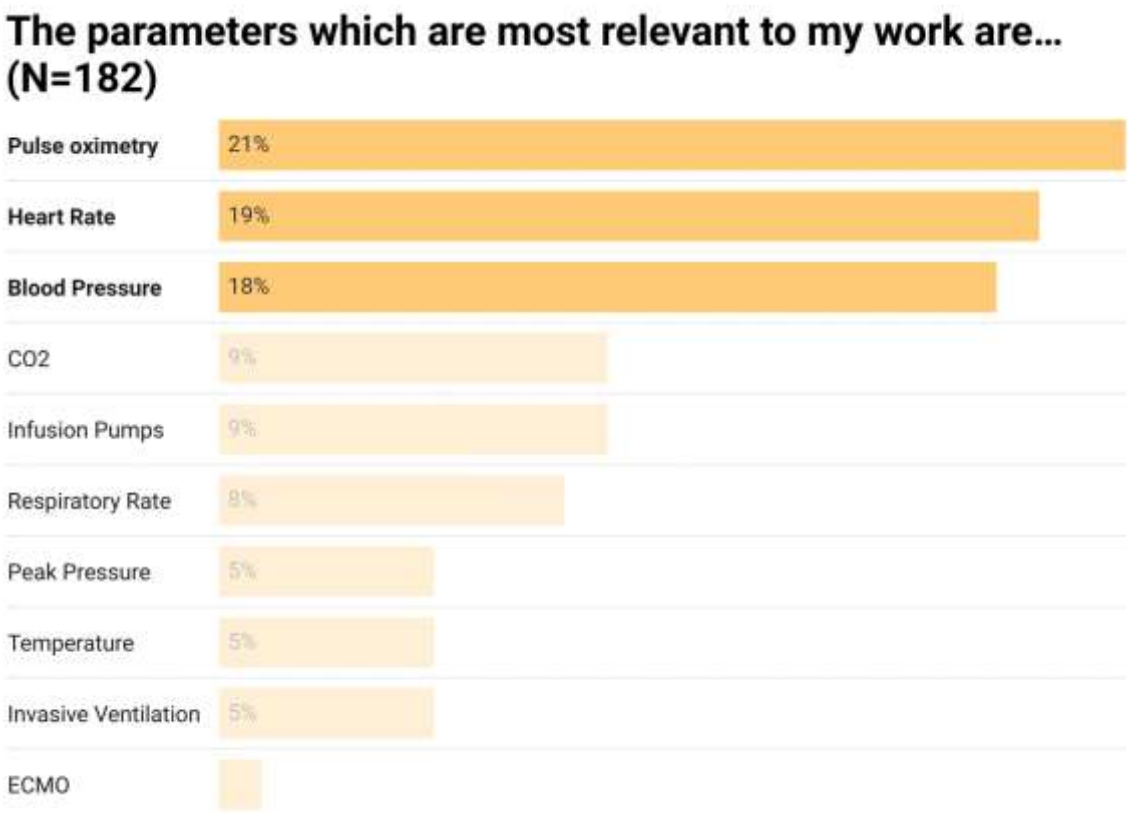


Figure 6-3. Most relevant parameters for their tasks according to online survey respondents.

One respondent added the Pulse Oximetry is the most important alarm, and professionals usually complement observations with the heart rate and blood pressure. Regarding Infusion Pumps, some added that pumps malfunctions cannot be easily identifiable without an alarm.

Types of surgery

When asked which surgery they considered the noisiest (Figure 6-4), 20 % of the answers (more than one could be selected) were attributed to Orthopaedic Surgery for the materials used, followed by Urology (13%) and General Surgery (12%). Some were considered noisy due to the teams, and Neurologic and Thoracic surgeries were considered noisy only in the beginning of the surgery.

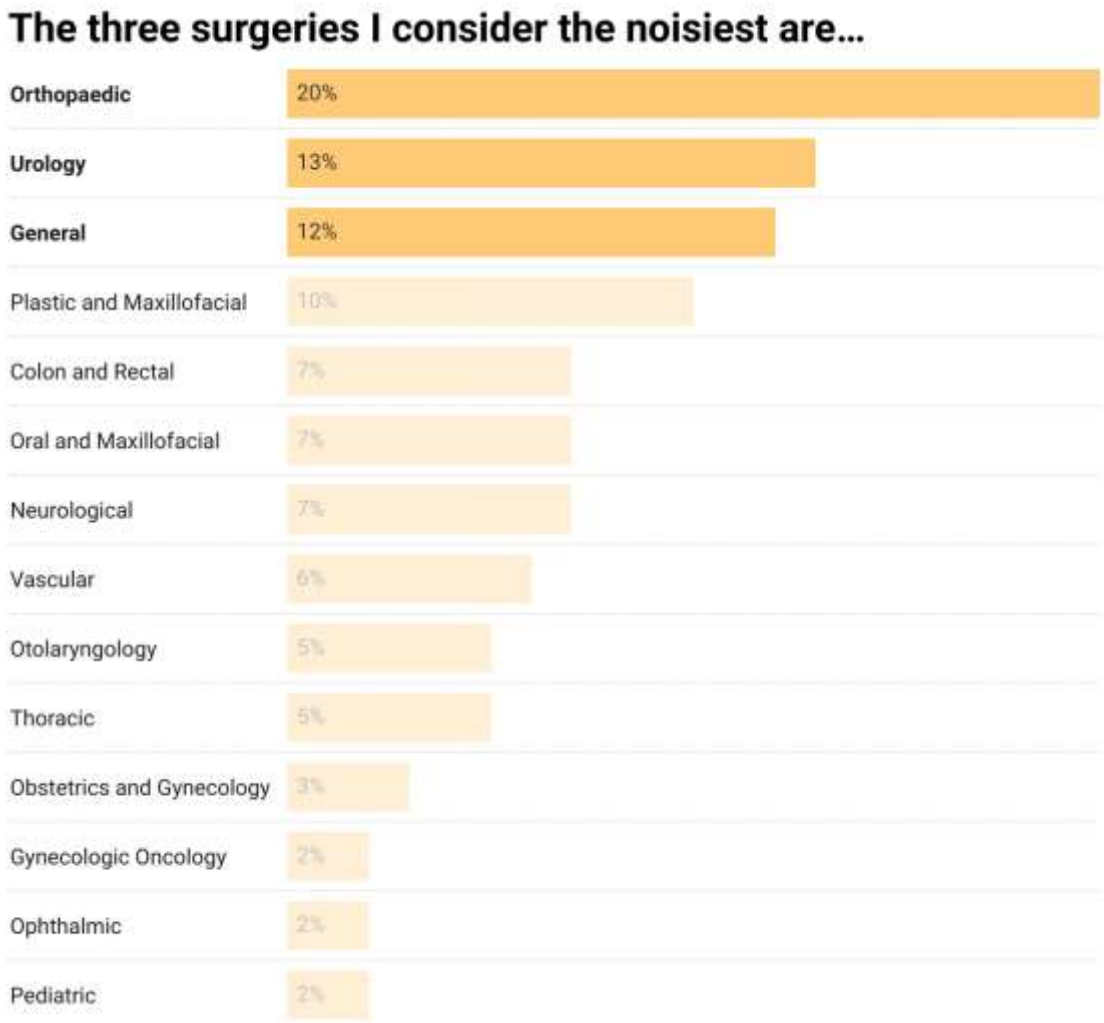


Figure 6-4. Ranking of the noisiest surgeries according to survey respondents.

Considering some surgeries are noisier than others, 74% of the respondents affirmed they always heard the auditory alarms.

Perception of Opinion of the patients

Regarding the perceptions respondents had about how the patients felt in relation to the auditory alarms inside the operating or recovery rooms, 78% of the respondents agreed that monitoring sounds generated anxiety on patients during surgery when they are away, and 70% agreed these alarms also generate anxiety on patients in recovery rooms. For 52% patients or family members have "Very Frequently" got scared with alarms which were signalling trivial information.

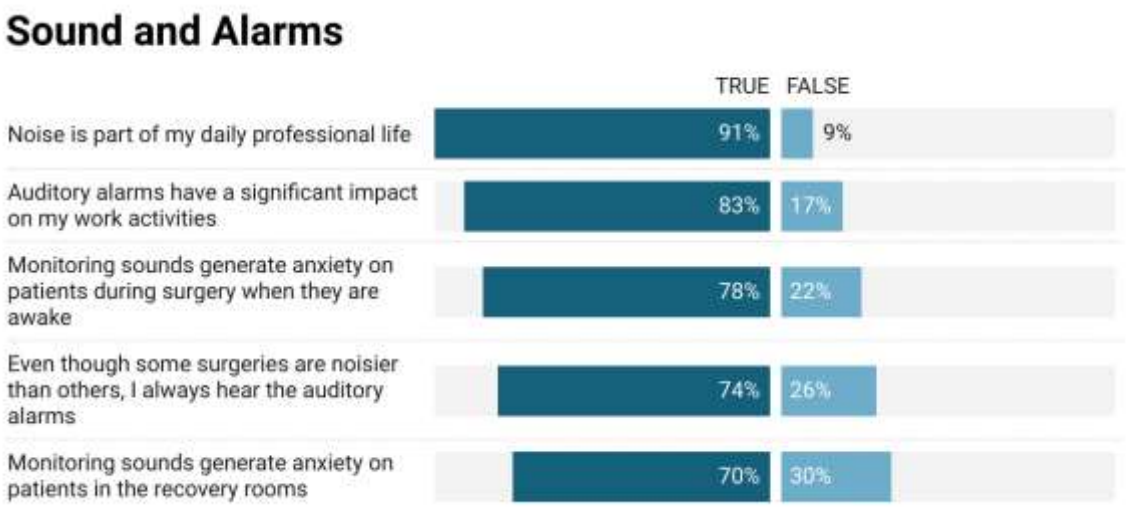


Figure 6-5. Summary of True/False answers for the Sound and Alarms dimension of the survey.

False alarms

In relation to false alarms, respondents can "Very Frequently "(46%) and "Occasionally" (28%) identify the source of a false alarm, like when there is an interference with equipment, or a sensor has fallen.

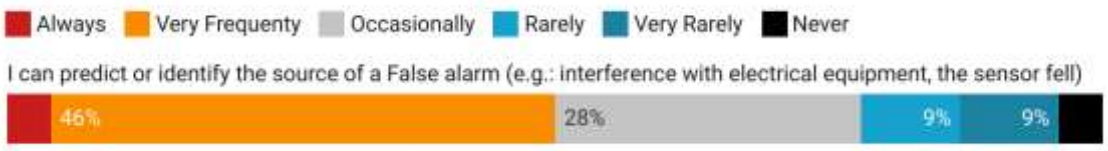


Figure 6-6. Summary of frequency answers for the Sound and Alarms dimension of the survey.

Sound Description

Although limited by the available multiple choice (respondents could choose more than one adjective), when asked how they would describe the alarm sounds from medical devices, 23% choices fell on the "repetitive" adjective, followed by "annoying" and "unpleasant". Although mostly negative words were chosen, 7% of the selections went to the "well-designed" adjective. Other proposed words were "fine", "safe if adapted to the patient" and "necessary" (Figure 6-7).

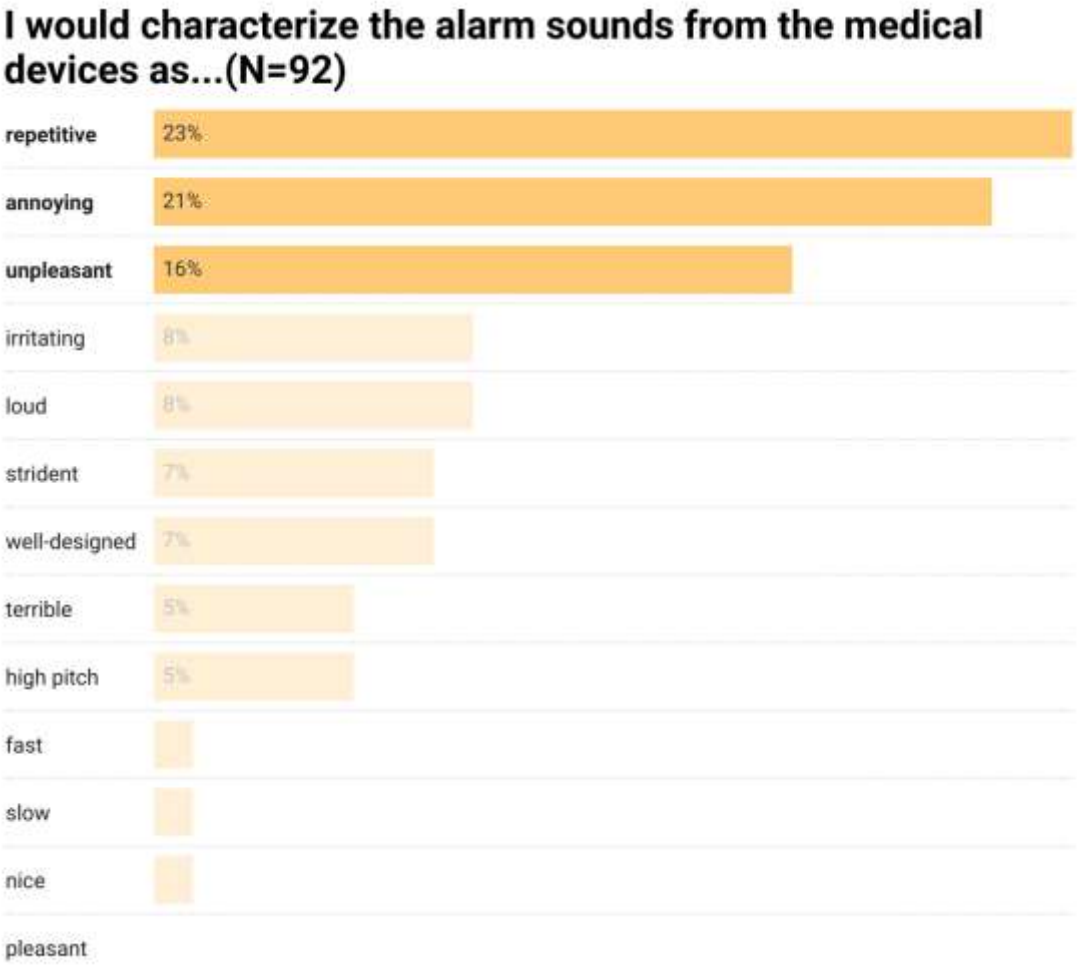


Figure 6-7. Words used to characterize alarm sounds according to survey respondents organized by order of frequency.

6.4.2.3 Alarm-triggered behaviour

The summary of all answers in the Alarm-triggered behaviour dimension is in Figure 6-8 and Figure 6-9.

Reactions to sound

When hearing an alarm, 41% of the respondents said their first reaction was to "Always" look at the monitor, 39% said they did this "Very Frequently".

Because during the interviews some professionals mentioned how they liked the pulse oximetry sound, we asked the same question in our survey and 74% of the respondents considered it a well-designed sound. One respondent corrected that it is not an alarm, but an auditory signal of the measure, and another added that pitch modulation was transformative.

Common practices

As for common practices, 76% of the respondents affirmed their first reaction when hearing an alarm was to silence it for 2-3 minutes (depending on the cause). Most (72%) stated that it is not usual to change the thresholds of the alarms, and when it is, it can be due to 1) critical alarms and thresholds needing to be changed (due to a predictable situation); 2) the cause (when the cause for the alarm is known); 3) the anaesthetist.

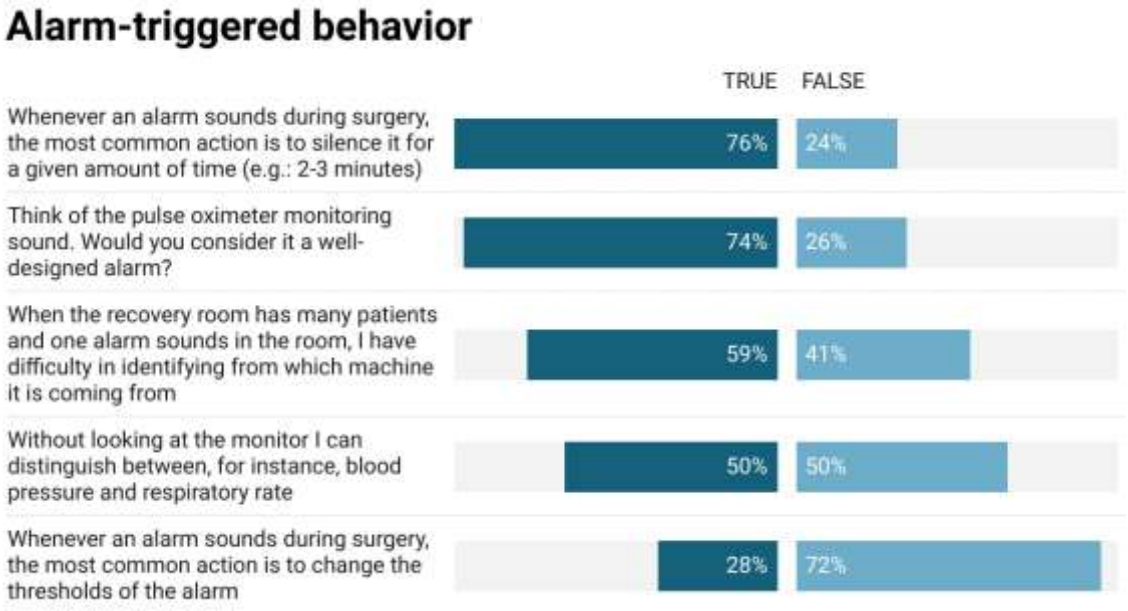


Figure 6-8. Summary of True/False answers for the Alarm-triggered behaviour dimension of the survey.

Listen without looking

Among our respondents, 37% mentioned they "Rarely" had difficulties in associating the alarm sound to the event it signals. Some added there were very different types of alarms but very few different sounds, or that they rarely had difficulties if the same type of sound comes from all monitors. 30% said they "Rarely" had difficulty in associating the alarm to the priority of the event.

In the context of the Recovery Room with several patients, 59% admitted having difficulties in identifying from which machine the alarm is coming from, even though some affirmed having a telemetry screen where they can see where the alarm is coming from. Considering the importance of knowing what is happening without having to look at the monitor, 74% referred that "Very Frequently" they knew what the event or problem was before looking. However, 50% said they could distinguish between, for instance, blood pressure and respiratory rate alarms while 50% said they could not (it depended on the monitor).

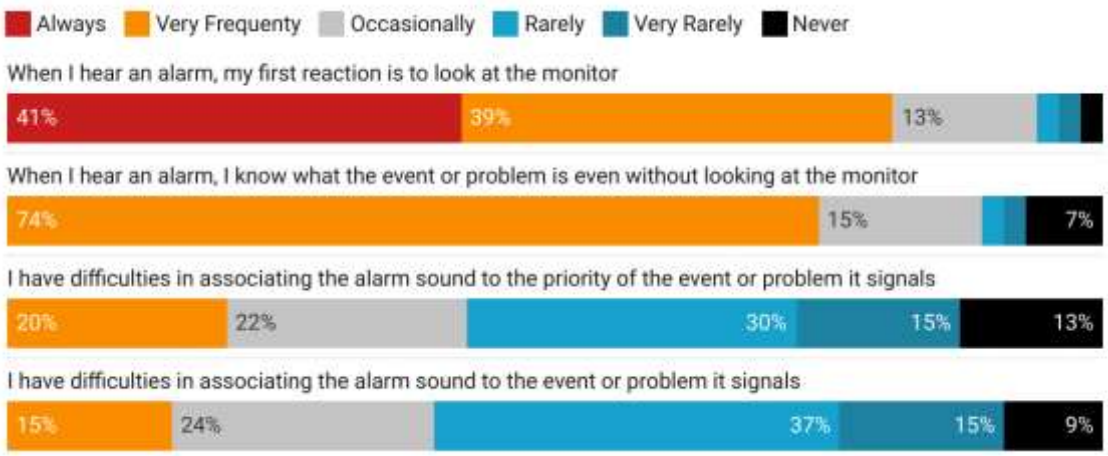


Figure 6-9. Summary of frequency answers for the Alarm-triggered behaviour dimension of the survey.

No correlation was observed between the answers to these questions and the question "*For how long you have been working in the operating room*?", indicating that the ease or difficulty with which professionals interpret alarm sounds is not related to years of experience.

6.4.2.4 Suggestions for improvements

Having equipment from one single manufacturer would be an advantage for 89% of respondents, in terms of visual and auditory interfaces. Regarding the types of auditory interfaces, 91% agreed that having distinct alarm sounds for the most important parameters would improve their work activities (although some mentioned it being hard to know without trying beforehand). The summary of all answers in the Suggestions for improvements dimension is in Figure 6-10.

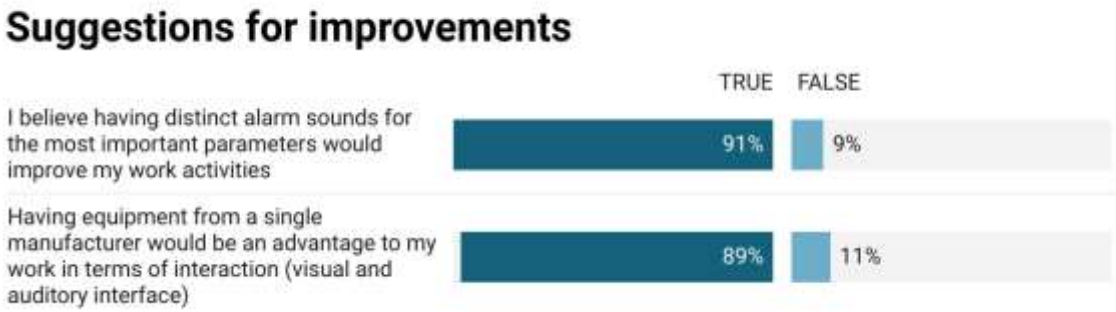


Figure 6-10. Summary of True/False answers for the Suggestions for improvements dimension of the survey.

Parameters which could be improved

Participants could select several parameters to have their sound redesigned. 26% of all choices fell on Pulse Oximetry, and 24% on Heart Rate and Blood Pressure (Figure 6-11Figure 6-11. Which vital sign parameter should be redesigned according to survey respondents, organized by order of).

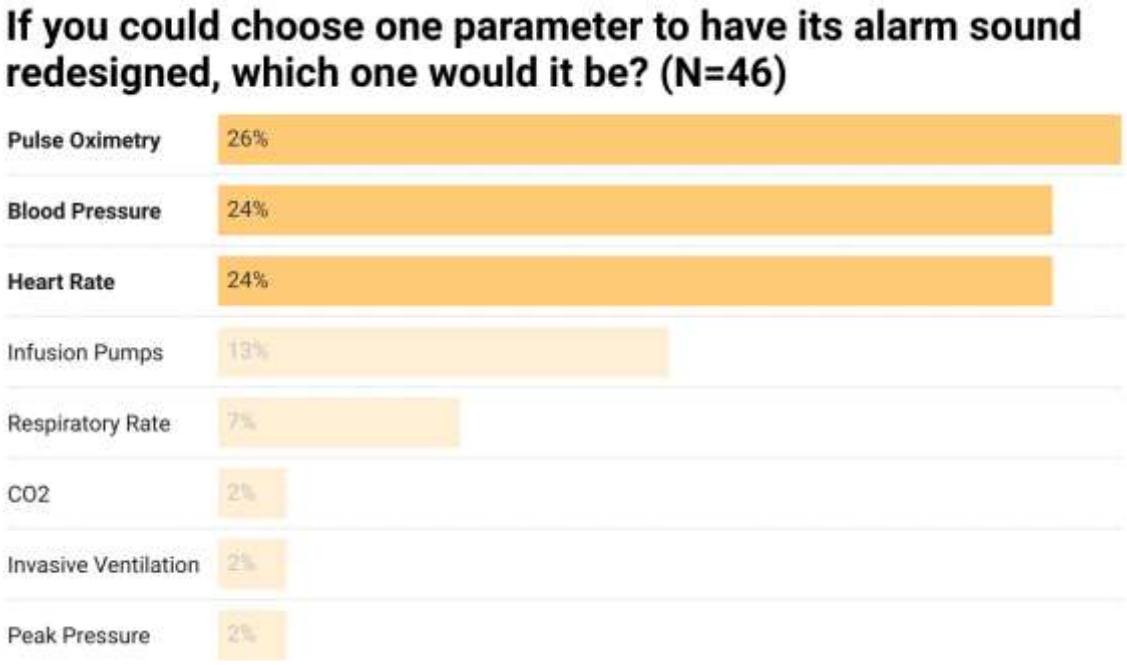


Figure 6-11. Which vital sign parameter should be redesigned according to survey respondents, organized by order of frequency.

6.5 Discussion

Interviews with six healthcare professionals from different Portuguese hospitals and cities allowed us to understand their specificities but most of all, the communalities of their daily practices inside the OR and RR regarding clinical auditory alarms. After an analysis of the interview data, a more structured survey was elaborated and spread online, to obtain feedback and experiences related to auditory alarms from a larger population of healthcare professionals. This survey was filled by forty-six operating room professionals who worked in the operating and/or recovery rooms from all over the world.

In what concerns physical space, we learned the position of nurses inside the operating room might differ according to room and surgical specialty or procedure. The survey results were inconclusive regarding the positioning of equipment per hospital. According to the interviewees, hospitals tend to maintain the same equipment distribution, and professionals must adapt to each spatial organization.

Concerning equipment, professionals interacted vital signs monitors from different manufacturers daily or weekly. In this case, not only are the visual interfaces different, but the auditory interfaces as well, as different brands have different auditory alarms, and this difference is evident for all participants of this study. This has, according to our survey respondents, a significant impact on their working activities. Interviewees stated that hospitals are adopting monitoring equipment from the same manufacturer for economic reasons, but that is also very positive for their work. Survey respondents also informed that this was becoming common practice in some of their hospitals.

This diversity in auditory alarms and difficulty in associating a sound to an event has, in our view, motivated a very strong behaviour in all interview and survey participants: whenever an alarm sounds, their first reaction is to look at the monitor to understand what parameters crossed a threshold. For this reason, and unlike survey respondents, interviewees stated that differences in sound do not have a significant impact in their work. Nevertheless, according to our own interpretation of the results, motivating the behaviour of looking at a monitor is by itself a significant impact of an alarm, as it moves their attention to a different area. Survey respondents mentioned they Rarely had difficulties in associating an alarm to and event or priority of an event, however, between 15-20% reported having this difficulty Very Frequently. Independently of being able or not to make the association of the event to the alarm sound, most participants mentioned it would have been advantageous to their work to have distinct sounds to distinct parameters, much in line with the results obtained with 250 nurses (Casey et al., 2018), in a study conducted with nurses and doctors about clinical alarm systems (Thangavelu et al., 2014), or a national survey performed in the USA (Korniewicz et al., 2008).

Another frequent event is the occurrence of false alarms during surgeries. Most often than not, professionals can predict a false alarm will be triggered, and once triggered, they can easily identify why it occurred. The examples referred included the monitor lacking an input for a given parameter from a disconnected sensor, or another equipment interfered with a different signal. Survey respondents also stated that Very Frequently and Occasionally they could identify one sound as false alarm.

Almost all participants in both data collections accepted noise as being part of their daily professional lives, distributed amongst auditory alarms, conversations, music, and personal devices. Survey respondents selected as the noisiest surgeries Orthopaedics, Urology and General Surgery, and stated they could always hear the alarms most of the times, independently of how noisy the surgery was. This slightly differs a more factual

statement from (Casey et al., 2018) in which 39% of nurses evaluated as true: "There have been frequent incidences where alarms could not be heard or were missed".

An interesting observation from the interviews was that surgeons, anaesthesiologists and nurses attributed different importance to alarms. For the latter, monitoring alarms were the most important, while for surgeons, their attention was mainly directed to sounds related to cut and coagulation. In the survey, monitoring alarms were considered more important for their work than laparoscopic alarms. The sources of these alarms are conveniently located in different places. The monitoring sounds and alarms come from the head of the patient and the laparoscopic equipment is by the patient's feet. For the interviewed anaesthesiologist, the most relevant parameter for her work is pulse oximetry, nevertheless, the sounds professionals will attend to differ from surgery to surgery. A Greek study on operating rooms obtained similar results, with anaesthetists being more sensitive to noise than surgeons (Tsiou et al., 2008). The three parameters evaluated as more important in the survey were Pulse oximetry, Heart Rate and Blood Pressure. Another interesting information captured was that during surgeries it is common to repeatedly silence alarms for 2-3 minutes, depending on its cause. More rarely, the thresholds of the alarm are changed. When these are changed, it may be due to 1) critical alarms and thresholds needing to be changed (due to a predictable situation); 2) the cause (when the cause for the alarm is known); or 3) the anaesthetist. One explanation might be that properly setting alarm parameters is overly complex in existing devices (Casey et al., 2018).

For a different perspective on auditory alarms, participants were asked about reactions from patients whether they were in surgery or in the recovery room. Participants referred the monitoring alarms often generated anxiety in the patients during surgery or in the recovery rooms, which is a very important information. Auditory alarms coming from medical devices are also heard by patients and sometimes their family members who, unknowingly, might make erroneous interpretations based solely on how they sound. These reports match those collected in an ICU study, in which patients reported that "*when they heard these unfamiliar sounds, especially at the beginning, many had frightening thoughts about the meaning of the sound*" (Johansson et al., 2012).

The situation in recovery rooms is different, as there are more patients being monitored, but also no instrumentation noise. Respondents had very different experiences in this aspect, with some participants working with 2 beds inside the recovery room, while others worked with 20 beds. Here, medical professionals can look at the patient and at the

monitor. Although survey respondents answered that alarms are equally important in the recovery room as in the operating room, according to our interviews, here sound is more important when, for instance, one nurse is with one patient and an alarm coming from a different patient is triggered. Although some participants stated that they have had difficulty in finding immediately from which patient monitor was the alarm coming from, most have referred that this is mitigated by a visual aid (like a red light) coming from the sounding monitor, however, this depended on the brand.

Finally, regarding the health monitoring parameter which would improve their work with a redesign, all interviewees mentioned the blood pressure parameter. The survey respondents also included this parameter in their selection, close to Pulse Oximetry and Heart Rate (more than half of the survey respondents considered Pulse Oximetry a well-designed alarm). Most stated that it would be important to have specific sounds for specific parameters. N1, the most experienced nurse with 25 years inside the OR stated that being able to identify the alarm before approaching the patient would probably mean saving one or two seconds and would be helpful in gathering the tools or necessary material beforehand. An individualized sound would help professionals being quicker in solving an emergency.

Throughout the six conversations, the most used adjectives were: annoying, unpleasant loud, fast, nice, strident, terrible, high pitch, repetitive and irritating. Only one positive note was made about the auditory alarms, and it was addressed to the pulse oximeter as it is "*an extremely well-designed alarm*" (A1). In the survey, the most frequent words used to describe auditory alarms were repetitive, annoying, and unpleasant.

The use of two sequential methods to understand the research problem was enriching, and consistency was observed in most of the questions. What is more, interviews allowed to reveal new problematic areas that were not initially the focus of our study, like the relevance given to alarms in recovery rooms, which is different from the scenario inside the operating rooms.

6.5.1 Limitations of the Study

Both studies are dominated by the opinions of nurses, and while the interviews have the opinions of two surgeons, the survey did not have the participation of surgeons. Nevertheless, most studies performed with healthcare professionals have a majority of nurses as participants (Korniewicz et al., 2008; Thangavelu et al., 2014), and in the end they are the ones which are most exposed to clinical auditory alarms throughout all their

professional lives. Another limitation is that having closed answers for all questions for the benefit of ease of filling may have affected the outcome of the survey, even though we have tried to minimize this limitation by using Likert Scales and True/False answers for most of the statements.

6.5.2 Design requirements and Needs

As the main aim of the current study was to gather human-centred needs and requirements regarding auditory alarms in the operating and recovery rooms, we summarize them in the following list:

i. Having unambiguously different sounds for different parameters would be advantageous.
ii. Similar types of devices should have the same audio alarms and not be brand dependent.
iii. Everyone should be involved in the design process of auditory alarms (as everyone is a potential patient).
iv. Pulse Oximeter as an example of a well-designed sound (even though it is not an alarm).
v. Pitch modulation is very informative.
vi. Alarms should be easily learned in training.

6.6 Conclusion

The interviews and online survey to healthcare professionals provided important information regarding their opinions, reactions, and behaviours towards the sounds inside the operating and recovery rooms. Understanding their work environment and needs is a fundamental stage at the human-centred design process, which is the goal of the present study. Main findings include:

- Nurse position differs according to room and surgical procedure.
- All professionals interact daily with equipment from different manufacturers.
- Most professionals notice different brands have different auditory alarms and this has a significant impact in their work activities.
- Hospitals are tending to adopt similar equipment from the same manufacturer.
- After hearing an alarm, most professionals look at the monitor.
- Some professionals have difficulty in associating the sound to an event or a priority.

- The surgeries considered the noisiest are Orthopaedics, Urology and General Surgery.
- The three parameters evaluated as more important in the survey were Pulse oximetry, Heart Rate and Blood Pressure.
- It is common to repeatedly silence alarms for 2-3 minutes.
- Surgeons, anaesthesiologists, and nurses attribute different importance to alarms.
- Monitoring alarms generate anxiety in the patients and their families.
- Alarms in recovery rooms are particularly important because there are more patients for less professionals.
- Some professionals have difficulty in finding immediately from which patient monitor is the alarm coming from in the recovery room.
- Some recovery rooms have monitors with red lights, but it is brand dependent.
- Regarding the health monitoring parameter which would improve their work with a redesign, participants mentioned the blood pressure, pulse oximetry and heart rate.

Importantly, this data has confirmed that individual alarms per parameter would add value to the workflow. Future work will include frequency and sound pressure level analysis inside the operating and recovery rooms, which will provide quantitative descriptors of the work environment inside the operating and recovery rooms.

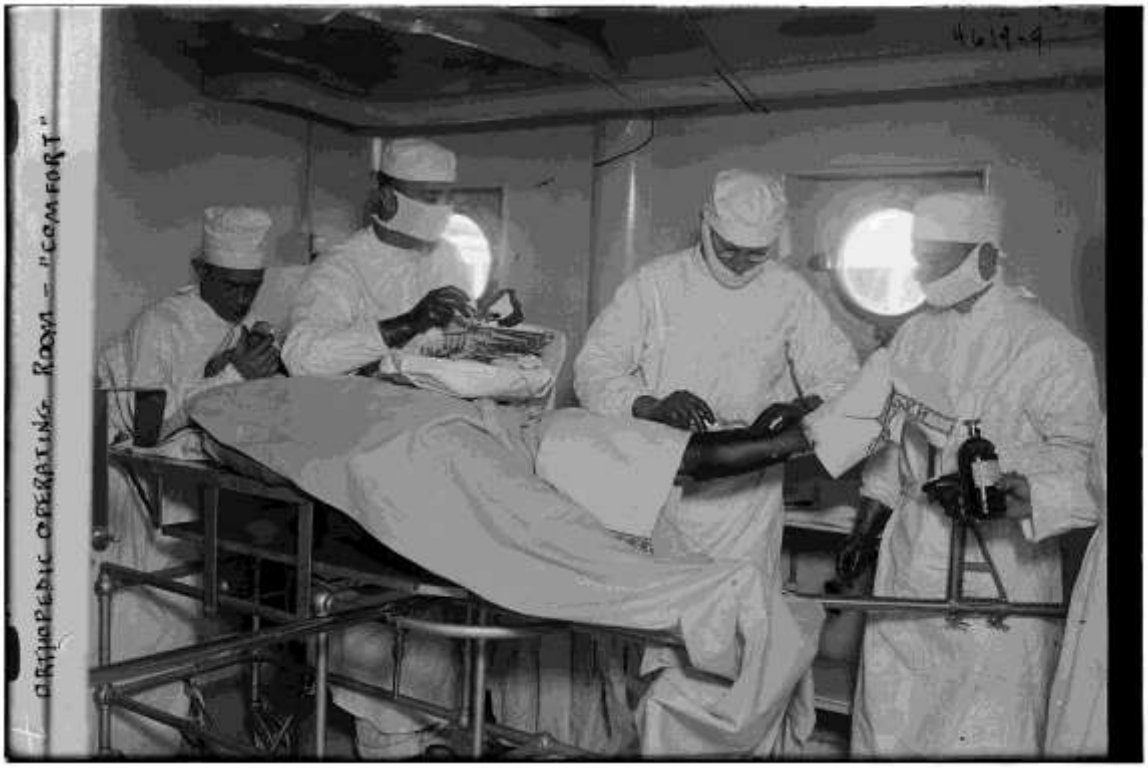
The current work has widespread implications for the design and implementation of alarms in healthcare. Having a thoughtful and people-centred redesign of clinical auditory alarms will create a more positive soundscape and environment inside the operating and recovery rooms.

7 Frequency Analysis in Operating and Recovery Rooms

Designing sound must consider not only the human operators, but also the context in which sound will propagate. Inside a hospital, some wards are noisier than others, and inside an operating room, some surgeries are louder than others.

Conversations, music, instrumentation, equipment, and other surgical artifacts happen within a certain range of frequencies. Characterizing them inside operating and recovery rooms is of the utmost importance. With the goal of designing alarm sounds outside that range of frequencies, 28 surgeries in two different hospitals were recorded. Visits occurred in a private and a public hospital, both in northern Portugal.

This chapter is based on the paper: Vieira, J., Pereira, F., Marques, R., Ferreira, V. Barker, T., Campos, A. G. Noise in Operating Rooms: an appraisal from two post-2010 Portuguese hospitals (submitted). *Journal of the Acoustic Society of America*.
The content of this chapter was sent as an environmental noise analysis report to both hospitals.



Bain News Service (between ca. 1915 and ca. 1920). Orthopaedic operating room – COMFORT. The Library of Congress. No known restrictions on publication.

7.1 Introduction

Noise in healthcare environments is a problem that affects not only professionals, but also patients and the quality of their recovery (Ecophon, 2018). One of the main reported problems is the large amount of auditory alarms triggered by the equipment, following a "*better safe than sorry*" premise (Patterson & Mayfield, 1990) which may unnecessarily disturb patients and family and, more importantly, contribute to desensitising professionals to their importance (Hravnak et al., 2018; Lorenz et al., 2017), thus increasing the probability of important alarms going unnoticed. A study by the Beryl Institute (Madaras & Apps, 2012) has categorized the primary sources of noise inside a hospital after interviewing professionals from 241 U.S. hospitals, and concluded that alarms accounted for 25% of the noise sources. Others include equipment noise, conversations, overhead paging, slamming doors, or cleaning noises. Medical devices' alarms can be heard in the most diverse spaces such as intensive care units, nurseries, patient rooms, recovery rooms or operating theatres. Patients, families and staff can be exposed to 700 or more physiologic monitor alarms per patient per day (Cvach, 2012)– Johns Hopkins Hospital recorded 771 alarms per bed in one day (ECRI, 2013).

Closely related to the characteristics of the auditory alarms are the characteristics of the background noise to which the alarms should overlay. For the specific case of hospitals, the World Health Organization (WHO) recommends exposure limits around 30-35 dB LAeq for interior spaces, and a 40 dB LAmax during the night (23h00 to 7h00) (World Health Organization, 1999). However, as complaints from staff, patients and visitors clearly indicate, the reality of hospitals' soundscapes is a long way from complying with these recommendations (Ryherd et al., 2008). A multitude of noise sources (the moving parts of medical equipment, synthetic sounds (namely "beeping alarms"), paging systems, floor cleaners, and air-conditioning systems, people talking…) make them "*far from restful*" (Ryherd et al., 2008). As the landmark research from Busch-Vishniac et al. (2005) showed, over the previous 45 years the Leq levels in hospitals rose from 57 dB(A) in 1960 to 72 dB(A) in 2005. The same applies to night-time measurements rising from 42 dB(A) Leq levels to 60 dB(A) in the same period. Noise levels inside hospitals have continuously increased since the 1960s. No published study has ever recorded a hospital soundscape which complied with WHO guidelines (Busch-vishniac et al., 2005; Chen, 2019; Darbyshire & Young, 2013; Lawson et al., 2010; White & Zomorodi, 2017). Also, standard IEC 61672-2:2013-2017 (IEC, 2017) provides recommendation values of exposure time (Table 7-1) to certain levels of sound pressure. For instance, in 8h it is not advisable an exposure to levels above 85 dB(A).

Table 7-1. Recommended sound pressure levels according to IEC 61672: 2013-2017

TIME OF EXPOSURE	RECOMMENDED SOUND PRESSURE LEVEL DB(A)
24 h	80
16 h	82
8 h	85
4 h	88
2 h	91
1 h	94

For reference, the north American Occupational Safety and Health Administration (OSHA) presents Figure 7-1 evaluating the typical sound pressure levels of familiar sounds. For instance, a conversation happens at around 60 dBA, while at 80 dBA one is exposed to the sound of a freight train.

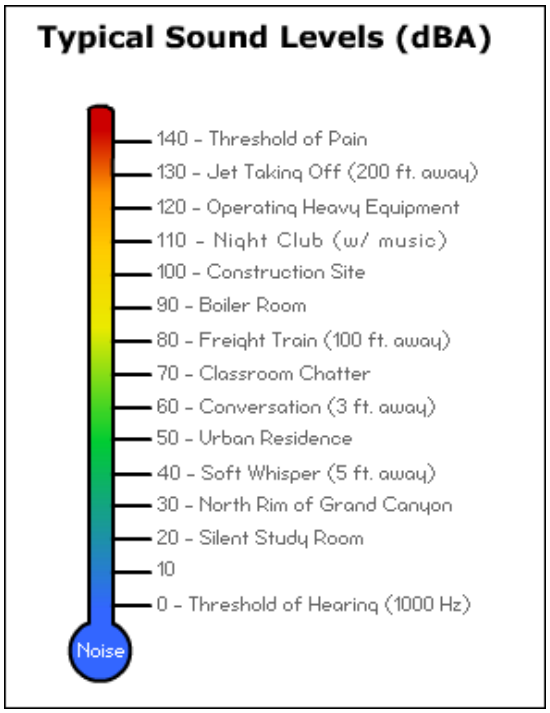


Figure 7-1. Evaluation of typical sound pressure levels according to the north American Occupational Safety and Health Administration (Osha.gov)

There are two variables that can be interpreted as noise mitigators or facilitators inside hospitals. One is the year of construction of the hospital's facilities, and the other is cultural background of the healthcare professionals. Concerning the latter, and to the best of our knowledge, there are no comparative studies that could support this statement. However, in informal conversations, surgeons with work experience in different countries,

have stated how they noticed a difference in conversation levels and general behaviour in countries such as The Netherlands and Portugal – Portuguese surgery rooms were louder and noisier than the ones in The Netherlands. The existent noise level studies characterize several rooms from hospitals in the same country. However, these studies do not always apply the same methodology and recording settings (fast/slow setting, integration time, recording duration) which makes comparisons between the noise inside hospital wards of different countries a hard task.

Considering the year of construction of the hospital's facilities, in the UK, five general inpatient hospital wards were surveyed in 2016 with continuous noise monitoring over several days in 31 locations. Daytime levels ranged from 50 to 61 dB(A) Leq and night-time levels from 41 to 51 dB(A) Leq. Noise levels in wards built after 2000 were lower than in older buildings, suggesting that construction year may be a relevant variable, thus bringing hope that the situation may be improving.

In the first study of the kind in New Zealand, one inpatient ward was evaluated in a 9-day period. The measured 1-hour average Leq levels, calculated for 24-hour periods, fluctuated between 52 and 55 dB(A) (Hulland et al., 2020). This study assessed the contribution of each noise source to the total noise. Staff produced noise was found to be the main contributor, followed by patient and machine noise.

In the U.S., several locations in the Johns Hopkins Hospital were evaluated for 24-hour periods, with 1-hour averaged Leq levels in the 55 dB(A) to 60 dB(A) interval. Corridors were found to be the noisiest areas (Busch-vishniac et al., 2005). Resulting from a spectral analysis, it was observed that energy was higher at low frequencies (< 63 Hz) due to Heating, ventilation, and air conditioning (HVAC) systems. At high frequencies (over 2.000 Hz) large contributions to noise energy came from alarms, medical equipment, and high velocity air flow from the HVAC system.

7.2 Operating Rooms

Ginsberg et al. (2013) measured sound levels during 23 cardiac surgeries with the aim of characterizing each phase of a surgery, including anesthesia induction and transport. The sound level meter recorded 2-minute periods and was placed flat over the anaesthesia machine in the operating room. Results have shown that all evaluated moments were louder than the baseline evaluation at room set up, which averaged 71.3 dB(A). Data points of all surgery moments (induction, incision, emergence, and transport) were close to or above 77 dB(A).

With the goal of characterizing how different surgeries could affect the hearing of patients and medical staff, Fritsch and Chacko (2010) evaluated the sound pressure level of different surgical instruments using a sound level meter with 200 ms integration time. It was placed 3 cm and 61 cm from the measured devices, which included drills, lasers, saws, and suction devices, to simulate the distance from the patient and the surgeon's ears, respectively. The loudest noise generators were the saw cuts commonly used in Orthopedy, peaking at 100 dB(A).

In a Greek public hospital with over 400 beds, sound measurements were acquired in 42 surgeries, all taking place in pre-1950 buildings (Tsiou et al., 2008). The sound level meter was set in 'slow' mode. Noise levels ranged between 46.7 and 106 dB(A). Similarly to Ginsberg et al. (2013), the highest levels of noise were observed in the pre-surgical or induction time. Again, noise levels differed significantly between orthopaedic and other kinds of surgeries, the main sources of noise in the former being surgical and suction tools. In non-orthopaedic surgeries the main sources of noise were attributed to anaesthesiologic and other machinery, people, and external noise (Tsiou et al., 2008) – probably associated with poor building infrastructure.

Finally, Kracht et al. (2007) have characterized several different surgeries, describing the soundscape of the operating rooms (ORs) of Johns Hopkins Hospital, including sound levels in octave bands for each OR and typical Leq and peak levels for each. Overall, surgeries spanned from 53 to 70.5 dB(A) Leq. Orthopaedic surgery had the highest level at 66.5 dB(A) Leq. According to the authors, from the measured levels, no standard auditory signature[10] could be found for any given type of surgery.

The issue of acoustic signatures, the ability to identify a surgery by its acoustic behaviour throughout time, is an experimental question that needs further research. While previous literature determines a lack of acoustic signature per surgery (Kracht et al., 2007), it should be made clear that these conclusions were taken after high-level analysis using sound pressure level or frequency analysis considering the surgery as a whole. More thorough analysis could have different results if researchers were given easier access to surgical logs.

With due awareness of the methodological differences and limitations of the mentioned studies, Table 7-2 summarizes their findings, aiming at characterizing the sound levels of

[10] The ability to identify a surgery by its acoustic behaviour throughout time. With everyday sounds, listeners learn recurring features or templates of complex sounds and associate them with specific sound sources, such as a piano, or a bird singing (Agus et al., 2010)

different surgical specialities. All values are presented in Leq dB(A) and include studies from Switzerland (Keller et al., 2016), Brazil (Lopes, 2005), Greece (Tsiou et al., 2008), and USA (Ginsberg et al., 2013; Kracht et al., 2007; Kuzmich et al., 2001).

Table 7-2. Sound levels according to surgical specialty (all Leq levels in dB(A)). Results from different studies.

	SWITZERLAND (Keller et al., 2016)	BRAZIL (Lopes, 2005)	GREECE (Tsiou et al., 2008)	USA (Ginsberg et al., 2013)	USA (Kracht et al., 2007)	USA (Kuzmich et al., 2001)
Elective open abdominal surgeries	59.54		63.86[11]		63	
General Surgery		78.3	63.86			
Urology		76.0			63,5	
Paediatric Urology					64	
Vascular		76.7				
Orthopaedics		72.0	68.31		66.5	65.8
Paediatric Orthopaedic					58	
Otorhinolaryngology		72.7			65	
Gynaecology		78.1				
Ophthalmology		70.2	63.86			
Cardiac		71.8	63.86	~77[12]	63.5	
Neurosurgery		73.6			64.5	
Thoracic					63	
Paediatric Plastic					65	
Plastic					67	

Current WHO noise levels recommendations for healthcare environments do not take into account their specific occupation (World Health Organization, 1999). Their main concern is background noise levels. Other significant metrics, such as spectral distribution or room acoustic properties, are not considered (Hsu, 2012). For the specific case of auditory alarm design, characterizing the spectral distribution of background noise is relevant. For maximum audibility, the frequency content of alarms should have minimal overlap with that of the typical background noise in their intended environments (AAMI, 2013).

Busch-Vishniac (2019) has recently reflected upon current research and metrics. Although Leq is the most common metric used in noise measurement studies inside hospitals, the author mentions that it is not the one which best represents its intricacies, failing to capture the relative 'peakiness' of noise. A new measure has been proposed, called the

11 Analysis performed comparing Non-Orthopaedic and Orthopaedic surgeries

12 Different moments of 23 cardiac operations

occurrence rate, OR(N), which represents the fraction of time that a level exceeds N decibels. Usually, these occurrences are local peaks or maximum sound levels. This measure correlates better with perceptions of hospital noise (Busch-Vishniac, 2019).

7.3 Aim

The aim of this study was to characterize the soundscape of the operating rooms of two recently built hospitals in the north of Portugal, and by that, enlarging the empirical dataset currently available. Considering the referred limitations of the measurements usually collected in similar studies, this study also intended to specifically consider the spectral content of different surgeries, to use this data to inform design guidelines for clinical auditory alarms. Results include sound levels of surgeries in octave bands, Leq and peak levels for each surgery speciality, as well as the fraction of time distribution corresponding to a given range of Leq intervals (helpful to calculate OR(N)). This will allow a comparison of sound levels and spectral content of each speciality, which should impact the design of clinical auditory alarms.

7.4 Method

Sound levels were measured in ten operating rooms of two Portuguese hospitals: Hospital A and Hospital B. Hospital A is a private hospital with five operating rooms inaugurated in December 2015. Hospital B is a public hospital, has thirteen operating rooms and was inaugurated in May 2011. Both have operating wards with modern construction, equipment, and technology.

The study comprised a total of 28 measurement sessions in different specialities: Vascular (5), Urology (11), Orthopaedics (6), Otorhinolaryngology (2), Neurosurgery (2), and Gynaecology (1). It also included one measurement session inside a recovery room. Measurements occurred in various operating rooms: 13 surgeries were measured at Hospital A and 15 were measured in Hospital B. Both studies were approved by the Ethics Committee associated with each Hospital. No personal information from patients was ever collected.

7.4.1 Frequency Analysis

The analysis of frequency components can provide important information for the characterization of a sound or noise. Most noises are not pure, but complex sounds. To have an exact understanding of these complex sounds it is necessary to determine the

sound pressure level of each frequency. This type of analysis is called spectral analysis or frequency analysis and is commonly represented in a graph with the frequencies in the x-axis and the sound pressure levels in the y-axis.

The energy of a sound is distributed in a spectrum of audible frequencies, and humans can hear between 20 to 20.000 Hz (Goldstein, 2007). The frequency unit is the Hertz (Hz) and it measures the cycles per second of the sound waves. The audible range is divided in 10 groups of frequencies which are called octaves. Each octave is subdivided into three groups of third octaves. The name of each octave corresponds to its central frequency, which is twice the frequency of the antecedent octave and the geometric mean of the limit frequencies (Miguel, 2000).

The frequency of a sound is the property perceived as pitch. A low frequency sound oscillates at a relatively slow rate, and a high frequency sound oscillates at a relatively high rate. For comparative purposes, the lowest note of a piano is approximately 32 Hz and middle C is 261Hz (Goldstein, 2007).

Frequency analysis typically examines 11 octaves starting in 16Hz (low frequencies) up to 16.000Hz (high frequencies). Since the human ear does not perceive all frequencies at the same level, sounds with different frequencies go through an adjustment using a filter. A-weighted filter is usually applied to compensate the response to frequencies of the human ear and is represented in dBA or dB(A).

7.4.2 Equivalent Continuous Sound Pressure Level (Leq)

Sound can be measured, modelled, and presented in several formats. The most common metric is the equivalent continuous sound pressure level (Leq). Leq is defined as the sound pressure level that, if constant during the measurement period, would contain the same sound energy as a real sound which is going through fluctuations during measurement (Everest & Pohlmann, 2015). Usually, equivalent continuous sound pressure level is presented in LAeq format. Just like with dBA, a filter is applied whose shape is like the human ear's response which listens better to central frequencies than high or low frequencies.

Another metric used to describe noise is LAF. The letter F indicates the integration time selected for the recording. F for Fast corresponds to the fast configuration with an integration time of 125ms. This configuration is recommended for signal recordings which are not impulsive, or environmental sounds with big variations in time. It is the

configuration usually selected for noise measurement. LAFmax and LAFmin correspond respectively to the highest and lowest values registered in the recording period.

7.4.3 Measuring Equipment

Measurements were performed using a set of Brüel & Kjaer instrumentation, a hand-held analyser type 2270, microphone capsule type 4189 and preamplifier type 2669 (Figure 7-2). Calibration was performed before the first (2019) and second (2020) recording sessions. The analyser serves as sound level meter and real-time frequency analyser. The template "Frequency Analysis" was used making third-octave band measurements with the time weighting parameter set to "Fast" (125 ms). The sound field correction setting was set to "free field". In all measurements, the sound level meter (SLM) was placed on a tripod to avoid vibrations and undesired movements. The SLM was placed away from the operating table to avoid occluding accesses or passages (room doors or material cabinets).

Figure 7-2. Brüel & Kjaer's Handheld Analyzer Type 2270 with a type 4189 microphone. Equipment used for the Frequency Analysis during surgeries.

In all measurement sessions, the surgical teams were aware of the study and instructed to keep conversation or play music during surgery as they normally would. No researcher was present inside the operating rooms during surgeries. The SLM was placed inside the operating room and set to measure as soon as the patient entered the room. When the surgery ended and the team started to clean the room, a nurse would inform the researcher who would enter the room to stop the measurement. For analysis, the first and last 10 minutes were removed to avoid preparation and cleaning periods. To preserve privacy, the Ethics committees did not allow collection of the surgery logs.

7.4.4 Measured rooms

In Hospital A, the operating ward responds to several surgical specialties, namely General Surgery, Paediatrics, Plastic surgery, Vascular surgery, Gynaecology, Orthopaedics, Otorhinolaryngology, Urology and Ophthalmology. The latter has two rooms exclusively dedicated, and those were not included in the current measurements. The other three rooms are next to each other, followed by the recovery room on the right of the main entrance. Rooms 1 and 2 have the same dimensions, and Room 3 is the largest room which is usually used for Orthopaedic surgeries which require bigger and heavier equipment. The red markers in Figure 7-3 indicate where the SLM was placed to perform the frequency analysis, after due sterilization before entering the operating room.

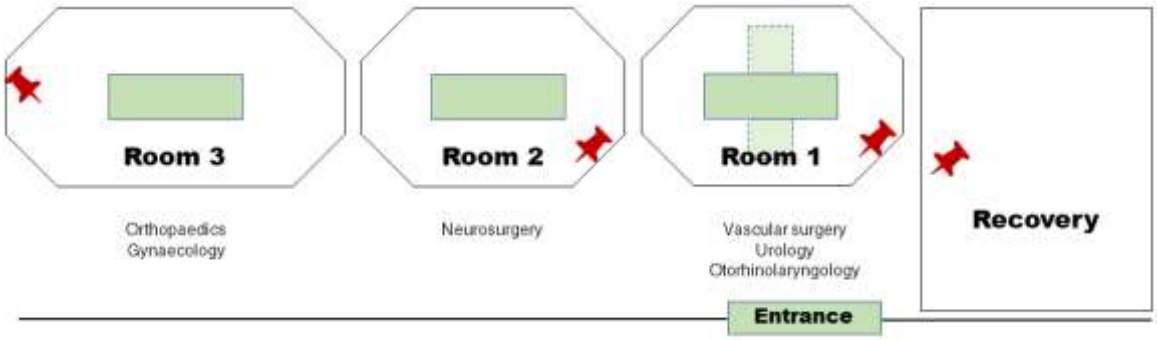


Figure 7-3. The three operating rooms and recovery room in Hospital A where measurements were performed. The red marker indicates the location of the sound level meter.

Figure 7-4 and Figure 7-5 depict the operating rooms before or after surgeries. They show the position of the sound level meter, which should not occlude passages.

Figure 7-4. Measurement in Vascular and Urology Surgeries in Hospital A

Figure 7-5. Measurement position inside the Recovery Room with one patient (not pictured) in Hospital A

Figure 7-6 represents a panoramic view of an operating room, to demonstrate the disposition of the equipment in all operating rooms. The central position, with a white mark on the floor, is reserved for the operation table. All equipment is portable, and next to the operation table is the equipment to monitor the patient (anaesthetic gases, intravenous fluids) and the instrumentation table. Usually, the intervenients of a surgery are the surgeon and assistant surgeon, anaesthetist, circulating nurse, instrumentalist nurse, anaesthetist nurse and technical health assistant, and a surgery can have up to seven people circulating on that space.

Figure 7-6. Panoramic view of an operating room in Hospital A

Considering the spaces' surfaces (walls, doors, ceilings, floor) affect the propagation and movement of the sound waves, Figure 7-7 demonstrates the surfaces of the recorded operating rooms, including the ceiling with the ventilation systems, the walls and floor, all of washable materials.

Figure 7-7. Surfaces of the operating room of Hospital A

In Hospital B, measurements were taken in rooms 1, 2, 5, 6, 7 and 8 (Figure 7-8). Rooms 1 to 5 have the same dimensions, and are smaller than rooms 6 to 9, usually used for Neurosurgeries and Orthopaedic surgeries for requiring larger equipment. The red markers in Figure 7-8 indicate the location of the SLM during measurements.

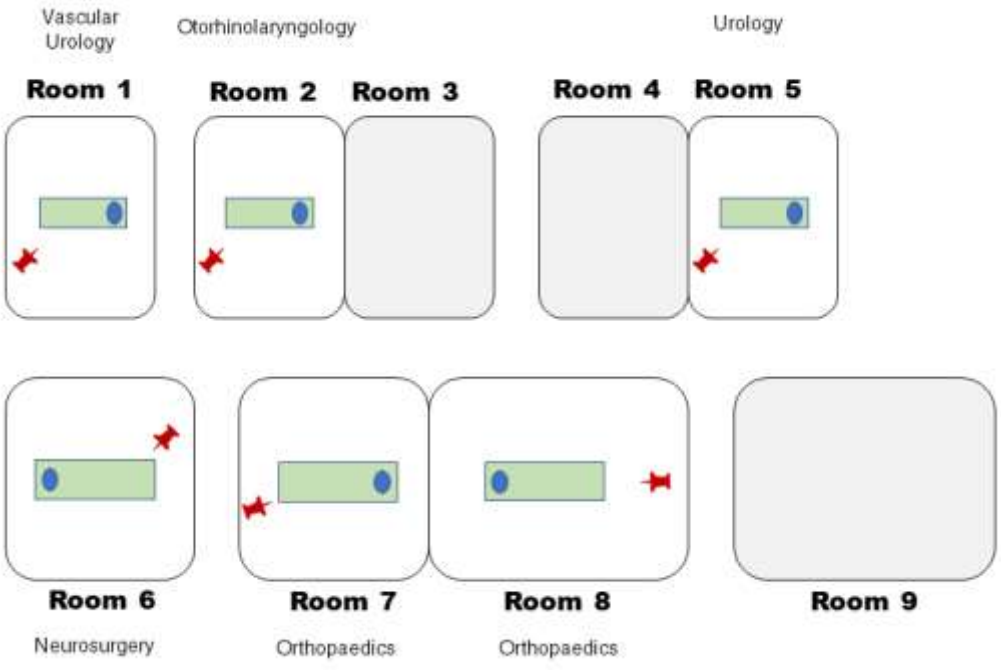


Figure 7-8. Outline of the surgical rooms where measurements were taken in Hospital B. The red marker signals the location of the SLM during measurements. Each room indicates which surgery was measured inside.

The following figures (Figure 7-9 to Figure 7-11) present some images taken from the operating theatres before or after surgeries. In all, it is possible to see the location of the SLM, which should be far from passage areas (way in, out or access to materials).

Figure 7-9. Measurement position in Hospital B's Room 1, where the Vascular and Urology surgeries took place. In rooms 2 and 5 the positioning was similar.

Figure 7-10. Measurement positioning during the Neurosurgery in Hospital B

Figure 7-11. Measurement position during Orthopaedic surgery in Hospital B

Figure 7-12 demonstrates an empty operating room.

Figure 7-12. View from an empty operating room inside Hospital B

7.5 Results

A-weighted broadband SPL measurements were obtained, as well as frequency analysis in third-octave bands from 25 Hz to 16.000 Hz. Maximum values were obtained in dB(A). Rooms were analysed using 10-minute intervals for the Leq. Leq values for each surgery were calculated by taking the logarithmic average of the multiple 10-minute measurements, totalling the duration of surgery procedures.

All results are presented in dB(A). For reference, the baseline Leq broadband level of an operating room was 54 dB(A) – empty room. Figure 7-13 compares the aggregate results for each hospital. All surgeries' measurements were averaged to give a single spectral characterisation per hospital.

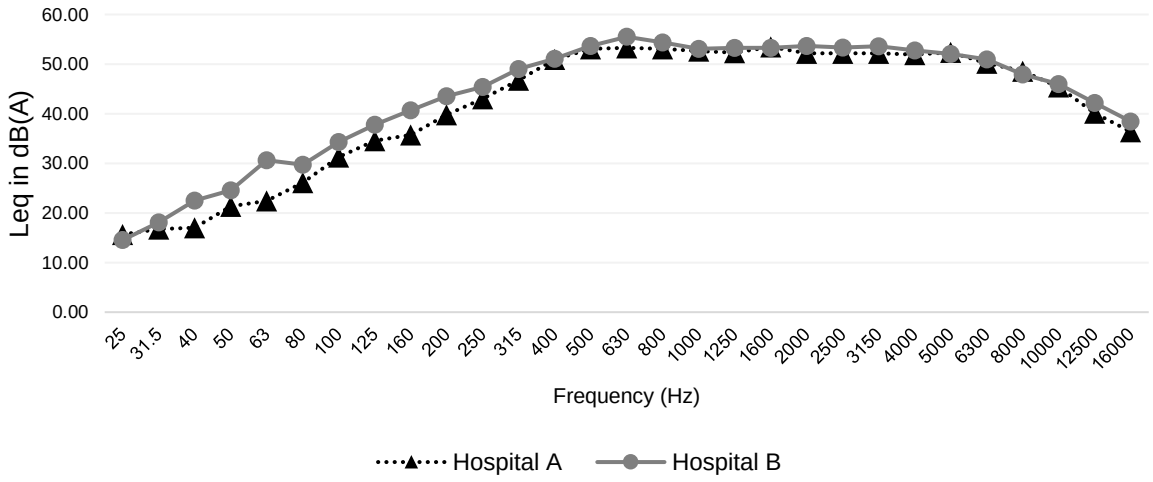


Figure 7-13. Average of all surgeries' level measurements in third octave bands. Results for each hospital.

The shape of the two spectral lines is very similar. Hospital A has slightly lower values than Hospital B, but the difference is statistically non-significant ($t(56) = 2.00$, $p>0.05$).

Table 7-3 summarises the results for surgical noise levels sorted by surgical speciality. The broadband Leq was obtained by first calculating the broadband level for each surgery, and then averaging all the results for each surgical specialty regardless of duration[13].

Table 7-3. Average broadband Leq dB(A) for surgeries according to surgical speciality

SURGERY	No. of Surgeries	Hospital A	Hospital B	Duration	Leq dB(A)
Vascular	5	2	3	00h33m – 3h30m	65,12
Urology	11	6	5	00h27m – 2h10m	63,92
Orthopaedics	6	2	4	00h50m – 1h47	66,51
Gynaecology	1	1	-	01h12m	64,05
Neurosurgery	2	1	1	2h28m – 5h54m	65,60
Otorhinolaryngology	2	-	2	1h40m – 2h40m	65,66
Recovery Room	1	1	-	1h37m	55,39

Figure 7-14 provides a breakdown of the data per surgical speciality regardless of hospital. The curves have similar shapes in every case, and there is no evident signature that allows distinguishing one surgery from the other regarding noise levels, except for the recovery room, which has Leq levels systematically lower than all surgeries. All surgeries have the

[13] An example of a dB calculator can be found in https://www.cesva.com/en/support/db-calculator/

lowest noise levels in the low frequency range. Sound amplitude increases with frequency, and the dominant range is between 315 Hz and 6.300 Hz.

The 55 dB(A) reference value (in red) was chosen based on the conversation level range (55-65 dBA). Sound amplitude typically decreases from around 2.000 Hz, except for the Orthopaedics speciality which had the uppermost levels at high frequencies. Gynaecology was the least energetic in the high end of the spectrum. From around 5.000 Hz on, all specialities revealed a continued decrease in noise levels.

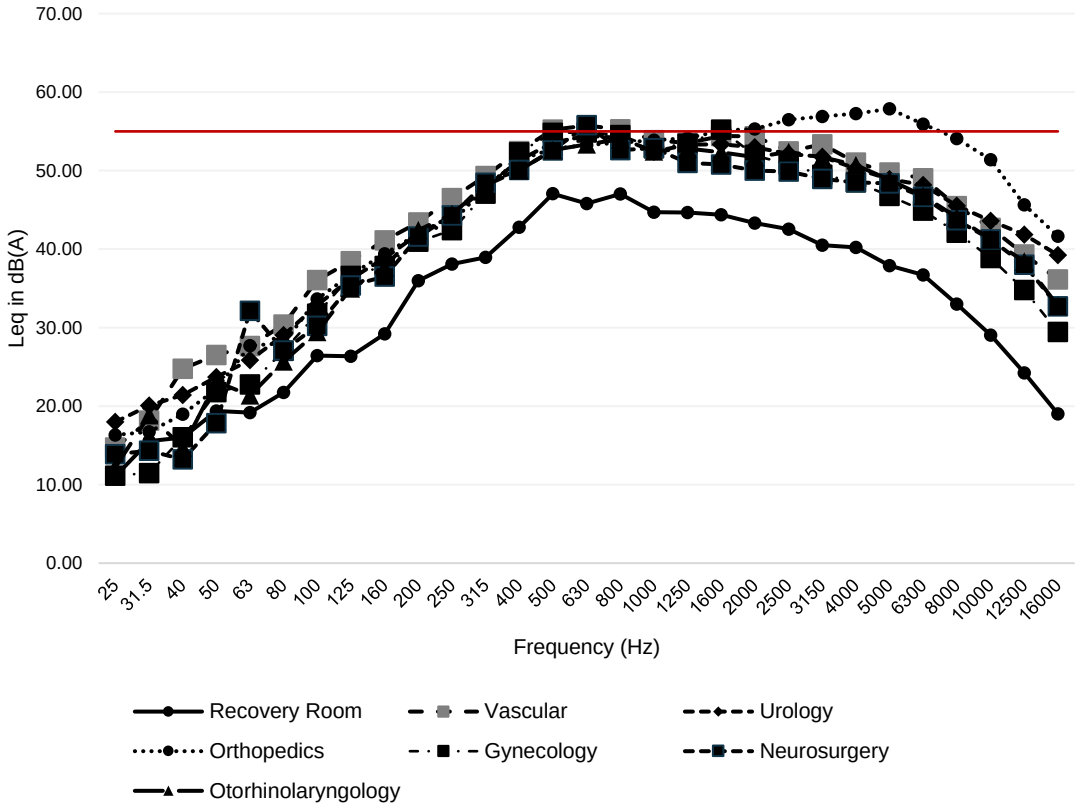


Figure 7-14. Average levels in third-octave bands for each type of surgery in both hospitals. The red line marks the 55 dB(A) conversation level (larger [14]).

At the lower end of the spectrum (25-250Hz), Vascular and Neurosurgery have the highest levels. Otorhinolaryngology was the least energetic in the low frequencies.

Another relevant acoustic metric is Lmax, which indicates the 'peakiness' of the sounds during surgeries (Table 7-4). This measure indicates noise consistency during time. As an

[14] https://figshare.com/articles/figure/Average_levels_in_third_octave_bands_and_Average_fraction_of_time/14925774?file=28741089

example, if Lmax is fairly higher than Leq, it may be inferred that it results from a brief time interval - a "peak".

Table 7-4. Average Lmax dB(A) for surgeries according to surgical speciality

	NO. OF SURGERIES	LMAX DB(A)
Vascular	5	103,9
Urology	11	98,3
Orthopaedics	6	92,8
Gynaecology	1	91,3
Neurosurgery	2	93,1
Otorhinolaryngology	2	100,6
Recovery Room	1	87,3

Figure 7-15 reveals the highest sounds observed in each surgery, all above 85 dB(A). The reference value of 85 dB(A) comes from IEC 61672-2:2017 international standard (IEC 61672-2:2013+AMD1, 2017) which recommends a limit of 8h exposure to such levels of noise. Hospital B had more variance in terms of sound peaks than Hospital A. The surgery with the highest Lpeak was Vascular, followed by Otorhinolaryngology, Urology and Orthopaedics.

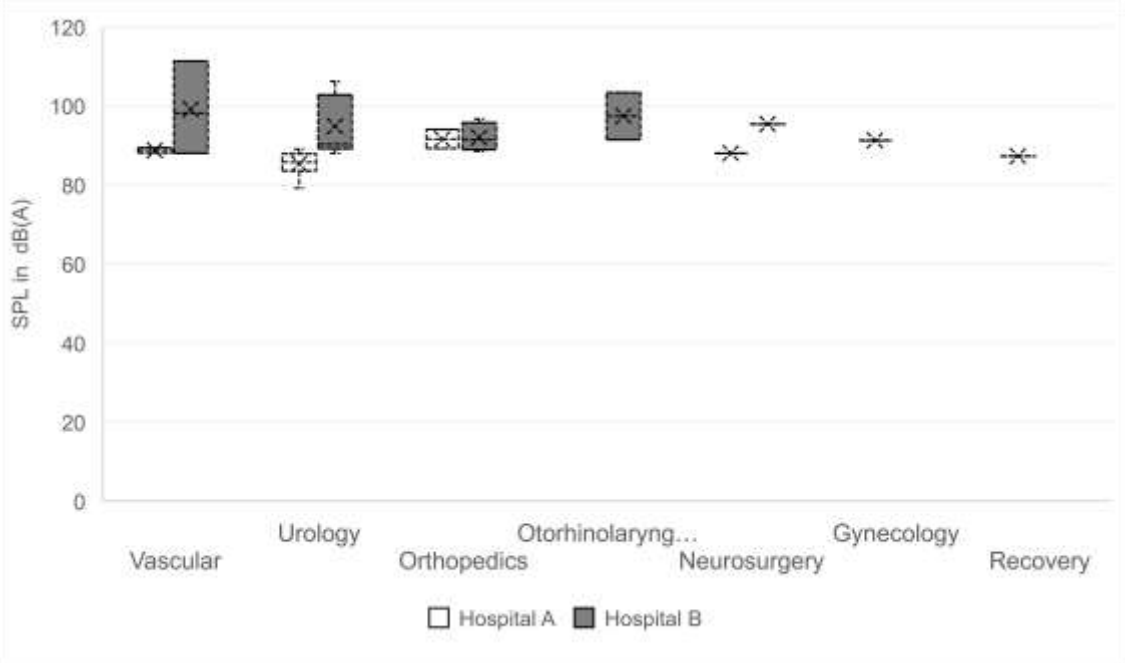


Figure 7-15. Lmax dB(A) distribution for each surgery

Figure 7-16 reveals the occurrence rate, that is, the time distribution of energy levels in the total duration of the surgeries in intervals of dB(A). The selected intervals were: Leq levels under 50 dB(A), between 51-70 dB(A), between 71-85 dB(A), 86-100 dB(A) and over 101

dB(A). On average, surgeries take up to 17,4% of duration time under 50 dB(A), most of the time (77%) in the 50-70 dB(A) interval, and 3,2% of the time between 71 and 85 dB(A).

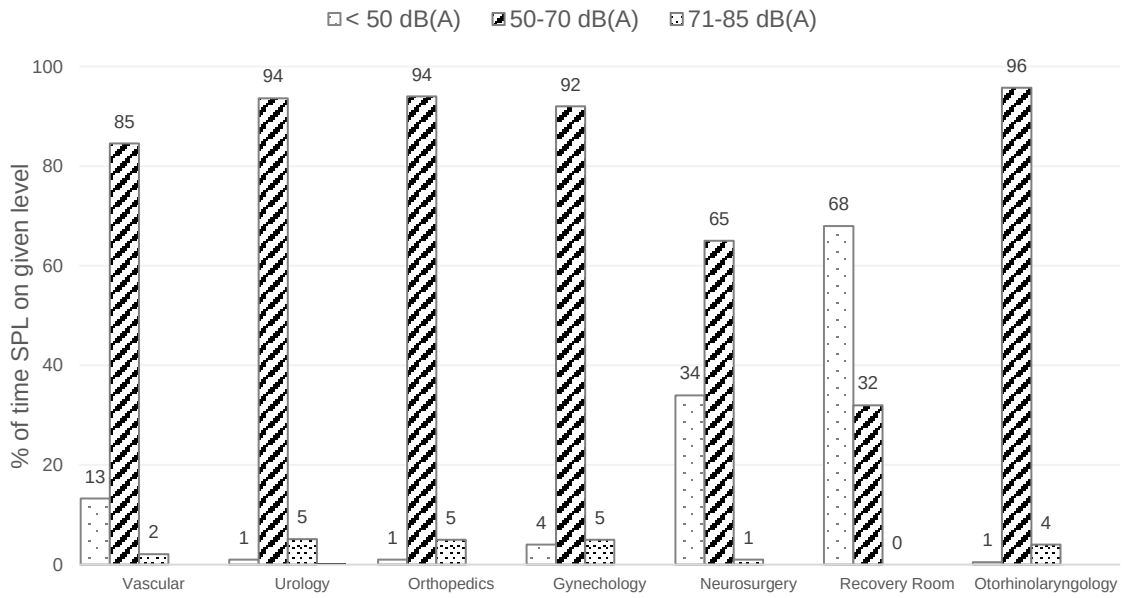


Figure 7-16. Average fraction of time, or occurrence rate (%), in given interval of SPL for each type of surgery (larger [15]).

Most surgeries ranged 65% to 95,8% of their time between the 50-70 dB(A) interval. The highest occurrence rate OR(70 dB(A)) was observed in Urology, Orthopaedics ad Gynaecology surgeries at 5% of the surgery time. Table 7-3 includes the durations of the analysed surgeries.

7.6 Clinical Alarm Design Implications

For alarm design, these results provide important indicators, suggesting intervals outside of which clinical auditory alarms should have their spectral components to assure their audibility (Bennett et al., 2019). To avoid masking and improve localizability of alarms (Edworthy et al., 2017), it is advised they have more than four harmonic components (AAMI, 2013; Patterson, 1982). Figure 7-17 and Figure 7-18 illustrate this requirement with real examples of alarms over surgical background noise. The example alarms in red are IEC 60601-1-8's 2020 amendment auditory icons for Cardiac and Oxygen-related problems

15 https://figshare.com/articles/figure/Average_levels_in_third_octave_bands_and_Average_fraction_of_time/14925774?file=28741089

(IEC, 2020). The background frequency analysis of surgical noise comes from a measured Orthopaedics surgery, and it was selected for being the surgery in the present study with the highest SPL levels on all frequency octave bands.

Figure 7-17 represents two overlaid octave band spectra. The red layer corresponds to the IEC 60601-1-8 High Priority Cardiac Alarm at an Leq 80 dB(A). The blue spectra correspond to the background noise of an Orthopaedic surgery (Leq 68,9 dB(A)).

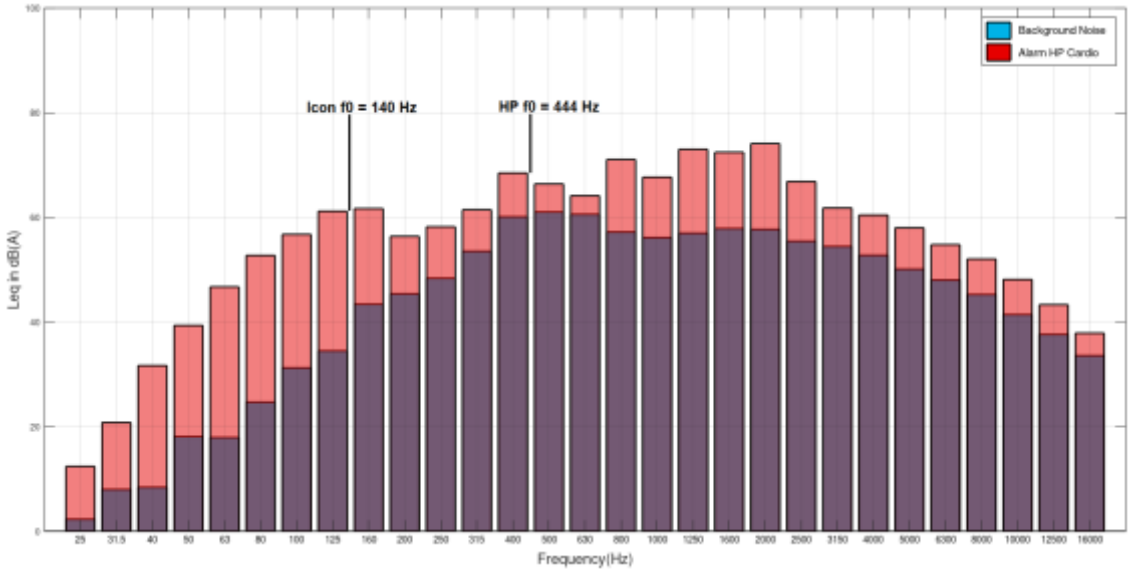


Figure 7-17. Third-octave band spectra of the High Priority Cardiac Alarm (red) over the background noise of an Orthopaedic surgery (blue). Overlap of frequencies is represented by the darkest tone.

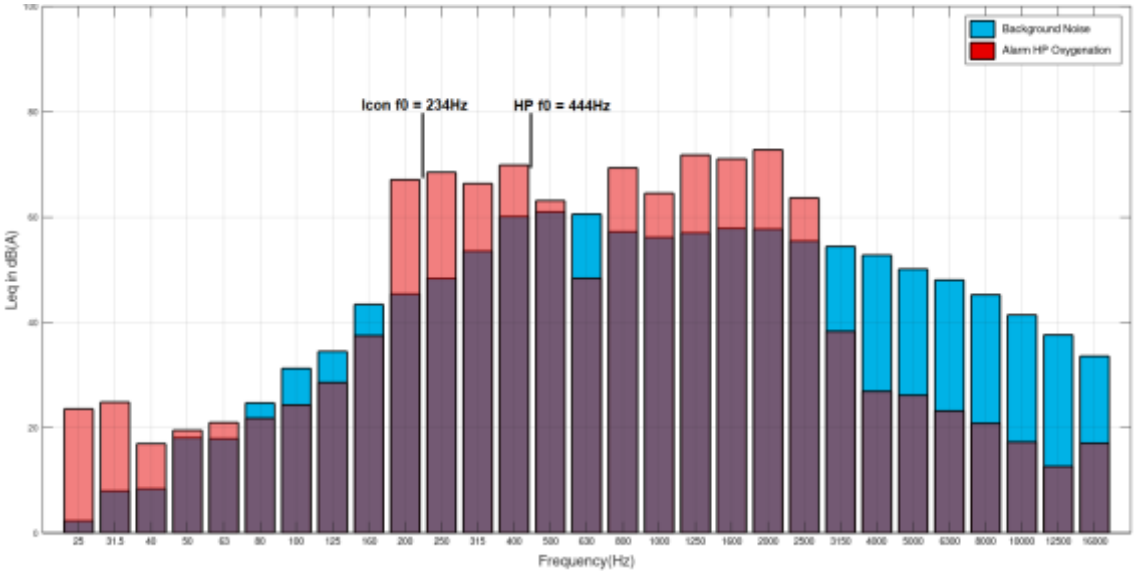


Figure 7-18. Third octave band spectra of the High Priority Oxygenation Alarm (red) over the background noise of an Orthopaedic surgery (blue).

From this exercise it is possible to understand that the alarm is well designed as its fundamental frequencies are not masked by the background noise. The example in Figure 7-18 uses the background noise of the same Orthopaedic surgery (blue) with the High Priority Oxygenation Alarm overlaid (red). The lower and medium frequencies are not masked, but there is a higher risk of masking in the higher frequencies.

Even though both alarms could have their highest frequency components masked during this Orthopaedics surgery, both would still be audible because they have several unmasked frequency components or harmonics. This makes the alarms more resistant to masking effects as there are always several components with energy above the background noise level.

Attending to our main aim with this study, which was of defining design guidelines for clinical auditory alarms, we came to understand that alarms should:

- Have several frequency components or harmonics;
- Have lower sound levels in lower frequencies;
- Have higher sound levels in medium frequencies;
- Grossly range between 50-85 dBA. Use the highest recorded value of approximately Leq 70 dBA (Orthopaedics) as threshold.

7.7 Discussion

This study aimed at going a step further in hospital noise analysis, usually focused on sound pressure levels, by analysing the distribution of surgical noises' spectral components, and also the occurrence rate of sound pressure levels considering the total duration of surgery. Its applied goal was to establish design requirements for clinical auditory alarms.

The sound pressure level and frequency analysis made on the course of 28 surgeries in two different hospitals indicate that, regarding average broadband equivalent sound levels, the highest level was observed for the Orthopaedics surgeries with 66,5 dB(A), following a similar pattern as found in studies of Greek (Tsiou et al., 2008) and U.S. hospitals' operating rooms (Kracht et al., 2007). The lowest level was observed in the Urology surgeries with 63,9 dB(A). The Recovery Room collected systematically lower Leq levels than any surgery, at an average of 55,4 dB(A).

The spectral third-octave bands average levels (A-weighted) have a similar shape across surgeries, with lower sound amplitude levels at low frequencies and the highest sound amplitudes occurring between 315 Hz and 8.000 Hz. The highest frequencies (10.000 Hz-

16.000 Hz) ranged between 35-45 dB(A). At the low end of the spectrum (25-250 Hz), Vascular and Neurosurgery have the highest levels (15-50 dB(A)). Otorhinolaryngology was the least energetic in the low frequencies (15-46 dB(A)). Orthopaedics had the most energy at high frequencies between 10.000Hz-16.000Hz (45-55 dB(A)) and Gynaecology was the least energetic in the high frequencies (30-40 dB(A)).

Comparing the observed levels with international guidelines, namely the standard IEC 61672-2:2013-2017, the values of the current study do not breach any of the established recommendations. According to this directive, the 8h-lower exposure values for an individual should not exceed 85 dB(A). This suggests that the observed sound levels are not high to the point of causing hearing loss. 77% of the time, most surgeries' levels were in the 50-70 dB(A) interval and, on average, only 3,2% of the time was spent with noise levels between 71-85 dB(A).

It could be stated that, similar to previous research (Kracht et al., 2007), there was no evident acoustic signature to any surgical speciality. A similar shape for the third-octave band levels of surgeries in two different hospitals serves as an indicator that the same design guidelines for clinical auditory alarms could be applied, independently of the surgery speciality. However, these results should consider the fact that surgeries were analysed as a whole, and not associated to surgical logs. Dividing recordings into surgeries' main moments (turnover, induction, procedure, closing, emergence) – as explained in Mhlaba et al., (2016), could point towards acoustical differences between them.

This proves to be one of the major limitations of this study. No access to logging files and no authorization for the presence of a researcher during surgical procedures by the Ethics Committees are reasons for which it is not possible to associate any peak levels with a specific event, or to understand if surgeries differ in their stages of higher noise or even spectral characteristics. For instance, it could be found that Vascular surgeries are noisier in the preparation stage while Orthopaedics are noisier at the procedure stage. It is also not possible to characterize which equipment contributed for higher LAeq in the low frequencies during Vascular surgeries, and in the high frequencies during Orthopaedics surgeries. We might speculate this includes suction instrumentation and equipment in the first case and some surgical hammering in the latter.

Another limitation includes the reduced number of some types of surgeries. The surgeries included are the ones we had access to on the course of one week in each Hospital. Regarding the number of surgeries per speciality, we can be more confident of the averaged results of 11 Urology surgeries than of 2 Neurosurgeries.

This study gave a good, applied contribution considering clinical alarm design guidelines. Looking at international standard IEC 60601-1-8 standard, it does not provide a design guideline regarding sound pressure level for detectability, due to the lack of constancy of the background noise in operating rooms. However, it suggests that "*values between 45 dB and 85 dB can be reliably detected without being too intrusive in most situations*". Patterson (1982) proposed a power spectrum model method (for civil aircraft) which suggested alarms levels to be +15dB above threshold - that is, a level in which 75% of detection answers are correct – and no more than +25dB to avoid issues of annoyance and interference with team communication. IEC 60601-1-8 suggests implementing the Hughson-Westlake 2-up-1-down method[16] as a detectability test.

This analysis allowed to understand different surgeries' average sound pressure levels, and their spectral components' distribution.

7.8 Conclusion

Sound levels were measured in the operating rooms of two Portuguese hospitals with the goal of characterizing the background noise of different surgeries regarding sound levels and spectral distribution analysis. A total of 28 measurements were acquired in surgeries comprising Vascular, Urology, Orthopaedics, Otorhinolaryngology, Neurosurgery, and Gynaecology. Results have shown all surgeries followed the same pattern of having low LAeq levels at the lower end of the spectrum and higher levels between 315 Hz and 6300 Hz. Vascular surgery had the highest LAeq levels on the low frequencies and Orthopaedics had the highest LAeq levels in the high frequencies. These results are comparable to previous studies performed in Greek and USA's hospitals, indicating some similarity in what concerns at least the noisiest surgeries. Broadband Leq levels ranged between 55,4 dB(A) for the Recovery Room and 66,5 dB(A) for the Orthopaedics surgeries. Maximum levels were also measured, and all observed values were under 100 dB(A).

[16] Hughson Westlake is an automatic pure tone test procedure. In this test method, the threshold of hearing is defined as 2 out of 3 (or 3 out of 5) correct responses during the ascending portion of the tone presentation. If the patient responds when the tone is ascending, the test will automatically decrease the level by 10 dB. The patient must respond to the same intensity 2 out of 3 or 3 out of 5 times for the threshold to be recorded.

Two case studies were presented in which two validated IEC 60601-1-8 standard alarms were overlayed to an Orthopaedic surgery background noise. One of the alarms had its spectral components consistently above the background noise, while the other had most of its components masked by the background noise. This exercise evidenced the need to have a spectrally rich alarm to avoid masking and improve detectability. Finally, clinical alarm design guidelines were systematized following the obtained results.

The application of this research consists of design guidelines that will help to develop clinical alarms which are audible, minimally masked by common background noise, and without the risk of interference with communication or causing annoyance to those who are exposed to them.

STEP 2. SPECIFY USER REQUIREMENTS

8 Design Guidelines for Clinical Auditory Alarms

The updated global medical safety standard IEC 60601-1-8 includes major changes in the auditory warning signals design guidelines. One of the most important is the fact that it allows stakeholders to design different versions of the same sound. It is thus relevant to systematize the thorough and robust body of knowledge developed in recent years regarding the design of clinical auditory alarms. All content of this chapter is based on standards and/or experimental studies on detection, learning, subjective perception, or spatial localization. Its aim is to provide a practical tool for designers to design better clinical auditory alarms.

This chapter is based on the paper: Vieira, J., Santos, J. A., & Noriega, P. (2019, July). A Review of Design Guidelines for Clinical Auditory Alarms. In International Conference on Healthcare Ergonomics and Patient Safety (pp. 325-333). Springer, Cham.

8.1 Introduction

The global medical device standard IEC 6060-1-8 and corresponding amendment was developed by IEC (International Electrotechnical Commission). Parts 1–8 of the medical standard include general requirements for basic safety and essential performance, and the alarms used with that equipment (Edworthy et al., 2013). The alarms proposed in the standard went under major updates and an amendment was released in 2020. It allows stakeholders to design their own alarm signals. This chapter intends to review design guidelines published in recent decades having as specific application the healthcare environment. The guidelines and recommendations were collected from the literature, spanning from standards to experimental studies. They are organized in three major groups: Spectral, Temporal, and Spatial aspects of clinical auditory alarms.

8.2 Spectral aspects of clinical auditory alarms

The spectral content of a sound refers to the overall frequency of the signal (ISO 7731 - Ergonomics - Danger Signals for Public and Work Areas — Auditory Danger Signals, 2003). We will approach the parameters of frequency, harmonics, timbre, and sound pressure level.

8.2.1 Frequency

The frequency content of an alarm should correspond to the maximal human sensitivity, between 200 Hz and 4000 Hz, with the fundamental frequency (f_0) between 300 Hz and 1000 Hz, suggest Begault and Godfroy (2007). Although stating that spatial localization is poor at low frequencies, IEC 60601-1-8 proposes an even lower limit for fundamental frequency, starting at 150 Hz. The upper limit for fundamental frequency is 1000 Hz as hearing impairment due to exposure or age usually affects higher frequencies(ANSI/AAMI/ IEC 60601- 1-8:2006 & A1:2012, 2013), and also, signals with this frequency or higher tend to be considered aversive (Haas & Edworthy, 1996).

As for the frequency range of an alarm signal, it should be between 200 Hz and 4000 Hz, with the preferred range being 500-3000Hz. Nevertheless, if the audio alarm is required to be heard around obstacles, frequency should be below 500 Hz (AAMI, 2013). Lower frequencies can contour solid objects (Begault & Godfroy, 2007) and help reducing the aversiveness of the alarm (Edworthy et al., 2017). Importantly, the frequency band should differ from the background frequencies of the place where the auditory signals are going to be heard (AMI, 2013). Figure 8-1 summarizes f_0 and frequency range recommendations.

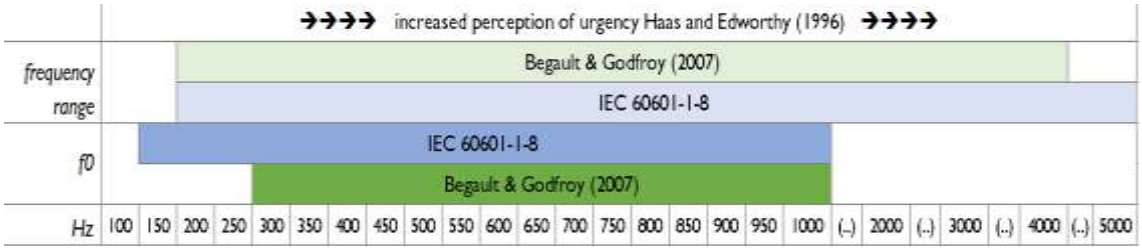


Figure 8-1. Summary of f_0 and frequency range recommendations

In auditory alarms, higher pitches are associated with increased urgency (AAMI, 2013; Haas & Edworthy, 1996; Patterson et al., 1986). Haas and Edworthy (1996) have compared auditory signals with a f_0 of 200 Hz, 500 Hz and 800 Hz and verified that the perception of urgency increased with the fundamental frequency (from 200Hz to 500Hz) at relatively high sound pressure levels above ambient, and the response time decreased with higher fundamental frequencies. A difference in pitch is perceived at 1/12 of a semitone (Weber (1834) as cited by Brown et al., 2015) in laboratorial conditions.

In healthcare environments, the oximeter is a representative example of the attempt of mapping pitch variations with real-life values. Pulse oximeters communicate the increase or decrease of arterial oxygen saturation (SpO2) through pitch variation, however, the direction of the pitch change is detectable only two thirds of the time by anaesthetists when the frequency changes are small (Brown et al., 2015). Pitch mapping, if uniform among equipment, is a powerful and informative manipulation of sound.

8.2.2 Harmonics

Regarding the number of harmonics, IEC 60601-1-8 suggested the presence of f_0 and at least three additional harmonics (AAMI, 2013). Other authors agree with the "four or more" harmonically related spectral components to the auditory alarm (Begault & Godfroy, 2007; Patterson & Mayfield, 1990). A warning signal with harmonics in the appropriate level range, spread across the spectrum is less likely to be masked (Patterson & Mayfield, 1990) and easier to localize (Edworthy et al., 2017).

As for the range of these harmonics, IEC60601-18-8 (IEC, 2020) proposes 150Hz to 4000 Hz at one meter of the intended operator's position, with some authors suggesting a range of 300 Hz to 4000 Hz (Edworthy et al., 2017) or 500 Hz to 4000 Hz (Patterson et al., 1986). These harmonics should not differ by more than 15 dB in amplitude (Semiconductors, 2009).

The harmonic content affects the timbre of the sounds which, if desired, allows a different tonal quality for each equipment. Sounds with an odd number of harmonics have a harsh quality, whereas even harmonics give a church organ type of sound (AAMI, 2013).

8.2.3 Timbre

Timbre has been used as a parameter to differentiate sounds between categories, with similar functions having sounds with the same timbre (Edworthy et al., 2011). Designing alarms with different timbre but maintaining its pitch and temporal pattern makes them confusing (Patterson, 1982). In fact, pitch and temporal patterns influence more the identity of an alert than its timbre (Edworthy et al., 2011).

8.2.4 Sound Pressure Level

ISO 60601-1-8 does not specify an absolute volume range, or range of levels for auditory alarm signals because the background noise levels may differ according to the environment. This standard suggests alarm designers to know the typical background noise level and respective variability when designing a sound. ISO 7731 provides a standardized method to calculate effective masked threshold (ISO 7731 - Ergonomics - Danger Signals for Public and Work Areas — Auditory Danger Signals, 2003). The sound level of the auditory alarms should be above background noise, nevertheless, achieving this "just right" level can be challenging when listeners are at varying distances of the loudspeakers (Begault & Godfroy, 2007). IEC 60601-1-8 states that, from clinical experience, values between 45 dB and 85 dB can be detected without being intrusive in-patient care environments. In operating rooms, noises range from 50 dBA to 85 dBA. Levels of 15-20 dBA are recommended as signal-to-noise ratios for alarms (Begault, 2012; Patterson & Mayfield, 1990) for cockpits, but these authors warn that often the frequency content of both the alarm and the noise are disregarded when considering these values. Two signals with the same LAeq can be perceived differently because of their spectral quality (Hsu, 2012).

Sound pressure levels are also related with the perception of urgency of a signal, with louder signals being perceived as more urgent (Haas & Edworthy, 1996). To communicate different levels of urgency, IEC 60601-1-8 proposed a difference of + 6 dB from medium priority to high priority – or a range going from an equal level (0 dB) to + 12 dB louder. Medium and low priority alarms should have the same level, but if they differ, medium alarms should not be more than 6 dB louder than low priority auditory alarm signals.

Other manipulations have been introduced in the loudness of a sound, such as the tremolo. This is produced by modulating the amplitude of a tone to produce a vibrating effect. It has been implemented to distinguish between oxygen saturation ranges in neonates (Paterson et al., 2016).

Interestingly, design guidelines are either written for unoccupied conditions or do not specify occupancy. The differences between occupied and unoccupied units average 13 dB LAeq (Ryherd et al., 2012). Metrics different from relative-loudness can influence the response to alarms such as potential for task interference and the quality of the background noise (Hsu, 2012).

8.3 Temporal aspects of clinical auditory alarms

The temporal aspects strongly affect the perception of urgency of an auditory alarm, and its learnability. Sharing similar temporal patterns is known to cause confusion between auditory alarms (Edworthy et al., 2013), regardless of differences in other structural features (Edworthy et al., 2011), even after considerable exposure to them (Edworthy et al., 2013). Discrimination between alarms works better when a unique temporal pattern is associated with each one (Begault & Godfroy, 2007).

In this section we will approach the parameters of inter-pulse interval, rhythm, and amplitude envelope.

8.3.1 Inter-pulse interval

ISO 7731 recommends the temporal distribution of the signal to be pulsating rather than continuous in time (ISO 7731, 2003). The pulse should have between 75 ms to 200 ms duration (Patterson, 1982). Considering this sort of alarms, Patterson et al. (1986) mention speed – determined by the inter-pulse interval, with faster bursts possessing shorter inter-pulse intervals - as the main variable for the perception of priority. IEC 60601- 1-8 proposes three different pulse duration patterns according to high, medium or low priority of the alarm, respectively 75 ms to 200 ms (high) and 125 ms to 250 ms (medium and low).

Several experimental manipulations managed to isolate this variable and correlate it with the perception of urgency, stating that a sound with a rapid inter pulse interval will be perceived as more urgent than a sound with a slower inter pulse interval (Haas & Edworthy, 1996; Mondor & Finley, 2003; Vieira et al., 2018).

8.3.2 Rhythm

IEC 60601- 1-8 suggests syncopated or "off-beat" rhythms for higher priority alarms and regular rhythms for medium and low priority alarms (AAMI, 2013). Vieira et al. (2018) obtained results in line with what was predicted by the standard, with syncopated rhythms being perceived as more urgent. Nevertheless, the inverse relation was found by Edworthy, Loxley, and Dennis, (1991), with syncopated rhythms being perceived as less urgent than regular ones.

8.3.3 Amplitude Envelope

Manipulating the beginning of an alarm is a common procedure to avoid startle reactions. The amplitude envelope refers to the shape of a sound over time, with the rise and fall time of an auditory warning. It is the interval over which the pulse increases from 10 % to 90 % of its maximum amplitude (AAMI, 2013), but shaping the envelope has little effect if the pulse in which it is applied is less than 100 ms (Patterson et al., 1986).

Amplitude envelope can be percussive – regularly heard, like in impact sounds – or flat – man-made, and not regularly heard. Amplitude envelope plays an important role in learning and memory (Schutz et al., 2017). Edworthy et al. (1991) have studied the perception of urgency as a consequence of the manipulation of the amplitude envelope. The authors found that a pulse with a standard envelope (20 ms onset or rise time, as suggested by Patterson (1982)) was perceived as more urgent that a slower onset, and a pulse with a slower onset was perceived as more urgent than a slower offset. For clinical auditory alarms, IEC 60601-1-8 (2020) recommends a rise time of pulses not to be so short as to create mechanical speaker noise. The fall time or offset can have any duration, if it does not overlap the next pulse (IEC,2020).

8.4 Spatial aspects of clinical alarms

Current technology allows sound to be designed considering other manipulations, such as the use of directional or spatial cues. 3D sound, also known as spatial audio, might enhance responsiveness to clinical auditory alarms. 3D sound takes advantage of the human's instinct to immediately look in the direction of loud, distinct sounds. Because of our orientation reflex, 3D sound has the potential to increase spatial awareness, reducing mental and visual workload in complex environments (McIntire et al., 2010). It also aids in the intelligibility of multiple sources (Begault, 2012), in target detection and identification, and navigation (Nelson et al., 1998; Nelson et al., 1999). In spatial audio systems, sound delivery can be binaural (based on headphones or earphones) or transaural

(based on loudspeakers). Although its implementation in healthcare settings is not as immediate as it is in other more entertaining applications, it will allow new uses of sound that will improve clinical soundscapes.

In this section we will approach the parameters of spatial modulation, number of sources and reverberation time.

8.4.1 Spatial modulation

Spatial modulation was proposed by Begault and Godfroy (2007) as a technique to increase the detectability of an alarm. It consists in laterally moving a tone at a rate of 2-10Hz. The spatial movement allows the sound to be more detectable than a stationary alarm. However, just like in the visual modality, acuity for moving targets decreases as the velocity increases, the same effect happens with auditory moving targets – auditory acuity gets worse as the velocity of the auditory image exceeds 10º of visual angle per second (McIntire et al., 2010).

8.4.2 Number of sources

An increase in the number of sources inside a room will result in an increase in response time to localize them, and a decrease in localization performance. Nelson et al. (1998) performed a study where several talkers were simultaneously presented to one listener on the horizontal pane. Response time increased between one and six talkers, and correct detection also decreased with the increase of the number of talkers. Similar results were found in studies using non-speech stimuli (Begault & Wenzel, 1993), and moving targets (McIntire et al., 2010).

8.4.3 Reverberation time

The reverberation time refers to the time interval required for the sound-pressure level to decrease by 60 dB, after the emission by the source is stopped. In the operating room, it should not exceed 0.6 second according to the Swedish standard 25268 (Quinn, 2017).

8.5 Conclusion

Having the adequate set of auditory alarms, means having alarms which will decrease clinical staff's workload, contribute to the patient's wellbeing, and comfort, and result in less errors because they will correctly map an event to an alarm. Achieving this is not trivial, as it requires a deep knowledge and understanding of the activities performed by people in the clinical setting (Edworthy, 2013), and validation in experimental studies. The

guidelines gathered in this paper come from literature, experimental studies, or standardizing bodies, and they intend to contribute to better auditory design for clinical settings by presenting an overview of spectral, temporal, and spatial parameters which affect the detection and learning of auditory alarms.

They still leave room for creativity and improvement. Abstract sounds are the hardest to learn (Edworthy et al., 2011). The alarms which are easier to learn are speech alarms and auditory icons or everyday sounds. However, there are practical issues like background noise masking speech alarms, workload or simple user acceptance of new alarms (Edworthy et al., 2011). In a several-phased alarm learning task, Edworthy et al., (2013) have found that when learning and recognizing a set of alarms is the primary task, people improve recognition with continued exposure, even if this exposure happens in increasing workload. In this study, a pair of commonly confused tones was one which shared temporal regularity and high pitch. However, even in unstressed situations, humans do not learn the significance of more than eight sounds (Meredith & Edworthy, 1995).

It is important to consider all factors when designing sounds for complex environments such as healthcare. Although the physical components are expressed in these guidelines (Table 8-1), user testing is important for further validation of association, learning and acceptance of the alarms.

Table 8-1. Summary of guidelines described in this paper

Spectral Aspects of Clinical Auditory Alarms	
Frequency	• The frequency content of an alarm should correspond to the maximal human sensitivity, between 200 Hz and 4000 Hz • The fundamental frequency (*f0*) between 300 Hz and 1000 Hz (Begault & Godfroy (2007)) • IEC 60601-1-8 (2012) proposes a *f0* starting at 150 Hz with the upper limit in 1000 Hz; • Signals with *f0 > 1000 Hz* tend to be considered aversive. • Preferred frequency range 500-3000Hz (IEC 60601-1-8,2012); • If alarm is required to be heard around obstacles, frequency should be below 500 Hz (IEC 60601-1-8, 2012); • Frequency band should differ from the background frequencies of the place where the auditory signals are going to be heard; • Higher pitches are associated with increased urgency
Harmonics	• IEC 60601-1-8 (2012) suggests $\boldsymbol{f_0}$ and at least three additional harmonics; • A warning signal with harmonics is less likely to be masked and is easier to localize; • IEC60601-1-8 (2012) proposes an harmonics range from 150Hz to 4000 Hz; • Some authors suggest a range of 300 Hz to 4000 Hz or 500 Hz to 4000 Hz ; • These harmonics should not differ by more than 15 dB in amplitude
Timbre	• Alarms with different timbre but same pitch and temporal pattern makes them confusing • Pitch and temporal patterns influence more the identity of an alert than its timbre.
SPL	• Values between 45 dB and 85 dB can be detected without being intrusive in patient care environments (IEC 60601-1-8); • Levels of 15-20 dBA are recommended as signal-to-noise ratios for alarms for cockpits; • Frequency content of both the alarm and the noise should be equally regarded; • Louder signals are perceived as more urgent • IEC 60601-1-8 (2012) proposes a difference of + 6 dB from medium priority to high priority – or a range going from an equal level (0 dB) to + 12 dB louder; • Medium and low priority alarms should have the same level, but if they differ, medium alarms should not be more than 6 dB louder than low priority auditory alarm signals (IEC 60601-1-8, 2012).

Temporal Aspects of Clinical Auditory Alarms	
Inter-pulse interval	• The temporal distribution of the signal to be pulsating rather than continuous in time (ISO 7731); • The pulse should have between 75 ms to 200 ms duration • Inter-pulse interval is the main variable for the perception of priority; • IEC 60601- 1-8 (2012) proposes three different pulse duration patterns according to high, medium or low priority of the alarm, respectively 75 ms to 200 ms (high) and 125 ms to 250 ms (medium and low).
Rhythm	• IEC 60601- 1-8 (2012) suggests syncopated or "off-beat" rhythms for higher priority alarms and regular rhythms for medium and low priority alarms
Amplitude Envelope	• Shaping the envelope has little effect if the pulse in which it is applied is less than 100 ms; • Amplitude envelope can be percussive – regularly heard, like in impact sounds – or flat – man-made, and not regularly heard. • A pulse with a standard envelope (20 ms onset) is perceived as more urgent that a slower onset; • A pulse with a slower onset is perceived as more urgent than a slower offset; • IEC 60601-1-8 (2012) recommends a rise time of pulses to not be so short as to create mechanical speaker noise. • The fall time or offset can have any duration, if it does not overlap the next pulse (IEC 60601-1-8, 2012).
Spatial Aspects of Clinical Auditory Alarms	
Spatial Modulation	• Spatial movement allows the sound to be more detectable than a stationary alarm; • Auditory acuity gets worse as the velocity of the auditory image exceeds 10º of visual angle per second.
Number of sources	• An increase in the number of sources inside a room will result in an increase in response time to localize them, and a decrease in localization performance.
Reverberation Time	• In the operating room, it should not exceed 0.6 second (Swedish standard 25268)

STEP 3. PRODUCE DESIGN SOLUTIONS

9 Subjective Perception of Auditory Warning signals (Lab vs. Web)

The relationship between physical acoustic parameters and the subjective responses they evoke is important to assess in audio alarm design. The perception of urgency has been thoroughly investigated, but the perception of other variables such as pleasantness has not. To characterize the psychological correlates of physical variables like frequency and speed, participants evaluated audio warning signals according to six different semantic differential scales. Regression analysis showed that speed predicted mostly the perception of urgency, preoccupation, and negativity; and frequency predicted the perception of pleasantness and irritability. The laboratory study was replicated online with 394 participants, with similar results, demonstrating the robustness of the findings.

This chapter is partly based on the paper: Vieira, J., Santos, J. A., & Noriega, P. (2018). Using semantic differential scales to assess the subjective perception of auditory warning signals. In Proceedings of the 21st International Conference on Digital Audio Effects (DAFx-18) (pp. 126-132).

Dabac, Tošo (1930). Young boy with headphones

9.1 Introduction

The study of the psychological correlates of physical parameters motivated early psychophysical research. This was considered a tool to better study and understand the mind (Gescheider, 1997). Early studies focused on sensory thresholds of humans, associating the human response to the systematic variation of a physical stimulus. Nowadays, this interest in human response is broader. Could we know more than sensory responses? Could similar methods be used to comprehend the relation between physical parameters and affective responses?

Several experimental methodologies attempt to understand the association of physical parameters with subjective perceptions and evaluations by humans. Mostly derived from these early works, it is common to have controlled laboratorial set-ups to understand how certain emotional states can be triggered. This happens because there is consensus and robustness in what a culturally similar group of participants finds pleasant, attractive, or annoying. For instance, semantic profiling stemmed from the wine tasting industry and is currently being applied in other areas such as acoustics (Lokki, 2014; Lokki et al., 2010). Here, the evaluators can taste and compare several samples of wines and then verbally create a vocabulary describing the perceptual differences between the wines. Later, consensus is achieved among all gathered vocabularies. Another technique, Kansei Engineering (Ishihara, 2011), originated in the automotive industry in Japan intending to quantitatively connect affective responses of the customers to physical design specifications. The evaluation method pairs representative samples of the product under evaluation with representative words usually presented in a semantic differential scale (a scale between two polar adjectives).

In the auditory modality, the semantic differential scale method is used to understand which variations in which acoustic parameters should be implemented to trigger the appropriate affective, attentional, or motor response. While the method is commonly applied in alarm design (Edworthy et al., 2004), disciplines such as sound design for products (Özcan-Vieira, 2008) or music theory (Von Helmholtz, 1912) are also interested in knowing exactly which acoustic structure originates which affective response.

In the auditory alarm design context, early work by Roy D. Patterson (1982;1986), Judy Edworthy and Elizabeth Hellier (1995; 1991; 1995) has set the fundamental work grounds to understand the perception of urgency. However, not all auditory warning signals are associated with urgent events, and thus the same work needs to be made to comprehend which acoustic parameters might trigger, for instance, irritability, preoccupation,

unpleasantness, or others - depending on their context and adequate response. This knowledge will allow designing more appropriate audio alarms or warning signals for environments heavily populated with alarming sounds, such as control rooms, intensive care units or operating theatres.

The purpose of this study is to use a semantic differential scale methodology to understand which acoustic parameters of simple computer-generated sounds influence perceived urgency, pleasantness, irritability, preoccupation, speed, and positiveness. The selection of these subjective parameters came from the stage of selection of a semantic space from the Kansei Engineering methodology, detailed in section 9.2.2.

9.2 Methodological considerations

It is known that research choices are based on trade-offs. All sorts of resources influence the method the researcher will use, and these might or might not influence the obtained results, depending on the robustness of the phenomena under study. For our research, we wanted to compare the results of the same study applied on laboratorial facilities and applied online. Nowadays, the possibilities of having large numbers of participants by performing remote tests are numerous, and the statistical power of large numbers provides more confidence in results. We have selected the current study to make this comparison, and the following sections, including results, will be presented either with the label “Lab” or “Web”.

9.2.1 Method

The selected methodology was based on previous studies of Kansei Engineering (Vieira et al., 2017) and semantic differential scales applied to psychoacoustic studies (Edworthy et al., 1991; Hellier & Edworthy, 1999; Hellier et al., 1995).

9.2.2 Selection of representative pairs of words

When using semantic differential scales, it is of extreme importance to select pairs of words that can adequately describe the object under evaluation (Ishihara, 2011). For this, in a pilot study, personal and professional contacts on the researcher were asked to suggest words they associated with artificial sounds, in all possible contexts. Any words, adjectives or not, were accepted. Examples of sounds were referred, such as sounds from house-hold devices, electronics, sounds from inside the vehicle, or alarms from queuing services. In total, 183 words were suggested that described sensations, emotions and perceptions evoked by sounds.

The most frequent words were: shrieking, loud, alert, irritating, deafening, confusing, noisy, pleasant, short, and sweet. Other words were related with a) physical properties (low, short, long, fast, vibrant, synchronous, slow, repetitive, harmonious); b) positive feelings (relief, calm, curiosity, fresh, gentle, positive, relaxing, pleasant, melodic, peaceful, soft); c) negative feelings (boring, anxious, unpleasant, strident, fiddly, nervous, stressful, irritating, intrusive, angry, frustrating, penetrating); d) other words (critical, strong, important, order, respect, safety, attention, artificial). All were grouped considering similitude of meaning and frequency. This resulted in 11 words and corresponding negation. Then, five human factors and acoustics researchers selected the most fitted pairs to describe artificial sounds/alarms, resulting in 6 pairs. The final six pairs of words are in Table 9-1.

Table 9-1. Pairs of words used for the evaluation

1	Not very - very Urgent
2	Unpleasant-Pleasant
3	Not very - Very Irritating
4	Not very - Very Preoccupying
5	Slow - Fast
6	Negative - Positive

All pairs of words were presented in a continuous visual scale, ranging from 0 to 100 mm without numbers (Figure 9-1).

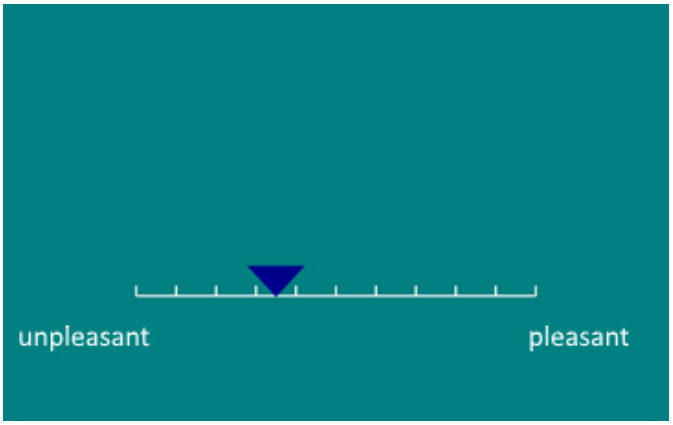


Figure 9-1. Image of the evaluation interface with the semantic differential scale under evaluation

9.2.3 Selection of acoustic parameters

This phase consisted in selecting the acoustic parameters to be manipulated, so in the evaluation phase they could be paired with the chosen pairs of words. Two types of

parameters were selected: spectral and temporal characteristics of sound. Four acoustic parameters were analysed in the present study:

1) **Frequency**: referred to by Hertz (Hz) where 1 Hz is one cycle per second.
2) **Amplitude Envelope:** the shape of a waveform's intensity throughout time. Rise (onset) and fall (offset) times were edited in milliseconds (ms).
3) **Speed:** determined by the inter-pulse interval with faster bursts possessing shorter inter-pulse intervals.
4) **Rhythm:** regular occurrence of an auditory event in time. This occurrence can have a given pattern that can be cyclic, thus having periodicity.

A total of three levels were defined for Frequency, Speed and Onset. Rhythm had two levels. The objective was to have three different levels of priority, High priority, Medium priority and Low priority. Table 9-2 depicts all levels for each parameter. These values were chosen after considering the design guidelines detailed in Chapter 8.

Table 9-2. Levels of each acoustic parameter

	HIGH	MEDIUM	LOW
Fo Frequency	2500 Hz	1500 Hz	500 Hz
Speed	x4	x2	x1
Rhythm	Regular	Regular	Regular
	Syncopated 0	Syncopated 5	Syncopated 10
Onset/Offset	Regular	Slow onset	Slow offset

Values and directionality of the variations were established after literature and international standards on the design of audio warning signals, detailed in the following sections.

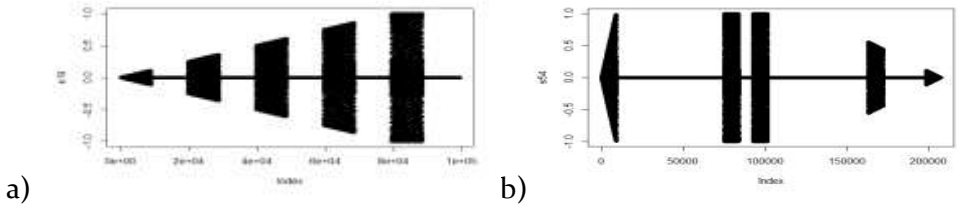
For this study, three levels of **frequency (Fo)** were chosen: 2500 Hz, 1500 Hz, and 500 Hz. All had four harmonics. Regarding **Speed**, the strategy adopted by Edworthy et al. (1991) was applied by creating three levels of speed with a systematic relationship: the faster speed was twice the speed of the moderate one, which was twice the speed of the slower one. The temporal distribution of the "pulse + silence" was repeated three times when the speed was x1 (slow) and x2 (moderate), and it was repeated five times when speed was in x4 (fast). However, the silence duration differed according to speed. In x1 it had 1 s, in x2 it had 0.5 s and in x4 it had 0.25 s.

As for **Rhythm**, standard IEC 60601- 1-8 suggests syncopated or "off-beat" rhythms for higher priority alarms and regular rhythms for medium and low priority alarms. Edworthy et al. (1991) have found the inverse relation with syncopated rhythms being perceived as

less urgent than a regular one. For this study, the rhythm was based on the syncopation index of Fitch and Rosenfeld (2007) (index 0, 5 and 10), and all stimuli were tested both with regular and syncopated rhythm. Concerning **onset and offset**, stimuli had either a slow onset (180 ms; offset of 20 ms), a regular onset and offset (20 ms) or a slow offset (180 ms; onset of 20 ms).

9.2.4 Auditory Stimuli

To test all variables, a combination of all parameters was performed, generating 54 stimuli (3 Frequency x 3 Speed x 2 Rhythm x 3 Onset/Offset). All audio stimuli were generated in R Studio using Seewave (Sueur et al., 2008) and TuneR (Ligges et al., 2016) packages. All were normalized to the same Leq(A) value and played at approximately 72.3 dB(A). A modular approach as first proposed by Patterson (1982) and used in Edworthy et al. (1991), was applied, where pulses were firstly created and then grouped into longer bursts of sound, which were then intercalated with periods of silence to form the full warning. All pulses were 200-ms long. Figure 9-2 depicts two warning signals. Similar depictions for all alarms are in Appendix 4[17] and can be heard here.



a) b)

Figure 9-2. Depictions of a) Stimuli with 180-ms onset, regular rhythm, speed level 4 (1500 Hz); b) Stimuli with 20-ms onset, 180-ms offset, syncopated rhythm, speed level 1 (500 Hz)

9.2.5 Participants

9.2.5.1 Lab

Twenty-six participants took part in the study (17 female, 9 male), from 20 to 50 years (M=33, SD=10), all with normal hearing and most (22) without formal musical education.

[17] Alarms available in: Vieira, Joana (2021): Using semantic differential scales to assess the subjective perception of auditory warning signals. figshare. Media. https://doi.org/10.6084/m9.figshare.14905200.v1

Data collections were carried out in two geographical locations to gather a higher number of participants, using the same equipment.

9.2.5.2 Web

Three hundred and ninety-four participants took part on the study (39% were female, 61% were male and), the average age was 34 years (SD=11). 54% had had musical training, and most participants were from the United States (N=156) and India (N=131). Other countries included Venezuela, Portugal, Italy, Philippines, UK, Canada, among 37 different countries.

9.2.6 Apparatus

9.2.6.1 Lab

The study took place in a quiet room, where the participant was seated in front of a display and made the sound evaluation using a computer mouse by clicking on the visual analog scale. Participants used AKG Pro Audio K271 MKII headphones and all stimuli were presented using PsychoPy (Peirce, 2007) software running on a Lenovo G500s with a 3rd generation Intel® Core™ i7-3612 processor and a Conexant Audio HD. Audio stimuli were presented in 77 dB SPL.

9.2.6.2 Web

The test was implemented on LimeSurvey (LimeSurvey GmbH, 2018), an online survey platform, with the sound clips hosted in Clyp.it (https://clyp.it). It was then distributed via Mechanical Turk (www.mturk.com), an Amazon service with two user profiles: requesters and workers. In the case of scientific research, the Requesters are researchers with online experiments, and workers are respondents from all over the work who receive a variable amount of money for their participation on the experiments or surveys. Mechanical Turk has measures to verify if respondents are consistent, providing them with a quality ranking.

Specific characteristics of the participant's equipment were beyond our control, and we requested them to perform the listening tests at a comfortable volume using head or in-earphones.

9.2.7 Procedure

9.2.7.1 Lab

Participants were welcomed and explained the main objective of the study, which consisted in evaluating several sounds according to a set of properties. They sat in front of a screen and placed the headphones. There was one participant per experimental session. After signing an informed consent and answering demographic questions, the instructions were given by the experimenter. These referred that after presenting a sound, an adjective was going to be presented, and participants should evaluate that sound according to that adjective. There was a total of six adjectives, and participants were told they should evaluate how much the sound was pleasant or unpleasant, irritating or not, preoccupying or not, slow or fast, urgent or not urgent and, finally, negative or positive (Table 9-3).

Table 9-3. Descriptors per pair of words. A sheet with this information was always near the participant

PAIR OF WORDS	DESCRIPTION
Unpleasant - Pleasant	I dislike the sound and it bothers me // I like the sound and it does not bother me
Not Irritating - Very Irritating	The sound does not make me feel irritated and impatient // The sound makes me feel irritated and impatient
Not Preoccupying - Very Preoccupying	The sound does not make me feel worried and alarmed// The sound makes me feel worried and alarmed
Slow -Fast	The sound has a slow pace// The sound has a fast pace
Not Urgent - Very Urgent	The sound communicates a need that may not be immediate// The sound communicates an immediate need
Negative - Positive	The sound communicates a negative information// The sound communicates a positive information

Participants were told they could only make the evaluation after hearing the entire sound once, which could last between 2 to 5 seconds. Participants could navigate with the mouse on the line of the scale, but after clicking with the mouse, it could not be changed. The scale was a continuous 100-mm scale. Before starting the experiment, all participants went through a training phase with the same scales and four different sounds (from (ANSI/AAMI/ IEC 60601- 1-8:2006 & A1:2012, 2013). After, the experimental session began, and the screen displayed one pair of words at the time (Figure 9-1). The presentation of sound files was randomized, as well as the presentation of the pairs of words. Due to the great number of stimuli to be evaluated, there were two intervals, which had the length the participant preferred. The total procedure lasted between 30 to 40 minutes. Each

participant evaluated each stimulus once for each pair of words. In total, each participant made 324 evaluations (54 stimuli x 6 pairs of words).

9.2.7.2 Web

For the online experiment, the same stimuli, scales, and randomization script from the Lab were transferred to the Web survey platform. Nevertheless, for participant engagement, of all 324 possible words and sounds combinations (54 sounds x 6 scales), each participant was randomly assigned 20 sounds to evaluate. After the conclusion of the online collection each sound was evaluated, on average, 25 times.

The instructions presented on the first screen of the survey (Figure 9-3) stated the goal of the study ("*the evaluation of different types of artificial sounds according to a set of properties*"). They required the use of headphones and normal hearing and added "*After listening to a sound - you will hear 20 in total - you will be required to evaluate that sound according to an adjective. There are six possible adjectives that will be randomly presented with the sounds, which are also randomly selected*". The same adjectives and descriptors as in Table 9-3 were used, and they were presented on a 0-100mm continuous scale with opposite ends, like the one on Figure 9-1. The first screen also included a test sound where the participant could adjust the volume to a comfortable level, and an example of the slider which would be used to provide an answer.

9.3 Results

Results will be presented by comparing Lab and Web results.

Because participants did not repeat the evaluation, it was important to assess the degree of agreement between participants as raters of a given stimuli, both on the laboratorial and web conditions. Outliers were removed from the sample using Tukey's method due to its independency from data distribution. This method ignores the mean and standard deviation, which are influenced by the outliers, by using an inter-quartile range approach (above and below the 1.5*IQR).

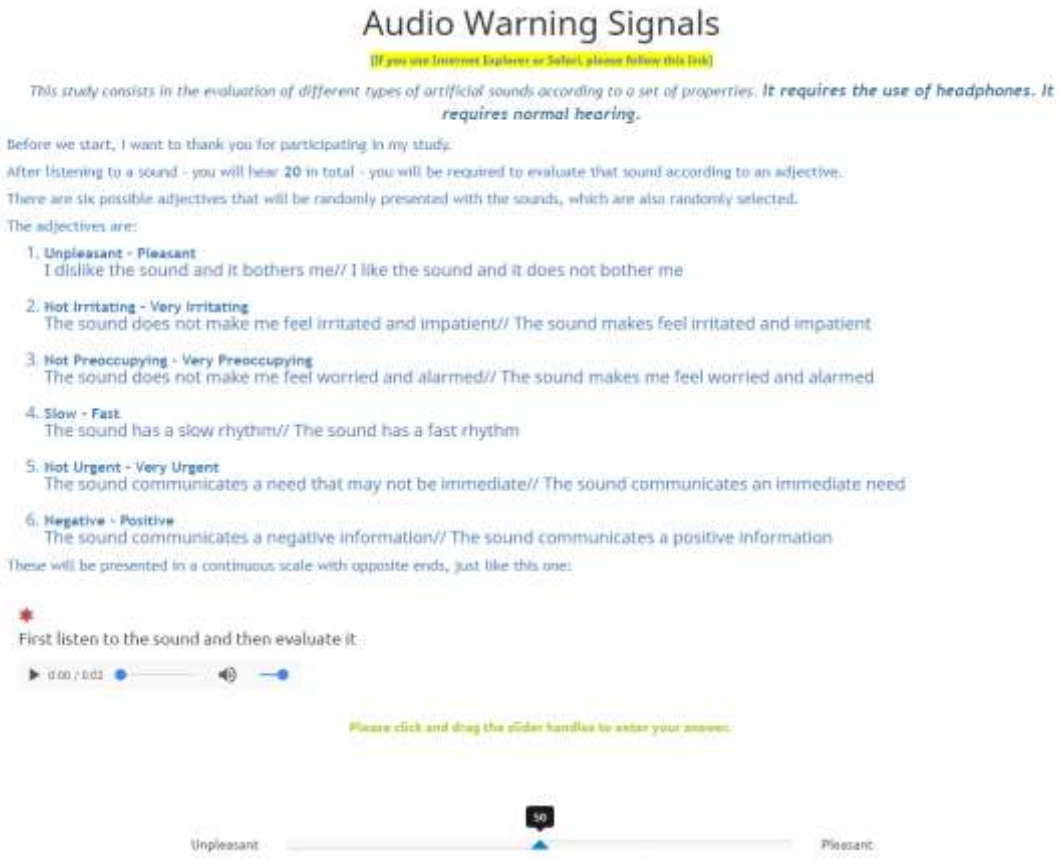


Figure 9-3. First screen of the LimeSurvey (LimeSurvey GmbH, 2018) listening test. It included instructions, description of the scales' adjectives, a test sound to adjust the volume and a trial question to test the 0-100 mm slider.

9.3.1 Inter-participant concordance

Kendall's W (also known as Kendall's coefficient of concordance) is a non-parametric statistic and can be used for assessing agreement among raters. Kendall's W ranges from 0 (no agreement) to 1 (complete agreement). The value of Kendall's W was calculated per pair of words to verify if the stimuli were rated in more or less the same order per participant. The results are in Table 9-4. All tests revealed a significant value of Kendall's W both in Lab and Web conditions. The values in the Web condition are very high, all significant and well above 0.70. In the Lab condition, the pair "Slow-Fast" obtained the highest value of concordance, followed by "Not Urgent – Urgent".

Table 9-4. Values of Kendall's W in Lab and Web conditions

KENDALL'S W	LAB (N=26)	WEB (N=394)
Slow – Fast	0.70 ***	0.76 ***
Not Urgent - Very Urgent	0.61 ***	0.76 ***
Not Irritating - Very Irritating	0.45 ***	0.77 ***
Not Preoccupying -Very Preoccupying	0.42 ***	0.78 ***
Unpleasant – Pleasant	0.36 ***	0.78 ***
Negative – Positive	0.12 ***	0.78 ***

*** Significant (p < .001) ** (p <.01) * (p < .05)

Kendall's W provides the agreement in the order of an evaluation, for instance, a 500Hz sound was consistently evaluated as more pleasant than a 2500 Hz sound. In this situation, the difference between values in both conditions may be due to the N power of the Web condition (26 vs 394 participants).

To ascertain the consistency between raters in both conditions, we ran an ICC (intra-class correlation) analysis. The ICC is used as a measure of association when studying the reliability of raters. The values presented on Table 9-5 result from the ICC3 value of the R (R Foundation for Statistical Computing, 2017) {psych} package.

Table 9-5. Values of ICC in Lab and Web conditions

ICC	LAB (N=26)	WEB (N=394)
Slow – Fast	0.69	0.48
Not Urgent - Very Urgent	0.60	0.39
Not Irritating - Very Irritating	0.43	0.18
Not Preoccupying -Very Preoccupying	0.40	0.20
Unpleasant – Pleasant	0.35	0.08
Negative – Positive	0.08	0.05

As expected, less rater consistency was found among the Web respondents. Nevertheless, the order was maintained with more consistency being observed in the "Slow-Fast" pair or words and less consistency on the "Negative-Positive" pair of words.

These analyses only show consistency and do not reveal the nature of the classification made by participants. For this purpose, correlational and linear regression analysis were performed after all data was pooled and averaged.

9.3.2 Correlation and Regression

The correlation results from both Lab and Web experiments are presented in Figure 9-4 and Figure 9-5 respectively. Both conditions have shown the same relations, with more significant *r* values present in the Lab condition.

9.3.2.1 Correlations (Lab)

The significant correlations found with Frequency were with Irritating ($r(52) = .69$, $p < .001$), Positive ($r(52) = -.27$, $p < .05$) and Pleasant ($r(52) = -.88$, $p < .001$); with Speed, the stronger correlations were with Positive ($r(52) = -.70$, $p < .001$), Preoccupying ($r(52) = .80$, $p < .001$), Urgent ($r(52) = .83$, $p < .001$), and Fast ($r(52) = .78$, $p < .001$); with Rhythm were Irritating ($r(52) = .27$, $p < .05$), Preoccupying ($r(52) = .31$, $p < .05$), Urgent ($r(52) = .37$, $p < .05$), and with Fast ($r(52) = .41$, $p < .001$).

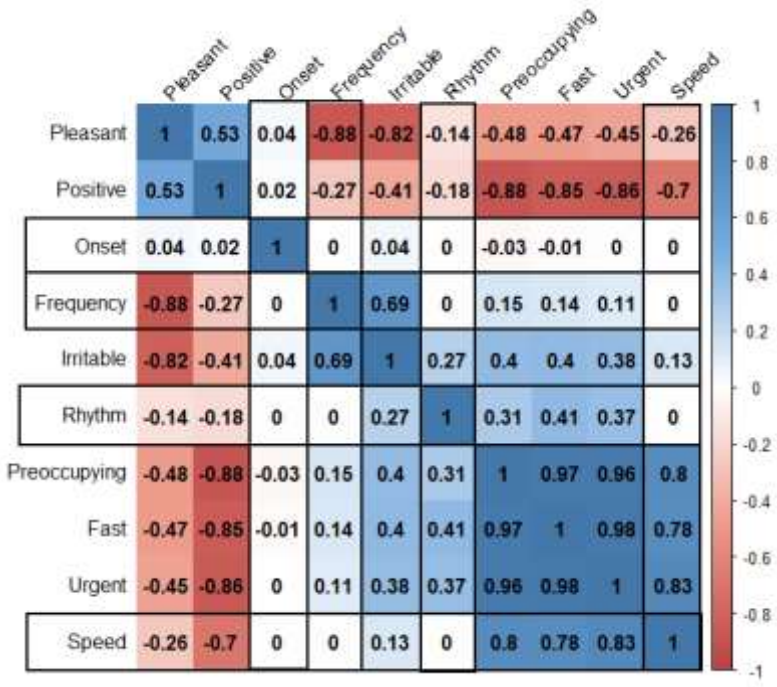


Figure 9-4. Correlations between the four acoustic parameters and the six pairs of words – Lab condition. Thick borders include the manipulated parameters.

No correlations were found with the acoustic parameter Onset-Offset. For this reason, this variable will not be used in further analysis.

Additionally, the pair of words Irritating correlated significantly with all other words, negatively with Positive and Pleasant. The pair of words Positive and Pleasant were negatively correlated with Preoccupying, Urgent and Fast. And Preoccupying was correlated with Urgent, and Fast. Following this, all relations between acoustic parameters and pairs of words were explored using linear regression models (Table 9-6 to Table 9-10).

9.3.2.2 Correlations (Web)

The significant correlations found with Frequency were with Pleasant ($r(52)$ = -.58, p < .05). Unlike the Lab condition, here Frequency did not have significant correlation with Irritating. With Speed, the stronger correlations were with Positive ($r(52)$ = -.64, p < .001), Pleasant ($r(52)$= -0.22, p<0.5, Preoccupying ($r(52)$ = .65, p < .001), Urgent ($r(52)$ = .75, p < .001), and Fast ($r(52)$ = .75, p < .001). Rhythm had significant correlations with Irritating ($r(52)$ = .41, p < .05) and Pleasant ($r(52)$ = -.33, p < .05). No correlations were found with the acoustic parameter Onset-Offset.

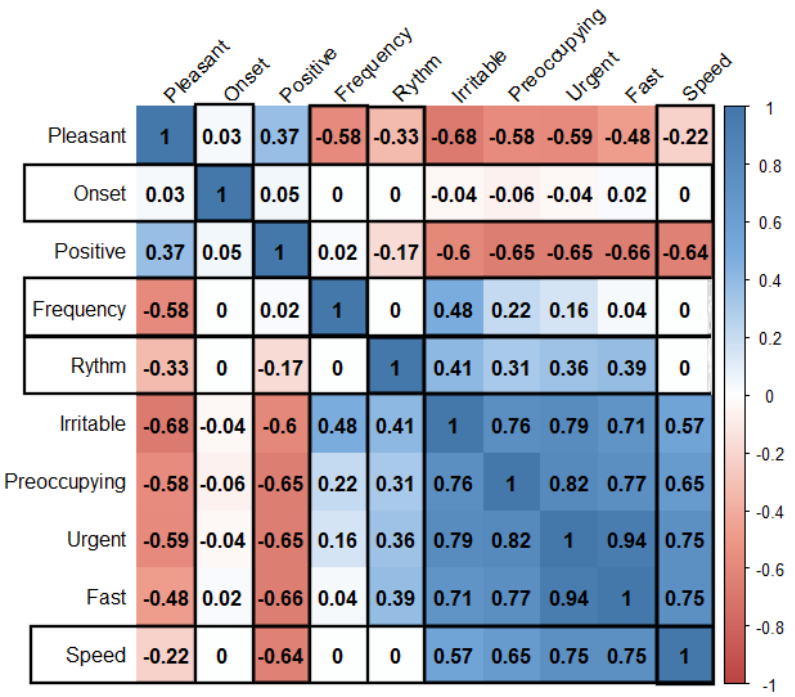


Figure 9-5. Correlations between the four acoustic parameters and the six pairs of words – Web condition. Thick borders include the manipulated parameters.

All relations between acoustic parameters and pairs of words were explored using linear regression models.

9.3.3 Frequency

9.3.3.1 Frequency - Lab

The Frequency (Table 9-6) variable had three levels, and each level increased the perception of unpleasantness of our participants, with 500 Hz (β = 53.96, F = 100.2, R^2= .80, p < .001) 1500 Hz (β = -15.55, p < .001) and 2500 Hz (β = -23.77, p < .001). A similar pattern was found regarding the perception of irritableness, with 500 Hz (β = 39.11, F = 29.43, R^2= .54, p < .001) 1500 Hz (β = 19.82, p < .001) and 2500 Hz (β= 24.75, p < .001). No significant relations with Frequency were observed among the other pairs of words.

Table 9-6. Results of linear regression by levels of Frequency (500 Hz, 1500 Hz and 2500 Hz). N = 54. 95% Confidence Interval (only R^2> 0.5 are depicted)

FREQUENCY	UNPLEASANT – PLEASANT		NOT - VERY IRRITATING	
	B	CI	B	CI
Intercept (500)	53.96 ***	51.54 – 56.38	39.11***	34.26 – 43.96
1500	-15.55 ***	-18.97 – -12.12	19.82***	12.96 – 26.68
2500	-23.77 ***	-27.19 – -20.34	24.75***	12.96 – 26.68
F	100.2		29.43	
R^2	.797		.536	

*** Significant (p < .001) ** (p <.01) * (p < .05)

These two regression models are plotted in Figure 9-6, with the y-axis depicting the 20-80 mm fraction of a 100-mm visual analog scale.

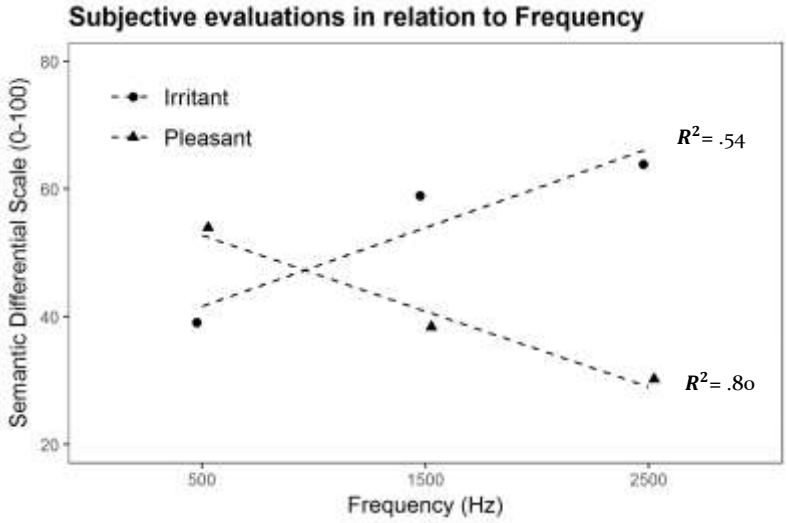


Figure 9-6. Significant regressions for Frequency as predictor. Relationship between participant's evaluation of a sound as Irritable (R^2= .54) or Pleasant (R^2 = .80) and three levels of increasing frequency (Lab).

According to these results, the higher the sound's frequency, the more irritant and the less pleasant the sound is evaluated.

9.3.3.2 Frequency - Web

In the Web condition, and like the Lab condition, Irritant and Pleasant were the only linear models with significant values. As with previous analysis, the response trend is the same in both conditions, with reduced quality of fit in the Web condition. In both conditions, as the Frequency increased, so did the perception of Irritability, and the perception of Pleasantness decreased. Figure 9-7 depicts these relations.

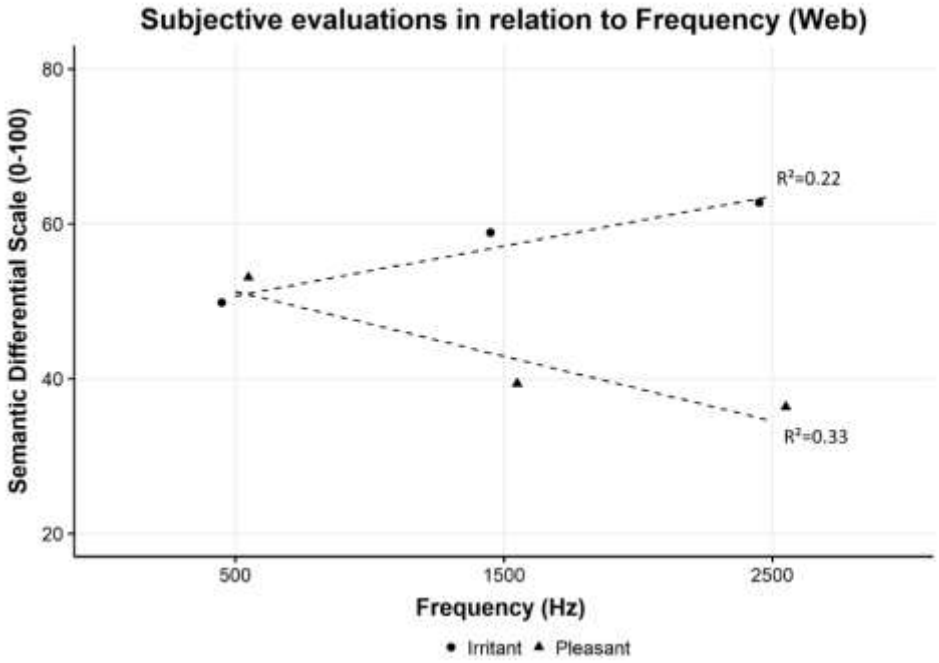


Figure 9-7. Significant regressions for Frequency as predictor. Relationship between participant's evaluation of a sound as Irritable (R^2= .22) or Pleasant (R^2 = .33) and three levels of increasing frequency (Web).

The results' tables will present the Mean, Standard Deviation, p-value and R^2 of both Lab and Web conditions. Table 9-7 presents the results from the Frequency significant linear models.

Table 9-7. Mean values and R^2 comparison between linear models of Lab and Web Conditions for the levels of Frequency (500 Hz, 1500 Hz and 2500 Hz)

FREQUENCY	IRRITANT		PLEASANT	
	Lab	Web	Lab	Web
500Hz | Mean (SD)	39,11(5,70)	49,81(9,51)	53,96 (4,38)	47,35 (4,16)
1500Hz | Mean (SD)	58,93 (8,22)	58,88 (10,10)	38,41 (5,14)	39,33 (6,19)
2500Hz | Mean (SD)	63,86 (14,05)	62,73 (8,69)	30,19 (5,35)	36,32 (7,56)
p-value	***	***	***	***
R^2	0.54	0.22	0.80	0.33

*** Significant (p < .001) ** (p <.01) * (p < .05)

9.3.4 Speed

9.3.4.1 Speed - Lab

The Speed (Table 9-8) variable also had three levels and it was the variable which better explained the variance of four pairs of words. As the speed increased, so did the perception of Urgency (Speed x1 β = 31.70, p < .001, Speed x2 β = 23.82, p < .001 and Speed x4 β = 39.68, p < .001, F = 76.11, R^2= .75): Preoccupation (Speed x1 β = 34.31, p < .001, Speed x2 β = 20.37, p < .001 and Speed x4 β = 30.78, p < .001 , F = 69.47, R2 = .73) and Speed (Fast) (Speed x1 β = 32.95, p < .001, Speed x2 β = 27.33, p < .001 and Speed x4 β = 40.85, p < .001, , F = 63.9, R^2= .72).

Table 9-8. Results of linear regression by levels of Speed (x1, x2, x4). N = 54, 95% Confidence Interval (only R^2 > 0.5 are depicted)

SPEED	NOT - VERY PREOCCUPYING		SLOW – FAST		NOT– VERY URGENT		NEGATIVE – POSITIVE	
	β	CI	β	CI	β	CI	β	CI
Interc. (x1)	34.3 ***	30.5 – 38.1	33.0 ***	27.7 – 38.2	31.7 ***	27.1 – 36.3	50.6 ***	48.6 – 52.6
x2	20.3 ***	15.0 – 25.7	27.3 ***	19.9 – 34.7	23.8 ***	17.3 – 30.3	-8.8 ***	-11.6 – -5.9
x4	30.7 ***	25.4 – 36.1	40.9 ***	33.5 – 48.2	39.7 ***	33.2 – 46.2	-12.0 ***	-14.8 – -9.1
F	69.47		63.9		76.11		37.9	
R^2	.731		.715		.749		.598	

*** Significant (p < .001) ** (p <.01) * (p < .05)

The reverse pattern was observed in the Negative-Positive pair of words, with the perception of positiveness decreasing as the speed increased (Speed x1 β = 50.60, p < .001, Speed x2 β = -8.75, p < .001 and Speed x4 β = -11.95, p < .001, F = 37.9, R^2 = .60). No significant relations with Speed were observed among the other pairs of words.

The four regression models are plotted in Figure 9-8.

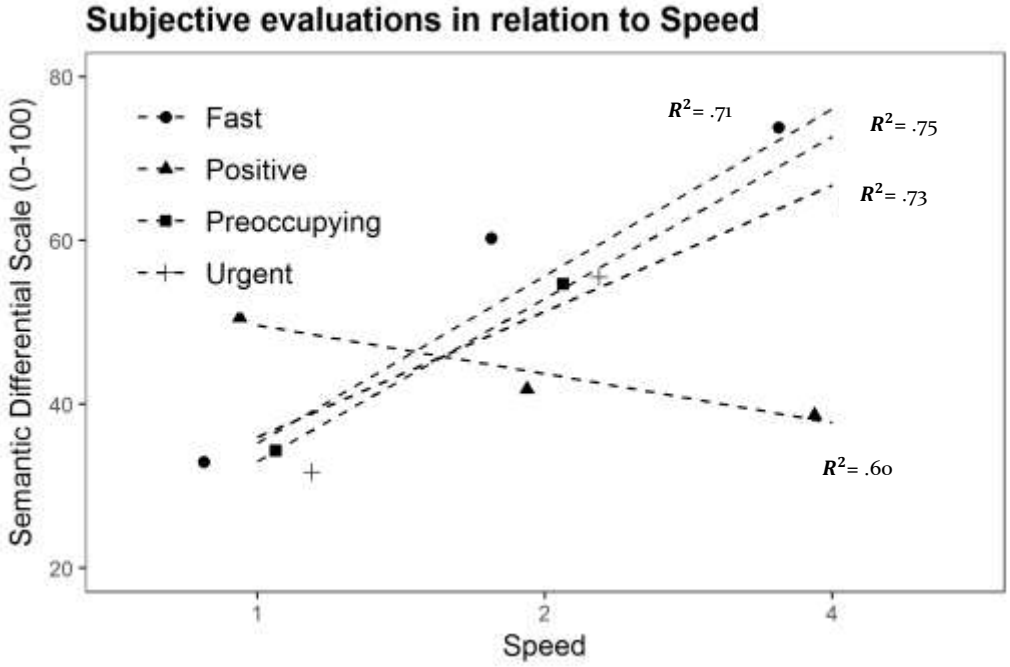


Figure 9-8. Significant regressions for Speed as predictor. Relationship between participant's evaluation of a sound as Fast (R^2=.71), Positive (R^2=.60), Preoccupying (R^2=.73), or Urgent (R^2=.75) and three levels of increasing speed (Lab)

According to these results, the higher the sound's speed, the more urgent, fast, and preoccupying and the less positive it is evaluated.

9.3.4.2 Speed - Web

In the Web condition, and like the Lab condition, Positive, Preoccupying, Urgent and Fast were the linear models with significant values, with Irritant being exclusive on the Web condition. As with previous analysis, the response trend is the same in both conditions, with reduced quality of fit in the Web condition. Table 9-9 presents the results from the Speed significant linear models.

Table 9-9. Mean values and R^2 comparison between linear models of Lab and Web Conditions for the levels of Speed (x1, x2 and x4)

SPEED	IRRITANT		POSITIVE		PREOCCUPYING		URGENT		FAST	
	Lab	Web	Lab	Web	Lab	Web	Lab	Web	Lab	Web
1 M (SD)	49,88 (14,4)	49,00 (10,98)	50,60 (6,59)	49,32 (6,93)	34,31 (15,08)	46,29 (10,71)	31,70 (18,95)	39,41 (16,37)	32,95 (20,30)	31,07 (18,00)
2 M (SD)	56,49 (14,16)	57,76 (10,77)	41,85 (6,47)	43,94 (6,82)	54,67 (14,86)	58,47 10,52)	55,52 (18,76)	61,19 (16,29)	60,28 (20,14)	57,19 (17,96)
4 M (SD)	55,54 (14,54)	64,66 (10,85)	38,65 (6,54)	38,51 (6,75)	65,08 (14,92)	64,13 (10,61)	71,38 (18,79)	71,20 (16,33)	73,80 (20,21)	66,18 (17,89)
p	-	***	***	***	***	***	***	***	***	***
R^2	-	0.31	0.60	0.39	0.73	0.41	0.75	0.56	0.72	0.56

*** Significant (p < .001) ** (p <.01) * (p < .05)

As the Speed increased, so did the perception of Irritability, Preoccupying, Fast, Urgent and the perception of Positiveness decreased. Figure 9-9 depicts these relations.

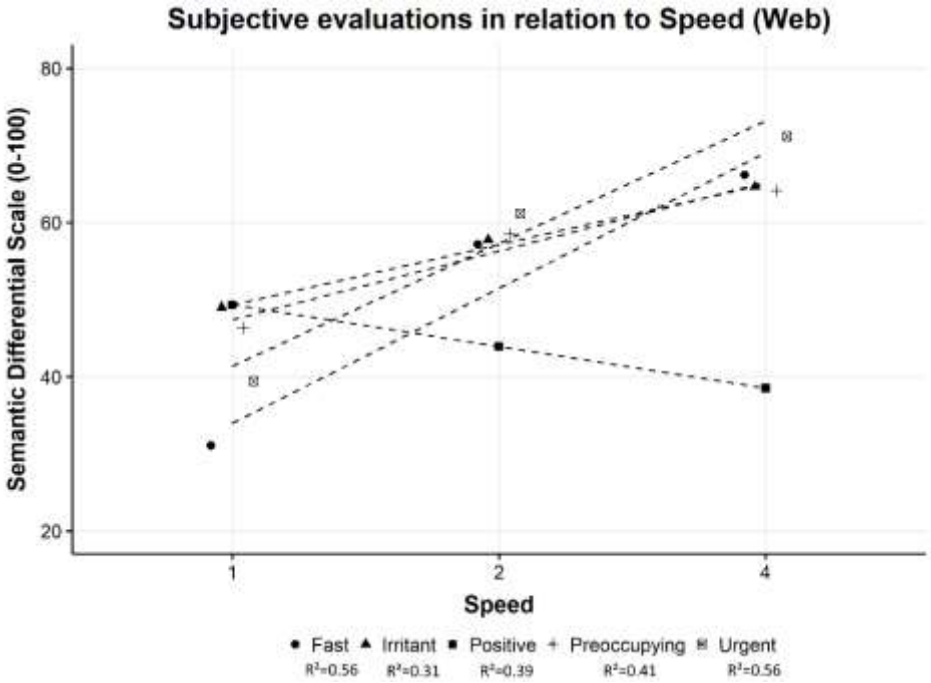


Figure 9-9. Significant regressions for Speed as predictor. Relationship between participant's evaluation of a sound as Fast (R^2= .56), Irritant (R^2 = .31), Positive (R^2 = .39), Preoccupying (R^2 = .41) or Urgent (R^2 = .56) and three levels of increasing Speed (Web).

9.3.5 Rhythm

9.3.5.1 Rhythm - Lab

The Rhythm (Table 9-10) variable had two levels and the perception of speed (word fast) (Regular β = 47.47, p < .001, Syncopated β = 16.41, p < .001, F=10.4, R^2 = .17) and urgency (Regular β = 45.89, p < .001, Syncopated β = 13.95, p < .001, F=8.26, R^2 = .14) increased when the rhythm was syncopated.

Table 9-10. Results of linear regression by levels of Rhythm (Regular, Syncopated). N = 54, 95% Confidence Interval (only R^2 > 0.5 are depicted)

RHYTHM	SLOW – FAST		NOT–VERY URGENT	
	β	CI	β	CI
Intercept (Reg.	47.47***	40.25 – 54.69	45.89***	39.00 – 52.78
Syncopated	16.41**	6.20 – 26.62	13.95**	4.21 – 23.69
F	10.4		8.26	
R^2	.167		.137	

*** Significant (p < .001) ** (p <.01) * (p < .05)

Having significant regression coefficients means the Rhythm is correlated with both subjective perceptions, nevertheless, the model does not account for the variability found among the data. The two regression models are plotted in Figure 9-10.

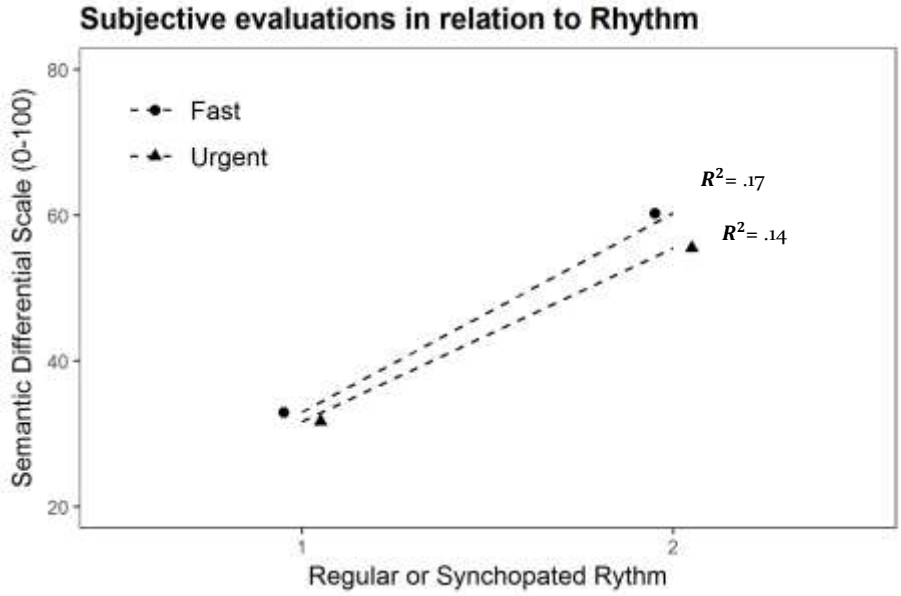


Figure 9-10. Significant regressions for Rhythm as predictor. Relationship between participant's evaluation of a sound as Fast (R^2=.17) and Urgent (R^2=.14) and two levels of Rhythm (regular or syncopated).

9.3.5.2 Rhythm - Web

In the Web condition, Rhythm had more significant linear models than in the Lab condition. Irritant, Pleasant, Preoccupying, Urgent and Fast were the linear models with significant values. The response trend is the same in both conditions, with reduced quality of fit in the Web condition. Table 9-11 presents the results from the Rhythm significant linear models.

Table 9-11. Mean values and R^2 comparison between linear models of Lab and Web Conditions for the levels of Rhythm (Regular and Syncopated)

RHYTHM	IRRITANT		PLEASANT		PREOCCUPYING		URGENT		FAST	
	Lab	**Web**	**Lab**	**Web**	**Lab**	**Web**	**Lab**	**Web**	**Lab**	**Web**
Regular | M (SD)	-	52,65 (10,9)	-	43,57 (7,81)	-	53,00 (10,67)	45,89 (18,79)	51,51 (16,49)	47,47 (20,29)	44,65 (18,02)
Syncopated | M (SD)	-	61,63 (11,01)	-	38,43 (7,9)	-	59,60 (10,53)	59,84 (18,88)	63,02 (16,39)	63,88 (20,04)	58,31 (18,06)
p-value	-	**	-	**	-	**	**	**	**	**
R^2	-	0.15	-	0.09	-	0.08	0.14	0.11	0.68	0.13

*** Significant ($p < .001$) ** ($p < .01$) * ($p < .05$)

As the Rhythm became syncopated, the perception of Pleasantness decreased, but the perception of Fast, Irritant, Preoccupying and Urgent increased. Figure 9-11 depicts these relations.

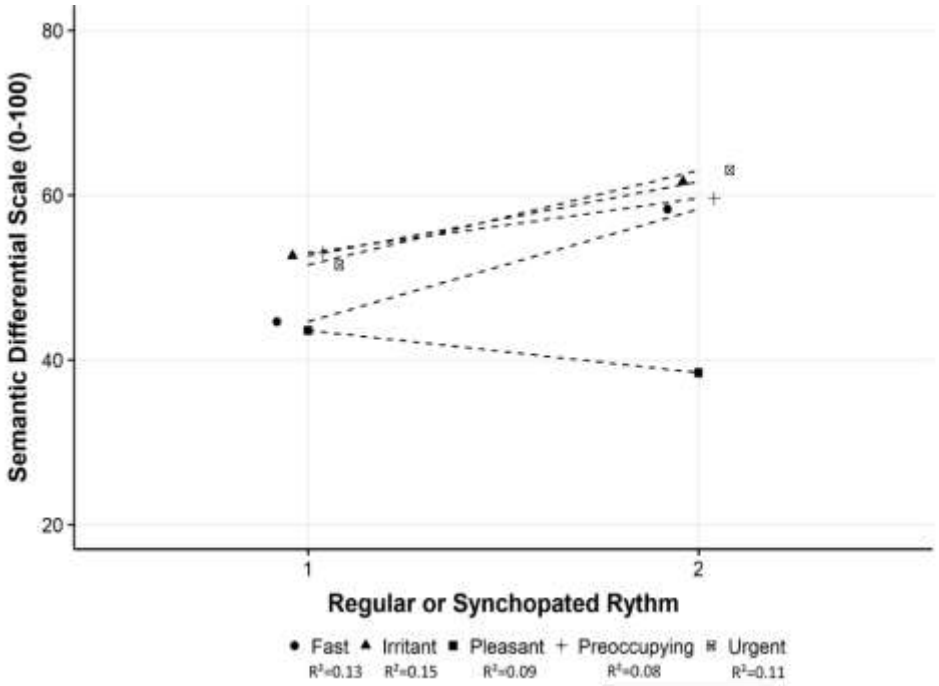


Figure 9-11. Significant regressions for Rhythm as predictor. Relationship between participant's evaluation of a sound as Fast (R^2= .13), Irritant (R^2 = .15), Pleasant (R^2 = .09), Preoccupying (R^2 = .08) or Urgent (R^2 = .11) and two levels of increasing Rhythm (Web).

9.4 Discussion

The obtained results demonstrate how the semantic differential scale methodology is robust and useful for the analysis of relations between subjective perceptions and physical acoustic parameters. Firstly, although some pairs of words were extremely subjective, like unpleasant – pleasant, or vague, like negative-positive, there was consistency among lab participants, revealed in the significant values of Kendall's W in all pairs of words.

This was an important result, as it allows to somewhat balance an obvious limitation of this study, which was the lack of repetitions of the evaluation sessions. At first, one could expect large interpersonal variability regarding such subjective perceptions, but these observations serve as an addition to the strengths of this simple method. It is important to add that during the data collection phase, some participants had informally mentioned they had trouble in classifying a sound as negative or positive, even though they had the definition sheet nearby. It is then somewhat surprising to understand that, although difficult, the classification was congruent among raters, later relating significantly to the manipulation of the Speed variable.

Regarding the associations between subjective and physical variables, with Frequency, it was observed that the subjective perceptions in which it had more effect were Pleasantness and Irritability. Again, although apparently a very personal evaluation, most participants found high-frequency audio signals as unpleasant and irritant. This is an important result that confirms that an alarm, to essentially fit its purpose of communicating an urgent event, does not need to increase its frequency. In fact, it should not, as it only affects the negative affective perception of the signal.

Also, and in agreement with Patterson's suggestions and standard norms, Speed is the variable which mostly affects an alarm's perception of urgency, communication of preoccupation or that "something" negative is happening.

Rhythm obtained results also aligned with the IEC 60601-1-8 standard, with participants evaluating as significantly more urgent those auditory signals with syncopation than those with regular rhythm. However, the association found was not robust, and no more elations can be made. One explanation can be that the irregularity of the rhythm might have been affected by the slow onsets and offsets, not allowing to hear the full structure of the auditory signal.

Contrary to the literature and standards, the onsets and offsets of the auditory signals had no effect on the perception of any pair of words. In the future, the variations of this parameter should be more numerous, and evaluations should consider this manipulation only. This would allow clarifying the effect this parameter has without interacting with other manipulations.

Importantly, another aim of this study was to analyse the differences between the same task performed on laboratorial settings versus online. The results were very interesting. It can be said that the lab tendencies were observed on the web, with reduced statistical strength when performed online. The direction of the evaluations demonstrated by Kendall's W was the same on both settings. However, and expectantly, with 394 participants the inter rater agreement reflected by the ICC was considerably lower. This was observed in the values of the correlations, and in the R^2 values of the linear regressions, lower for the web condition. Concerning the latter, the web condition has pointed more parameters as being influenced by Speed (Irritant, Positive, Preoccupying, Urgent and Fast) and Rhythm (Irritant, Pleasant, Preoccupying, Urgent and Fast).

As stated by Crowston (2012), with careful task design, online data collection resorting to Mechanical Turk "*offers an interesting new capability to recruit research subjects or labour for a research project, providing useful research data*". In our case, it allowed us to collect

an impressive number of 394 respondents (after removing outliers) which would be impossible for laboratorial settings. Of course, the experiment needed to be adjusted to the conditions, which affected its robustness by having fewer evaluations per participant. Nevertheless, it allowed us to verify that the directions of the evaluations were similar in both conditions, which strengthen the value of the any design guidelines following this study.

With this study, it was possible to understand which acoustic features trigger what affective state when designing for auditory warning signals. For instance, that a signal to be understood as urgent should have shorter and irregular inter-pulse intervals, preferably with lower frequencies. However, these sound design recommendations must co-exist with other requirements such as the ability to localize audio warning signals in an open space, and the ability to recognize it among other devices with similar spectral and temporal patterns.

9.5 Conclusion

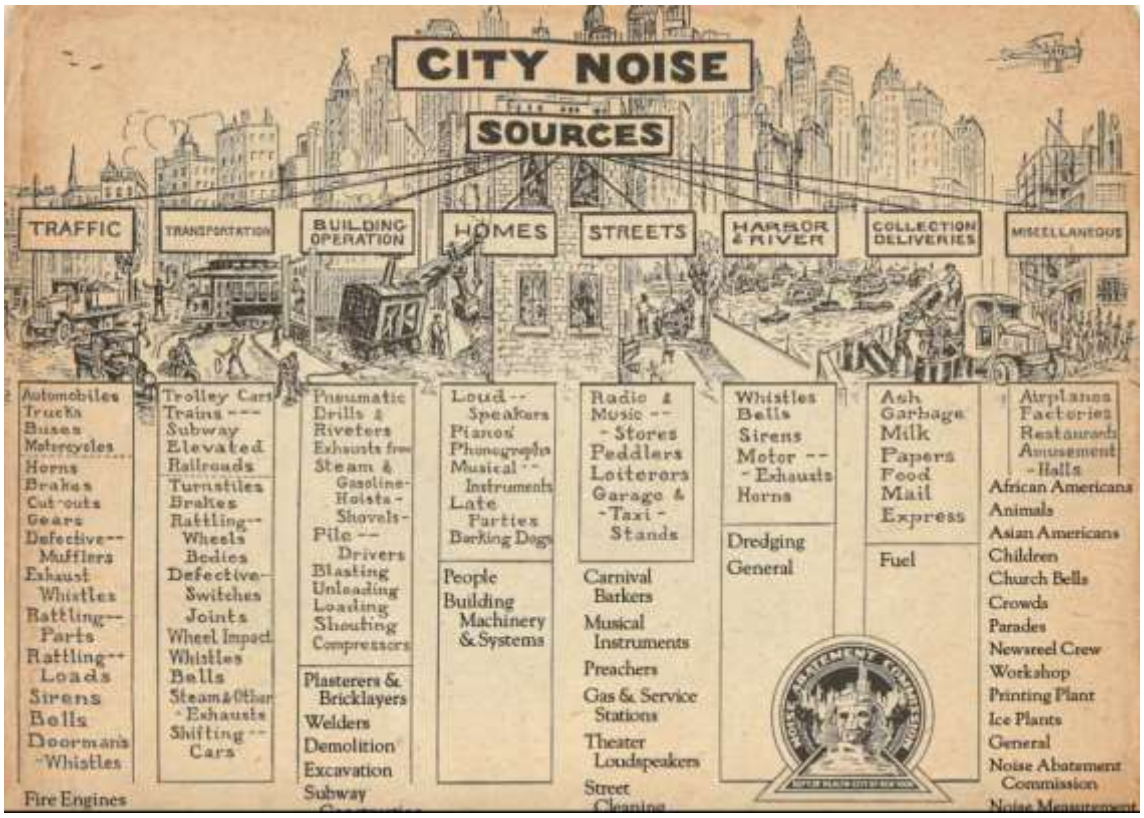
A study was performed to better understand the psychological correlates of acoustic parameters. The study was implemented on laboratorial settings and on web settings alike. Fifty-four stimuli were created manipulating frequency, speed, rhythm and onset and offset times. Twenty-six participants listened to each stimulus six times on the lab, each time considering a different pair of words presented in a visual analog scale. Three hundred and ninety-four participants made the experiment online. These words were selected among more than a 100 sound-related words. The applied methodology consisted in using semantic differential scales. The findings allowed to consolidate this method as a good evaluator of subjective perceptions. Results have demonstrated that the acoustic features which most contribute to the perception of these states in audio stimuli are frequency (pleasantness and irritability) and speed (urgency, preoccupation, and negativity). Rhythm also affected the perception of urgency, although to a lesser extent, with irregular rhythms obtaining higher ratings for the perception of urgency.

After these fundamental associations between acoustic parameters and subjective perception have been established, the next step will be to understand the semantic associations between everyday sounds and the alarm categories we intend to design. Afterwards, both fundamental and semantic knowledge will be used to create clinical alarms.

10 CATEGORIZATION OF SOUNDS

An online sound categorization study was carried out to assess the association of everyday sounds regarding medical equipment audio alarms. Participants categorized 51 sounds into one of eight categories. They were asked to rate the quality of fit of their evaluation (Poor, Fair, Perfect) after sorting each sound. A consensus was achieved for the categories of Blood Pressure, Cardiac, Power Down and Ventilation. For the categories of Drug Administration, Temperature, Perfusion and Oxygen, lower consistency was observed, highlighting the difficulty of sound design for specific contexts. Evidence of agreement among participants is an important output for the general goal of designing a library of informative sounds for medical devices.

This chapter is partly based on the paper: Vieira, J., Marques, R., Pereira, F., Ferreira, V., Santos, J. A., & Noriega, P. (2020). Categorization of sounds for the design of clinical auditory alarms. In Proceedings of the International Conference eForum Acusticum 2020 (pp. 2979-2986).



Edward Brown et al. (1930), Soundscape of the modern city, New York Department of Health.

10.1 Introduction

Clinical auditory alarms are a tool in healthcare facilities to alert and inform caregivers of patient or medical equipment state changes (McNeer et al., 2018). Nevertheless, both healthcare professionals and patients face auditory alarms as a necessary hindrance which affects not only their mental health and professional capabilities (I. Busch-Vishniac, 2019; ECRI, 2013), but which also, in the case of patients, affects their effective recovery (Birdja & Özcan, 2019; Ecophon, 2018) 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 (Edworthy et al., 2018; Edworthy et al., 2017; Hellier et al., 1993; Patterson et al., 1986), but the implementation of changes in a complex sociotechnical context such as healthcare is famously slow (Norman & Stappers, 2015).

The current chapter 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 environmental or 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.

10.1.1 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 (Houix et al., 2012) and is part of a “cognitive economy” where the environment is simplified and information is retrieved with the least cognitive effort

(Bones et al., 2018). 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*" (Gygi et al., 2007).

There are several everyday sounds taxonomies (Bones et al., 2017), based on an ongoing effort to understand how these sounds are interpreted. These include sorting exercises of ambient urban noise (Guastavino, 2007), human activity sounds (Dubois, 2000), or road traffic noise (Morel et al., 2012). Category formation generally follows the strategy of similarity between sound sources, places, and actions (Bones et al., 2018). 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 (Vanderveer, 1979). This behaviour matches the statement from Giordano, McDonnell, and McAdams (2010) 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 (Bones et al., 2017). In 1979, Vanderveer 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 (Giordano et al., 2010).

Importantly, Ballas (1993) 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, Houix, Misdariis, and Susini (2010) 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 pay more attention (Lemaitre et al., 2010). 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 (Lemaitre et al., 2010) 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 (Lemaitre et al., 2010).

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 (Bones et al., 2017). 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 (Bones et al., 2017). 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 (Bones et al., 2017; Lemaitre et al., 2010). Most studies do not present participants with predefined categories.

A call for attention has been made Aldrich, Hellier, and Edworthy (2009), 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.

10.1.2 Clinical Auditory Alarms

In the context of alarm design for healthcare settings, to our knowledge, only Bennet et al. (2011) have performed similar studies, although strictly focused on the auditory features of alarms. The authors compared the IEC tonal 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.

10.2 Method

In usability and 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 (Hudson, 2014). 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.

10.2.1 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, and 63% were female while 38% were male. Further demographic data is displayed Table 10-1.

Table 10-1. Demographic data of participants

Experience with medical devices	**N**	**Profession**	**N**
Yes	30	Physician	14
No	33	Nurse	9
Age	**N**	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	**N**	Consultant	2
Yes	20	Industrial Engineer	2
No	43	Acoustic Engineer	2
Location	**N**	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		

10.2.2 Stimuli

We presented fifty-one stimuli. Previous literature (Bennett et al., 2019) 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.). 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 IEC 60601-1-8 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. Acoustic characterization of the sounds can be found in Appendix 5, and sounds can be heard here[18].

All sound files were normalized to the same Leq(A) value. This process consisted in adding a digital gain to the sound file so that all frequencies reached their maximum amplitude. The Leq(A) of each sound was calculated to obtain an average value for all files, in this case -32dB(A) which was the lowest level found in all sound files.

10.3 Procedure

The study was conducted via the online survey platform LimeSurvey (LimeSurvey GmbH, 2018). Initially, participants were asked demography questions including whether they were healthcare professionals or had musical training. Before the categorization task, participants performed a listening screening test.

10.3.1 Listening Screening Test

To assure that all participants performed the listening test while using headphones, we implemented the intensity-discrimination task designed by (Woods et al., 2017). 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. The task consists of a three-alternative forced-choice (3-AFC) which asks participants '*Which tone is quietest?*'. All two-channel tones were 200 Hz pure tones with a duration of 1000 ms, with 100ms Tukey's on- and off-ramps. The three signals differed from that in the following characteristics:

- Signal 1: channel 1 and channel 2 in-phase, Leq -3.59 dB Full Scale (dBFS);
- Signal 2: channel 1 and channel 2 phase reversed (180 degrees), Leq -3.59 dBFS;
- Signal 3: channel 1 and channel 2 in-phase, Leq -9.59 dBFS

The three tones were presented in random order with an interstimulus interval of 500 ms on each trial. The listener was asked to pick the interval containing the quietest tone, by selecting one of three buttons labelled 'FIRST sound is SOFTEST', 'SECOND sound is SOFTEST", and 'THIRD sound is SOFTEST". Each participant responded to six trials. 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.

[18] Vieira, Joana (2021): Categorization of sounds for the design of clinical auditory alarms. figshare. Media. https://doi.org/10.6084/m9.figshare.14905113

10.3.2 Categorization Task

The main task consisted in grouping the 51 sounds into nine categories. Figure 10-1 represents the categorization interface. 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.

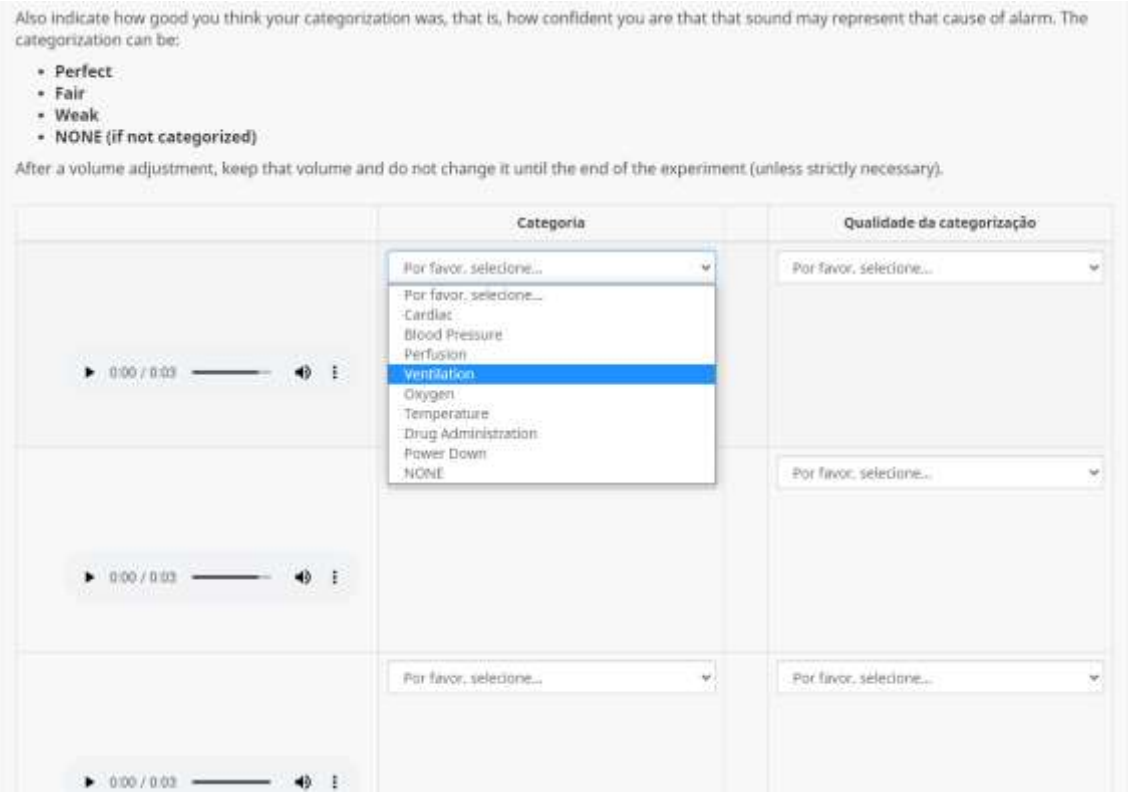


Figure 10-1. Categorization Interface on LimeSurvey

They were also provided a short description of what could be included in each category:

> *"You will hear 51 different types of sounds (3 to 4 seconds). You should categorize each sound in terms of the category it may represent. The categories are related to the medical causes that usually activate audible alarms in medical equipment. These categories are already defined, and relate to the main causes of audible alarms in hospitals:*
>
> ***Heart Rate** (Anaesthetist Workstations such as Heart Rate Monitors, Plethysmographs, Echo transoesophageal, Foetal Heart Rate Monitors).*
>
> ***Blood pressure** (invasive or non-invasive blood pressure monitors).*
>
> ***Artificial perfusion** (cardiopulmonary perfusion pumps and associated equipment, intra-aortic balloon pumps, renal dialysis systems).*
>
> ***Ventilation** (lung ventilators, spirometers, CO_2 monitors, airway pressure.*

Oxygen *(pulse oximeters, transcutaneous or tissue oxygen monitors, oxygen analysers, oxygen concentrators, oxygen gas suppression pathways).*

Temperature *(temperature monitors, heated air humidifiers, children's radial heaters, neonatal incubators, patient heating or cooling systems, blood, or fluid heaters, electrocauterizes, ultrasound, laser, X-ray, or electromagnetic resonance systems, nerve stimulators).*

Drug or fluid administration *(volume infusion pumps, syringe guides, anaesthetic delivery systems or analysers).*

Equipment Failure *(any equipment that has a power loss or other major or significant failure)*

NONE *(if you think the sound doesn't fit in any 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.

10.4 Analysis

The two groups (H - Healthcare and non-H – non-Healthcare professionals) were analysed 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 (Figure 10-2). 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.

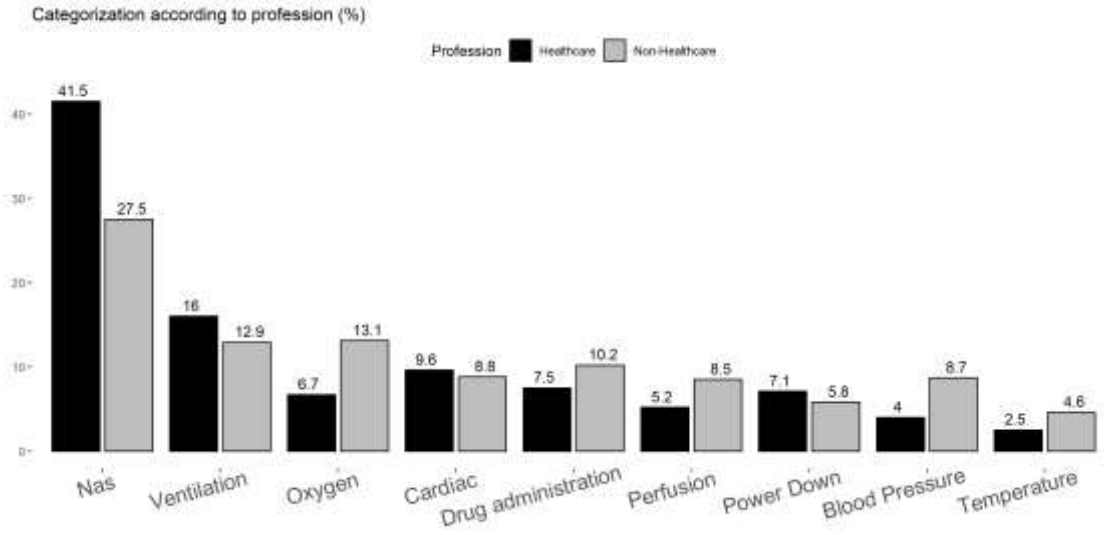


Figure 10-2. Percentage of sounds placed in each Category according to participants' profession

Analysing the categorization of each group, the similarities regarding categorization are noteworthy. Table 10-2 includes the main similarities (green) and differences between (red) of the categorization of each group in percentage. These percentual values correspond to the number of times each stimulus was placed in a category.

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.

Table 10-2: Main differences and similarities 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.

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

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. The Perfusion and Temperature categories had a residual number of classifications.

To understand whether both groups' data could be pooled, a similarity analysis was performed using SynCaps V3 (Syntagm, 2018). 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 10-3) in which the weight corresponded to the average quality of fit (from 1 – Fair to 3 - Perfect) of each pairing. The highest values were attributed to "sound_koroktoff" (2,4 QoF), followed by "bellows", "mask", "darthvader" (2,1 QoF). 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. Due to the reduced number of categorizations in the Temperature category, it was removed from analysis.

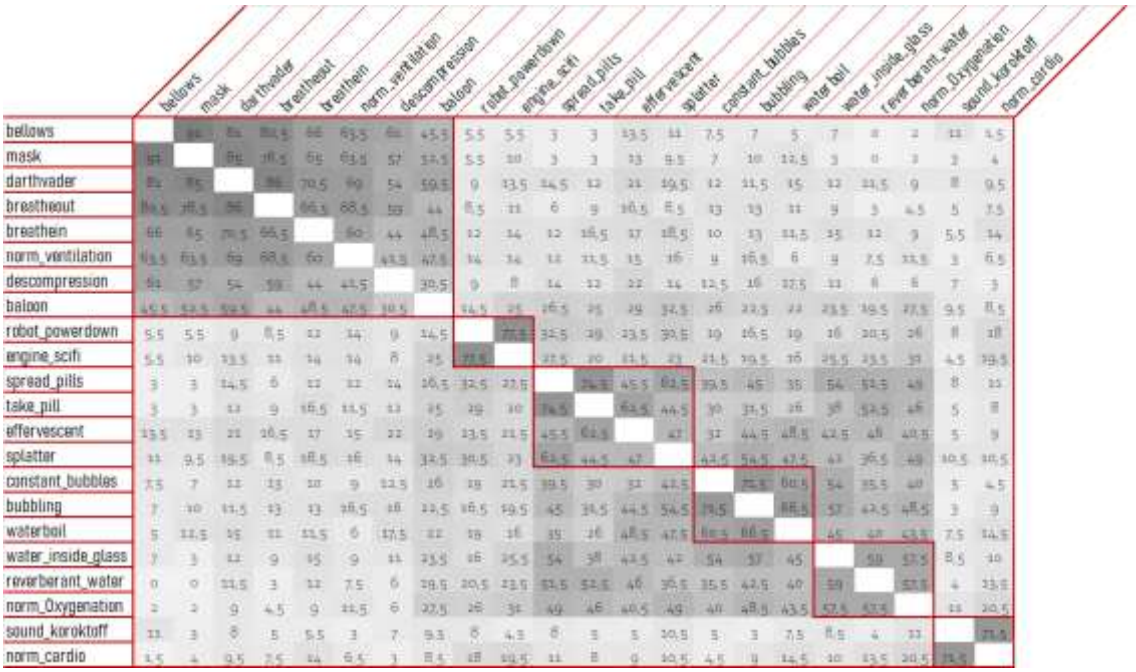


Figure 10-3. Weighted similarity matrix (item proximity and Quality of Fit) of the final pool of 22 sounds

10.5 Results

The weighted matrix of the final 22 sounds (Figure 10-3) represents the closest items in darker colour, and six main clusters can be observed, with breathing and ventilation sounds representing the biggest and strongest group. Regarding categorization, the preliminary results are presented in Figure 10-4. From this depiction it is possible to

observe that some sounds were almost unequivocally categorized, such as "sound_koroktoff" which was placed inside the Cardiac and Blood Pressure categories. Similarly, a group of sounds was placed both in the Oxygen and in the Ventilation category, such as "bellows", "breathein", "breatheout", "darth vader", "mask" or "norm_ventilation". The sounds of "engine_scifi" and "robot_powerdown" were often placed under the Power Down category. The Drug administration category was selected for sounds such as "spread_pills", "take_pill", and "water inside glass".

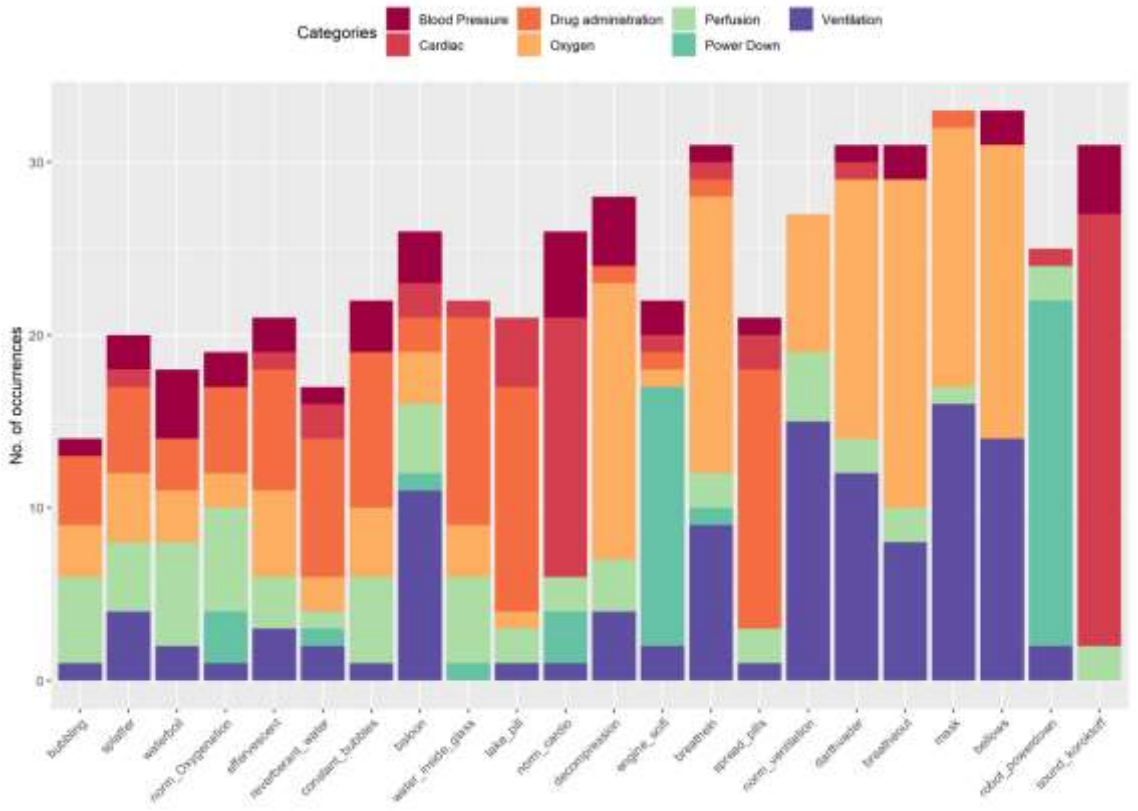


Figure 10-4. Categories in which 22 final sounds (x-axis) were placed

To confirm the number of dimensions of the categorization we have 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 objects, in this case, sounds. A preliminary analysis (cluster dendrogram) to understand the number of clusters that should be inputted for the MDS proposed K=7 (Figure 10-5). 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 10-5. "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 10-5. 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.

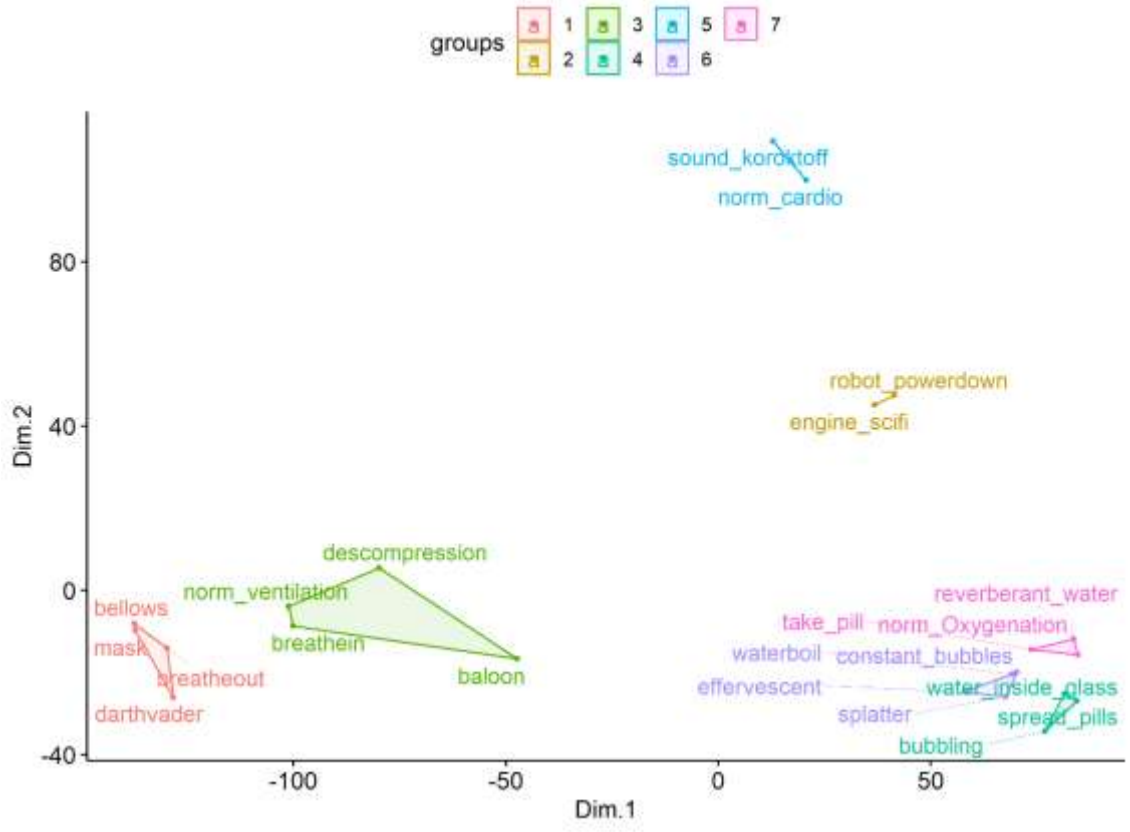


Figure 10-5. Multidimensional scaling with k=7

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.

10.6 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 analysed 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 sci-fi) were put together as a single category, and so were the Cardiac sounds (norm cardio; sound Korotkoff). 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 their 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 Lemaitre et al. (2010). From our results, it seems the followed strategy was based on similarity between sound sources (air, water, percussion) as proposed by Bones (2018). Similar to Aldrich et al. (2009), it seems the grouping task encouraged the use of categorical information instead of acoustic information.

10.7 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:

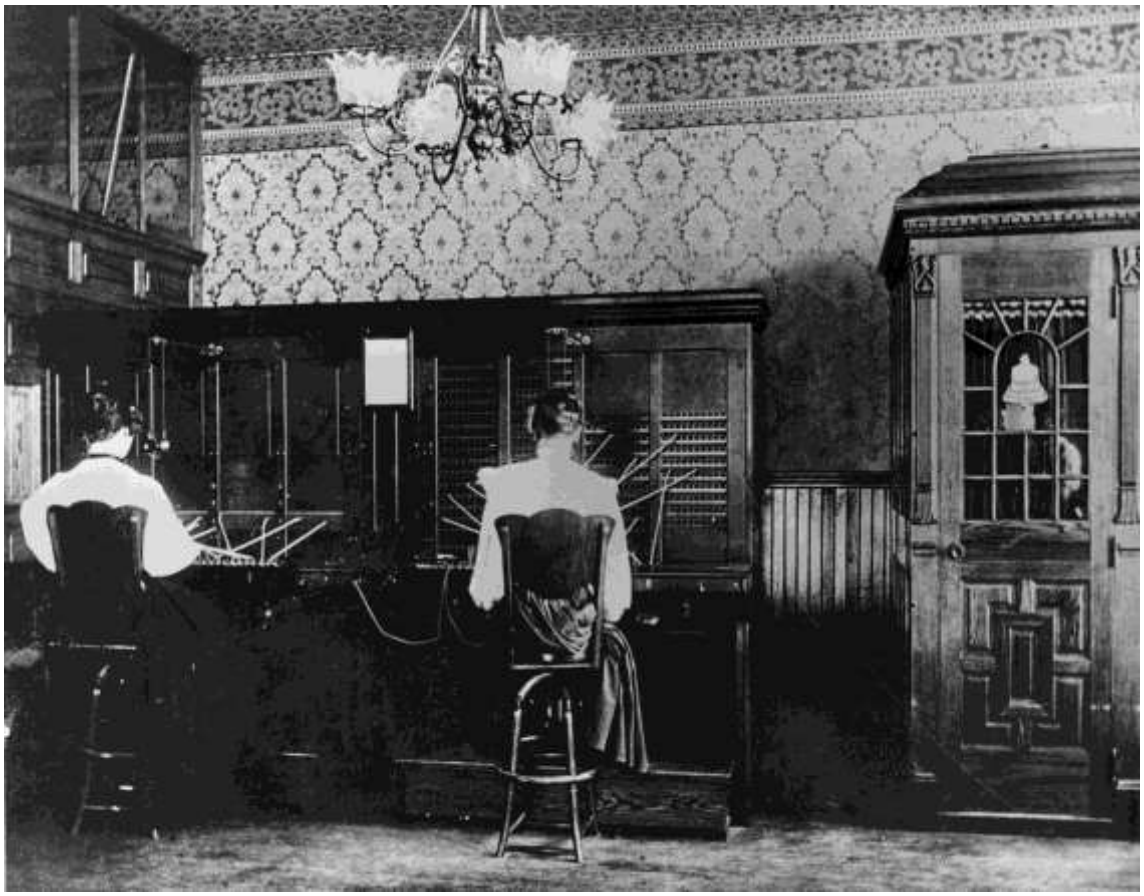
- Cardiac – Sound Korotkoff
- Blood Pressure – Cuff Inflate
- Power down – Robot Power Down
- Ventilation – Mask/Darth Vader
- Drug administration – Spread pills/Take Pill/ Water inside glass

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.

11 Design Strategy of a set of Clinical Auditory Alarms

This chapter includes a description of the strategy that guided the design of our alarms. It is divided in terms of expected Functionality, triggered Perception, and manipulated Physical parameters of the sound.

Finally, a test Alarm Set is described. This Alarm Set will be used in the next human centred design step of Evaluating the Design.



U.S. Department of Agriculture (1895). Telephone switchboard operators answer calls at the South telephone exchange in Minneapolis, Minnesota in 1895. Photo courtesy National Archives and Records Administration. Public Domain.

11.1 Introduction

Strongly based on literature, visits to hospitals, interviews with healthcare professionals and preliminary fundamental research, we set out to define our design strategy for a set of clinical auditory alarms.

We have decided to focus our design on assuring that the alarm is detected, its priority understood, its cause identified, and the tendency of the inherent vital sign comprehended. The goal was to provide the healthcare worker with as much information as possible in no more than 4 seconds – the duration of an alarm. This would allow previous preparation regarding equipment, material, and necessary procedures.

This chapter includes a description of the strategy that guided the design of our alarms. It is divided in terms of expected Functionality, triggered Perception, and manipulated Physical parameters of the sound. Its work is the result of several design workshops and meetings with the Paterson team project (CCG, 2018), and includes their vision and strategy for better clinical alarms.

Finally, a test Alarm Set is described. This Alarm Set will be used in the next human centred design step of Evaluating the Design.

11.2 Our Alarm Sound Design Strategy

After all research stages with different stakeholders, it was time to transform all insights into actionable design requirements. Figure 11-1 represents a hierarchical structure of the clinical alarm as we interpreted it. Most important, the alarm should be **detected** over the background noise. Then, the healthcare workers should be able to understand how **quick their response action** should be: delayed, prompt, or immediate. During our interviews we understood that being able to **identify** what vital sign triggered the alarm was very important and so, also following IEC 60601-1-8 amendment from 2020 design choices, we have opted for auditory icons as a source of identification. Finally, an innovative design requirement would be understanding the alarm's **tendency:** whether it was increasing or decreasing.

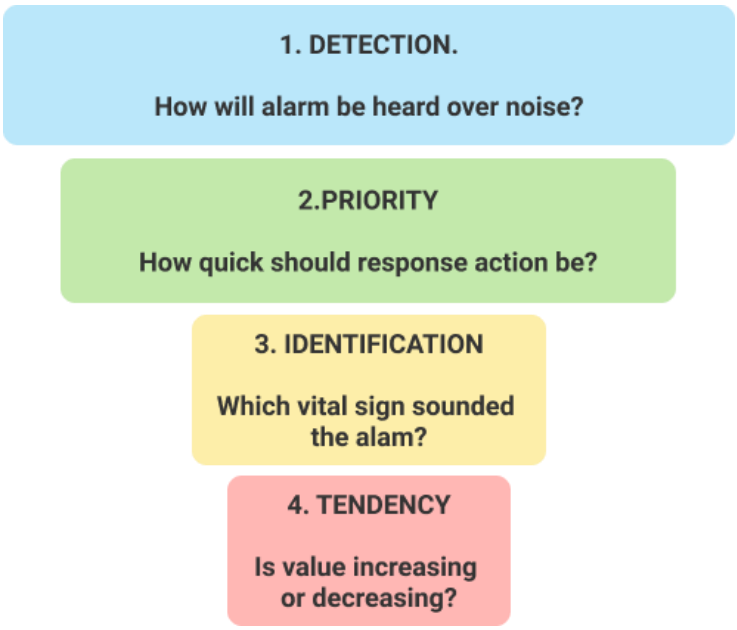


Figure 11-1. Adopted Clinical Alarm Design Strategy

Each step of the alarm design strategy will be detailed in the following sections.

11.2.1 Detectability

The detectability of an alarm is strongly connected with the signal to noise ratio (SNR) concept, and peak to average SNR. Other factors affecting detectability include Loudness, Roughness and Pitch (Figure 11-2).

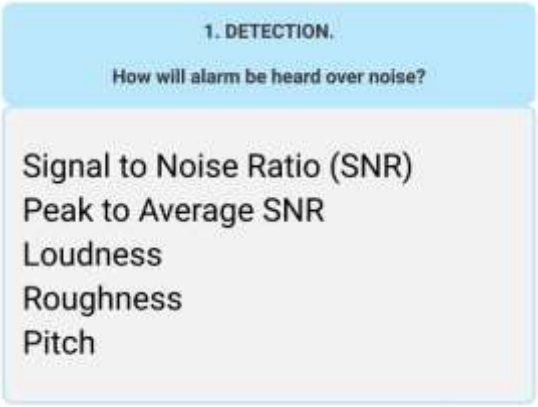


Figure 11-2. Perceptual correlates of the detection of an alarm.

What this means for sound designers is that for Loudness, they should manipulate Fundamental Frequency, Power, Amplitude Modulation, Frequency Modulation and Bursts/Interval Pattern. For Roughness they should edit Fundamental Frequency,

Amplitude and Frequency Modulation. For Pitch perception, they should edit Fundamental Frequency and the Harmonics (Figure 11-3).[19]

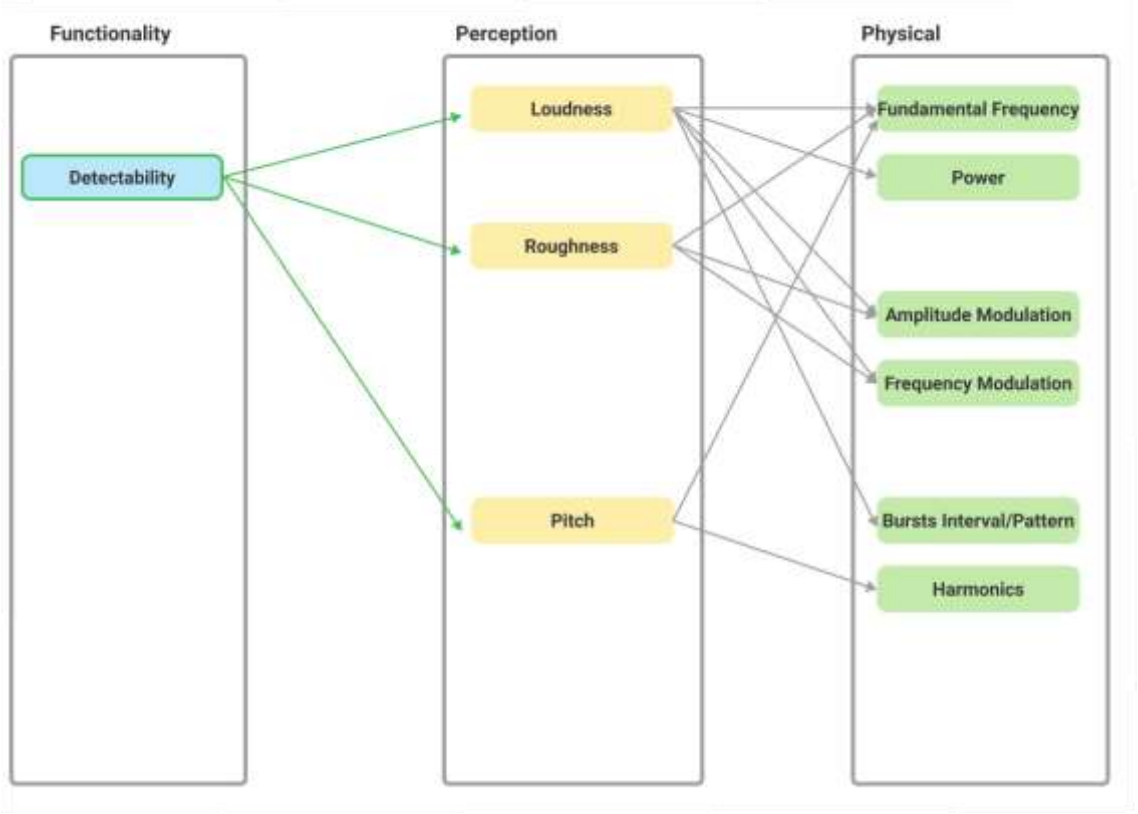


Figure 11-3. Functionality – Perception – Physical relation. With the objective of Detectability, certain percepts need to be triggered and for that, physical components of the sound should be manipulated.

11.2.2 Priority

To understand the type of response action to the alarm, the listener needs to comprehend its communicated priority. For that, perceptual variables include velocity, pitch, harmonicity and the subjective perception of urgency (Figure 11-4).

[19] All visualizations in this chapter of the relations between *Functionality – Perception – Physical* were proposed by Audio and Acoustic Researcher Frederico Pereira, from the Paterson project team (https://www.linkedin.com/in/frederico-pereira-953ba17/).

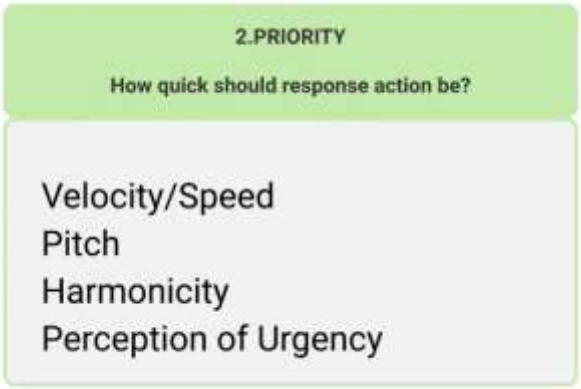


Figure 11-4. Perceptual correlates for the comprehension of the Priority of an alarm.

One of the most important variables for understanding priority is the perception of velocity, which is affected by the Bursts Interval or Pattern. Pitch is manipulated via the Fundamental Frequency and Harmonics; Harmonicity is manipulated through Harmonics and the perception of Urgency is manipulated via Fundamental Frequency, Amplitude Modulation and Bursts Interval and Patterns. Roughness and Sharpness may also play an important role in understanding Priority.

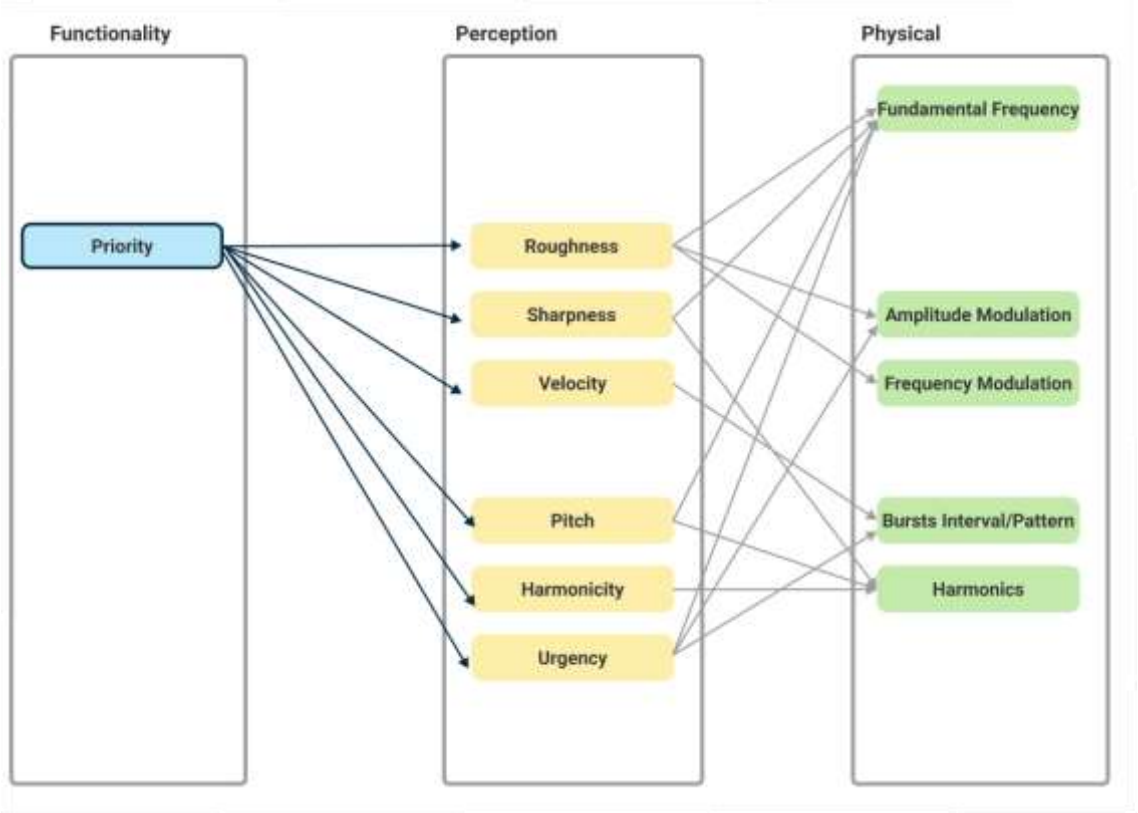


Figure 11-5. Functionality – Perception – Physical relation. With the objective of understanding Priority, certain percepts need to be triggered and for that, physical components of the sound should be manipulated.

11.2.3 Identification

To understand whether what triggered an alarm was a Cardiac or a Ventilation problem, it is important to have a perceptual dimension that allows us to identify the alarm source. We have chosen to use Auditory Icons as main items of an alarm after interviews with healthcare professionals and because that was part of the IEC 60601-1-8 2020 amendment proposed design strategy (Figure 11-6).

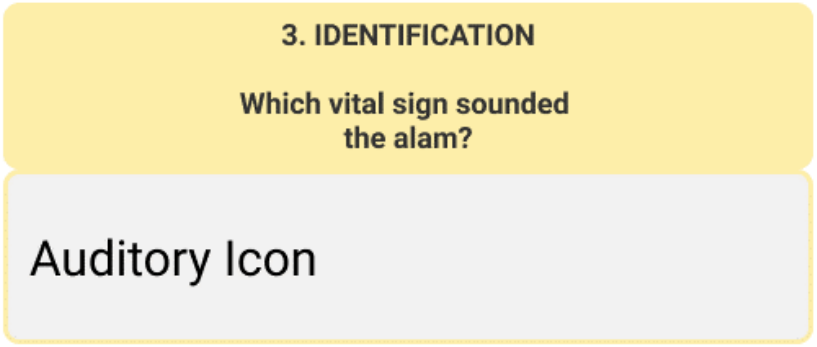


Figure 11-6. Perceptual correlates of the identification of the source of an alarm.

The Auditory Icon can be a real recording of an environmental sound but can also have its spectral-temporal parameters varied to be more easily detected (Figure 11-7). Manipulations such as *cartooning*[20] allow to limit real sounds to their minimal associative contents by being synthetically created. In this manner, they can still be associated to familiar processes (Haverkamp, 2009), like a heart pumping. In the case of operating rooms, it is also an important refinement to avoid potential confusion with real sounds of the clinical environment, in this case, the real sound of a pumping heart (Bennett et al., 2019, Sandau et al., 2017).

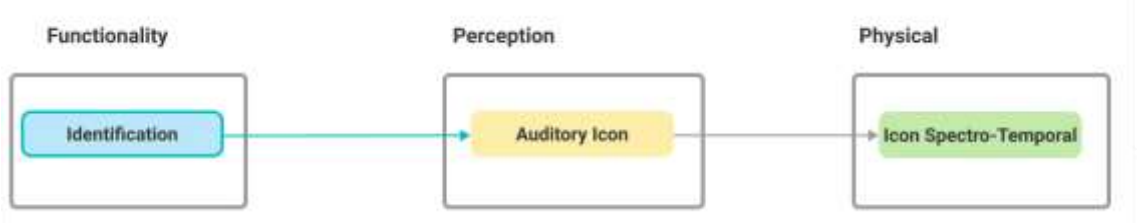


Figure 11-7. Functionality – Perception – Physical relation. With the objective of Identification, certain percepts need to be triggered and for that, physical components of the sound should be manipulated.

[20] This technique simplifies and exaggerates the inherent physics to improve the computation efficiency and perceptual clarity.

11.2.4 Tendency

In an interview with an anaesthesiologist, she has given the example of the pulse oximeter as a good sound. Pulse oximetry is not an alarm, but rather a continuous sonification of the SpO2 levels of a patient. When the value is increasing, the pitch of the alarm ramps up, and when the value is decreasing, the pitch of the alarm ramps down. We have decided to use that powerful affordance of the sound to map the tendency of the values triggering the alarm. The perceptual correlate for this ramping up or down is the Pitch Envelope (Figure 11-8).

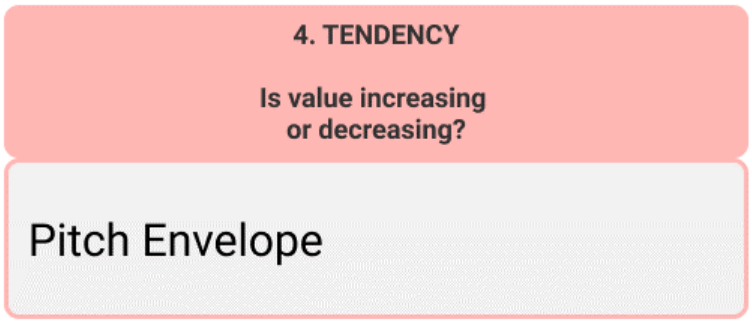


Figure 11-8. Perceptual correlates of the understanding of the Tendency of an alarm.

To manipulate the Tendency of an alarm, the sound designer should manipulate its Frequency Envelopes (Figure 11-9).

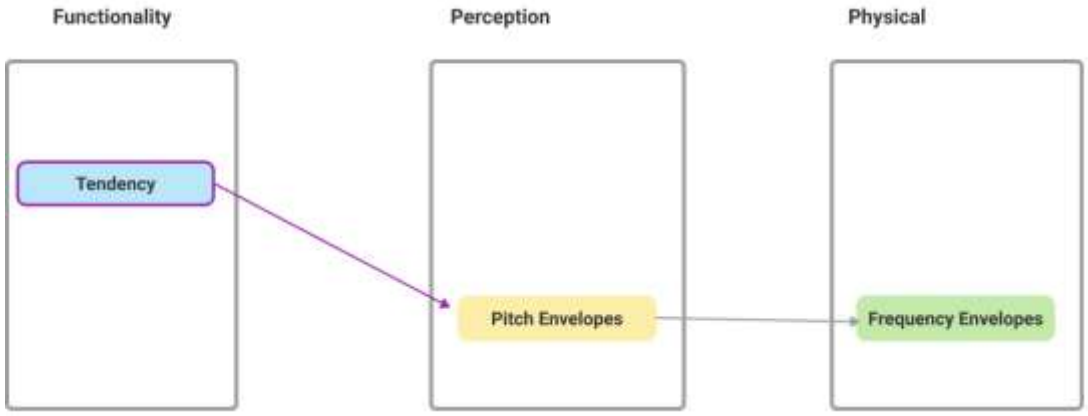


Figure 11-9. Functionality – Perception – Physical relation. With the objective of identifying Tendency, certain percepts need to be triggered and for that, physical components of the sound should be manipulated.

This is not trivial, and previous studies with Pulse Oximetry made relevant observations regarding the increase or decrease of the envelope. If it should be linear or logarithmic is one of the many possible variables (Brown et al., 2015).

11.3 Our Alarm Sound Design Structure

The temporal distribution of the referred variables is represented in **Figure 11-10**.

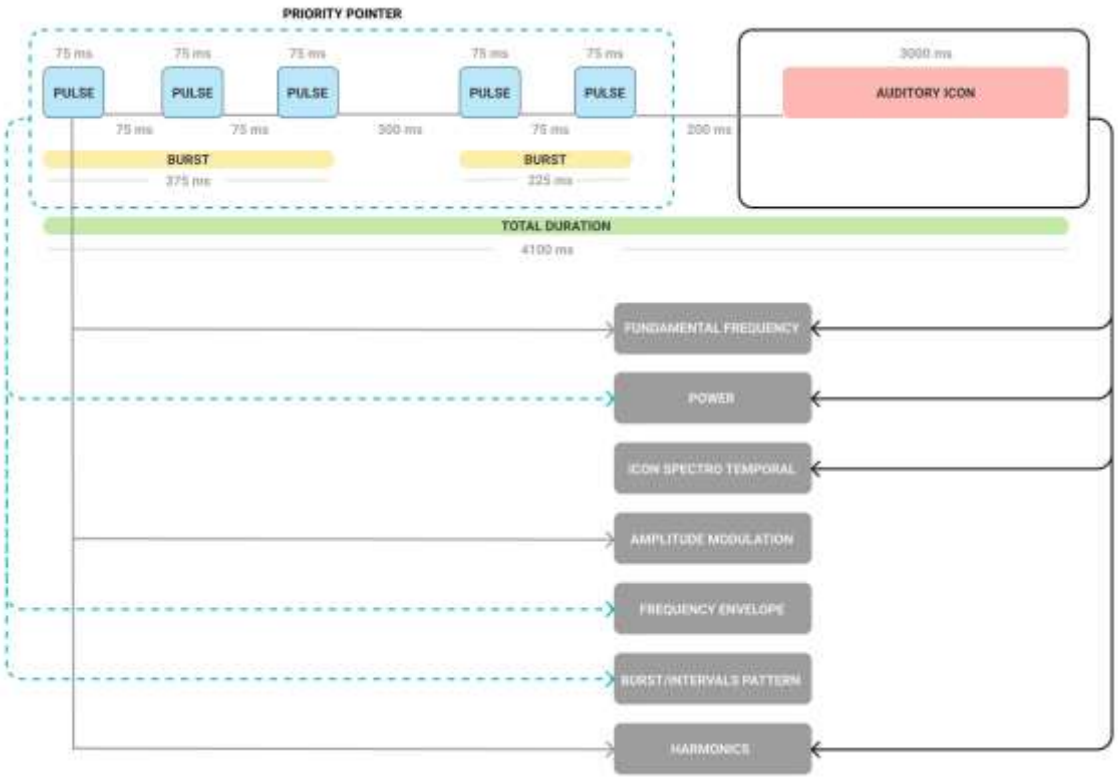


Figure 11-10. Anatomy of an Alarm Signal

The alarm is divided in two main parts: a priority pointer and an auditory icon. To optimize the information provided in relation to alarm time, they may be more or less overlapped.

The priority pointer contains information that should make the listener understand the priority of the alarm. The number of Bursts, number of Pulses, duration of the Pulses and the interval between Pulses will vary. Longer durations are related to a lower perception of urgency and shorter durations are related to a higher perception of urgency. The auditory icon represents the alarm category and is a synthesized sound representing a real event.

The manipulated variables go from the smallest unit, the Pulse, to the largest, the Auditory Icon. Within the Pulse, the designer can manipulate Fundamental Frequency, Amplitude Modulation and Harmonics. Within the Priority Pointer section, the Power, Frequency envelope and the Burst Intervals/Pattern can be edited. Finally, in the Auditory Icon, the Fundamental Frequency, Power, Spectral-Temporal characteristics, and Harmonics can be modified.

11.4 Test Alarm Set

For evaluation and validation of users, a test alarm set was generated following the established alarm sound design structure. Figure 11-11 represents the three main parts of the alarms: 1) auditory icon; 2) priority pointer and 3) tendency of the vital sign.

Although the combination of Low Priority + Auditory Icon is made available, IEC 60601-1-8 (2020) states its usage is optional as to reduce auditory burden on staff.

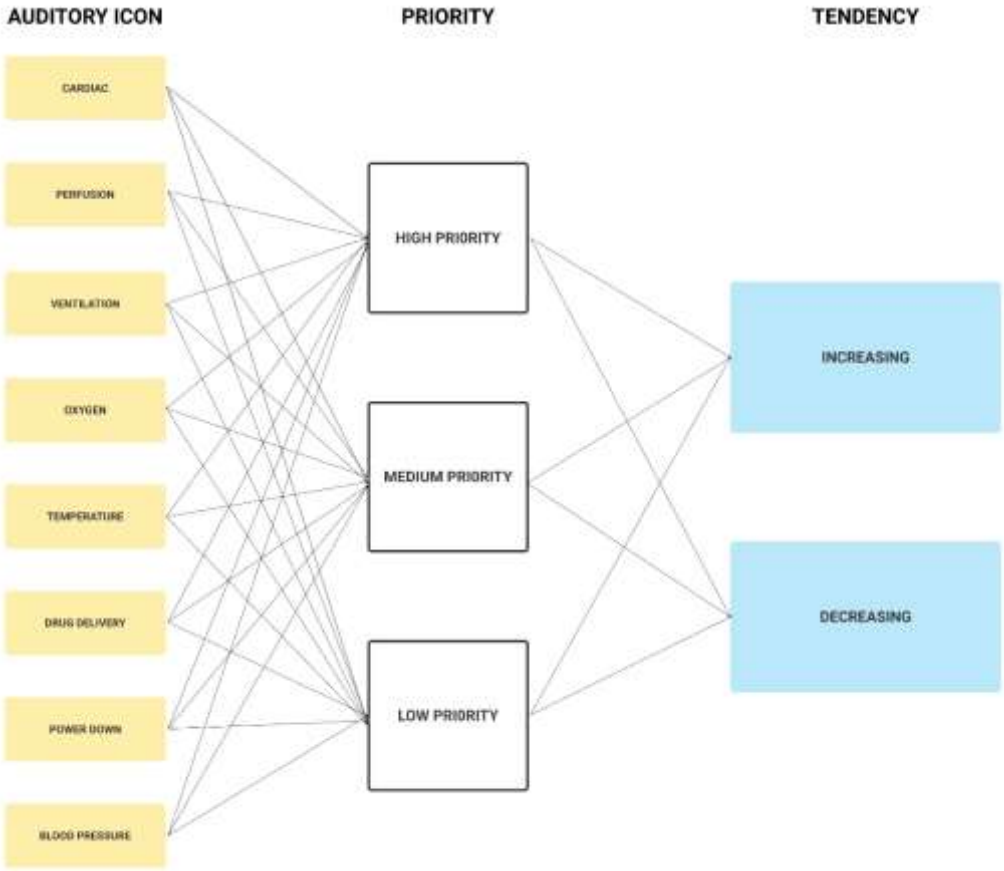


Figure 11-11. Modular structure of an alarm

All sounds of the test set were designed by sound designer Rui Marques, and can be heard on this link[21]. Further information regarding the alarm set can be found in Appendix 7.

[21] https://figshare.com/articles/media/Final_set_of_clinical_alarms_with_tendency_indicators_/16575002

11.5 Conclusion

A modular design system allows for scalability and adaptability. Our alarm structure allows the design of an entire family of alarms having the same structure and basis (the auditory icon) but with different priority and tendencies. The alarm philosophy of IEC 60601-1-8 allows the inclusion of more alarm causes if a manufacturer finds it necessary. Thus, it will be simple to create new alarms following this design guidance.

All previous work has led to the creation of a test set of alarms which refer to 8 categories. The following stages will include the evaluation of the alarms. The priority pointers will be evaluated for their perception of urgency, the comprehension and adequacy of the auditory icons will be assessed, and conclusions will be obtained for further refinement.

STEP 4. EVALUATE THE DESIGN

12 VALIDATION OF CLINICAL AUDITORY ALARMS WITH USERS: PERCEPTION OF URGENCY

The aim of this study was to validate the perception of urgency of a set of urgency pointers designed to precede the auditory icons. These pointers are 2-second sounds that considered all literature and design recommendations. Two levels of priority were chosen to be compared: Medium and High. With this study we intend to validate whether listeners can differentiate between both levels. The low priority urgency pointer was not included to keep the study short to be performed online. The chosen variables to differentiate priority levels were five: Fundamental Frequency, Speed, Rhythm, Inharmonicity and Pulse Length.

Library of Congress. (ca. 1921-1925), Ernest Hare, US star of Vaudeville, phonograph records, and radio, listens to the radio with headphones. No known copyright restrictions.

12.1 Introduction

The main goal of an auditory warning is to alert a given individual of an event in need of an intervention. Importantly, this warning should convey the degree of urgency with which this individual should respond, by communicating different priorities. This is particularly useful when several warnings are triggered at the same time (Hellier & Edworthy, 1989).

Alarm signals interrupt duties associated with the healthcare professionals' tasks. A high priority alarm signal requires professionals to immediately stop what they are doing to address the cause of the alarm condition. A medium priority alarm signal also interrupts duties but allows a minute or a few minutes for the operators to finish a task before addressing the cause of the alarm condition, or to find another person to address it. A low priority alarm signal should not interrupt professionals, and they should be able to address the cause of the alarm condition at a convenient time. However, this alarm should escalate to medium priority and then high priority if professionals fail to address the alarm condition (AAMI, 2013).

For the particular case of alarms in operating rooms, Bennett et al. (2011) evaluated four emotional descriptors that could be related with an auditory warning. These were:

1. Urgency – A sense of requiring immediate action.
2. Anxiety – A sense of apprehension due to uncertainty or doubt.
3. Attention – A sense of drawing your focus or observation.
4. Severity – A sense of harshness or intensity.

In their study, participants evaluated each emotional descriptor in a nine-point Likert-scale after hearing a sound. Results have shown that urgency was collinear with all other descriptors. Urgency can be described as what requires immediate action or attention (Suied et al., 2008). In its turn, perceived urgency may refer to the sense of urgency conveyed by the warning sound's parameters (Burt et al., 1995). Perceived urgency affects how quickly an individual will recognize and respond to an alarm, and people are quicker to respond to warnings sounding more urgent (Edworthy et al., 2000; Haas & Edworthy, 1996).

Perceived urgency is generally interpreted as the subjective response to a warning, and this match should be quantified and understood. Great efforts have been put into understanding the nature and power of spectral and temporal parameters on perceived urgency (Hellier & Edworthy, 1989). Other parameters have also been proposed, such as predictability. Unpredictable temporal or harmonic events can get more attention and

increase the perception of urgency (Edworthy et al., 1991). In general, the following modifications induce different degrees of urgency: the faster the rate of an alarm, the higher its pitch and the more irregular the harmonics, the greater the perceived urgency.

The effect of manipulating temporal parameters in reaction times was studied by Suied et al. (2008). In two visuomotor tracking tasks simulating driving conditions, the inter-onset interval (IOI, corresponding to the time elapsed from the onset of a pulse to the onset of the next) was varied, leading to two design guidelines for warning urgency: a) IOI can be used to modulate warnings' urgency and b) temporal irregularity of the IOIs can capture the listener's attention (Suied et al., 2008).

Other parameters analysed include pulse duration, interpulse interval, alert duty cycle and sound type (Marshall et al., 2007). Duty cycle represents the proportion of time the sound is present. This last study focused on in-vehicle warnings and studied the perception of annoyance in auditory warnings along with their perception of urgency. The referred parameters affected the perception of urgency substantially more than annoyance.

Of course, these relations are the result of laboratorial and controlled psychophysical experiments (Hellier et al., 1993). Other conditions, such has high workload, may greatly impair these impressions (Guillaume et al., 2002). In fact, in most realistic conditions, an urgent situation occurs under high workload (Suied et al., 2008). Another important observation is that most of these experiments used synthesized alarms, and most obtain results according to Edworthy et al.'s predictions in which higher values of frequency and speed increase the perception of urgency.

In 2012, Gonzalez et al. revisited Edworthy's studies and results. The premise was that on the twenty years that followed her seminal study, the general soundscape has increasingly complexified, which could have affected the perception of auditory parameters. Their results indicated that sensitivity to changing pulse rates has decreased with a general increase in exposure to technology. Pulse rate continues to be the most reliable and robust parameter to manipulate the perception of urgency, but increased levels of change in pulse rate may be needed to convey the same urgency as in the original study.

Guillaume et al. (2003) have designed a framework for the perception of urgency based on early studies by Edworthy and Hellier. The prevalent approach considered that the perceived degree of urgency could be predicted from a detailed analysis of acoustic characteristics, especially pitch and rate. Guillaume et al. (2003) have confirmed the role of these characteristics, particularly melodic and tone structure. However, the authors' interpretation is that such a simplistic explanation may not "*convey the whole picture*"

(pp.210). They suggest that the alarms used in their study (for instance, one resembled French police siren) evoke a mental representation regarding the source of the sound. This representation is judged as having a high or a low degree of urgency, and not the sound. The authors suggest "*an alternative representational processing mode short circuits the previously described exhaustive processing mode: It would be highly automatic, require little attention, be considerably faster (because identification is immediate and does not require a continuous monitoring of the sequence), and lead to immediate, appropriate behavioral responses*" (pp.211). With this shortcut in consideration, only a minimal analysis of acoustic parameters is necessary to evoke a mental representation. When this representation does not exist, the exhaustive processing mode would take place, leading to a slower and less appropriate reaction, because it would not be linked to the cause from the alarm, requiring more attentional resources (Guillaume et al., 2003). The ecological perception of alarms assumes a faster interpretation of an existent mental representation, when compared to an unfamiliar sound which needs to be thoroughly processed.

12.1.1 Urgency Mapping

One of the first identified problems concerning auditory alarms in medical equipment was the lack of urgency mapping between the acoustic properties of the sounds and the real medical urgency of the cause of the alarm (Finley & Cohen, 1991; Momtahan et al., 1993; Momtahan, 1991). Correctly mapping the perceived urgency of a signal with the hazard level is a very important aspect of alarm design. If not, alarms can be too intense, abrasive, startling, they can annoy or distract providing no performance benefit, and reduce trust in the system (Baldwin & Lewis, 2014).

To proceed with this mapping, it is necessary to understand the perceived urgency of different sounds and what variables affect it. Once this is known, systematic variations of certain parameters will have predictable effects on the perception of urgency (Hellier & Edworthy, 1999).

Some guidelines were already validated on numerous studies:

- As the fundamental frequency of a sound increases, it is perceived as increasingly urgent (Edworthy et al., 1991; Gonzalez et al., 2012; Hellier & Edworthy, 1999; Hellier et al., 1993)
- As the time interval between pulses of sounds decreases, it is perceived as increasingly urgent (Edworthy et al., 1991; Hellier & Edworthy, 1999; Hellier et al., 1993,)

These guidelines are important because they indicate ordinal information. However, an immediate follow-up question should be "by how much"? "*How much does pitch have to change to double perceived urgency? How much does speed have to change to double perceived urgency?*" (Hellier & Edworthy, 1999, p.168).

12.1.1.1 Steven's Power Law

The power of the relationship between an objective and a subjective change can be described within Steven's Power Law (Hellier & Edworthy, 1999; Hellier et al., 1993; Hellier & Edworthy, 1989):

$$S = kO^m$$

S refers to the subjective value, *O* is the objective value, and *k* and *m* are the intercept and the slope of the line of best fit when the logarithmic values of objective and subjective values are plotted as xy coordinates (Hellier & Edworthy, 1989). For the case of urgency perception, *m* refers to the ability of an objective parameter to affect it. The steeper the slope, the more effective is the parameter.

Steven's Power Law can be used as a tool in urgency mapping. Participants listen to sounds varying in physical parameters and judge the urgency of the sounds. Power functions allow comparisons to be made between different warning parameters and establish which affect the perception of urgency the most. A parameter with a large exponent would require a small change to cause a unit change in perceived urgency, and vice versa. Knowing this allows a designer to select the most economical method and to change the perception of urgency without changing much of its structure, maintaining the sounds qualitatively similar (Hellier & Edworthy, 1999).

Table 12-1. Power function exponents for five warning parameters according to Hellier and Edworthy (1999)

WARNING PARAMETER	EXPONENT
Pitch	0.38
Speed	1.35
Repetition	0.50
Inharmonicity	0.12
Length	0.49

The largest exponent represented in Table 12-1, after studies by Hellier and Edworthy (1999) is for speed, which indicates that changes in speed have the most powerful

relationship with changes in perceived urgency. Small changes in the speed of a warning sound are necessary to produce a unit change in urgency perception.

The same authors also predict that the effects of the different warning parameters on the perception of urgency would be additive, and that stimuli with the same combination of levels in their parameters would be judged as equally urgent.

12.2 Aim

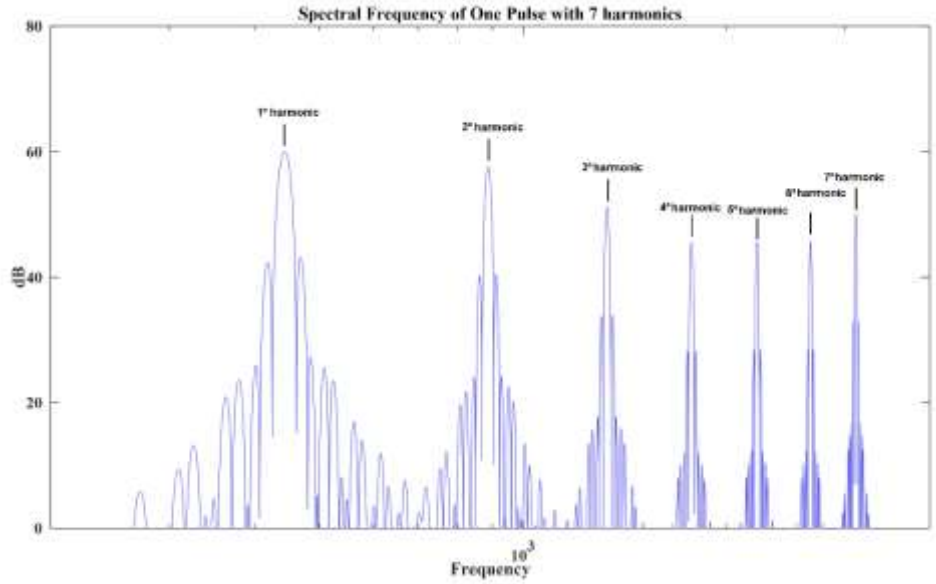
The aim of this study was to validate the perception of urgency of a set of urgency pointers designed to precede the auditory icons. These pointers are 2-second sounds that considered all literature and design recommendations. Two levels of priority were chosen to be compared: Medium and High. With this study we intend to validate whether listeners can differentiate between both levels. The low priority urgency pointer was not included to keep the study short to be performed online. The chosen variables to differentiate priority levels were five: Fundamental Frequency, Speed, Rhythm, Inharmonicity and Pulse Length.

Fundamental Frequency and Speed are established predictors for the perception of urgency. Rhythm, Inharmonicity and Pulse Length are temporal and spectral variables whose effect on urgency is either understudied or inconsistent. Table 12-2 presents the definition of each of the selected independent variables.

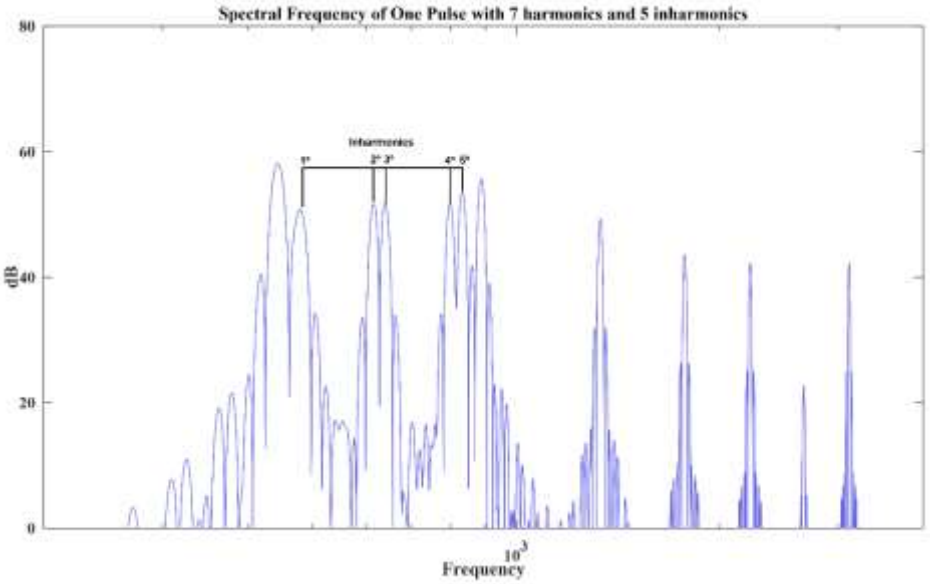
Table 12-2. Independent Variables of the current study

Fundamental Frequency	Lowest frequency of a periodic waveform
Speed (no. Pulses)	Determined by the inter-pulse interval. In this study, it also affected the number of times a small unit of sound repeats
Rhythm	Regular occurrence of an auditory event in time.
Inharmonicity (Figure 12-1)	Number of harmonics in the pulse which are not integer multiples of the fundamental frequency
Pulse Length	Duration of a pulse

Figure 12-1 a) shows an example of a pulse with a spectral content of seven harmonics and Figure 12-1 b) shows a pulse with the same seven harmonics plus five inharmonics between two harmonics, which are not integer multiples of the fundamental frequency.



A)



B)

Figure 12-1. Example of a pulse with a) seven harmonics and b) seven harmonics and five inharmonics

12.3 Method

Selecting the appropriate method to make evaluations is a difficult compromise between its accuracy and the participant's engagement with the task. While as researchers we would like to select a method as precise as possible, that may come at the risk of losing all participants on the course of the experiment, as that usually means longer and more repetitive evaluation tasks.

Parizet et al. (2005) evaluated the pleasantness of in-car ventilation noises using six different methods (Figure 12-2). In T2, called mixed evaluation of comparison technique (Chevret & Parizet, 2007), all sounds are evaluated on the same scale. The difference is that all sounds are presented in the same page. Beside each scale is the audio file with the corresponding sound. In this way, the listener can compare the different stimuli before giving an answer for each. This technique led to shorter experiments with 9 sounds (6 minutes) when compared to paired comparison (15 minutes), reducing the risk of fatigue (Hellier & Edworthy, 1989) and there was a reduced risk of habituation to the sounds. However, confidence intervals are greater than those obtained in the paired comparison.

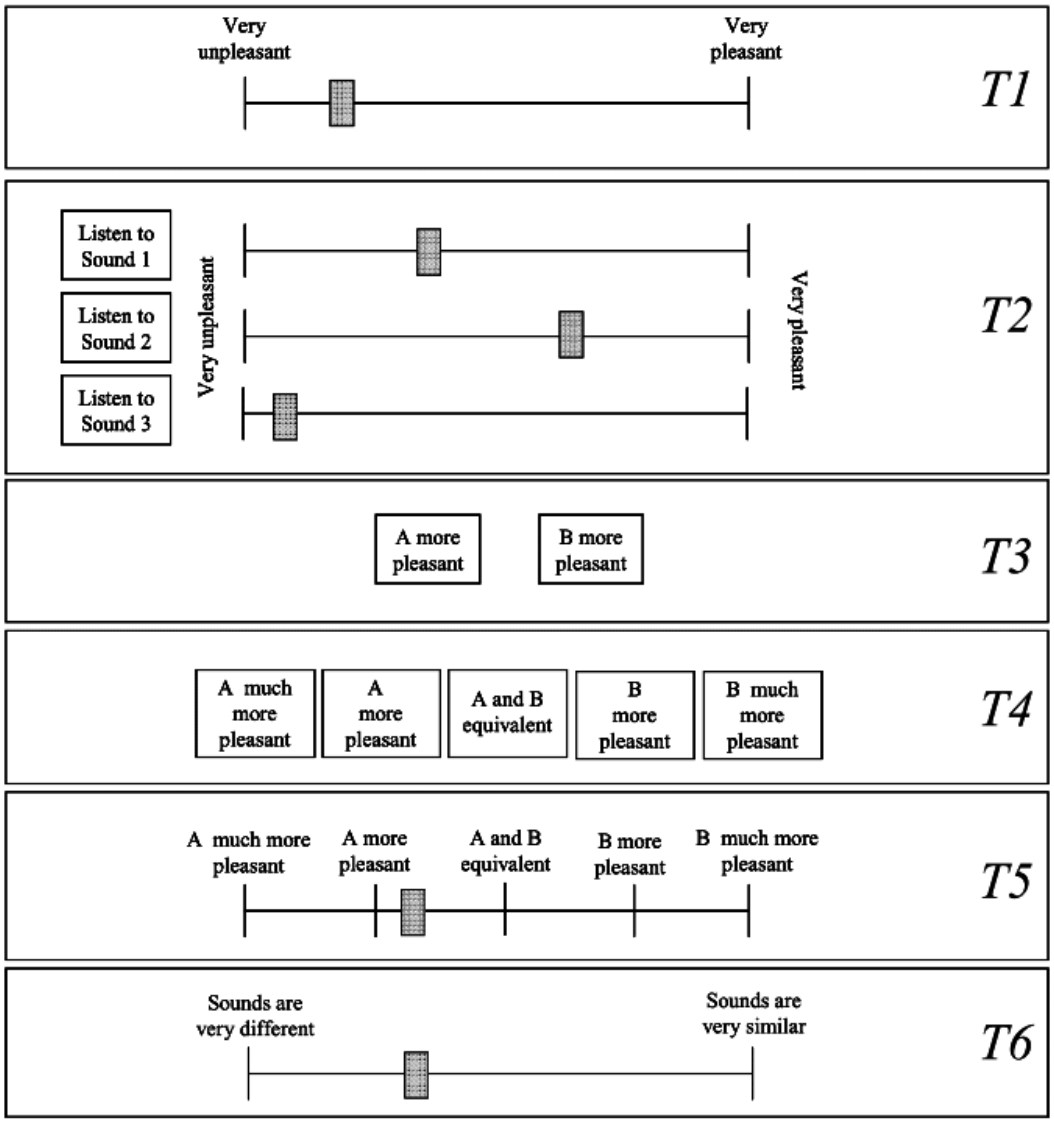


Figure 12-2. Answering scales of the six listening tests (T1 to T6) (Parizet et al., 2005)

As recommendation, the authors suggest this technique for the evaluation of sound pleasantness as it offers a good compromise between the accuracy of the results and the time needed by each subject to complete the test.

The present study for the evaluation of Perceived Urgency used the mixed evaluation/comparison technique as proposed by Parizet et al., (2005). Due to COVID-19 constraints this study was performed online using Qualtrics platform (Qualtrics, 2021). It followed a between subject experimental design with five two-level independent variables (Medium priority and High priority for Fundamental Frequency, Speed, Rhythm, Inharmonicity and Pulse Length).

12.3.1 Stimuli

Thirty-two sounds were designed with two levels of five independent variables: Fundamental Frequency, Speed, Rhythm, Inharmonicity and Pulse Length (2 x 2 x 2 x 2 x 2 = 32). Reference values were obtained from Hellier & Edworthy (1999) using Steven's Power Law. The considered values of k ($S = kO^m$) for each level of priority are in Table 12-3. Sounds can be heard here[22].

Table 12-3. k values used as reference for the creation of the two levels priorities

	MEDIUM PRIORITY	HIGH PRIORITY
Fundamental Frequency	150 Hz	243 Hz
Speed	1,66	2
Rhythm	Regular	Syncopated
Inharmonicity	0,95	1,58
Pulse Length	150 ms (pulse) + 50ms(IPI)	75 ms (pulse) + 25ms(IPI)

The envelope shape of the pulses was constant over time, with a 20 ms onset and offset. The synthesis model used for the harmonic constitution was the additive synthesis with sinusoidal waves, with a total of 7 whole harmonics and a variation of 1 to 5 inharmonics depending on the k values. Regarding inharmonics, the values ranged between f1*25/24 (major 2nd), f1 * 25/18 (augmented 4th), f1 * 36/25 (diminish 5th), f1 * 9/5 (minor 7th), f1 * 15/8 (major 7th). These intervals were chosen because they are the intervals with greater

[22] Vieira, Joana (2021): Validation of clinical auditory alarms with users: Perception of Urgency. figshare. Media. https://figshare.com/articles/media/Validation_of_clinical_auditory_alarms_with_users_Perception_of_Urgency/14905512

dissonance in relation to the fundamental. More details about the stimuli are presented in Appendix 6.

12.3.2 Procedure

The study was conducted via the online survey platform Qualtrics (2021). The survey was distributed among personal contacts, acoustics-related newsletters, and social media. After a period, the survey was distributed using Mechanical Turk, (www.mturk.com). Mechanical Turk has measures to verify if respondents are consistent, providing them with a quality ranking. Before the main task, participants performed a listening screening test (cf. 10.3.1 - Listening Screening Test).

The main instructions to the Evaluation of Urgency task stated: "*You will now proceed to the evaluation of urgency of the sounds. Imagine the most urgent sound possible. Let this represent "Very Urgent" (100) on an urgency rating scale. Let "Not Urgent" (0) represent the other end of the continuum (the most non-urgent sound that you can imagine). You are required to assign a number between 0 and 100 to each of the following sounds*".

Each evaluation page had 7 sounds presented simultaneously. The platform randomly selected seven sounds from the pool of 32 sounds. Participants then evaluated the same seven sounds 4 times on continuous scales with Not Urgent 0 and Very Urgent 100 on each extremity (Figure 12-3). In each scale, participants could move the pointer and the respective value would temporarily show.

At every page, instructions referred: "*You will have to listen to 2-seconds sounds and evaluate their urgency according to your perception. Urgency is what requires immediate action or attention. You can adjust your evaluation at any time*" (Figure 12-3.

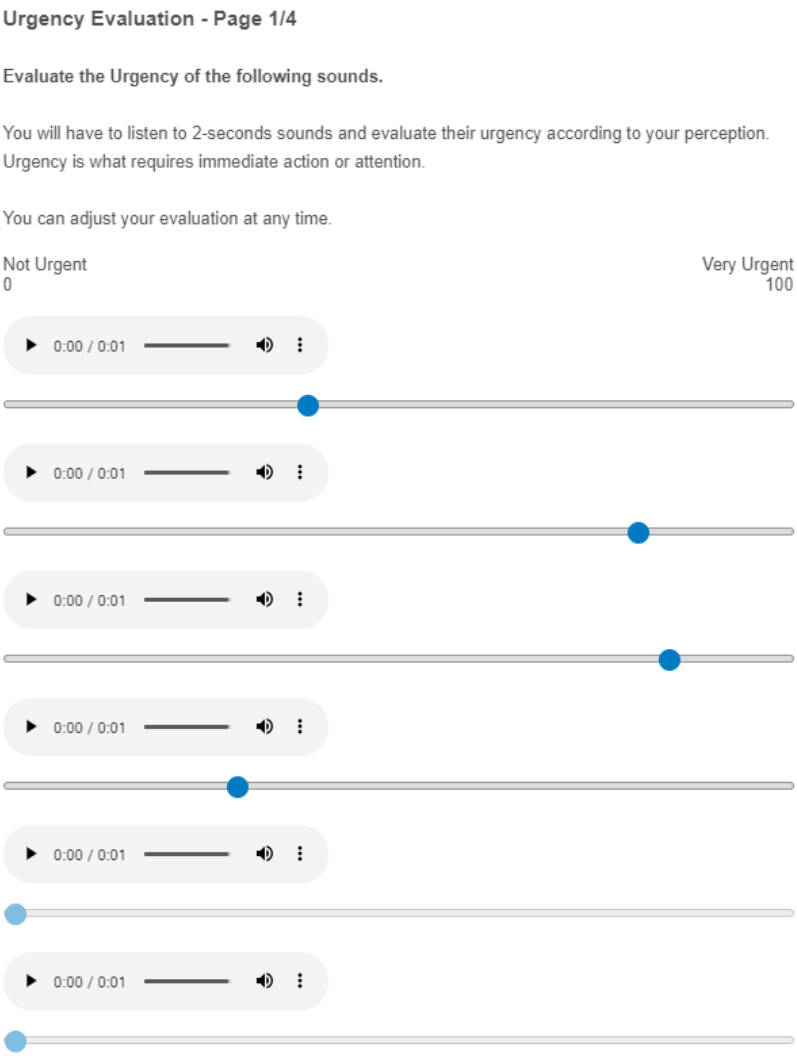


Figure 12-3. Interface for the Evaluation of Urgency of sounds

After the evaluation of urgency, participants were invited to provide demographic information and the experiment concluded. It took on average 11 minutes to complete.

12.3.3 Results

12.3.3.1 Participants

325 participants took part in the experiment and correctly performed the screening test. 22 participants were eliminated for not having submitted a total of 28 evaluations, corresponding to four evaluations of seven sounds. The final sample was composed of 303 participants. 65% were male and 35% were female participants. 22% had musical training. Further characterization of the participants is presented in Table 12-4.

Table 12-4. Characterization of the participants

AGE	
< 18 years	3%
18-29 years	28%
30-39 years	40%
40-49 years	21%
50-59 years	7%
60-69 years	1%

COUNTRY OF ORIGIN	
United States	56%
Portugal	17%
India	16%
Brazil	4%
Italy	3%
Others	4%

EDUCATION	
Primary/Elementary school	1%
Secondary/Technical/Business School	9%
Bachelor's degree	55%
Master's degree	33%
PhD	2%

Each of the 32 sounds was evaluated, on average, 67 times, and each participant made 28 evaluations.

12.3.3.2 Inter-participants Concordance

To understand if all participants rated sounds in similar order, we ran Kendall's W, also known as Kendall's coefficient of concordance. It is a non-parametric statistic and can be used for assessing agreement among raters. Kendall's W ranges from 0 (no agreement) to 1 (complete agreement). The value of Kendall's W was calculated for ratings obtained through Mechanical Turk (MT) and through other sources (non-MT) - Table 12-5.

Table 12-5. Kendall's W values for the MT evaluations, non-MT Evaluations, and all evaluations together

MT (N=234)	NON-MT (69)	TOGETHER (303)
0.38***	0.39***	0.38***

*** Significant ($p < .001$)

All tests revealed a significant value of Kendall's W, meaning that there were no random rankings. Kendall's W provides the agreement in the order of an evaluation, for instance, a higher frequency sound was consistently evaluated as more urgent than a lower frequency sound.

Even though the Kendall's W value was not particularly high, it was very similar for both samples, and they were grouped for further analysis.

Another important step was to observe whether participants used the 100mm scale on the same range. For that, minimum values were subtracted to the maximum value of each

participant to obtain their evaluation range. These ranged from using the entire 100 points to using 12. Figure 12-4 presents a histogram which represents the use of the 100mm scale. Most participants used between 34 to 44 points of the scale to perform their evaluations.

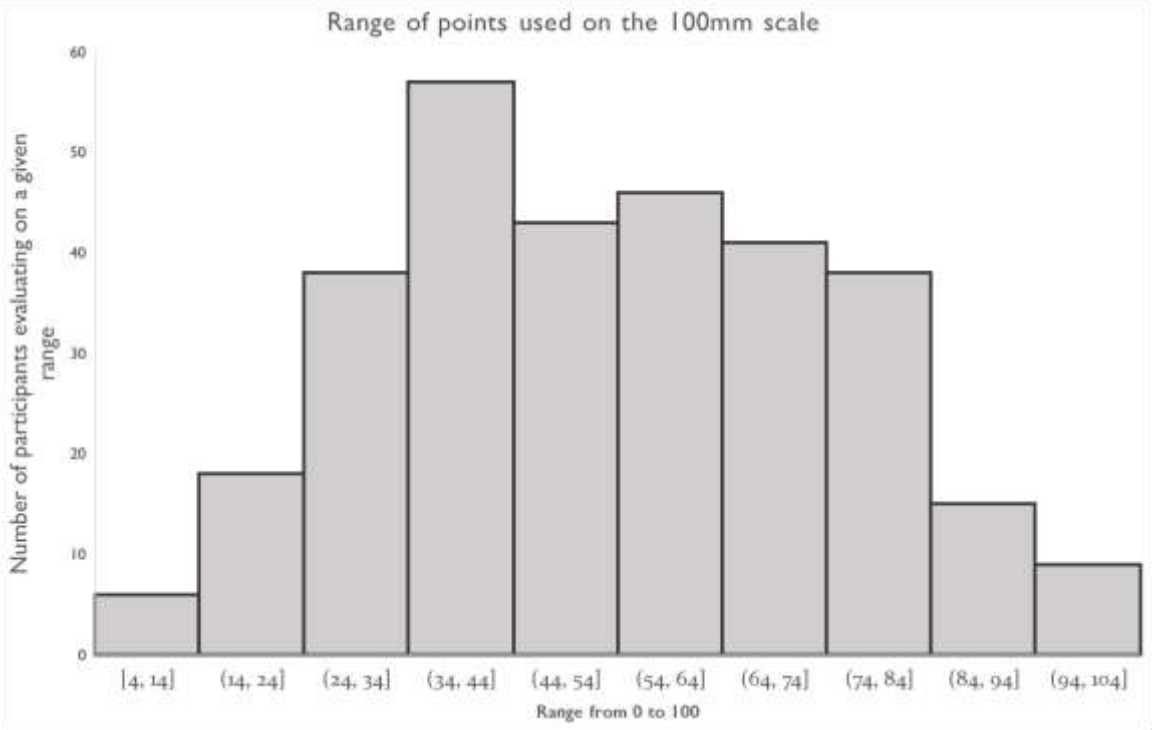


Figure 12-4. Histogram representing the range of the 100mm scale used by participants. Each bin corresponds to 10 points of the scale.

To reduce this variability, all evaluations were transformed from 0-100 to 0-1 evaluations. All 4 rounds of evaluation of the same 7 sounds were averaged into a single number. The form (MEAN-MIN)/RANGE was applied to all mean evaluations of all participants to convert from 0-100 numbers into 0-1 decimals, where 1 would correspond to the highest value used by a given participant, and 0 would correspond to the lowest.

12.3.3.3 Differences in Perception of Urgency

The distribution of the data was not significantly different from normal distribution. Assuming the normality of the distribution, we performed t-tests per each independent variable to verify if there were differences in the evaluation of perceived urgency between the hypothesized medium and high priority sounds.

Participants evaluated sounds with a fundamental frequency of 441 Hz (M=0,53, SD=0,04) as more urgent than sounds with 295 Hz (M=0,43, SD= 0,04), ($t(602)=-7{,}37$, $p<0{,}001$) - Figure 12-5.

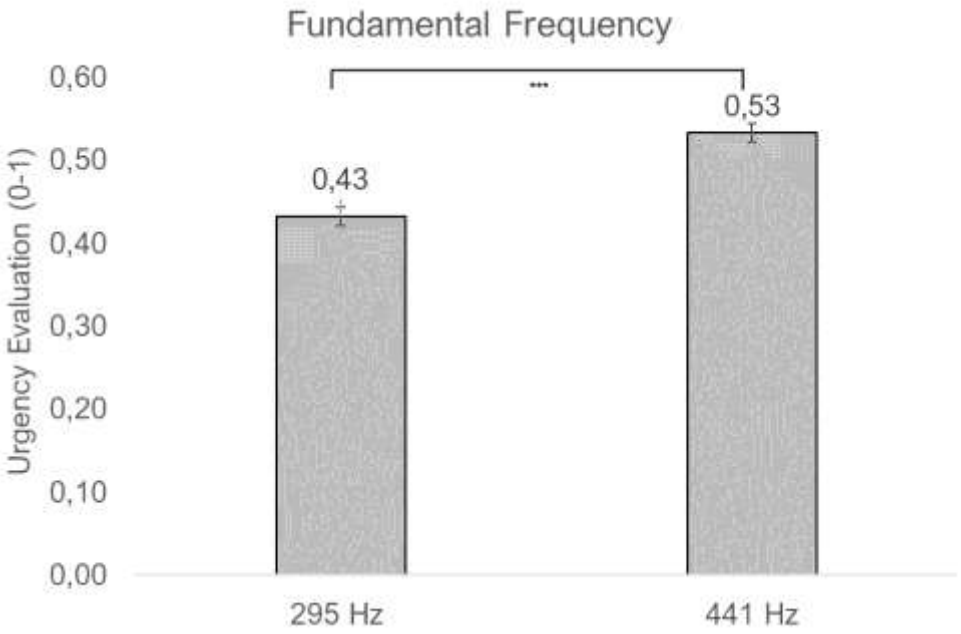


Figure 12-5. Differences in Perception of Urgency between the two levels of the variable Fundamental Frequency.

Participants evaluated sounds with a regular rhythm (M=0,50, SD=0,03) as more urgent than sounds with a syncopated rhythm (M=0,46, SD= 0,04), (t(602)=3,36, p<0,001) - Figure 12-6.

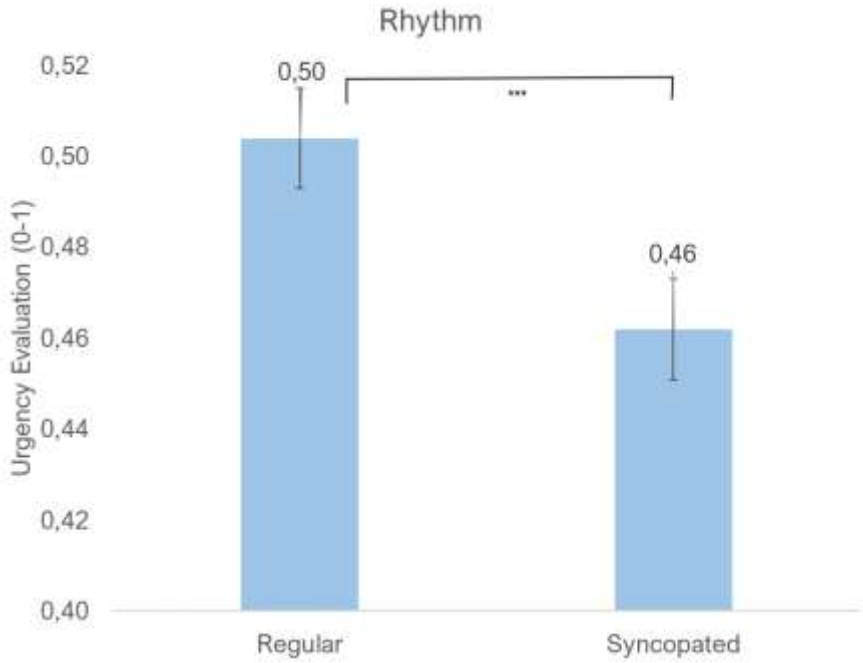


Figure 12-6. Differences in Perception of Urgency between the two levels of the variable Rhythm.

Participants evaluated sounds with varying levels of inharmonicity as equally urgent (M=0,48, SD= 0,03), (t(604)=-0,23, *p*>0,5) -Figure 12-7.

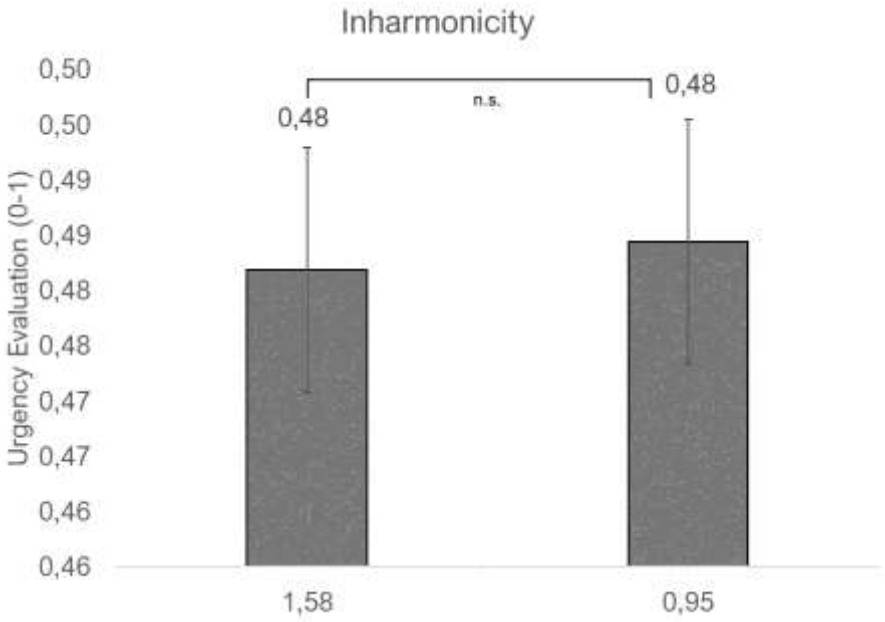


Figure 12-7. Differences in Perception of Urgency between the two levels of the variable Inharmonicity.

Participants evaluated sounds with a shorter pulse length (M=0,60, SD=0,05) as more urgent than sounds with a longer pulse length (M=0,36, SD= 0,06), (t(602)=13,25, *p*>0,001) - Figure 12-8.

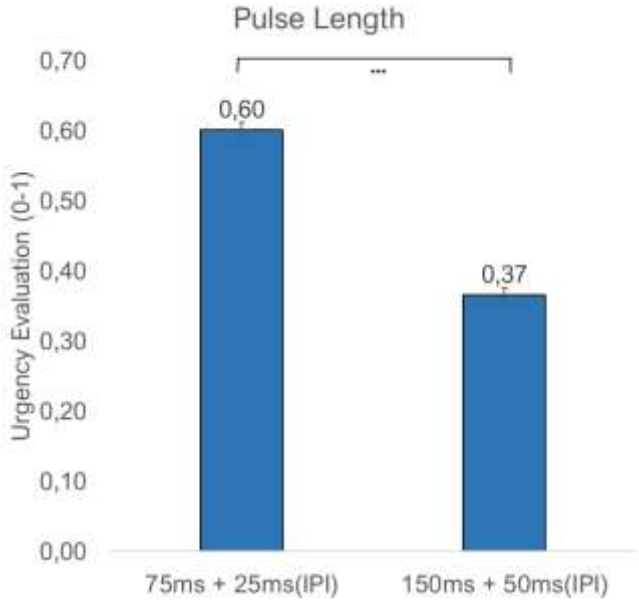


Figure 12-8. Differences in Perception of Urgency between the two levels of the variable Pulse Length.

For the No. of Pulses independent variable, we ran an analysis of variance. There was a significant effect of No. of Pulses on the perception of urgency at the p<.001 level F (3, 1208) = 81,50, p = 0.000. Post hoc comparisons using the Tukey HSD test indicated that the mean score for 3-pulse sound (M = 0,36, SD = 0,07) was significantly different than the 5-pulse sound (M=0,58, SD= 0,08) and the 6-pulse sound (M=0,62, SD=0,07). The mean score for a 4-pulse sound (M = 0,37, SD = 0,07) was significantly different than the 5-pulse sound (M=0,58, SD= 0,08) and the 6-pulse sound (M=0,62, SD=0,07). However, the 3 and 4 pulse sounds, and 5 and 6 pulse sounds did not significantly differ from each other (Figure 12-9).

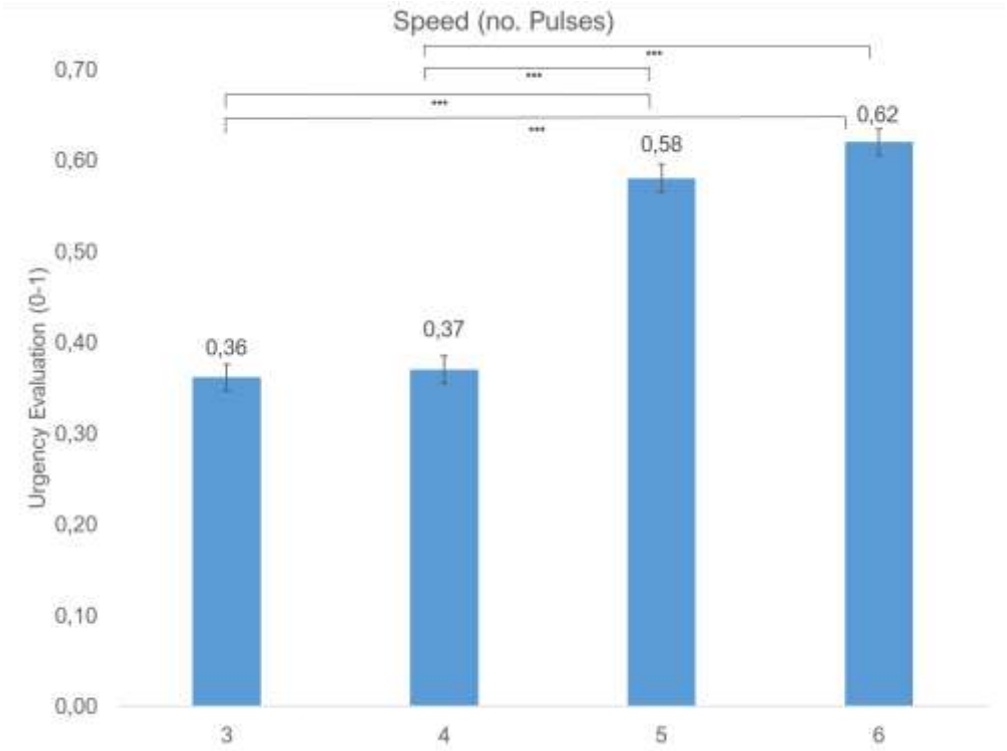


Figure 12-9. Differences in Perception of Urgency between the two levels of the variable Speed (No. Pulses).

To understand the contribution of each independent variable on the general variance of the perception of urgency, we ran a multiple regression considering all categorical independent variables with 2-levels which showed significant differences between both levels: frequency, rhythm, and pulse length.

We did a hierarchical style of regression, in which predictors are selected based on previous knowledge with the most important contributors in predicting the outcome added first. To obtain this order, simple linear regressions were performed for all variables.

- Length of Pulse explained a significant amount of this variance, F(1, 31)= 82,39, p= 0.00, R^2=0,72;
- Speed or No. of Pulses explained a significant amount of this variance, F(1, 31)= 26,78, p= 0.00, R^2=0,71;
- Fundamental frequency did not explain a significant amount of the variance in perception of urgency, F(1, 31)= 4,39, p= 0.05, R^2=0,10;
- Rhythm did not explain a significant amount of the variance in perception of urgency, F(1, 31)= 1,08, p= 0.30, R^2=0,003;

The Speed or No. of Pulses variable had 4 levels, and for that reason it was not included in the multiple regression model (all other variables have 2 levels)[23]. The order of variables included was as reported in Table 12-6.

Table 12-6. Results of a hierarchical multiple regression

	R^2	B	SE B	B	P
Length of Pulse	0.72	0.37	0.02	0.86	***
Length of Pulse + Frequency	0.85	0.32	0.02	52.18	***
Length of Pulse + Frequency + Rhythm	0.88	0.39	0.03	-0.19	**

Multiple linear regression was carried out to investigate the relationship between Length of Pulse, Fundamental Frequency and Rhythm in the perception of Urgency. There was a significant relationship between perception of urgency and Length of Pulse ($p < 0.001$), Fundamental Frequency ($p < 0.001$) and rhythm ($p < 0.01$). Perception of urgency increased with shorter pulses with higher frequencies. This perception reduced when the rhythm changed from regular to syncopated. The adjusted R^2 was 0.88 so 88% of the variation the perception of urgency can be explained by the model containing length of pulse, fundamental frequency, and rhythm.

The two variables explaining most variance are depicted in Figure 12-10.

[23] A hierarchical multiple regression with the extreme values of the Speed or No. of Pulses variable (3 and 6 pulses) resulted in exact collinearity with Length of Pulse variable and was excluded from this exercise.

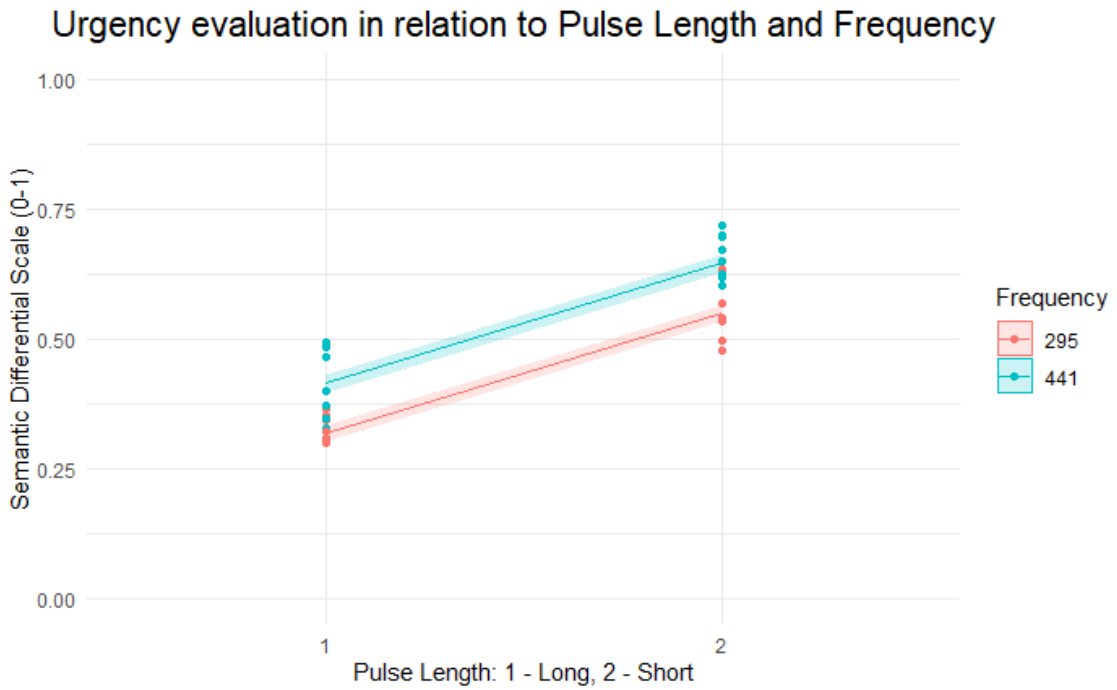


Figure 12-10. Relation between the variables Pulse Length and Frequency in the Perception of Urgency

The Perception of Urgency increased as the sounds became faster and higher pitched. Additionally, the rhythm also played a part with more urgency being perceived from regular sounds, as opposed to syncopated sounds.

12.4 Discussion

This study allowed to validate our hypothesis in several fronts. Having previously designed 32 of 2-seconds sounds varying in two levels of priority (medium and high) based on literature and previous research, it was our intention to validate whether listeners were able to make this distinction. The short answer is that yes, there were distinctions between two levels of perceived urgency. The longer answer is that not all variables contributed to this difference as expected.

As observed in the chapter "Subjective Perception of Auditory Warning signals (Lab vs. Web)", temporal variables contributed the most to this distinction. First, Pulse Length was the main predictor of variance in the perception of urgency, followed by the Number of Pulses. The two variables are related, because shorter pulses were repeated more to occupy the same duration of a sound. As such, sounds with six 75 ms pulse and 25 ms inter-pulse intervals (IPI) were evaluated as more urgent then sounds with three 150 ms pulses and 50ms IPIs. As confirmed by Gonzalez et al. (2012) it appears that pulse rate continues to be a very reliable and robust parameter to manipulate the perception of urgency. This was

already observed in a number of previous studies (Edworthy et al., 1991; E. Hellier & Edworthy, 1999; E. J. Hellier et al., 1993), nevertheless, for our goal it was important to understand how each independent variable related with each other and in which order of importance.

A last temporal variable consisted of Rhythm, which again strikes us with its inconsistency. While in our previous study (Chapter 9) increased evaluations of urgency were observed with syncopated sounds, the reverse happed in the current study. This inconsistency is also stated in the literature. According to Judy Edworthy et al. (1991), unpredictable temporal events can get more attention and increase the perception of urgency. According to Suied et al. (2008), syncopated rhythm is perceived as slightly less urgent than a regular one. The latter statement matched exactly our observations in the present study. The comparison of rhythms is quite complicated. While the regularity is easy to comprehend, syncopated or irregular rhythms are difficult to compare, for the possibilities of variation are numerous. In our case, a previous validation of syncopation should have been performed.

Regarding spectral parameters, the paper of fundamental frequency was again observed as expected. Sounds with higher fundamental frequency were evaluated as more urgent. According to numerous research, as the fundamental frequency of a sound increases, it is perceived as increasingly urgent (Edworthy et al., 1991; Gonzalez et al., 2012; Hellier & Edworthy, 1999; Hellier et al., 1993).

Inharmonicity was another variable which we intended to validate as affecting the perception of urgency. No relation was observed between the two levels of inharmonicity we have introduced in our sounds. This goes against previous research stating that the more irregular the harmonics, the greater the perceived urgency (Edworthy et al., 1991; Guillaume et al., 2003). It could be that the two levels of inharmonicity used in our study were not different enough. It was observed in previous research that changes in inharmonicity have the least powerful relationship with changes in perceived urgency (Hellier & Edworthy, 1999), and in our case, it could be that we should have set them more apart to be more easily distinguished, although values were based on the power function exponents of Table 12-1.

All variables except Inharmonicity contributed to the Perception of Urgency. The variable which affected perception of urgency the most was Pulse Length, followed by Speed (No. Pulses), Frequency and Rhythm.

12.5 Conclusion

An online study was performed to evaluate the Perception of Urgency of a set of 32 sounds. Its goal was to validate the two levels of priority (Medium and High) embedded in the 2-second sounds. The study had five independent variables, which were Fundamental Frequency, Speed (no. Pulses), Rhythm, Inharmonicity and Pulse Length. The experiment consisted in evaluating seven random sounds in a 0-100mm continuous scale, ranging from not urgent to extremely urgent. The final sample was composed of 303 participants, the majority of whom came from the Mechanical Turk platform. After evaluating the sample's consistency, all evaluations were pooled and averaged.

Significant differences were observed between the two levels in the variables of Fundamental Frequency, Rhythm, Pulse Length and Speed (no. Pulses). No difference between the two levels were observed in the Inharmonicity variable. The variables which affected the Perception of Urgency the most were Pulse Length, Speed (no. Pulses) and Frequency.

What this means in practical terms is that we can select these three variables to design the priority pointers of our own set of alarms, for it was validated that 303 participants were able to distinguish between these two levels.

13 Validation of clinical auditory alarms with users: Comprehensibility and referent association

This chapter describes the last study validating the comprehensibility of the designed auditory icons. It had three main aims: 1) Test the comprehensibility of the designed auditory icons. The comprehensibility provides a measure of the extent to which an auditory icon communicates its intended message; 2) Test the referent association of the designed auditory icons; 3) Evaluate the general adequacy of the designed auditory icons for their referents. A final set of auditory icons is proposed.

Keene Public Library and the Historical Society of Cheshire County. (1900-1920?), Nurses of the Elliot City Hospital in Keene New Hampshire. No known copyright restrictions

13.1 Introduction

Visual icons and symbols have a long history of passionate opinions and controversy (Engber, 2014). The standardized methodology that has been proposed for the design of visual icons assures that the icon represents correctly its referent, and that the majority of the population that will be exposed to it understands it (ISO 9186-1, 2014a, ISO 9186-32014b). The same does not occur for auditory icons. The examples of what is a good auditory icon are the same for decades – the sound of crushing paper to represent the trashcan in Windows operating systems (W. Buxton et al., 1994; Graham, 1999), and there is not an established methodology to validate them.

The early 1990's saw the first use of auditory icons as warnings, and a natural question regarded their design method. Mynatt (1994) first proposed a six-step method consisting of: 1) choosing sounds; 2) evaluate their identifiability using free-form answers; 3) evaluate their learnability; 4) test possible conceptual mappings for the sounds; 5) evaluate masking, discriminability and conflicting mapping of the auditory icons and 6) conduct usability experiments with interfaces using the auditory icons. Shortly after, Edworthy and Stanton (1995) proposed the method already described in Chapter 4. This proposal was strongly based on the current standard for developing and evaluating public information symbols.

ISO 9186-1/2014 - Graphical symbols - Test methods - Part 1: Method for testing comprehensibility (ISO, 2014a) presents a general method for the analysis of any type of icon, and it assumes several variants for the same referent which will be used and compared. Another standard used for the validation of icons is SAE J2830 - Process for Comprehension Testing of In-Vehicle Icons (2008). However, this one was elaborated specifically for the automotive context. The SAE standard has a very rich and detailed evaluation methodology, with cut-off values and a finer scale to characterize the population's answers.

To evaluate auditory icons, all these references were considered appropriate, with due adaptation to the concerned modality. To achieve the aims of this study, a compromise between ISO 9186-1/2014, ISO 9186-3/2014 and SAE J2830/07-2008 was put into practise (Figure 13-1).

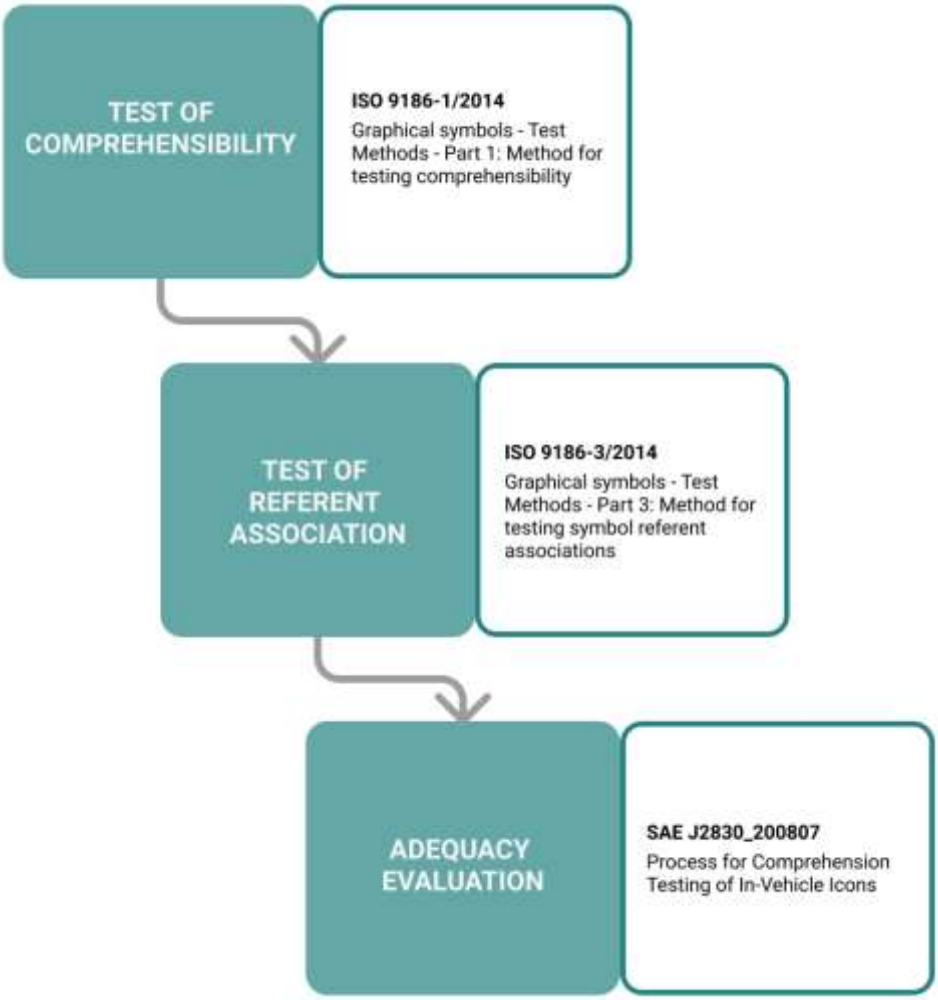


Figure 13-1. The three stages of evaluation of the designed auditory icons with reference to the source of each method

This study had three main aims: 1) Test the **comprehensibility** of the designed auditory icons. The comprehensibility provides a measure of the extent to which an auditory icon communicates its intended message; 2) Test the **referent association** of the designed auditory icons; 3) Evaluate the general **adequacy** of the designed auditory icons for their referents.

13.2 Method

13.2.1 Participants

A total of 102 participants took part in the current online experiment. 17% were healthcare workers and 83% were non-healthcare workers.

Table 13-1 presents further demographic details.

Table 13-1. Demographic description of participants

HEALTHCARE WORKERS			
Gender		**Profession**	
Female	82 %	Nurse	31 %
Male	18 %	Assistant	13 %
Age		Anesthesiologist	6 %
18-29	53 %	Surgeon	6 %
30-39	35 %	Hospital Pharmacist	6 %
40-75	12 %	Medical Doctor	6 %
Musical training		Surgical Tech.	6 %
Yes	35 %	Support Technician	6 %
No	65 %	Student	6 %
Location		Other	13 %
Portugal	67 %		
USA	7 %		
Spain	27 %		
NON-HEALTCARE WORKERS			
Gender		**Profession**	
Female	72 %	Student	48 %
Male	27 %	Researcher	18 %
Other	1 %	Musician	5 %
Age		UX Designer	5 %
18-29	67 %	Unemployed	5 %
30-39	18 %	Psychologist	3 %
40-75	15 %	Architect	2 %
Musical training		Geriatric assistant	2 %
Yes	37 %	Cognitive Scientist	2 %
No	63 %	Design Thinking	2 %
Location		Developer	2 %
Portugal	74 %	Electronic Engineer	2 %
UK	4 %	Restoration	2 %
Brazil	3 %	IT and Communication	2 %
Canada	3 %	Theoretical scientist	2 %
Switzerland	3 %	Salesperson	2 %
USA	3 %	IT in Health	2 %
Others	11 %		
Experience with medical devices			
Yes	12 %		
No	88 %		

Non-healthcare workers included alarm designers and researchers. Of those, 12 % referred having experience with medical devices from family needs, from using them as musical instruments, having them as subject of research and from taking part of a project for musicians inside hospitals.

13.2.2 Stimuli

Twenty-six auditory icons were synthesized following the insights collected from the previous categorization experiment. All had 2 seconds of duration and were calibrated to the same Leq of -20 dB. All categories were represented with three sounds except for Temperature which had five auditory icons. The Temperature category did not obtain relevant feedback from earlier user-studies and for that reason more options needed to be tested. Table 13-2 includes a description of each sound (as perceived by the author). All stimuli are available to be heard here[24].

Table 13-2. Descriptions of the sounds used in the comprehension and referent association experiment.

ID	CATEGORY	DESCRIPTION
1	Cardiac_cardio_1	A rhythmic, electronic thumping beating in pairs, repeating five times
2	Cardiac_cardio_2	Natural sounding heart beat in pairs, repeating five times
3	Cardiac_cardio_4	A rhythmic, beat-boxy low-pitch sound beating in pairs, repeating five times
4	Blood Pressure_BP_2	Pumping air to a sphygmomanometer and hearing the release of the air
5	Blood Pressure_BP_3	Pumping air to a sphygmomanometer and hearing the release of the air along with the sound of a heart beating
6	Blood Pressure_ BP_4	Air being sucked by a straw with the sound of a beating heart
7	Oxygen_Oxy_1	Synthezised bubbles going up
8	Oxygen_Oxy_2	Videogame-y sound of liquid falling and of bubbles going up
9	Oxygen_Oxy_3	Videogame-y sound of liquid being shot. A more aggressive sound than the above.
10	Ventilation_v_1	A heavy breathing. A lot of roughness.
11	Ventilation_v_2	Sound of breathing in and breathing out
12	Ventilation_v_3	Gasping for air and slowly releasing it
13	Drug Administration_D_A_1	A single drop falling in liquid followed by the sound of something effervescing
14	Drug Administration_ D_A_2	A cup being filled with liquid

[24] Vieira, Joana (2021): Validation of clinical auditory alarms with users: Comprehensibility and referent association. figshare. Media. https://figshare.com/articles/media/Validation_of_clinical_auditory_alarms_with_users_Comprehensibility_and_referent_association/14905230

	ID (CONT.)	CATEGORY (CONT.)
15	Drug Administration_ D_A_3	Pills falling on hard surface
16	Power Down_powerdown1	A giant robot losing energy at increasing rate
17	Power Down_powerdown3	A small robot losing energy at increasing rate
18	Power Down_powerdown4	Fast rotating turbine losing strength
19	Perfusion_Perfusion_1	Liquid running down a closed bambu pipe
20	Perfusion_ Perfusion_2	Constant stream of water running
21	Perfusion_ Perfusion_3	Inconstant stream of water running down a hole or drain
22	Temperature_Flame	Wood setting on fire very fast
23	Temperature_ Frig_1	A drop of water falling on a burning fryer
24	Temperature_Frig_2	Grains falling on a fryer
25	Temperature_ Frig_Flame	Grains falling on a fryer whith the sound of burning wood
26	Temperature_Water_boiler	Kettle with water boiling and whistle blowing

13.2.3 Procedure

The study was conducted via the online survey platform Qualtrics (2021). It was distributed among personal contacts, acoustics-related newsletters, and social media. The survey was available in Portuguese, English, and Spanish languages. It consisted in two steps: 1) Step 1: Comprehension Test; 2) Step 2: Icon-referent association test and Adequacy test. Before the main task, participants performed a listening screening test (cf. 10.3.1 - Listening Screening Test).

13.2.3.1 Step 1: Comprehension Test

The comprehension test followed closely the guidelines presented in ISO 9186-1/2014 (2014a). As recommended in the standard, participants were presented no more than 15 auditory icons to evaluate. These were randomly selected by the platform out of the pool of 26 sounds. Instructions to participants were the following:

"***This is the first part of this experiment. This experiment has two parts.*** *We would like you to help us evaluate some proposed sounds that will be later worked on to be clinical auditory alarms. We want to see how well these sounds are understood. You will be helping us do this by* ***writing down what you think each sound refers to****. On each page we will show you a sound. All sounds are to be heard coming from medical equipment such as vital signs monitors, infusion pumps, ventilators, among many others.*

It is very important that you write down what, according to you, the sound refers to. — There is a blank space for you to write down what you think the sound refers to.

Some of the sounds may be easy and some may be difficult to understand. Just do the best you can for each of the sounds. If you can't figure out what a sound means, write down "don't know". Please do not just leave a blank. There is no time limit but try not to spend too much time on each sound. Once you finish a page do not go back over it."

Figure 13-2 shows the example page that was shown to participants after reading the instructions, and before starting the comprehension test.

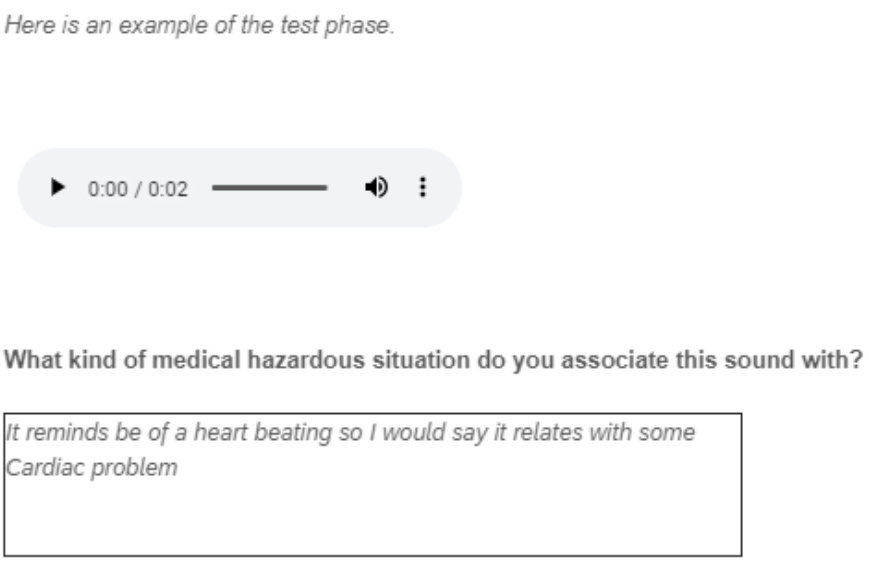


Figure 13-2. Example page before the comprehension test for 15 auditory icons

After this step was concluded, participants were presented a second page of instructions, for the second step and last of the evaluation.

13.2.3.2 Step 2: Referent association test and Adequacy test

The second stage of the evaluation was inspired in ISO 9186-3/2014 for the referent-association test and in SAE J2830/07-2008 for the Adequacy test. In this round, the same 15 sounds were presented to participants. However, this time they would have to choose a single category in which they believed the sound fitted. All categories related to clinical hazards, has determined in IEC 60601-1-8, plus the Blood Pressure category. After each evaluation, they would be provided with feedback stating the correct category of the sound. In this feedback page, they should evaluate on a 7-point Likert scale how adequate they thought that sound was to that referent. Figure 13-3 shows the interface where

participants chose the category better represented by the auditory icon they were listening.

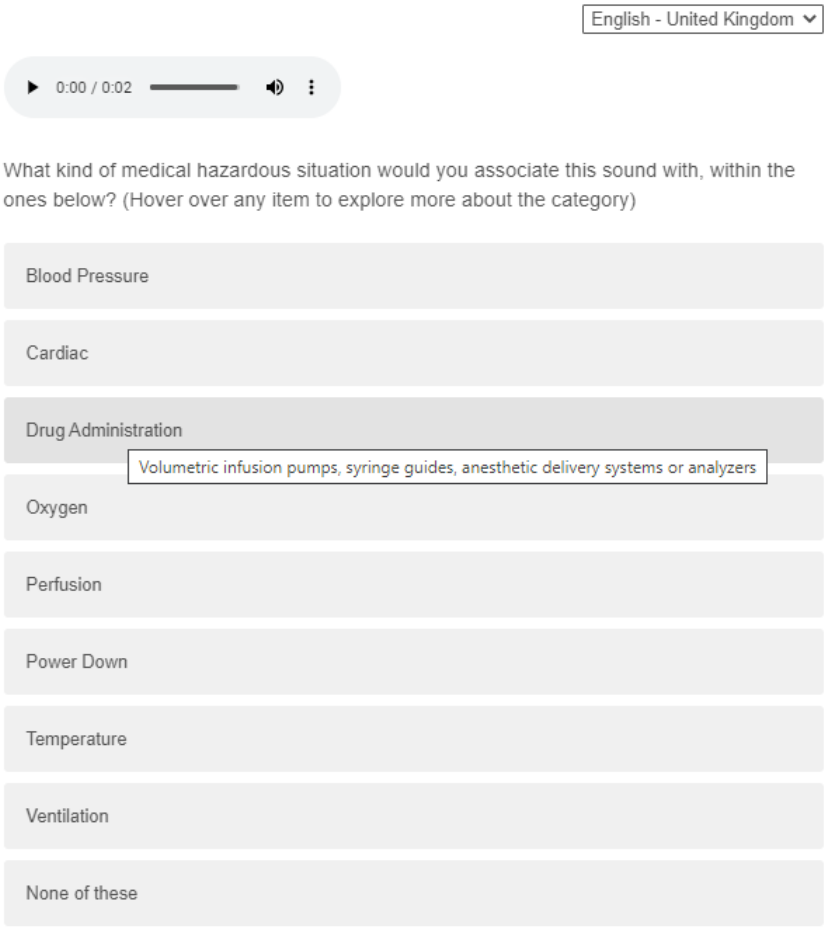


Figure 13-3. Page of the referent association test for 15 auditory icons. When participants hovered over each category, they were presented with examples of equipment associated with that category.

The instructions were the following:

"*Now you will hear the same sounds and indicate what you think each sound means from a list of possible meanings.* ***All sounds are to be found in medical equipment.***

There will be a list of possible meanings for that sound. There is also the opportunity to show that you think none of the meanings in the list is the meaning of the icon. If you think that one of the meanings given is the meaning for that icon, please select that meaning.

If you think none of the meaning in the list is the one for that icon, select "none of the above" at the bottom of the list.

***After each evaluation**, we will tell you what the sound is supposed to mean and will ask you to rate the adequacy of the sound-meaning pair. The scale of adequacy is from 1-7: Not at all adequate 1 - 2 - 3 - 4 - 5 - 6 - 7 Totally adequate.*

After this stage, we will ask you some demographic questions, and the experiment ends."

Figure 13-4 shows the interface where participants rated the adequacy of the auditory icon to its predetermined category.

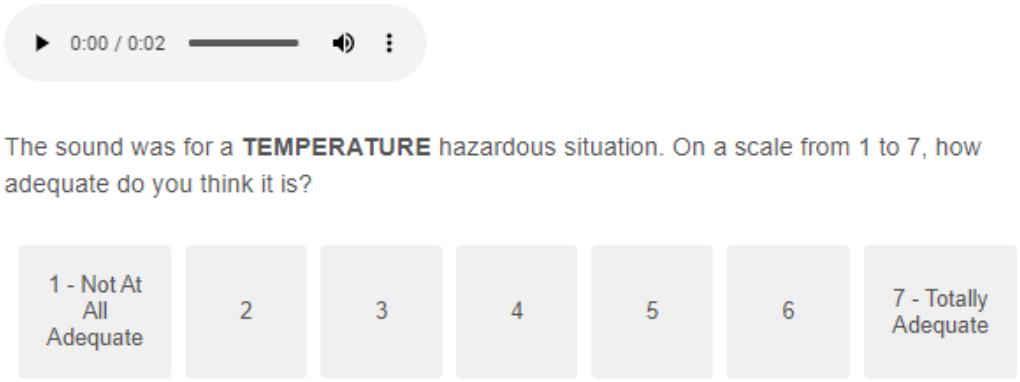


Figure 13-4. Feedback page presented after each referent-association evaluation. The participant should state how adequate the sound was for that category on a 7-point Likert Scale from "Not adequate at all" to "Totally adequate"

The online experiment took on average 17 minutes.

13.3 Analysis

The analysis of the results poses some questions regarding up to which point should a sound be adequate or not for a give referent. For the case of graphical symbols, SAE requirements for an acceptable symbol are a 66% correct comprehension level. Figure 13-5 represents the evaluation methods for each stage of the evaluation, and where the decision criteria was taken from.

13.3.1 Comprehension

A list of all responses for every variant of a referent was generated. According to SAE J2830, the analysis of the Comprehension test distinguishes between "Major" and "Minor" components of the message. "Major" elements are those words or phrases within the definition or message associated with an auditory icon that are critical to the meaning of the auditory icon. "Minor" elements are those words of phrases within the message associated with an icon that are less critical. As an example, for a "Tire pressure warning –

low", the "tire" and "warning" parts of the message are "major", while "pressure" and "low" are "minor".

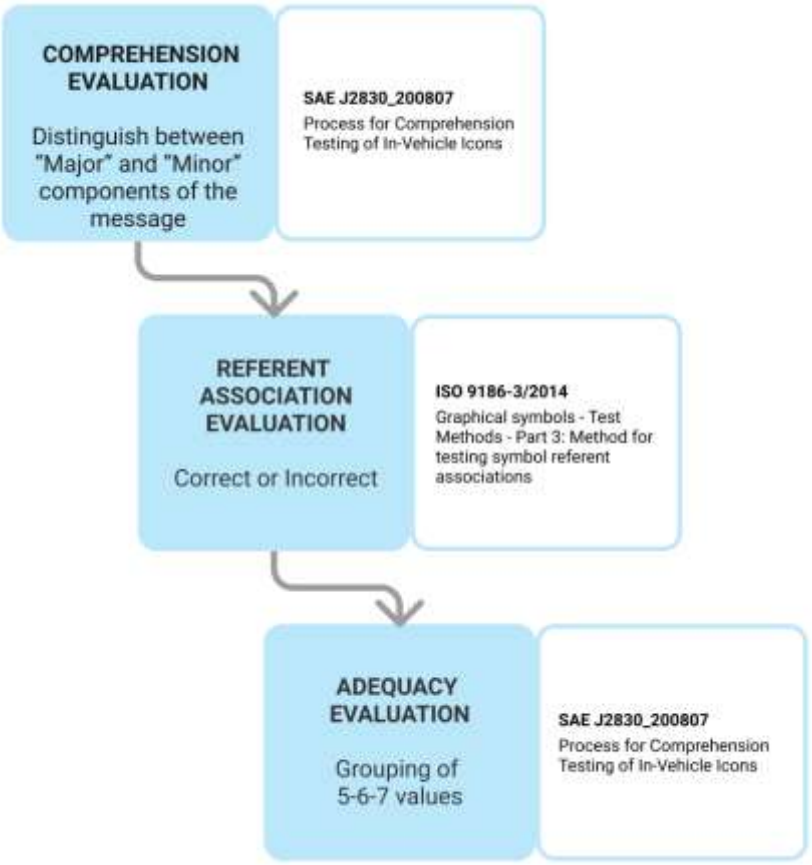


Figure 13-5. The three methods of analysis of each evaluation stage with reference to the source of each evaluation method

Of the nine scoring categories presented in Table 13-3, scores of 1 or 2 were considered the most important, as they defined overall comprehension rates. The remaining scores were most useful for diagnostic purposes. They were useful in identifying problems with candidate auditory icons and possible methods for improving comprehension.

Each item was scored independently. For each icon, the total number of responses in each category was converted into percentages. For an auditory icon to be considered acceptable, it had to reach 66% correct comprehension level (combined 1 and 2 categories).

Table 13-3. Table of evaluation for the Comprehension test (SAE J2830)

COMPREHENSION SCORE		DESCRIPTION
High Comprehension	1	The response matches the intended meaning of the auditory icon exactly.
	2	The response captures all major informational elements of the intended meaning of the auditory icon, but is missing one or more minor informational elements.
Low Comprehension	3	The response captures some of the intended meaning of the auditory icon, but it is missing one or more major informational elements.
	4	The response does not match the intended meaning of the icon but it captures some major or minor informational elements.
None	5	The response does not match the intended meaning of the auditory icon but it is somewhat relevant.
	6	Participant's response is in no way relevant to the intended meaning of the auditory icon.
	7	Participant indicated he or she did not understand the auditory icon.
	8	No answer.
Risk	9	For safety-critical icons, identify the number and percentage of critical confusion or errors. Critical confusions or errors reflect responses that indicate that the participant perceived the message to convey a pottentially unsafe action.

13.3.2 Referent Association and Adequacy

The "correct identification" score for each auditory icon was the number of times the correct category was selected by the participants. The auditory icon with the largest mean percentage of "the correct category was selected" was considered as the most recognizable auditory icon for that category. This analysis method was taken from ISO 9186-3/2014.

Regarding the Adequacy data, Likert scores were converted into percentages. Icons with higher percentages of scores in categories 5 – 7 of the Likert scale had higher levels of perceived appropriateness. This analysis method was taken from SAE J2830/07-2008.

13.3.3 Summary of the decision criteria

For this study, the following criteria was used to determine four levels of auditory icon approval (Table 13-4):

- Cut-off point: 66%
    - Comprehension: High (Categories 1-2)
    - Referent Association: 66% or more on the Correct category selected.
    - Adequacy: 66% or more on the 5-6-7 values of the Likert scale

Table 13-4. Criteria for the definition of auditory icon approval levels

	APPROVED			REJECTED	
	Level 1	**Level 2**	**Level 3**	**Level 4**	**Level 5**
Comprehension	≥ 66%	< 66%	< 66%	≥ 66%	< 66%
Referent Association	≥ 66%	≥ 66%	< 66%	< 66%	< 66%
Adequacy	≥ 66%	≥ 66%	≥ 66%	< 66%	< 66%

- Icons on **Level 1** were icons with a comprehension, referent association and adequacy rate above 66%, meaning that participants correctly identified the icons, their category and agreed with the description of the correct category.
- **Level 2** reported to icons with a comprehension rate below 66%, meaning that participants did not always immediately understand the meaning of the icon but, after being faced with an option and a given description, they agreed that the category was appropriate for the icon (adequacy > 66%).
- **Level 3** reported icons with a comprehension rate below 66%, the choice of the correct referent association below 66%, but once shown the correct answer they thought it appropriate (adequacy > 66%).
- Icons on **Levels 4 and 5** should be rejected because even if the participants guessed the meaning of the icon, they did not agree with it. As a principle, Level 4 should not occur often, as it would be somewhat surprising to have participants guess a meaning with which they would not agree. Occurrences of this kind should be further investigated.
- **Level 5** represents the absolute level of rejection, in which participants do not comprehend the icons' meanings, nor do they agree with the given description.

13.4 Results

The following tables present the results of the three evaluation stages according to each one of the 26 sounds, divided by category. Results are divided between Healthcare and non-Healthcare participants, and in case of difference, Healthcare participants' results determined the winning sound.

Each sound was evaluated, on average, 92 times.

13.4.1 Cardiac

1. Cardiac_cardio_1	Healthcare	Non-Healthcare
Comprehension Evaluation	54 %	39 %
Referent Association	100 %	80 %
Adequacy Evaluation	88 %	95 %
Result	Level 2	Level 2
2. Cardiac_cardio_2	**Healthcare**	**Non-Healthcare**
Comprehension Evaluation	60 %	56%
Referent Association	88 %	93%
Adequacy Evaluation	89 %	98%
Result	Level 2	Level 2
3. Cardiac_cardio_4	**Healthcare**	**Non-Healthcare**
Comprehension Evaluation	60 %	56%
Referent Association	88 %	93%
Adequacy Evaluation	89 %	98%
Result	Level 2	Level 2

13.4.2 Blood Pressure

4. Blood Pressure_BP_2	Healthcare	Non-Healthcare
Comprehension Evaluation	0%	9%
Referent Association	29%	30%
Adequacy Evaluation	57%	46%
Result	**Level 5**	**Level 5**
5. Blood Pressure_BP_3	**Healthcare**	**Non-Healthcare**
Comprehension Evaluation	6%	12%
Referent Association	50%	36%
Adequacy Evaluation	63%	53%
Result	**Level 3**	**Level 5**

6. Blood Pressure_BP_4	Healthcare	Non-Healthcare
Comprehension Evaluation	15%	4%
Referent Association	14%	22%
Adequacy Evaluation	43%	60%
Result	**Level 5**	**Level 5**

13.4.3 Oxygen

7. Oxygen_Oxy_1	Healthcare	Non-Healthcare
Comprehension Evaluation	7%	3%
Referent Association	43%	18%
Adequacy Evaluation	71%	31%
Result	**Level 3**	**Level 5**
8. Oxygen_Oxy_2	**Healthcare**	**Non-Healthcare**
Comprehension Evaluation	7%	1%
Referent Association	8%	8%
Adequacy Evaluation	17%	29%
Result	**Level 5**	**Level 5**
9. Oxygen_Oxy_3	**Healthcare**	**Non-Healthcare**
Comprehension Evaluation	0%	0%
Referent Association	0%	10%
Adequacy Evaluation	20%	14%
Result	**Level 5**	**Level 5**

13.4.4 Ventilation

10. Ventilation_v_1	Healthcare	Non-Healthcare
Comprehension Evaluation	18%	34%
Referent Association	100%	45%
Adequacy Evaluation	100%	71%
Result	**Level 2**	**Level 3**
11. Ventilation_v_2	**Healthcare**	**Non-Healthcare**
Comprehension Evaluation	46%	49%
Referent Association	88%	51%
Adequacy Evaluation	75%	86%
Result	**Level 2**	**Level 3**

12. Ventilation_v_3	Healthcare	Non-Healthcare
Comprehension Evaluation	43%	51%
Referent Association	90%	41%
Adequacy Evaluation	80%	76%
Result	**Level 2**	**Level 3**

13.4.5 Drug Administration

13. Drug Admin_D_A_1	Healthcare	Non-Healthcare
Comprehension Evaluation	0%	4%
Referent Association	44%	31%
Adequacy Evaluation	22%	45%
Result	**Level 5**	**Level 5**
14. Drug Admin_D_A_2	**Healthcare**	**Non-Healthcare**
Comprehension Evaluation	0%	5%
Referent Association	50%	30%
Adequacy Evaluation	38%	39%
Result	**Level 5**	**Level 5**
15. Drug Admin_D_A_3	**Healthcare**	**Non-Healthcare**
Comprehension Evaluation	7%	1%
Referent Association	22%	35%
Adequacy Evaluation	33%	40%
Result	**Level 5**	**Level 5**

13.4.6 Power Down

16. PD_powerdown1	Healthcare	Non-Healthcare
Comprehension Evaluation	29%	19%
Referent Association	100%	82%
Adequacy Evaluation	100%	83%
Result	**Level 2**	**Level 2**
17. PD_powerdown3	**Healthcare**	**Non-Healthcare**
Comprehension Evaluation	27%	25%
Referent Association	73%	85%
Adequacy Evaluation	91%	93%
Result	**Level 2**	**Level 2**

18. PD_powerdown4	Healthcare	Non-Healthcare
Comprehension Evaluation	31%	21%
Referent Association	100%	88%
Adequacy Evaluation	100%	91%
Result	**Level 2**	**Level 3**

13.4.7 Perfusion

19. Perfusion_Perfusion_1	Healthcare	Non-Healthcare
Comprehension Evaluation	15%	7%
Referent Association	22%	11%
Adequacy Evaluation	56%	24%
Result	**Level 5**	**Level 5**
20. Perfusion_Perfusion_2	**Healthcare**	**Non-Healthcare**
Comprehension Evaluation	0%	3%
Referent Association	50%	19%
Adequacy Evaluation	58%	36%
Result	**Level 5**	**Level 5**
21. Perfusion_Perfusion_3	**Healthcare**	**Non-Healthcare**
Comprehension Evaluation	0%	4%
Referent Association	36%	36%
Adequacy Evaluation	36%	34%
Result	**Level 5**	**Level 5**

13.4.8 Temperature

22. Temperature_Flame	Healthcare	Non-Healthcare
Comprehension Evaluation	7%	0%
Referent Association	33%	4%
Adequacy Evaluation	11%	15%
Result	**Level 5**	**Level 5**
23. Temperature_ Frig_1	**Healthcare**	**Non-Healthcare**
Comprehension Evaluation	8%	3%
Referent Association	14%	15%
Adequacy Evaluation	14%	23%
Result	**Level 5**	**Level 5**

24. Temperature_Frig_2	Healthcare	Non-Healthcare
Comprehension Evaluation	0%	7%
Referent Association	20%	18%
Adequacy Evaluation	10%	22%
Result	**Level 5**	**Level 5**
25. Temp_ Frig_Flame	**Healthcare**	**Non-Healthcare**
Comprehension Evaluation	0%	0%
Referent Association	0%	7%
Adequacy Evaluation	11%	20%
Result	**Level 5**	**Level 5**
26. Temp_Water_boiler	**Healthcare**	**Non-Healthcare**
Comprehension Evaluation	0%	3%
Referent Association	0%	18%
Adequacy Evaluation	11%	13%
Result	**Level 5**	**Level 5**

13.4.9 Final selection for next design stage

We were able to select several auditory icons for the following design stage, including categories where all icons were approved (Cardiac, Ventilation and Power Down). To determine the approval and the order of approval, special attention was given to the Healthcare professionals' evaluations.

The categories of Drug Administration, Perfusion and Temperature had no auditory icon approved according to participants. The sounds referred on Table 13-5 are those which obtained the least amount of negative feedback.

Table 13-5. Final selection of auditory icons for each category.

CATEGORY	ORDER OF APPROVAL	NAME OF FILE
Cardiac	1st	1. Cardiac_cardio_1
	2nd	2. Cardiac_cardio_2
	3rd	3. Cardiac_cardio_4
Drug Administration	Not approved	14. Drug Admin_D_A_2
Ventilation	1st	10. Ventilation_v_1
	2nd	12. Ventilation_v_3
	3rd	11. Ventilation_v_2
Blood Pressure		5. Blood Pressure_BP_3
Perfusion	Not approved	20. Perfusion_Perfusion_2
Oxygen		7. Oxygen_Oxy_1
Temperature	Not approved	22.Temperature_Flame
		23. Temperature_ Frig_1
Power Down	1st	18. PD_powerdown4
	2nd	16. PD_powerdown1
	3rd	17. PD_powerdown3

In total, 15 sounds were selected from the initial pool of 26 sounds.

According to IEC 60601-1-18 (2020) new performance metrics regarding recognition, these 15 sounds perform as represented in Table 13-6 . The standard's reference values report to 10 trials per participant while in this study each sound was evaluated once per participant, with a total of 102 participants. This means that the reference values from the standard report to learnability and our results cannot be interpreted as such. They refer to immediate recognizability after a single exposure.

Table 13-6. Comparison between learnability performance metrics form IEC 60601-1-8 (2020) and current study's results

CATEGORY	IEC 60601-1-8 REFERENCE VALUE	REFERENT ASSOCIATION (HEALTHCARE)	NAME OF FILE
Cardiac	90% to 95%	100%	1. Cardiac_cardio_1
		88%	2. Cardiac_cardio_2
		88%	3. Cardiac_cardio_4
Drug Administration	90% to 95%	50%	14. Drug Admin_D_A_2
Ventilation	80% to 85%	100%	10. Ventilation_v_1
		90%	12. Ventilation_v_3
			11. Ventilation_v_2
Blood Pressure	NA	50%	5. Blood Pressure_BP_3
Perfusion	75% to 85%	50%	20. Perfusion_Perfusion_2

CATEGORY (CONT.)	IEC 60601-1-8 REFERENCE VALUE	REFERENT ASSOCIATION (HEALTHCARE)	NAME OF FILE (CONT.)
Oxygen	70% to 80%	43%	7. Oxygen_Oxy_1
Temperature	85% to 95%	33%	22.Temperature_Flame
		14%	23. Temperature_ Frig_1
Power Down	90% to 95%	100%	18. PD_powerdown4
		100%	16. PD_powerdown1
		73%	17. PD_powerdown3

13.5 Discussion

The present study adapted an evaluation methodology used for visual icons to the auditory modality. The main question is to understand whether the direct adaptation is adequate or if it should suffer any modification. It is our understanding that the methodology is adequate for the evaluation of auditory icons. The stepwise evaluation with more and more information for the participant resulted in incremental positive evaluations.

One debatable step refers to the free text comprehensibility evaluation. The fact is that, from all approved sounds, not one obtained more than 66% of approval in the comprehensibility stage. This could mean that auditory icon's interpretation is extremely hard and subjective, and thus more subject to disparity in the participants' free text input. It could mean that participants needed more contextual information or training on the alarm categories. It could also mean than a 66% cut-off value for this stage is too ambitious, and a lower value should be considered. The feedback obtained from the comprehensibility evaluation was relevant, nevertheless, some answers were as explicit as "Trouble breathing, related with ventilation" while others were "Giant spider running away". Researchers should consider that this stage is the most demanding to participants. They should also be prepared to receive relevant and not so relevant feedback, and only afterwards make the decision of incorporating or not this comprehensibility phase.

One important limitation of this study is the reduced number of healthcare professionals that took part. However, they were determinant in approving sounds for the categories of Blood Pressure and Oxygen. Some participants who identified as non-healthcare professionals happened to be sound designers and researchers in the healthcare context,

which indicates some knowledge of the area even from non-healthcare professionals. Additionally, 12% of this group referred having had experience with medical devices.

Finally, some categories are consistently hard to design for. Again, Drug Administration, Perfusion and Temperature were not approved by our participants. Temperature is extremely difficult to represent in a different modality, and attempts will continue to be made to improve sounds related with warm or cold objects (wood burning, flames, ice scrapping, etc.). Drug Administration and Perfusion are both liquid sounds and some confusion can arise. We believe, however, that it is possible to refine these sounds and improve them. An example is that at this stage the Oxygen (bubbles in water) sound was approved when in the previous study it was put together with Drug Administration and Perfusion. As the pool of sounds gets reduced, the association to a referent will become easier, and learnability will be facilitated by heterogeneity between sounds (Edworthy et al., 2011).

The main outcome of this study was the validation of several auditory icons to represent the IEC 60601-1-8 categories (plus Blood Pressure). These can be further refined by sound designers knowing that they are adequate representatives of those categories of alarms.

13.6 Conclusion

A set of 26 auditory icons was subject of an evaluation process by 102 participants who were later divided into Healthcare and non-Healthcare professionals. This evaluation methodology was custom-adapted from two standardized frameworks for the validation of visual icons and symbols - ISO 9186-1&3/2014 and SAE J2830/07-2008. It consisted in three steps where 1) participants wrote their free interpretation of the sound, knowing beforehand it was to be heard in a medical context; 2) participants selected one option from a list of options according to what they believed that sound represented and 3) after feedback of the correct option, participants gave their opinion regarding adequacy of that sound for that category.

The procedure allowed to select 15 auditory icons with an order of preference according to our participants. The Cardiac, Ventilation and Power Down categories had all their three auditory icons approved by all participants. Blood Pressure and Oxygen had one auditory icon selected for each by the Healthcare professionals' group. The categories of Drug Administration, Perfusion and Temperature did not have an auditory icon approved, but for further design refinement the auditory icons with the best evaluation were selected. Although the methodology worked well for the initial purposes of the study,

considerations were made regarding the effort/outcome relation of the first phase of the evaluation (comprehensibility) and regarding its cut-off value of 66% proposed by SAE's standard for in-vehicle visual icons. As the free interpretation of sounds is significantly more subjective, more difficult, and more diverse than with visual icons, future use of the method should consider a lower cut-off value. Evidence for this choice stands on the fact that not a single auditory icon was approved with a "Level 1" approval, which includes approval from participants at all three stages of the evaluation process. The best approvals in this study were "Level 2" with approval in the second and third evaluation stages.

The outcome of this study consists in a pool of auditory icons associated with all IEC60601-1-8 alarm categories plus Blood Pressure which were evaluated by several participants, including healthcare professionals, regarding their context of use. These sounds will be made available for future work on sound design of clinical alarms.

14 General Discussion and Conclusions

Bain News Service, publisher. (between ca. 1915 and ca. 1920), Bertha M. Furman (LOC). No known copyright restrictions

This thesis serves as a small contribution to the complex and systemic problem of noise in healthcare. Particularly, it focused on the clinical auditory alarms' issues, and proposed a redesign of these alarms.

The history of auditory alarms is long and eventful, and that by itself is an indicator of the challenge that is designing alarms. On an extremely short amount of time – between two to five seconds - we need to place the most relevant information for an operator to adequately respond to the cause of the alarm.

Our main goal was to propose a method to design alarms that went from concept to evaluation and validation, using psychoacoustics, usability, and user experience techniques. Considering healthcare's known and documented problems, this environment served as use case. However, the method is applicable and scalable to any context using auditory alarms.

This work followed the structure of human-centric design. As such, it first tried to understand and specify the context of use. The visit, interviews and survey were initial activities of extreme importance. Not only did they allow to empathize with the nurses and professionals who dealt with the alarms problem daily, it also allowed to identify another context which was going unnoticed until then. Recovery Rooms are a relevant scenario for alarm research, since they have many patients, many devices and many caregivers circulating. It also became evident from these activities how alarms affect different stakeholders, from healthcare professionals to patients and visitors.

Also interesting was observing nurses, surgeons and anaesthetists seriously reflecting on auditory alarms for what seemed the first time. The level of acceptance that noise is part of their work is high, and that is an indicator that intervention in their acoustic space is needed. The first moment of silence after leaving the hospital lulls them because it feels comforting. The noise levels between the interior and exterior of a hospital should not be so dramatically different.

Regarding noise levels, the frequency analysis made inside two hospitals, on the course of 28 different surgeries allowed to understand that levels of background noise are, in average, between the conversational range of 60 dB(A) and classroom chatter at 70 dB(A) (Figure 7-1). Table 14-1 compares our results with the same surgeries measured in other countries – and often, other measurement conditions. Our results bear close resemblance with Kracht et al. (2007) and Kuzmich et al. (2001), both from the USA. The first study included measurements inside old and post-2000 construction buildings. The second study has no information regarding year of construction. In our situation, both hospitals

were built after 2010, and we believe that that is a relevant information to be added in future noise level and frequency analysis studies.

Table 14-1. Comparison between operating rooms' LAeq between other countries and our study

	Brazil (Lopes, 2005)	Greece (Tsiou et al., 2008)	USA (Kracht et al., 2007)	USA (Kuzmich et al., 2001)	Portugal
Urology	76.0		63.5		63.92
Vascular	76.7				65.12
Orthopaedics	72.0	68.31	66.5	65.8	66.51
Otorhinolaryngology	72.7		65		65.66
Gynaecology	78.1				64.04
Neurosurgery	73.6		64.5		65.60

Even though most surgeries were in between the 60-70 dB(A) range, in terms of frequency it was possible to distinguish some surgeries from others. Orthopaedics surgeries had higher sound pressure levels at high frequencies. In terms of noise maximum Leq levels, it was also possible to indicate surgeries with more peaks than others, namely Vascular, Urology and Otorhinolaryngology surgeries. However, because we are unable to associate the frequency analysis records with real events occurred during surgery, it is not possible to understand what caused these peaks and whether it is a frequent occurrence in that type of surgery. This would be relevant information for it would allow to sketch a kind of acoustic signature for each type of surgery.

A second step of the human-centred approach included specifying the user requirements. Having collected several needs, requirements, and concerns from the previous stage, in this moment we completed the set of field-requirements with those from standards and literature. An effort was made to span all previous knowledge, and a set of alarm design requirements were presented in table form, divided by spectral, temporal, and spatial acoustic parameters.

On the third step of producing design solutions, we first needed to understand which temporal and spectral parameters of sound could be associated with clinical alarms' information. We made simple beeps with four harmonics in which we manipulated the onset/offset, rhythm, speed, and frequency. We wanted to understand how the manipulation of these variables affected the perception of pleasantness, urgency, positiveness, speed, annoyance, and preoccupation. Our results regarding speed and frequency matched the literature. Speed correlated with more subjective perceptions than

any other variable, thus pointing towards the importance of temporal manipulations when designing alarms. Frequency correlated with annoyance and pleasantness. Our results were inconclusive regarding the onset and offset of the alarms. Our interpretation is that there were too little variations, and the 20ms manipulation was imperceptible. As for rhythm, results matched IEC 60601-1-8 which stated that a syncopated rhythm was perceived as more urgent than a regular rhythm. However, the literature is not unanimous with this variable, and for that reason we have chosen to study the effect rhythm again on a subsequent urgency study.

For mere methodological curiosity, this same study was replicated online on Amazon's Mechanical Turk platform. Unbeknownst to us, remote testing ended up being the only testing method available on the course of the development and validation stages of this work, due to the global COVID-19 pandemic. This first study was performed with more than 300 participants and although consistency between raters was smaller (each participant only evaluated 20 sounds, unlike the 324 evaluations made by laboratorial participants), results were very similar. We interpreted this similarity by means of the robustness of the relations between subjective perceptions and physical parameters, and the robustness of the evaluation method.

Following evidence of the efficacy of auditory icons, we have decided to use them as the main core of our alarms. An exercise of categorization ensued, to define which everyday sound could be used for each one of the predefined categories. This experiment allowed to select auditory icons, some with more certainty than others. The cardiac category was attributed the Korotkoff sound, the blood pressure was attributed the sound of a cuff inflating, power down was strongly associated with the sound of a machine going down, ventilation was associated with the sound of breathing through a mask and drug administration was associated with liquid sounds like water pouring inside a glass, or with the sound of spreading pills over a table. The categories of oxygen, perfusion and temperature had dispersed results and we were not able, with this exercise, to select a single auditory icon.

After having evidence of the way perceptual variables affect the subjective perception, and after selecting the basic unit of our alarms, which was the auditory icon, we established a Design Strategy. This strategy should be modular and scalable. Our goal was to create a useful tool for designers to use in any context using auditory alarms. On the course of this work, we understood there were not a lot of resources specifically guiding sound designers in an integrated fashion.

With an established design strategy, known effects of acoustic parameters on subjective perception, and which auditory icons to use, we have created a set of prototypical alarms with two separated components: the priority pointer and the auditory icon. This set of sounds was to be used in the following human centred design stage of evaluating the designed sounds with users.

The last stage included user tests to evaluate those two components of the alarm, the priority pointer, and the auditory icon. For the auditory pointer, we wanted to assure that our three priority levels mapped with the perception of urgency. For experimental reasons, we have selected only two levels, Medium and High, to validate our designed priority pointers. Results have shown that of all five variables tested (Fundamental Frequency, Speed (No.Pulses), Pulse Length, Inharmonicity and Rhythm), four significantly weighted on the evaluation of urgency. In order of importance for the prediction of urgency we found that Pulse Length, Speed (No. Pulses), Fundamental Frequency and Rhythm affected the perception of urgency. In sum, a short pulse with a short inter-pulse interval, repeated six times in a regular fashion, with a fundamental frequency of 441 Hz was evaluated with the highest level of perceived urgency. This meant that these four variables, could be transferred to the design priority pointers of our alarms.

For the auditory icons, we adapted standardized methodologies used for the validation of visual symbols. Participants helped us select an auditory icon for oxygen, however, the perfusion and temperature categories were again challenging sounds. Having auditory icons approved in this experiment meant that, even out of context, our alarms were associated with a category or a medical event.

After testing the two main parts of the alarm, we were able to propose a final library of alarms for clinical use with identifiable alarms regarding their meaning and their priority. This library follows IEC 60601-1-8 categorization system, plus a blood pressure alarm after understanding it was a generalized need. It is presented in Appendix 7.

IEC 60601-1-8 (2020) states that "*the two main requirements of an alarm signal are that signals are audible and recognizable. Almost everything else is a matter of taste and preference*". Although the recognizability tests were performed, no detectability methods were implemented. Nevertheless, having our own measurement of different surgeries allow us to estimate whether our own alarms will be detected over Orthopaedics, the surgery with the highest LAeq SPL. Figure 14-1 represents two auditory icons, Cardiac and Ventilation, both signalling the value of the vital sign is decreasing. While the Cardiac

sound is quite percursive, with a cartoonized sound of a beating heart, the Ventilation sound could be considered easier to mask, as its icon includes someone breathing.

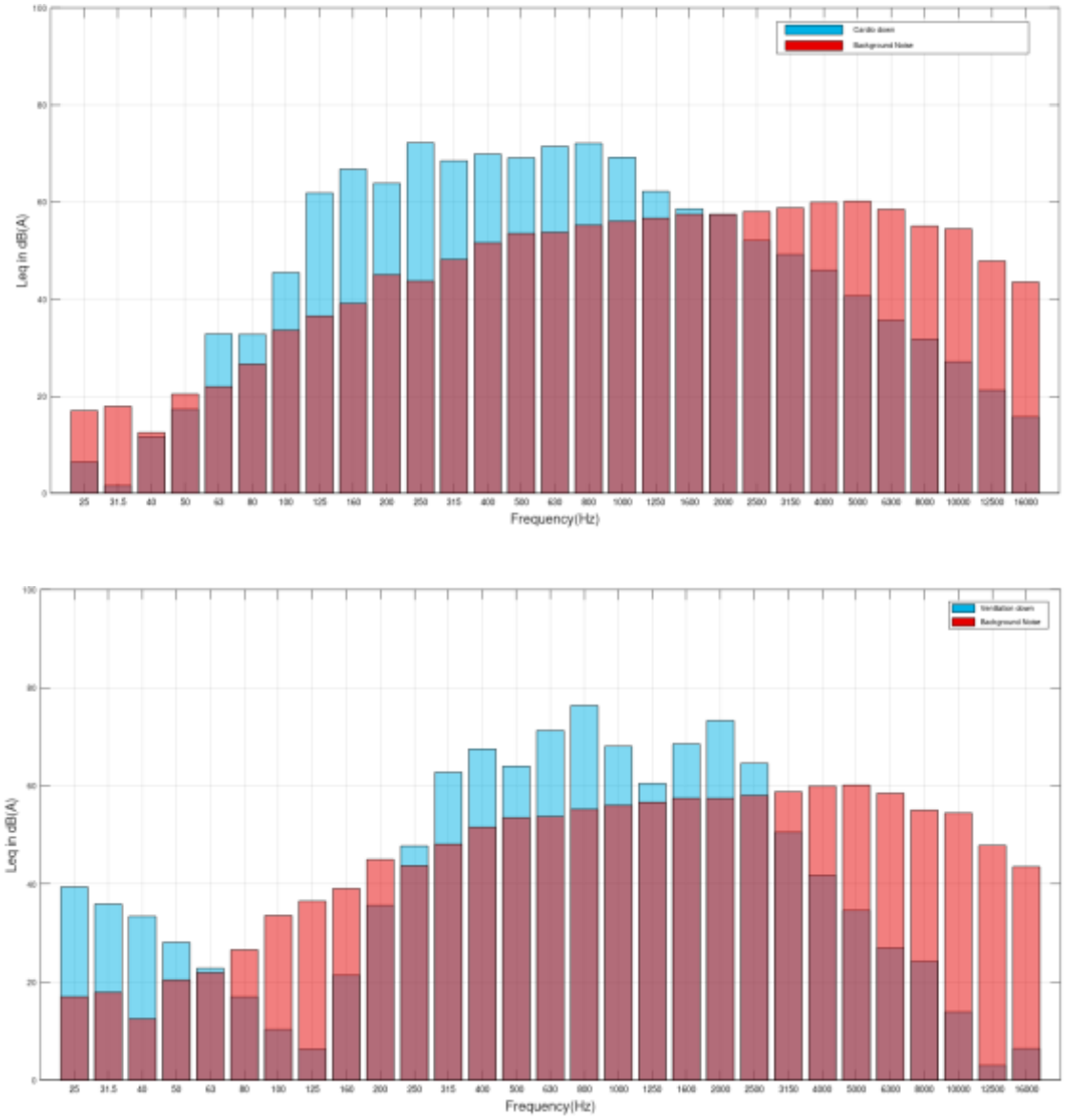


Figure 14-1. (Top) Cardiac auditory icon (Bottom) Ventilation auditory icon represented over the background noise of an Orthopaedic surgery.

As depicted in the simulations, both auditory icons have frequency bands which would be masked by the background noise, but both alarms would have components heard over it.

The library of alarms' files is destined for sound designers, however, they were not validated with users in simulated or real settings. Before implementation, we advise running all alarms on a predictive masking tool (Pereira et al., 2018); and perform user testing for detection, localizability, learnability, as suggested in IEC 60601-1-8, amendment 2.

14.1 Methodological considerations

Our main goal was to propose a method to design alarms that went from concept to evaluation and validation, using psychoacoustics, usability, and user experience techniques applied to sound design. It should be applicable and scalable to any context using auditory alarms.

On the course of this work, professor Judy Edworthy led other human-factors researchers in the release of their findings, which would afterwards result in the current 2020 IEC 60601-1-8 amendment. The results were very positive regarding the auditory icons they designed. The masking, learning, and localization experiments where validation of the new alarms was performed were also robust and very well-designed (Bolton et al., 2018; Edworthy, McNeer et al., 2018; Edworthy et al., 2017; Edworthy et al., 2017, 2018; McNeer et al., 2018). This left us in an ambivalent position of understanding where we could direct our contribution towards, as apparently, we were addressing the same human factors problem. We have decided to provide a hands-on guide to sound designers at the early stages of ideation and prototyping of alarms. As such, we tried to establish methods for these stages. For instance, the other studies were not very clear about how the sounds for the auditory icons were selected, and so we tried to use a sound sorting strategy with quality of fit involving both healthcare and non-healthcare professionals.

We also used a test for the comprehensibility and referent association with evaluation of adequacy to select the most understandable sound. This sort of test was not performed, or at least not published in all IEC standard-associated publications. It was thus possible to understand that this method, translated from the visual modality (ISO and SAE standards for validation of visual symbols - ISO, 2008; ISO, 2014a, 2014b), could be simplified when applied to the auditory modality.

While the standard-related publications were not associated with a particular design methodology, we understood it followed years and years of research, and privileged the validation stages of the methodology proposed by Edworthy and Stanton (1995) – depicted in Figure 4-2. Figure 14-2 represents the stages followed on the course of our work and is our proposed sound design methodology. Blue boxes represent all the steps we implemented, and the red box represent the steps we were not able to implement. Looking from afar, it looks as if both approaches (IEC committee led by Judy Edworthy and ours) are complementary.

Another complement to the ongoing work resulted in our proposal of including tendency cues on the alarm, indicating whether the monitored sound was increasing or decreasing. This variable was not validated.

Our validation steps were very compromised by the global situation of the last year and a half due to COVID-19. Ironically, our first study with participants compared laboratorial results with online results out of sheer curiosity. Although with different statistical strength, both studies revealed the same tendencies. Then, as the world confined, online research became the only option, and new tools and screening tests became better and more popular. Most of our results come from online studies, which poses some limitations, but also benefits. Online studies allowed us to reach people from all over the world. However, the experiments needed to be adapted because online tests should not be very long to avoid experimental mortality. For this reason, the number of presented stimuli and respective repetitions were reduced. As described in Chapter 10 about the Categorization of sounds, the selected method always affects the outcome and researchers need to be aware of this sensitivity when designing experiments, and when generalizing results.

Finally, a word about the multidisciplinary involved in alarm design. Schafer, (1977) explains it perfectly (sic): "*The true acoustic designer must thoroughly understand the environment he is tackling; he must have training in acoustics, psychology, sociology, music, and a great deal more besides, as the occasion demands. There are no schools where such training is possible, but their creation cannot long be delayed, for as the soundscape slumps into a lo-fi state, the wired background music promoters are already commandeering acoustic design as a bellezza business.*" Indeed, it took a small village to make this research. Experimental psychology helped design experiments and select testing frameworks. Cognitive Ergonomics and Human-centred design knowledge helped design the entire process, understand, and characterize the underlying human phenomena. The technical knowledge of acoustics and sound design allowed us to actually have sound files with the most wonderful and minutia modifications. It is quite impressive the work condensed in 4 seconds of sound. The investment on this type of development should always be balanced with the weight of a potential error.

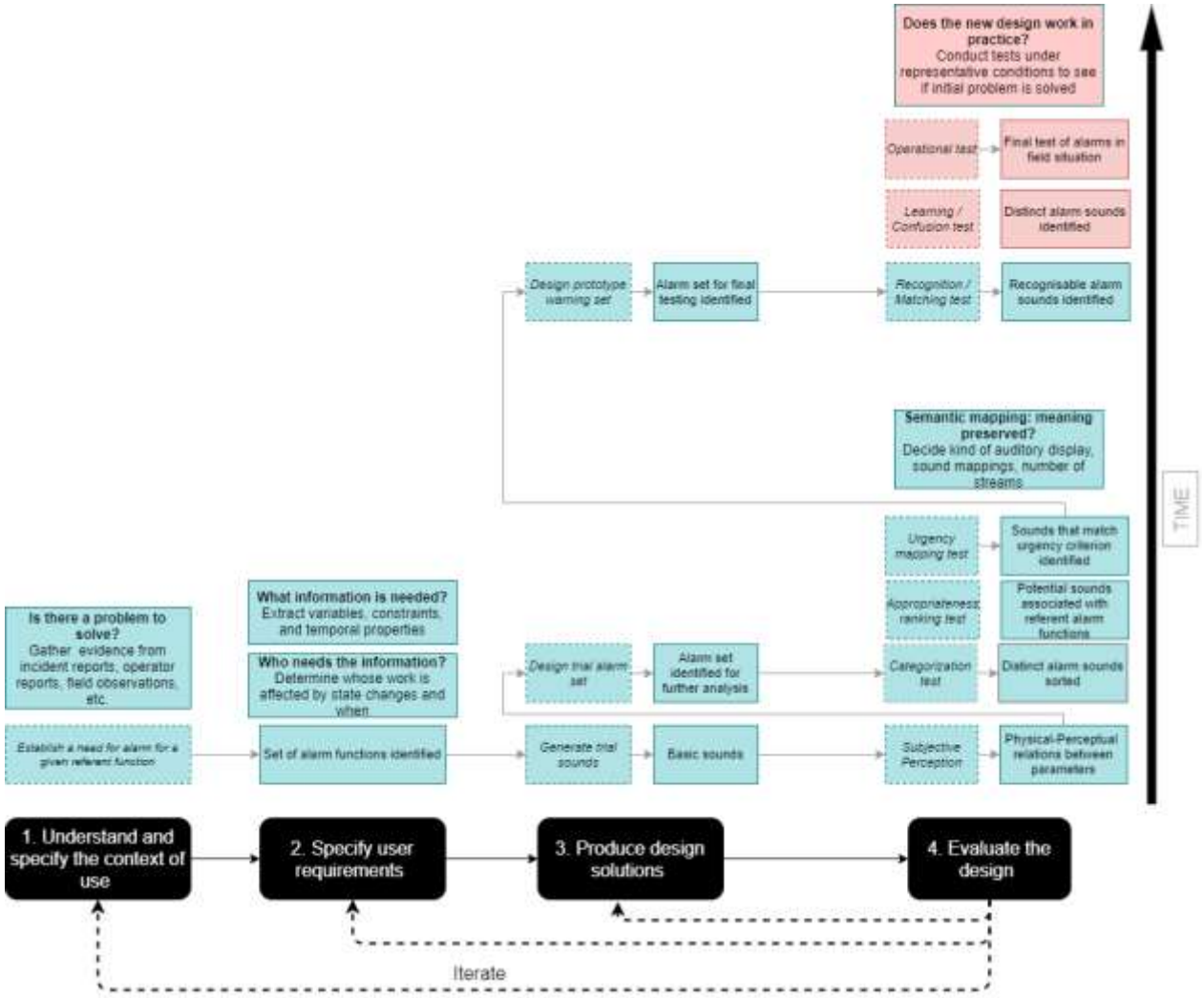


Figure 14-2. Methodological steps followed on this work. The blue squares were implemented, the red squares were not implemented.

14.2 What are the limitations?

A major limitation concerned the difficult access to healthcare environments. The bureaucratic weight of the permissions, of the ethics committees, consumed so much energy that it might have motivated even more the use of online studies for our research purposes. We believe these protocols are dependent on culture, and our Portuguese context is still not very open to human factors research in healthcare – we believe for fears of researchers finding opportunities for improvement, which on the hospital's perspective, if disclosed, might also represent opportunities for errors. For all this, we are particularly grateful to the facilities and people who allowed our visits and data collection inside hospitals.

This institutional obstacle also contributed to another important limitation of this research. Amongst our participants, we always have healthcare participants, but their number is reduced when compared to non-healthcare participants. We would like to be able to recruit more, and we put our efforts into it, but were quickly faced with new procedures. To spread a survey in a hospital or academic courses' email contacts we needed new documentation. All our healthcare participants answered our online tests voluntarily and with interest of improving a recognized problem. For this reason, whenever results were different between healthcare and non-healthcare participants in the selection of auditory icons, we privileged the healthcare participants' choices, even though their number was reduced.

There were experiments which revealed new limitations in methodology, but also on the design choice of using auditory icons. The comprehensibility stage was maybe too difficult and unguided – even people who work daily with sound have difficulties in explaining the meaning of some sounds. This study also allowed to uncover one limitation of auditory icons, which has been stated since the early days of auditory icon research (Graham, 1999). It is extremely difficult to design auditory icons to represent events which have no associated sound. The example of the Temperature category is paramount. Temperature sounds like what? We have tried a kettle, wood and paper burning, scraping, or melting ice – all to very low levels of identifiability. Despite this limitation, we were able to select one sound for each category, some with really good results like the Cardiac, Blood Pressure and Power Down categories. One thing we do know, is that if the alarm set is small, learning will be easier. The association between one category and its corresponding sound will be facilitated.

The online testing decision was affected by the general tiredness of people in answering online tests and surveys. If in the early days of confinement due to the COVID-19 global pandemic people were very helpful, our last studies captured less participants. For this reason, our urgency study resorted to Mechanical Turk to quickly obtain responses.

Another limitation of this work regards the implementation stage, not covered by any of the design methodologies as it goes beyond their scope. An alternative would be simulation in virtual reality scenarios or healthcare training facilities, where the manipulation of tasks and of mental workload variables would be easier. The implementation of research findings in complex socio technical systems is always a long and bumpy process (Norman & Stappers, 2015). However, following a thorough literature review and envisioning the implementation step, Bach et al. (2018) have proposed a guide to optimise the implementation of alarms improvement elements in healthcare (Figure 14-3). According to this guide, a first step would include the formation of a multidisciplinary team. This team should consist of managers, front-line staff, clinical engineers, biomedical technicians, information technology and information systems experts, patient safety experts, researchers, human factors engineers, administrative and facility staff, and representatives of manufacturers of alarming medical devices. Step 2 includes four implementation actions, namely Promote alarm safety culture, Alarm inventory and prioritisations, Alarm assessment and evaluation and Co-create and implement alarm protocols and procedures. Step 3 includes Continuous staff training on alarm management and Configure alarm settings. The authors suggest Step 4 Improvement in alarm safety to be used as checkpoint to assess if all the process has had a positive influence on alarm safety.

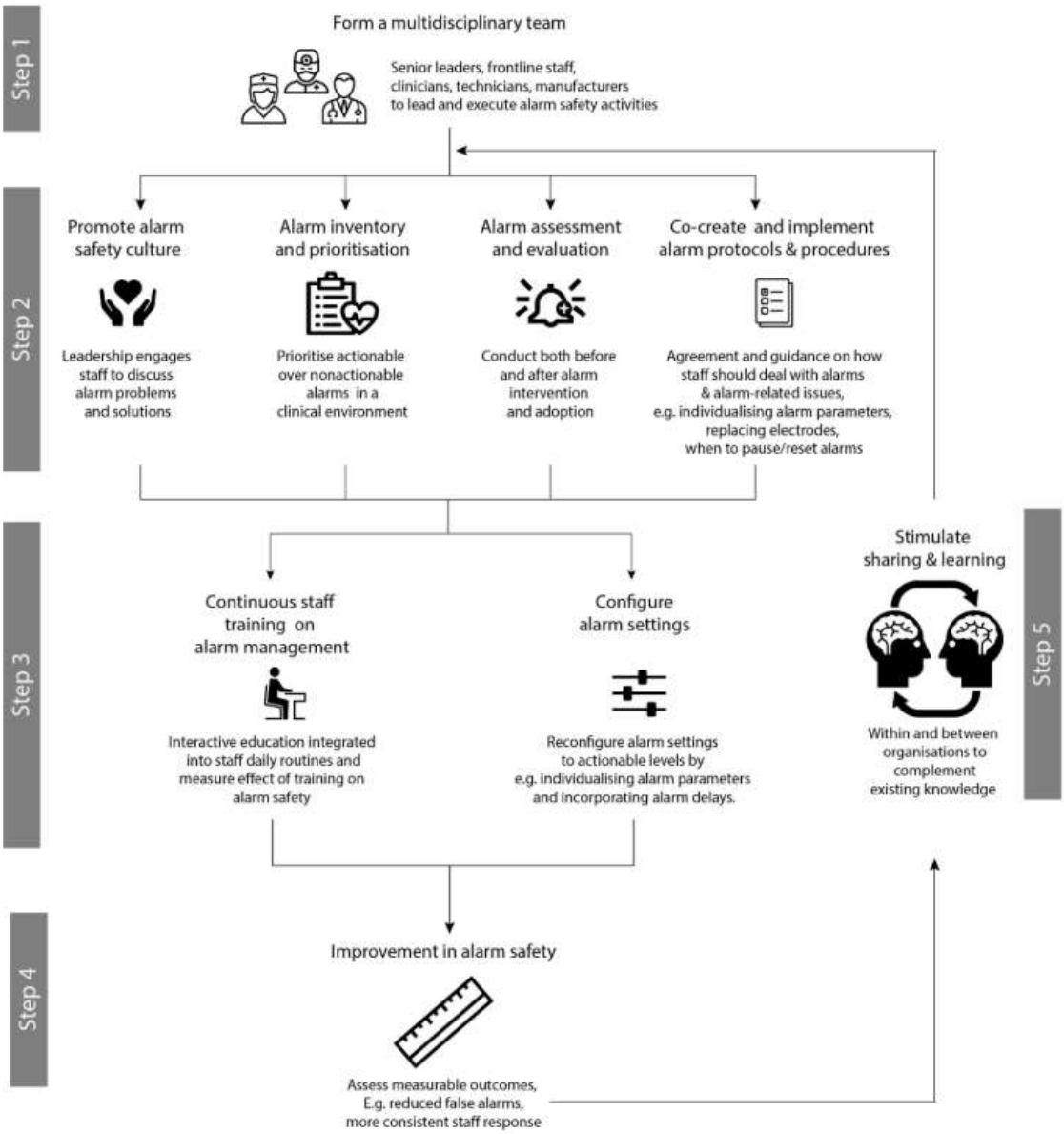


Figure 14-3. Guide to optimise the implementation of alarms improvement elements in healthcare (Bach et al., 2018)

Finally, Step 5 refers to Stimulate sharing and learning envisioning a culture of continuous quality improvement.

14.3 Future research perspectives

We need to be aware that any design change, as small as how an alarm sounds, will have an impact on a complex system. For this reason, a long-term strategy should be thought of to accommodate the changes imposed by these small modifications. What if providing more information to anaesthesiologists during a surgery interferes with the communication protocols with surgeons? What if more information to nurses embedded in alarms results in more anxiety in patients?

Our research has answered the needs of healthcare professionals, and we are positive they will improve their workflow as it was described to us. But a certain amount of consequence scanning should anticipate any major change which might affect a lot of people. For instance, we know that ICU delirium (Johansson et al., 2012) is precipitated by the particular sound environment of ICUs, and it affects not only patients but also professionals. ICU delirium has been found in association with Post-Traumatic Stress Disorder (PTSD) (Schlesinger et al., 2017). Symptoms include agitation, restlessness and emotional instability or apathy and lack of responsiveness (Johansson et al., 2012). What if auditory icons, which are similar to everyday sounds and thus more prevalent then beeping alarms, end up triggering delirium or PTSD symptoms in professionals? Could the sound of water pouring inside a glass, or the sound of Darth Vader breathing in a Star Wars movie trigger unwanted feelings and emotions?

The discussion should move towards the best soundscape for professionals and healthcare professionals, and that appears to be a much more silent soundscape. This dissertation started with Sen Sound's proposal for improving hospital soundscapes, which was three staged. First, we should make incremental changes by improving existent sounds. Second, create new sounds with healthcare professionals and patients. This dissertation is included in this sort of effort. The third stage is one of transformative innovation, and that includes designing the entire environment with healing soundscapes.

We agree with this future perspective and research plan. Alarm fatigue, ICU delirium, blunt tiredness are affected not only with better sounds for alarms, but with less or no alarms. The technology inside hospitals and caregiving spaces should be integrated enough and intelligent enough to distinguish actionable from false alarms, and to address alarms to specific and not all workers.

On another front, presenting alarms in different modalities such as haptics might be an alternative. Not only would haptic alarms be more inclusive, they would also be more silent. Alirezaee et al. (2020) compared the performance of nurses in perceiving and identifying the specific type and level of an alarm event, using several designs of a vibrotactile display. They have concluded that these displays are competitive with graphical and auditory alarms. Other studies used tactile cues to convey detailed physiological information about more than one patient, sent through a tactile interface placed on the lower leg. Results showed a successful interpretation rate of 85%, although taking its toll on the parallel task (Katzman et al., 2019). These results are much in line with the use of vibrotactile display to provide continuous pulse oximetry information. The actuators were placed on the participants' arms, and changes in SpO_2 levels were

identified with more than 90% accuracy (McNulty et al., 2016). Combining auditory with haptic displays is also a solution with promising results (Burdick et al., 2018).

Either way, in complex contexts all recommendations and plans should be taken with due care. Norman and Stappers (2015) list all properties which characterize the difficulty of designing for complex systems. Some of these include:

1) System design that does not take into account Human Psychology.
2) Human Cognition: The human tendency to want simple answers, decomposable systems, and straightforward linear causality.
3) Multiple disciplines and perspectives.
4) Mutually incompatible constraints.
5) Dynamically changing operating characteristics.

Importantly, the authors point that designers currently focus on the ideation and development stages of services, leaving implementation to others. Nowadays, designers must continue through the implementation stage. They also propose a process to improve complex systems, which matches very well with Sen Sound's roadmap for better hospital soundscapes: Incrementalism. With this modus operandi, improvements are achieved in small steps instead of tacking the entire problem. This is advisable because major projects have effects on culture, work practices and are thus more likely to overlook important emerging issues. Small steps also ignite less passion and are more likely to be approved.

As such, future steps should include what was not concluded on the course of this research: learning and confusion tests using our proposed library of sounds, and validation studies in simulated or real settings, preferably with double tasks paradigms to simulate the high cognitive workload that healthcare professionals face on a daily basis.

14.4 Conclusion

Many things influence our health and happiness, and the environment is one of the main drivers of societal wellbeing, as underlined in OECD's "Wellbeing in 2030" report. Often neglected, sound can play a very important role in stimulating life in a health-promoting environment. Although we have been discussing the role of alarms in healthcare, the noise problem is prevalent in other domains. Work and living spaces routinely disregard acoustics, despite ample evidence on how environmental noise and poor acoustic design impact wellbeing both directly (sleep loss, stress) and indirectly (poor communication, loss of productivity and performance). To promote soundscape design, researchers should investigate the need for sound scaping interventions in a holistic way.

Soundscapes can be modified through passive interventions (e.g., by installing sound-absorbing materials) but also actively, by adding and controlling new sound sources. This active technique operates by masking unwanted sounds with pleasant ones, such as natural sounds (e.g., running water) or abstract minimalist music like Brian Eno's functional music or Satie's *musique d'ameublement*. Hospitals have pioneered the use of generative soundscapes designed by artists and researchers to be soothing to patients and informative to professionals.

This sort of disruptive and transformative innovation should be well weighted with its consequences and contextual acceptance. One suggestion to implement disruptive changes is that of "dividing and conquering", that is, making small, successive improvements.

We proposed a strategy to develop alarms adapted to their context of use; visited and talked to nurses, anesthetists, and surgeons; iterated over different sets of sounds and studied their psychoacoustical effects. There is a hospital soundscape transformation strategy being developed and being agreed upon by human factors researchers, designers, and acousticians. We believe this dissertation is a relevant, even though small, contributor to that global wellbeing goal.

15 References

99percentinvisible.org. (2016). *Separation Anxiety - 99% Invisible*. https://99percentinvisible.org/episode/separation-anxiety/

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, Pub. L. No. IEC 60601-1-8:2006 & A1:2012 (2013).

AAMI - Association for the Advancement of Medical Instrumentation. (2009). Human factors engineering – Design of medical devices Objectives and uses of AAMI standards and recommended practices. In *Medical Instrumentation* (Issue 877).

AAMI - Association for the Advancement of Medical Instrumentation. (2013). *How to Identify the Most Important Alarm Signals to Manage*.

Agus, T. R., Thorpe, S. J., & Pressnitzer, D. (2010). Rapid formation of robust auditory memories: Insights from noise. *Neuron*, *66*(4), 610–618. https://doi.org/10.1016/j.neuron.2010.04.014

Aldrich, K. M., Hellier, E. J., & Edworthy, J. (2009). What determines auditory similarity? The effect of stimulus group and methodology. *Quarterly Journal of Experimental Psychology (2006)*, *62*(1), 63–83. https://doi.org/10.1080/17470210701814451

Alirezaee, P., Duflos, A. W., Schlesinger, J. J., & Cooperstock, J. R. (2020). Exploring the Effectiveness of Haptic Alarm Displays for Critical Care Environments. *2020 IEEE Haptics Symposium (HAPTICS)*, 948–954.

Atyeo, J., & Sanderson, P. M. (2015). Comparison of the identification and ease of use of two alarm sound sets by critical and acute care nurses with little or no music training: A laboratory study. *Anaesthesia*, *70*(7), 818–827. https://doi.org/10.1111/anae.13020

Bach, T. A., Berglund, L.-M., & Turk, E. (2018). Managing alarm systems for quality and safety in the hospital setting. *BMJ Open Quality*, *7*(3), e000202. https://doi.org/10.1136/bmjoq-2017-000202

Baldwin, C. L., & Lewis, B. A. (2014). Perceived urgency mapping across modalities within a driving context. *Applied Ergonomics*, *45*(5). https://doi.org/10.1016/j.apergo.2013.05.002

Ballas, J. A. (1993). Common Factors in the Identification of an Assortment of Brief Everyday Sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(2), 250–267.

Begault, D., & Godfroy, M. (2007). Auditory Alarm Design for NASA CEV Applications. *13th International Conference on Auditory Displays*, 131–138.

Begault, D. R., & Wenzel, E. M. (1993). Headphone localization of speech. *Human Factors*, *35*(2), 361–376.

Begault, Durand R. (2012). Guidelines for NextGen Auditory Displays. *Journal of the Audio Engineering Society*, *60*(7/8), 519–530.

Begault, Durand R. (1993). *3-D Sound for Virtual Reality and Multimedia*. *April*.

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

Bennett, Christopher, Dudaryk, R., Crenshaw, N., Edworthy, J., & McNeer, R. (2019). Recommendation of New Medical Alarms Based on Audibility, Identifiability, and Detectability in a Randomized, Simulation-Based Study. *Critical Care Medicine*, *47*(8), 1. https://doi.org/10.1097/ccm.0000000000003802

Birdja, D., & Özcan, E. (2019). *Better Sleep Experience for the Critically Ill : A Comprehensive Strategy for Designing Hospital Soundscapes*. *Multimodal Technologies and Interaction*. 3(2):36. https://doi.org/10.3390/mti3020036

Block, F. E. (1992). Evaluation of users' abilities to recognize musical alarm tones. *Journal of Clinical Monitoring*, *8*(4), 285–290. https://doi.org/10.1007/BF01617910

Block, F. E. (2008). "For if the trumpet give an uncertain sound, who shall prepare himself to the battle?" (I Corinthians 14:8, KJV). *Anesthesia and Analgesia*, *106*(2), 357–359. https://doi.org/10.1213/ane.0b013e3181606927

Block, F. E., Rouse, J. D., Hakala, M., & Thompson, C. L. (2000). A proposed new set of alarm sounds which satisfy standards and rationale to encode source information. *Journal of Clinical Monitoring and Computing*, *16*(7), 541–546. https://doi.org/10.1023/A:1011471607530

Boersma, P., & Weenink, D. (2019). *Praat: doing phonetics by computer* (Version 6.1.03). http://www.praat.org/

Bogers, K. (2018). *CareTunes. Music as a nurses' monitoring tool - Critical Alarms Lab Lab*. https://delftdesignlabs.org/projects/caretunes/

Bolton, M. L., Edworthy, J., & Boyd, A. D. (2018). A Formal Analysis of Masking Between Reserved Alarm Sounds of the IEC 60601-1-8 International Medical Alarm Standard.

Proceedings of the Human Factors and Ergonomics Society Annual Meeting, *62*(1), 523–527. https://doi.org/10.1177/1541931218621119

Bolton, M. L., Zheng, X., Li, M., Edworthy, J. R., & Boyd, A. D. (2019). An Experimental Validation of Masking in IEC 60601-1-8:2006-Compliant Alarm Sounds. *Human Factors*. https://doi.org/10.1177/0018720819862911

Bones, O., Cox, T. J., & Davies, W. J. (2018). Sound Categories: Category Formation and Evidence-Based Taxonomies. *Frontiers in Psychology*, *9*(July), 1–17. https://doi.org/10.3389/fpsyg.2018.01277

Bones, O., Cox, T. J., & Davies, W. J. (2017). An Evidence-Based Soundscape Taxonomy. *24th International Congress on Sound and Vibration*, 1–7.

Bottalico, P., Piper, R. N., & Legner, B. (2020). LOMBARD EFFECT , AMBIENT NOISE AND WILLINGNESS TO SPEND TIME AND MONEY IN A RESTAURANT AMONG. *EForum Acusticum*, 3055–3062.

Brown, M. (2013). *Surgeon prescribes Brian Eno to patients | Brian Eno | The Guardian*. The Guardian. https://www.theguardian.com/music/2013/apr/18/surgeon-prescribes-brian-eno-to-patients

Brown, Z., Edworthy, J., Sneyd, J. R., & Schlesinger, J. (2015). A comparison of linear and logarithmic auditory tones in pulse oximeters. *Applied Ergonomics*, *51*, 350–357. https://doi.org/10.1016/j.apergo.2015.06.006

Burdick, K. J., Jorgensen, S. K., Holmberg, M. O., Kultgen, S. P., Combs, T. N., & Schlesinger, J. J. (2018). Benefits of sonification and haptic displays with physiologic variables to improve patient safety. *Proceedings of Meetings on Acoustics*, *35*(1), 020001. https://doi.org/10.1121/2.0000941

Burt, J. L., Bartolome, D. S., Burdette, D. W., & Comstock, J. R. (1995). A psychophysiological evaluation of the perceived urgency of auditory warning signals. *Ergonomics*, *38*(11), 2327–2340. https://doi.org/10.1080/00140139508925271

Busch-Vishniac, I. (2019). Hospital Soundscapes: Characterization, Impacts, and Interventions. *Acoustics Today*, *15*(3), 11. https://doi.org/10.1121/at.2019.15.3.11

Busch-Vishniac, I. J. (2019). Next steps in hospital noise research. *PROCEEDINGS of the 23rd International Congress on Acoustics*, 7797–7802.

Busch-vishniac, I., West, J. E., Orellana, D., Barnhill, C., Hunter, T., & Chivukula, R. (2005). Noise levels in Johns Hopkins Hospital. *Journal of the Acoustical Society of America*,

118(6), 3629–3645. https://doi.org/10.1121/1.2118327

Buxton, O. M., Ellenbogen, J. M., Wang, W., Carballeira, A., O'Connor, S., Cooper, D., Gordhandas, A. J., McKinney, S. M., & Solet, J. M. (2012). Sleep disruption due to hospital noises: A prospective evaluation. *Annals of Internal Medicine*, *157*(3), 170–179. https://doi.org/10.7326/0003-4819-156-12-201208070-00472

Buxton, W., Gaver, W., & Bly, S. (1994). Chapter 6 : Auditory icons. In *Audio Interfaces* (Unpublished).

Cabral, J. P., & Remijn, G. B. (2019). Auditory icons: Design and physical characteristics. *Applied Ergonomics*, *78*, 224–239. https://doi.org/10.1016/j.apergo.2019.02.008

Casey, S., Avalos, G., & Dowling, M. (2018). Critical care nurses' knowledge of alarm fatigue and practices towards alarms: A multicentre study. *Intensive and Critical Care Nursing*. https://doi.org/10.1016/j.iccn.2018.05.004

Catchpole, K. R., McKeown, J. D., & Withington, D. J. (2004). Localizable auditory warning pulses. *Ergonomics*, *47*(7), 748–771. https://doi.org/10.1080/00140130310001629739

CCG. (2018). *PATErSoN – Prototyping medical AcousTic Environments for the deSign of New auditory warning signals – CCG*. Www.Ccg.Pt. http://www.ccg.pt/my-product/paterson-prototipagem-acustica-de-ambientes-medicos-para-o-desenho-de-novos-sinais-de-aviso-auditivos/?lang=en

Chan, T., Mills, M., & Sayers, J. (2018). *Optophonic reading, prototyping optophones*. Amodern.Net. http://amodern.net/article/optophonic-reading/

Chen, C. Y. (2019). Characterizing Community Noise in Hospital. *23rd International Congress on Acoustics*, 7847–7854.

Chevret, P., & Parizet, E. (2007). An efficient alternative to the paired comparison method for the subjective evaluation of a large set of sounds. *Proceedings of the 19th International Congress on Acoustics (ICA 2007), Madrid, September*, 1–5.

IEC 61672-2:2013+AMD1:2017 CSV Consolidated version. Electroacoustics - Sound level meters - Part 2: Pattern evaluation tests, (2017).

Council, D. (2007). Eleven lessons: Managing design in eleven global companies-desk research report. *Design Council.*, *0*.

Creswell, J. W., Fetters, M. D., & Ivankova, N. V. (2004). Designing a mixed methods study in primary care. *Annals of Family Medicine*, *2*(1), 7–12. https://doi.org/10.1370/afm.104

Crowston, K. (2012). *Amazon Mechanical Turk: A Research Tool for Organizations and*

Information Systems Scholars. 210–221. https://doi.org/10.1007/978-3-642-35142-6_14

Cvach, M. (2012). Monitor Alarm Fatigue: An Integrative Review. *Biomedical Instrumentation & Technology*, *46*(4), 268–277. https://doi.org/10.2345/0899-8205-46.4.268

Cycling '74. (n.d.). Retrieved July 8, 2021, from https://cycling74.com/

Dalton, P., & Fraenkel, N. (2012). Gorillas we have missed: sustained inattentional deafness for dynamic events. *Cognition*, *124*(3), 367–372. https://doi.org/10.1016/j.cognition.2012.05.012

Darbyshire, J. L., & Young, J. D. (2013). An investigation of sound levels on intensive care units with reference to the WHO guidelines. *Critical Care*, *17*(5), R187. https://doi.org/10.1186/cc12870

Dehais, F., Causse, M., Vachon, F., Regis, N., Menant, E., & Tremblay, S. (2013). Failure to Detect Critical Auditory Alerts in the Cockpit: Evidence for Inattentional Deafness. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *56*(4), 631–644. https://doi.org/10.1177/0018720813510735

Dubois, D. (2000). Categories as acts of meaning: The case of categories in olfaction and audition. *Cognitive Science Quarterly*, *1*(1), 35–68.

Eaton, J. W., Bateman, D., & Søren Hauberg, R. W. (2020). *GNU Octave version 5.2.0 manual: a high-level interactive language for numerical computations*. https://www.gnu.org/software/octave/doc/v5.2.0/

Ecophon. (2018). Noise causes sleep disruption for critical care patients. *Ecophon Group*.

ECRI. (2013). Ecri Institute Honors The Johns Hopkins Hospital For Innovations In Alarm Management. *TechNation*, *March 2013*.

ECRI. (2014). The Path to Alarm Safety. *TechNation*, *March*, 32–33.

ECRI Institute. (2012). *Top 10 Technology Hazards for 2012*. www.ecri.org

ECRI Institute. (2013). *Top ten health technology hazards for 2013*. www.ecri.org

ECRI Institute. (2014). *Top 10 Health Technology Hazards for 2014*. www.ecri.org

ECRI Institute. (2015). *Top 10 Health Technology Hazards for 2015*. www.ecri.org

ECRI Institute. (2016). *Top 10 Health Technology Hazards for 2016*. www.ecri.org

ECRI Institute. (2017). *Top 10 Health Technology Hazards for 2017*. www.ecri.org

ECRI Institute. (2018). *Top 10 Health Technology Hazards for 2018*. www.ecri.org

ECRI Institute. (2019). *Top 10 Health Technology Hazards 2019*. www.ecri.org/2019hazards,

Edworthy, Judy Reed McNeer, R. R., Bennett, Christopher L. Dudaryk, R., McDougall, S. J. P., Schlesinger, J. J., Bolton, M. L., Reed Edworthy, J. . ., Vieira, E. O., Boyd, A. D., Reid, S. K. J., Rayo, M. F., Wrigh, M. C., & Osbor, D. (2018). *Getting alarm sounds into a global standard: a case study with reflections*.

Edworthy, J. R., McNeer, R. R., Bennett, C. L., Dudaryk, R., McDougall, S. J. P., Schlesinger, J. J., Bolton, M. L., Edworthy, J. D. R., Özcan, E., Boyd, A. D., Reid, S. K. J., Rayo, M. F., Wright, M. C., & Osborn, D. (2018). Getting Better Hospital Alarm Sounds Into a Global Standard. *Ergonomics in Design*, 26(4), 4–13. https://doi.org/10.1177/1064804618763268

Edworthy, J., & Hellier, E. (2005). Fewer but better auditory alarms will improve patient safety. *Quality & Safety in Health Care*, *14*(3), 212–215. https://doi.org/10.1136/qshc.2004.013052

Edworthy, J., Hellier, E., Walters, K., Weedon, B., & Adams, A. (2000). The Relationship between Task Performance, Reaction Time, and Perceived Urgency in Nonverbal Auditory Warnings. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *44*(22), 674–677. https://doi.org/10.1177/154193120004402247

Edworthy, J. R. (2015). Audible medical alarms. *Anaesthesia*, *70*(10), 1215–1215. https://doi.org/10.1111/anae.13205

Edworthy, Judy. (1994). The design and implementation of non-verbal auditory warnings. *Applied Ergonomics*, *25*(4), 202–210. https://doi.org/10.1016/0003-6870(94)90001-9

Edworthy, Judy. (2013a). Alarms are still a problem. *Anaesthesia*, *68*, 791–794. http://www.respond2articles.com/ANA/forums/thread/1437.aspx

Edworthy, Judy. (2013b). Medical audible alarms: a review. *Journal of the American Medical Informatics Association : JAMIA*, *20*(3), 584–589. https://doi.org/10.1136/amiajnl-2012-001061

Edworthy, Judy, & Baldwin, C. (2016). Medical audible alarms and IEC 60601-1-8. *Proceedings of the Human Factors and Ergonomics Society*, 634–635. https://doi.org/10.1177/1541931213601145

Edworthy, Judy, & Hards, R. (1999). Learning auditory warnings: The effects of sound type, verbal labelling and imagery on the identification of alarm sounds. *International*

Journal of Industrial Ergonomics, *24*(6), 603–618. https://doi.org/10.1016/S0169-8141(98)00066-3

Edworthy, Judy, Hellier, E., Aldrich, K., & Loxley, S. (2004). Designing Trend-Monitoring Sounds for Helicopters: Methodological Issues and an Application. *Journal of Experimental Psychology: Applied*, *10*(4), 203–218. https://doi.org/10.1037/1076-898X.10.4.203

Edworthy, Judy, Hellier, E., & Hards, R. (1995). The semantic associations of acoustic parameters commonly used in the design of auditory information and warning signals. *Ergonomics*, *38*(11), 2341–2361. https://doi.org/10.1080/00140139508925272

Edworthy, Judy, Hellier, E., Titchener, K., Naweed, A., & Roels, R. (2011). Heterogeneity in auditory alarm sets makes them easier to learn. *International Journal of Industrial Ergonomics*, *41*(2), 136–146. https://doi.org/10.1016/j.ergon.2010.12.004

Edworthy, Judy, Loxley, S., & Dennis, I. (1991). Improving auditory warning design: relationship between warning sound parameters and perceived urgency. *Human Factors*, *33*(2), 205–231. https://doi.org/10.1177/001872089103300206

Edworthy, Judy, Meredith, C., Hellier, E., & Rose, D. (2013). Learning medical alarms whilst performing other tasks. *Ergonomics*, *56*(9), 1400–1417. https://doi.org/10.1080/00140139.2013.819448

Edworthy, Judy, Page, R., Hibbard, A., Kyle, S., Ratnage, P., & Claydon, S. (2014). Learning three sets of alarms for the same medical functions: Aperspective on the difficulty of learning alarms specified in an international standard. *Applied Ergonomics*, *45*(5), 1291–1296. https://doi.org/10.1016/j.apergo.2013.10.003

Edworthy, Judy Reed, McNeer, R. R., Bennett, C. L., Dudaryk, R., McDougall, S. J. P., Schlesinger, J. J., Bolton, M. L., Edworthy, J. D. R., Özcan, E., Boyd, A. D., Reid, S. K. J., Rayo, M. F., Wright, M. C., & Osborn, D. (2018). Getting Better Hospital Alarm Sounds Into a Global Standard. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *26*(4), 4–13. https://doi.org/10.1177/1064804618763268

Edworthy, Judy Reed, Schlesinger, J. J., McNeer, R. R., Kristensen, M. S., & Bennett, C. L. (2017). Classifying alarms: Seeking durability, credibility, consistency, and simplicity. *Biomedical Instrumentation and Technology*, *51*(Horizons), 50–57. https://doi.org/10.2345/0899-8205-51.s2.50

Edworthy, Judy, Reid, S., Edworthy, J., & Hall, S. (2016). The design and testing of novel IEC 60601-1-8 audible alarms. *Human Factors and Ergonomics Society*, *61*(2006), 641–

642.

Edworthy, Judy, Reid, S., McDougall, S., Edworthy, J., Hall, S., Bennett, D., Khan, J., & Pye, E. (2017). The Recognizability and Localizability of Auditory Alarms: Setting Global Medical Device Standards. *Human Factors*, *59*(7), 1108–1127. https://doi.org/10.1177/0018720817712004

Edworthy, Judy, Reid, S., Peel, K., Lock, S., Williams, J., Newbury, C., Foster, J., & Farrington, M. (2018). The impact of workload on the ability to localize audible alarms. *Applied Ergonomics*, *72*(October 2017), 88–93. https://doi.org/10.1016/j.apergo.2018.05.006

Edworthy, Judy, & Stanton, N. (1995). A user-centred approach to the design and evaluation of auditory warning signals: 1. Methodology. *Ergonomics*, *38*(11), 2262–2280. https://doi.org/10.1080/00140139508925267

Engber, D. (2014). *Who Made That Weather Icon?* The New York Times. https://www.nytimes.com/2014/05/25/magazine/who-made-that-weather-icon.html

Everest, F. A., & Pohlmann, K. C. (2015). Mater Handbook of Acoustics. In *Aging* (Vol. 7, Issue 11). https://doi.org/10.1017/CBO9781107415324.004

Fallis, A. . (2009). Acoustics and Psychoacoustics. In *Focal Press* (Vol. 53, Issue 9). https://doi.org/10.1017/CBO9781107415324.004

Finley, G. a, & Cohen, a J. (1991). Perceived urgency and the anaesthetist: responses to common operating room monitor alarms. *Canadian Journal of Anaesthesia = Journal Canadien d'anesthesie*, *38*(8), 958–964. https://doi.org/10.1007/BF03008611

Fitch, W. T., & Rosenfeld, A. J. (2007). Perception and Production of Syncopated Rhythms. *Music Perception: An Interdisciplinary Journal*, *25*(1), 43–58. https://doi.org/10.1525/mp.2007.25.1.43

Frimalm, R., Fagerlönn, J., Lindberg, S., & Sirkka, A. (2014). How Many Auditory Icons, in a Control Room Environment, Can You Learn? *The 20th International Conference on Auditory Display (ICAD-2014)*. http://smartech.gatech.edu/handle/1853/52084

Fritsch, M. H., Chacko, C. E., & Patterson, E. B. (2010). Operating Room Sound Level Hazards for Patients and Physicians. *Otology & Neurotology*, *31*, 715–721. papers2://publication/uuid/AF04AD30-F6A7-43CD-8898-7C378CD7D35D

Gaver, W. W. (1987). Auditory Icons. *ACM SIGCHI Bulletin*, *19*(1), 74.

https://doi.org/10.1145/28189.1044809

Gescheider, G. A. (1997). *Psychophysics: The Fundamentals* (Third Edit). Lawrence Erlbaum Associates.

Ginsberg, S. H., Pantin, E., Kraidin, J., Solina, A., Panjwani, S., & Yang, G. (2013). Noise levels in modern operating rooms during surgery. *Journal of Cardiothoracic and Vascular Anesthesia*, *27*(3), 528–530. https://doi.org/10.1053/j.jvca.2012.09.001

Giordano, B. L., McDonnell, J., & McAdams, S. (2010). Hearing living symbols and nonliving icons: Category specificities in the cognitive processing of environmental sounds. *Brain and Cognition*, *73*(1), 7–19. https://doi.org/10.1016/J.BANDC.2010.01.005

Goldstein, E. B. (2007). *Sensation and Perception* (7th ed.). Thomson Wadsworth.

Gonzalez, C., Lewis, B., & Baldwin, C. (2012). *Revisiting pulse rate, frequency and perceived urgency: have relationships changed and why?* 44–51. http://smartech.gatech.edu/handle/1853/44420

Graham, R. (1999). Use of auditory icons as emergency warnings: evaluation within a vehicle collision avoidance application. *Ergonomics*, *42*(9), 1233–1248. https://doi.org/10.1080/001401399185108

Gross, B., Dahl, D., & Nielsen, L. (2011). Physiologic monitoring alarm load on medical/surgical floors of a community hospital. *Biomedical Instrumentation and Technology*, *45*(SPRING), 29–36. https://doi.org/10.2345/0899-8205-45.s1.29

Guastavino, C. (2007). Categorization of environmental sounds. *Canadian Journal of Experimental Psychology*, *61*(1), 54–63. https://doi.org/10.1037/cjep2007006

Guillaume, A, Drake, C., Rivenez, M., Pellieux, L., & Chastres, V. (2002). Perception of urgency and alarm design. *Proceedings of the 8th International Conference on Auditory Display (ICAD2002)*. Proceedings/2002/GuillaumeDrake2002.pdf

Guillaume, Anne, Pellieux, L., Chastres, V., & Drake, C. (2003). Judging the urgency of nonvocal auditory warning signals: perceptual and cognitive processes. *Journal of Experimental Psychology. Applied*, *9*(3), 196–212. https://doi.org/10.1037/1076-898X.9.3.196

Gygi, B., Kidd, G. R., & Watson, C. S. (2007). Similarity and categorization of environmental sounds. *Perception and Psychophysics*, *69*(6), 839–855. https://doi.org/10.3758/BF03193921

Haas, E., & Edworthy, J. (1996). Designing urgency into auditory warnings using pitch,

speed and loudness. *Computing & Control Engineering Journal*, *1*, 193–198. http://digital-library.theiet.org/content/journals/10.1049/cce_19960407

Hasanain, B., Boyd, A. D., Edworthy, J., & Bolton, M. L. (2017). A formal approach to discovering simultaneous additive masking between auditory medical alarms. *Applied Ergonomics*, *58*(2008), 500–514. https://doi.org/10.1016/j.apergo.2016.07.008

Haverkamp, M. (2009). Synesthetic Design – Building Multi-sensory Arrangements. *Audio Branding*, 164–181. https://doi.org/10.5771/9783845216935-164

Hellier, E., & Edworthy, J. (1999). On using psychophysical techniques to achieve urgency mapping in auditory warnings. *Applied Ergonomics*, *30*(2), 167–171. https://doi.org/10.1016/S0003-6870(97)00013-6

Hellier, E. J., Edworthy, J., & Dennis, I. (1993). Improving auditory warning design: quantifying and predicting the effects of different warning parameters on perceived urgency. *Human Factors*, *35*(4), 693–706. https://doi.org/10.1177/001872089303500408

Hellier, Elizabeth, & Edworthy, J. (1989). Quantifying the perceived urgency of auditory warnings. *Canadian Acoustics*, *17*(4), 3–11.

Hellier, Elizabeth, Edworthy, J., & Dennis, I. (1995). A comparison of different techniques for scaling perceived urgency. *Ergonomics*, *38*(4), 659–670. https://doi.org/10.1080/00140139508925138

Hermann, T., Hunt, A., & Neuhoff, J. (2011). *The Sonification Handbook*. Berlin: Logos Verlag.

Ho, C., & Spence, C. (2005). Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. *Journal of Experimental Psychology. Applied*, *11*(3), 157–174. https://doi.org/10.1037/1076-898X.11.3.157

Hospitals struggle to address terrifying "ICU delirium." (n.d.). Retrieved July 10, 2019, from https://www.statnews.com/2016/10/14/icu-delirium-hospitals/

Houix, O., Lemaitre, G., Misdariis, N., Susini, P., & Urdapilleta, I. (2012). A lexical analysis of environmental sound categories. *Journal of Experimental Psychology: Applied*, *18*(1), 52–80. https://doi.org/10.1037/a0026240

Hravnak, M., Pellathy, T., Chen, L., Dubrawski, A., Wertz, A., Clermont, G., & Pinsky, M. R. (2018). A call to alarms: Current state and future directions in the battle against alarm fatigue. In *Journal of Electrocardiology* (Vol. 51, Issue 6, pp. S44–S48). Churchill

Livingstone Inc. https://doi.org/10.1016/j.jelectrocard.2018.07.024

Hsu, T., Ryherd, E. E., Waye, K. P., & Ackerman, J. (2012). Noise pollution in hospitals: Impact on patients. *Journal of Clinical Outcomes Management*, *19*(7), 301–309.

Hsu, T. Y. (2012). *Relating Acoustics and Human Outcome Measures in Hospitals*. Georgia Institute of Technology.

Hudson, W. (2014). Card Sorting. In R. F. Soegaard, Mads and Dam (Ed.), *The Encyclopedia of Human-Computer Interaction* (2nd ed.). The Interaction Design Foundation. https://www.interaction-design.org/encyclopedia/card_sorting.html

Hulland, T., Su, A., & Kingan, M. (2020). Noise in an inpatient hospital ward in New Zealand. *Building Acoustics*. https://doi.org/10.1177/1351010X20916120

IEC 60601-1-8 - 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 equipment and medical electric, Pub. L. No. IEC 60601-1-8, 62 (2020).

International Organization for Standardization. (2010). Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems. In *International Standard ISO 9241-210*. https://doi.org/10.5100/jje.30.1

SAE J2830 Process for Comprehension Testing of in-Vehicle Icons, Pub. L. No. J2830 (2008).

International Sound Awards. (2019). *Helsinki Childrens Hospital - ISA - International Sound Awards*. https://www.international-sound-awards.com/helsinki-childrens-hospital/

Ishihara, S. (2011). Psychological Methods of Kansei Engineering. In M. Nagamachi (Ed.), *Kansei/Affective Engineering* (pp. 31–38). CRC Press.

ISO 7731 - Ergonomics - Danger signals for public and work areas — Auditory danger signals, 2003 (2003).

ISO. (2014a). Graphical symbols - Test methods - Part 1: Method for testing comprehensibility. *ISO 9186-1/2014*, *2014*.

ISO. (2014b). Graphical symbols - Test methods - part 3: Method for testing symbol referent association. *ISO 9186-3/2014*, *2014*.

Johansson, L., Bergbom, I., Waye, K. P., Ryherd, E., & Lindahl, B. (2012). The sound environment in an ICU patient room-A content analysis of sound levels and patient experiences. *Intensive and Critical Care Nursing*, *28*(5), 269–279.

https://doi.org/10.1016/j.iccn.2012.03.004

Katzman, N., Gellert, M., Schlesinger, J. J., Oron-Gilad, T., Cooperstock, J. R., & Bitan, Y. (2019). Evaluation of tactile cues for simulated patients' status under high and low workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 658–662. https://doi.org/10.1177/1071181319631285

Keller, S., Tschan, F., Beldi, G., Kurmann, A., Candinas, D., & Semmer, N. K. (2016). Noise peaks influence communication in the operating room. An observational study. *Ergonomics*, *0139*(May), 1–12. https://doi.org/10.1080/00140139.2016.1159736

Kerr, J. H. (1985). Warning devices. *British Journal of Anaesthesia*, *57*(7), 696–708. https://doi.org/10.1093/bja/57.7.696

Konkani, A., Oakley, B., & Bauld, T. J. (2012). Reducing hospital noise: A review of medical device alarm management. In *Biomedical Instrumentation and Technology* (Vol. 46, Issue 6, pp. 478–487). https://doi.org/10.2345/0899-8205-46.6.478

Koreimann, S., Gula, B., & Vitouch, O. (2014). Inattentional deafness in music. *Psychological Research*, *78*(3), 304–312. https://doi.org/10.1007/s00426-014-0552-x

Korniewicz, B. D. M., Clark, T., & David, Y. (2008). A National Online Survey On The Effectiveness Of Clinical Alarms. *American Journal of Critical Care*, *17*(1), 36–42.

Kracht, J. M., Busch-Vishniac, I. J., & West, J. E. (2007). Noise in the operating rooms of Johns Hopkins Hospital. *The Journal of the Acoustical Society of America*, *121*(5), 2673–2680. https://doi.org/10.1121/1.2714921

Kristensen, M. S., Edworthy, J., & Özcan, E. (2016). Alarm fatigue in the ward: An acoustical problem? *Sound Effects*, *6*(1), 89–104.

Kuzmich, G. a, Phillips, M., & Rojas, L. (2001). Noise Levels in the Operating Room. In *Noise Levels in the Operating Room*.

Lacherez, P., Donaldson, L., & Burt, J. S. (2016). Do Learned Alarm Sounds Interfere With Working Memory? *Human Factors*, 0018720816662733. https://doi.org/10.1177/0018720816662733

Lacherez, P., Limin Seah, E., & Sanderson, P. (2007). Overlapping Melodic Alarms Are Almost Indiscriminable. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*(4), 637–645. https://doi.org/10.1518/001872007X215719

Lawson, N., Thompson, K., Saunders, G., Saiz, J., Richardson, J., Brown, D., Ince, N., Caldwell, M., & Pope, D. (2010). Sound intensity and noise evaluation in a critical care

unit. *American Journal of Critical Care*, *19*(6). https://doi.org/10.4037/ajcc2010180

Lazarevic, G. (2015). *SONIC ENVIRONMENTS FOR HEALING – A HOOU project at the HfMT Hamburg*. https://hss.hfmt-hamburg.de/?page_id=24&lang=en

Lee, J. D., McGehee, D. V., Brown, T. L., & Reyes, M. L. (2002). Collision Warning Timing, Driver Distraction, and Driver Response to Imminent Rear-End Collisions in a High-Fidelity Driving Simulator. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *44*(2), 314–334. https://doi.org/10.1518/0018720024497844

Lemaitre, G., Houix, O., Misdariis, N., & Susini, P. (2010). Listener expertise and sound identification influence the categorization of environmental sounds. *Journal of Experimental Psychology: Applied*, *16*(1), 16–32. https://doi.org/10.1037/a0018762

Lemmens, P., Bussemakers, M., & Haan, A. De. (2001). Effects of auditory icons and earcons on visual categorization: the bigger picture. Georgia Institute of Technology.

Ligges, U., Krey, S., Mersmann, O., & Schnackenberg, S. (2016). *tuneR: Analysis of music*. http://r-forge.r-project.org/projects/tuner/

LimeSurvey GmbH. (2018). *LimeSurvey: An Open Source survey tool*.

Logan, M. K. (2011). Signal to noise. *Horizons*, *Spring: 1*.

Lokki, T. (2014). Tasting music like wine: Sensory evaluation of concert halls. *Physics Today*, *67*(1), 27–32. https://doi.org/10.1063/PT.3.2242

Lokki, T., Vertanen, H., Kuusinen, A., Pätynen, J., & Tervo, S. (2010). Auditorium acoustics assessment with sensory evaluation methods. *Proceedings of International Symposium on Room Acoustics, Melbourne, Australia*, *August*, 1–10.

Lopes, S. G. (2005). *Avaliação dos níveis de pressão sonora em salas de operação de um centro cirúrgico geral* [Universidade do Vale do Paraíba]. http://biblioteca.univap.br/dados/000001/00000155.pdf

Lorenz, B., Peters, J., & Frey, U. H. (2017). How Much Alarm Can the Human Being Tolerate? In *Anasthesiologie Intensivmedizin Notfallmedizin Schmerztherapie* (Vol. 52, Issues 7–8, pp. 564–570). Georg Thieme Verlag KG. https://doi.org/10.1055/s-0042-118618

MacDonald, A., & Schlesinger, J. (2018). Canary in an operating room: integrated operating room music. In D. Waard, F. Di Nocera, D. Coelho, J. Edworthy, K. Brookhuis, F. Ferlazzo, T. Franke, & A. Toffetti (Eds.), *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2017 Annual Conference* (pp. 79–83).

Macdonald, J. S. P., & Lavie, N. (2011). Visual perceptual load induces inattentional deafness. *Attention, Perception & Psychophysics*, *73*(6), 1780–1789. https://doi.org/10.3758/s13414-011-0144-4

Madaras, G. S., & Apps, N. (2012). *2012 Webinar Series The Hospital Noise Project: Lessons on Addressing Noise from 241 U.S. Hospitals*. https://cdn.ymaws.com/www.theberylinstitute.org/resource/resmgr/webinar_pdf/final_-_noise_webinar.pdf

Marshall, D. C., Lee, J. D., & Austria, R. A. (2007). Alerts for in-vehicle information systems: annoyance, urgency, and appropriateness. *Human Factors*, *49*(1), 145–157. https://doi.org/10.1518/001872007779598145

McIntire, J. P., Havig, P. R., Watamaniuk, S. N. J., & Gilkey, R. H. (2010). Visual Search Performance With 3-D Auditory Cues: Effects of Motion, Target Location, and Practice. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *52*(1), 41–53. https://doi.org/10.1177/0018720810368806

McKinney, M. (2014). Hospital's simple interventions help reduce alarm fatigue. *Modern Healthcare*, *44*(5), 26–27.

McNeer, R. R., Horn, D. B., Bennett, C. L., Edworthy, J. R., & Dudaryk, R. (2018). Auditory Icon Alarms Are More Accurately and Quickly Identified than Current Standard Melodic Alarms in a Simulated Clinical Setting. *Anesthesiology*, 1–9. https://doi.org/10.1097/ALN.0000000000002234

McNulty, E., Brown, D., Santomauro, C., McLanders, M., Tran, J., & Sanderson, P. (2016). Vibrotactile Displays of Pulse Oximetry: Exploratory Studies of Three Novel Designs. *Proceedings of the Human Factors and Ergonomics Society*, *60*(1), 1556. https://doi.org/10.1177/1541931213601359

Meredith, C., & Edworthy, J. (1995). Are there too many alarms in the intensive care unit? An overview of the problems. *Journal of Advanced Nursing*, *21*(1), 15–20. https://doi.org/10.1046/j.1365-2648.1995.21010015.x

Mhlaba, J. M., Christianson, L. W., Davidson, S. J., Graves, S. N., Still, B. R., Silas, M. R., Fong, A. J., Nassiri, A. M., Pariser, J. J., & Langerman, A. J. (2016). Field Research in the Operating Room. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *October*, 10–19. https://doi.org/10.1177/1064804616642916

Miguel, A. S. (2000). *Manual de Higiene e Segurança do Trabalho* (5th ed.). Porto Editora.

Mills, M. (2015). *Optophones and Musical Print*. Sounding Out! https://soundstudiesblog.com/2015/01/05/optophones-and-musical-print/

Momtahan, K., Hétu, R., & Tansley, B. (1993). Audibility and identification of auditory alarms in the operating room and intensive care unit. *Ergonomics*, *36*(10), 1159–1176. https://doi.org/10.1080/00140139308967986

Momtahan, K. L. (1991). *Mapping of psychoacoustic parameters to the perceived urgency of auditory warning signals* (Doctoral dissertation, Carleton University).

Mondor, T. a, & Finley, G. A. (2003). The perceived urgency of auditory warning alarms used in the hospital operating room is inappropriate. *Canadian Journal of Anaesthesia = Journal Canadien d'anesthesie*, *50*(3), 221–228. https://doi.org/10.1097/01.sa.0000108474.38017.92

Morel, J., Marquis-Favre, C., Dubois, D., & Pierrette, M. (2012). Road traffic in urban areas: A perceptual and cognitive typology of pass-by noises. *Acta Acustica United with Acustica*, *98*(1), 166–178. https://doi.org/10.3813/AAA.918502

Mynatt, E. D. (1994). Designing with auditory icons. *Conference Companion on Human Factors in Computing Systems - CHI '94*, 269–270. https://doi.org/10.1145/259963.260483

Nelson, W., Bolia, R., Ericson, M., & McKinley, R. (1998). Monitoring the simultaneous presentation of spatialized speech signals in a virtual acoustic environment. In *(No. ASC-98-1409). AIR FORCE RESEARCH LAB WRIGHT-PATTERSON AFB OH HUMAN EFFECTIVENESS DIRECTORATE.* (Vol. 166, Issue June).

Nelson, W. T., Bolia, R. S., Ericson, M. A., & McKinley, R. L. (1999). Spatial Audio Displays for Speech Communications: A Comparison of Free Field and Virtual Acoustic Environments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *43*(22), 1202–1205. https://doi.org/10.1177/154193129904302207

Nightingale, F. (1859). Notes on Nursing. *The American Journal of Nursing*, *47*(7), 508. https://doi.org/10.2307/3456987

Norman, D. A., & Stappers, P. J. (2015). DesignX: Complex Sociotechnical Systems. *She Ji: The Journal of Design, Economics, and Innovation*, *1*(2), 83–106. https://doi.org/10.1016/j.sheji.2016.01.002

Osha.gov. (n.d.). *How loud is too loud?* Retrieved May 10, 2020, from https://www.osha.gov/SLTC/noisehearingconservation/loud.html

Özcan-Vieira, E. (2008). *Product Sounds - Fundamentals & Applications*. http://repository.tudelft.nl/view/ir/uuid:8144d042-021f-4fa0-809b-1e5b3b72647b/

Ozcan, E., Birdja, D., & Edworthy, J. R. (2018). A holistic and collaborative approach to audible alarm design. *Biomedical Instrumentation and Technology*, *52*(6), 422–432. https://doi.org/10.2345/0899-8205-52.6.422

Özcan, E., Birdja, D., Simonse, L., & Struijs, A. (2018). Alarm in the ICU ! Envisioning Patient Monitoring and Alarm Management in Future Intensive Care Units. In *Service Design and Service Thinking in Healthcare and Hospital Management* (pp. 421–446).

Özcan, E., Chou, C., Bogers, K., & Van Der Helm, A. (2021). CARETUNES: TURNING PATIENT VITALS INTO MUSIC. *SMC Conference*.

Parizet, E., Hamzaoui, N., & Sabatié, G. (2005). Comparison of some listening test methods: A case study. *Acta Acustica United with Acustica*, *91*(2), 356–364.

Paterson, E., Sanderson, P. M., Paterson, N. A. B., Liu, D., & Loeb, R. G. (2016). The effectiveness of pulse oximetry sonification enhanced with tremolo and brightness for distinguishing clinically important oxygen saturation ranges: A laboratory study. *Anaesthesia*, *71*(5), 565–572. https://doi.org/10.1111/anae.13424

Patnaik, D. (2009). *Forget Design Thinking and Try Hybrid Thinking*. Fast Company. https://www.fastcompany.com/1338960/forget-design-thinking-and-try-hybrid-thinking

Patnaik, D., & Becker, R. (1999). Needfinding: The Why and How of Uncovering People's Needs. *Design Management Journal (Former Series)*, *10*(2), 37–43. https://doi.org/10.1111/j.1948-7169.1999.tb00250.x

Patterson, R D, Mayfield, T. F. (1990). Auditory warning sounds in the work environment. *Philosophical Transactions of the Royal Society of London*, *327*(1241), 485–492. https://doi.org/10.1098/rstb.1990.0091

Patterson, R. D. (1982). *Guidelines for Auditory Warning Systems on Civil Aircraft*. Civil Aviation Authority.

Patterson, R., Edworthy, J., & Shailer, M. (1986). *Alarm sounds for medical equipment in intensive care areas and operating theatres*. University of Southampton, Institute of Sound and Vibration Research. .

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/J.JNEUMETH.2006.11.017

Pereira, F., Marques, R., & Vieria, J. (2020, May). Auditory Alarms Design Tool: Spectral Masking Estimation Based on a Psychoacoustic Model. In *Meeting of Research in Music, Arts and Design* (pp. 621-639). Springer, Cham..

Public Domain Review. (2020). *Charles Babbage's Observations on Street Nuisances (1864) – The Public Domain Review*. Public Domain Review. https://publicdomainreview.org/collection/observations-on-street-nuisances-charles-babbage

Qualtrics. (2021). *Qualtrics*. https://www.qualtrics.com

Quinn, M. (2017). Modular operating rooms – how do they really sound? *Acoustic Bulletin by Ecophon*.

R Foundation for Statistical Computing. (2017). *R: A language and environment for statistical computing*. https://www.r-project.org/

Rueb, E. S. (2019). To Reduce Hospital Noise , Researchers Create Alarms That Whistle and Sing. *The New York Times*. https://www.nytimes.com/2019/07/09/science/alarm-fatigue-hospitals.html?fbclid=IwAR3hUB8sm_jnxqK4T29PT-flG-HdUayLjS2lkX099Ql8TZyAABD8g5rE6ek

Ruiter, E. De. (2015). Healing soundscape: hospital acoustics 2.0. *Euronoise2015*, 2439–2444.

Ryherd, E. E., West, J. E., Ackerman, J., Zimring, C., & Waye, K. P. (2012). Evaluating and Improving Hospital Soundscapes. *Inter.Noise2012*, 2512–2517.

Ryherd, E. E., West, J. E., Busch-Vishniac, I., & Waye, K. P. (2008). Evaluating the Hospital Soundscape. *Acoustics Today*, *4*(4), 22. https://doi.org/10.1121/1.3058487

Sandau, K. E., Funk, M., Auerbach, A., Barsness, G. W., Blum, K., Cvach, M., Lampert, R., May, J. L., McDaniel, G. M., Perez, M. V., Sendelbach, S., Sommargren, C. E., & Wang, P. J. (2017). Update to Practice Standards for Electrocardiographic Monitoring in Hospital Settings: A Scientific Statement From the American Heart Association. In *Circulation* (Vol. 136, Issue 19). https://doi.org/10.1161/CIR.0000000000000527

Sanderson, P. M. (2009). Auditory alarms for medical equipment: How do we ensure they convey their meanings? *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *53*(4), 264–268. https://doi.org/10.1518/107118109X12524441080380

Sanz-Segura, R. (2021). *Safe & Sound: Laying the Foundation for Alarm Design in Critical*

Contexts. Universidad de Zaragoza.

Schafer, R. M. (1977). *The soundscape: Our sonic environment and the tuning of the world*. Destiny Books.

Schlesinger, J. J., Reynolds, E., Sweyer, B., & Pradhan, A. (2017). *Frequency-Selective Silencing Device for Digital Filtering of Audible Medical Alarm Sounds to Enhance ICU Patient Recovery*. 95–100. https://doi.org/10.21785/icad2017.062

Schutz, M., Stefanucci, J. K., H. Baum, S., & Roth, A. (2017). Name that percussive tune: Associative memory and amplitude envelope†. *Quarterly Journal of Experimental Psychology*, *70*(7), 1323–1343. https://doi.org/10.1080/17470218.2016.1182562

Science Daily. (2012). Major study links chronic noise exposure to risk of heart attacks. *Choice Reviews Online*, *50*(02), 50-0848-50–0848. https://www.sciencedaily.com/releases/2005/11/051124111916.htm

Semiconductors, N. (2009). *AN10875 IEC 60601-1-8 audible alert generator using the LPC1700* (Patent No. AN10875).

Sen, Y. (2020). *The Beepers and the Beeped*. www.vimeo.com. https://vimeo.com/479496240

Sen, Y. K., & Sen, A. (2020). The Future of Hospital Sound: Transforming Healthcare Through Sound Experience. *Music and Medicine*, *12*(1), 60–64.

Similowski, C., Souppart, V., Azoulay, E., Meli, A., Coppola, S., Chiumello, D., Gommers, D., Roepke, R., Ranzani, O., Deprivation, S., Unit, I. C., Darbyshire, J., & Greig, P. (2019). The Night in the ICU. *ICU Management & Practice*, *20*(2).

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059–1074. http://www.ncbi.nlm.nih.gov/pubmed/10694957

Smith, S., Stephan, K., & Parker, S. (2004). *Auditory warnings in the military cockpit: A preliminary evaluation of potential sound types*. http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA444013

Spence, C., & Keller, S. (2019). Medicine's Melodies: On the Costs & Benefits of Music, Soundscapes, & Noise in Healthcare Settings. *Music and Medicine*, *11*(4), 211. https://doi.org/10.47513/mmd.v11i4.699

Srinivasa, E., Mankoo, J., & Kerr, C. (2017). An Evidence-Based Approach to Reducing

Cardiac Telemetry Alarm Fatigue. *Worldviews on Evidence-Based Nursing*, *14*(4), 265–273. https://doi.org/10.1111/wvn.12200

Stanton, N.A. (Ed.) (1994). *Human Factors in Alarm Design.* CRC Press.

Sueur, J., Aubin, T., & Simonis, C. (2008). SEEWAVE, a free modular tool for sound analysis and synthesis. *Bioacoustics*, *18*(2), 213–226. https://doi.org/10.1080/09524622.2008.9753600

Suied, C., Susini, P., & McAdams, S. (2008). Evaluating warning sound urgency with reaction times. *Journal of Experimental Psychology. Applied*, *14*(3), 201–212. https://doi.org/10.1037/1076-898X.14.3.201

Syntagm. (2018). *SynCaps V3*. https://syntagm.co.uk/design/cardsortdl.shtml

Thangavelu, S. D., Ifeachor, E., Edworthy, J., Yunus, J., & Chinna, K. (2014). Challenges and recommendation of clinical alarm system in intensive care units from user perspective. *2014 Ieee Region 10 Symposium*, 366–369. https://doi.org/10.1109/TENCONSpring.2014.6863059

Toor, O., Ryan, T., & Richard, M. (2008). Auditory masking potential of common operating room sounds: A psychoacoustic analysis. *Anesthesiology*, 109(A1207), A1207.

Trzeciak, S., Gaughan, J. P., Bosire, J., & Mazzarelli, A. J. (2016). Association Between Medicare Summary Star Ratings for Patient Experience and Clinical Outcomes in US Hospitals. *Journal of Patient Experience*, *3*(1), 6–9. https://doi.org/10.1177/2374373516636681

Tsiou, C., Efthymiatos, G., & Katostaras, T. (2008). Noise in the operating rooms of Greek hospitals. *Journal of the Acoustical Society of America*, *2*(February), 757–766. https://doi.org/10.1121/1.2821972

Ulfvengren, P. (2003). *Design of natural warning sounds in human-machine systems* [Stockholm Royal Institute of Technology]. http://www.diva-portal.org/smash/record.jsf?pid=diva2:9515

Ulrich, R. S. (1984). View through a window may influence recovery from surgery. *Science*, *224*(4647), 420–421.

Us, T., Jensen, N., Lind, M., & Jørgensen, S. (2011). Fundamental principles of alarm design. *International Journal of Nuclear Safety and Simulation*, *2*(1), 44–51. http://orbit.dtu.dk/fedora/objects/orbit:27022/datastreams/file_5537608/content

Vanderveer, N. J. (1979). Confusion errors in identification of environmental sounds.

Journal of the Acoustical Society of America, *65*(S60).

VERBI Software. (2010). *MAXQDA 10 [computer software]*. Available from maxqda.com

Vicente, K. J. (2002). Ecological interface design: Progress and challenges. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *44*(1), 62–78. https://doi.org/10.1518/0018720024494829

Vieira, J., Osório, J. M. A., Mouta, S., Delgado, P., Portinha, A., Meireles, J. F., & Santos, J. A. (2017). Kansei engineering as a tool for the design of in-vehicle rubber keypads. *Applied Ergonomics*, *61*. https://doi.org/10.1016/j.apergo.2016.12.019

Vieira, Joana, Santos, J. A., & Noriega, P. (2018). Using semantic differential scales to assess the subjective perception of auditory warning signals. *Proceedings of the 21st International Conference on Digital Audio Effects (DAFx-18)*, 126–132.

Von Helmholtz, H. (1912). *On the Sensations of Tone as a Physiological Basis for the Theory of Music.* Longmans, Green.

Walker, B., & Nees, M. (2011). Theory of Sonification. *Principles of Sonification: An Introduction to Auditory Display and Sonification*, 1–32.

Wallace, J. S., & Fisher, D. L. (1998). Sound localization: information theory analysis. *Human Factors*, *40*(1), 50–68. http://www.ncbi.nlm.nih.gov/pubmed/9579104

Watson, M. O., & Sanderson, P. (2007). Designing for attention with sound: challenges and extensions to ecological interface design. *Human Factors*, *49*(2), 331–346. https://doi.org/10.1518/001872007X312531

Wee, A. N., & Sanderson, P. M. (2008). Are melodic medical equipment alarms easily learned? *Anesthesia and Analgesia*, *106*(2), 501–508. https://doi.org/10.1213/01.ane.0000286148.58823.6c

White, B. L., & Zomorodi, M. (2017). Perceived and actual noise levels in critical care units. *Intensive and Critical Care Nursing*, *38*, 18–23. https://doi.org/10.1016/j.iccn.2016.06.004

WHO (World Health Organization). (1999). Guidelines for community noise. *World Health Organization, Geneva. ISBN: 9780203882597*, 1–161. https://doi.org/10.1260/0957456001497535

Wiegner, M. B. (1991). Proposed new alarm standards may make a bad situation worse (letter). *Anesthesiology*, *74*, 732–793. https://doi.org/10.1167/8.5.1.

Wilken, M., Hüske-Kraus, D., Klausen, A., Koch, C., Schlauch, W., & Röhrig, R. (2017).

Alarm fatigue: Causes and effects. *Studies in Health Technology and Informatics*, *243*, 107–111. https://doi.org/10.3233/978-1-61499-808-2-107

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, and Psychophysics*, *79*(7), 2064–2072. https://doi.org/10.3758/s13414-017-1361-2

Zwicker, E., & Fastl, H. (2013). *Psychoacoustics: Facts and models (Vol. 22)*. Springer Science & Business Media.

16 APPENDICES

Appendix 1 - Glossary.....285

Appendix 2 –Interview with healthcare professionals: Script.....288

Appendix 3 – Online questionnaire for healthcare professionals.....290

Appendix 4 - Stimuli used in Subjective perception of auditory warning signals 293

Appendix 5 – Stimuli used in Categorization of Sounds.....305

Appendix 6 – Stimuli used in Perception of Urgency 307

Appendix 7 – Final library of Clinical Auditory Alarms 309

Appendix 8 – General organization of studies, timeframe and no. of participants.....316

APPENDIX 1 - GLOSSARY

Amplitude envelope - changes in the amplitude of a sound over time and is an influential property as it affects our perception of timbre (https://maplelab.net/overview/amplitude-envelope).

Burst - group of pulses with a distinctive rhythm or pattern (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).

Duty cycle - the proportion of time the sound is present.

Fall Time or Offset - interval over which the pulse amplitude decreases from 90 % to 10 % of its maximum (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).

Frequency (f) - the number of times that a periodic function or vibration occurs or repeats itself in a specified time, often 1 second - cycles per second. It is usually measured in Hertz (Hz) (www.acoustic-glossary.co.uk).

Frequency Analysis - analysing an overall broadband noise to identify the different contributions in different parts of the audio spectrum. Typically the analysis in made using octave, one-third octave or narrow band (FFT) Analysis (www.acoustic-glossary.co.uk).

Frequency band - Continuous range of frequencies between two limiting frequencies (www.acoustic-glossary.co.uk).

Frequency range - A continuous range or spectrum of frequencies that extends from one limiting frequency to another. The frequency range of 20 Hz to 20 kHz is conventionally referred to as the human audible range.

Fundamental frequency (f_0) - first harmonic of a pulse (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).

Harmonics - A harmonic is a wave with a frequency that is a positive integer multiple of the frequency of the original wave, known as the fundamental frequency. The original wave is also called the 1st harmonic, the following harmonics are known as higher harmonics. As all harmonics are periodic at the fundamental frequency, the sum of harmonics is also periodic at that frequency. For example, if the fundamental frequency is 50 Hz, a common AC power supply frequency, the frequencies of the first three higher harmonics are 100 Hz (2nd harmonic), 150 Hz (3rd harmonic), 200 Hz (4th harmonic) and any addition of waves with these frequencies is periodic at 50 Hz. (Wikipedia)

Interburst interval - period of time between the end of the last pulse of a burst and the start of the first pulse of the next burst of the same alarm signal (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).

Inter-onset interval (IOI) - the time elapsed from the onset of a pulse to the onset of the next

Interpulse Interval (IPI) – Interval between pulses, measured in time units (e.g.: ms).

Maskee – a sound that renders other inaudible to a human listener is called masker and the inaudible one is the maskee (Fallis, 2009).

Masker - a sound that renders other inaudible to a human listener (Fallis, 2009).

Masking (spectral) - Noise can have much more power than a target signal, but if the spectral regions of noise and target are sufficiently distant, the target will be audible. On the other hand, if both regions coincide, noise needs only to have residual more power than the target to mask it. This is the fundamental of spectral masking of two simultaneous sounds (Zwicker & Fastl, 2013).

Modulation - a process by which a spectral component is modified by another component and sidebands result. Sidebands are pairs of frequencies with similar amplitude that appear equally spaced on either side of a centre frequency (www.acoustic-glossary.co.uk).

Offset or Fall Time - interval over which the pulse amplitude decreases from 90 % to 10 % of its maximum (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).

Onset or Rise Time - interval over which the pulse increases from 10 % to 90 % of its maximum amplitude (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).

Power (Sound) - the total sound energy radiated by a sound source and measured in watts (W) (www.acoustic-glossary.co.uk).

Pulse - brief continuous sound having a specific spectral content (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).

Pulse spacing – see Interpulse Interval

Rise Time or Onset - interval over which the pulse increases from 10 % to 90 % of its maximum amplitude (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).

Shape of onset – changes in the amplitude of a sound at onset.

Soundscape - The sonic environment. Technically, any portion of the sonic environment regarded as a field for study. The term may refer to actual environments, or to abstract constructions such as musical compositions and tape montages, particularly when considered as an environment (Schafer, 1977).

Spectral regions - range of frequencies contained by a signal.

Third Octaves - sub-divide the audio spectrum typically into 33 bands with constant percentage bandwidth filtering, three times more detailed than octave band filters. The cut-off frequencies have a ratio of 21/6 for example a 1 kHz third-octave band filter has a center frequency of 1000 Hz with lower and upper frequencies of 891 Hz and 1122 Hz respectively (www.acoustic-glossary.co.uk).

APPENDIX 2 –INTERVIEW WITH HEALTHCARE PROFESSIONALS: SCRIPT

Demography and career path

1. Date of Birth, Gender

2. Profession

3. How long ago

4. Have you worked in other areas of the hospital, or were you always in the OR?

5. What is your opinion about general BO noise (equipment, people, A/C)

6. In the OR, can you list the different auditory alarms that exist?

7. Did you or do you have any difficulty in learning what they correspond to, that is, how to associate the sound with the event it intends to announce? (If you had or have, specify the difficulty)

8. During your experience, have you ever had an accident or critical event that was related to hearing alarms emitted by medical equipment? What happened? Can you report the event as it happened?

Daily life at the OR (Explore tasks associated with their main activity)

9. Please describe a typical day at the OR, explaining in some detail what kind of tasks you perform.

10. For these tasks, which equipment do you consider the most important or essential (e.g., cardiac monitoring machine)?

11. (Show plan of an empty operating room/recovery room) Can you tell me/draw where this equipment is located? And where are you (you)?

12. Does this equipment have audible alarms? How are they? (beep, buzzers - vocalize)

13. What is your opinion about these audible alarms?

14. I imagine that you have to deal with several versions of the same equipment located in different rooms, from different manufacturers/brands. In that case, are the alarms also different?

15. (If yes) Does this cause you confusion? How do you handle it?

16. Can you always detect auditory alarms?

17. When you don't listen, under what circumstances are you in? (e.g. most critical patient)

18. Are there surgeries that are noisier than others? Can you give examples? Are the alarms adjusted to these particular surgeries, or is the volume always the same?

19. In noisier surgeries, can you describe to me how you perceive alarms? How is your experience of perceiving these alarms in these surgeries?
20. In all the tasks that you are involved in in your profession, what are the critical events that can happen and that are usually announced by audible alarms?
21. How do you react when you hear these alarms? And what do you feel or think at those times?
22. What kind of response do you give in these cases?

Alarm design

23. I will now list some parameters that are monitored and I would like you to tell me which is the most important for you and your tasks?

- Blood pressure
- Heartbeat
- Pulse Oximeter
- Respiratory frequency
- Temperature
- CO_2
- Other

24. They all have different sounds, that is, without looking at the monitor, can you distinguish, for example, if what is crossing a threshold is blood pressure or respiratory rate?
25. How are these thresholds entered - manually, always for each new patient?
26. Taking the parameter you mentioned to be the most important, what does a high priority event sound like? And average? And low?

Finish

We are now at the end of our conversation, is there anything else you would like to add? Do you have a question you would like to clarify?
Acknowledgment.

APPENDIX 3 – ONLINE QUESTIONNAIRE FOR HEALTHCARE PROFESSIONALS

Thank you for your participation on this survey. This survey is addressed at health professionals who currently work at the operating and/or recovery rooms of a hospital.

To proceed, please confirm you fulfil the following conditions:

- I work at an operating room and/or recovery room
- I have normal audition

This survey is part of a research project intending to design more informative auditory alarms for the equipment inside operating and recovery rooms.

The project intends to adopt a people-centric design strategy, and several health professionals have been interviewed regarding their experience with sounds in their day-to-day professional activities.

The following questions are based on statements collected on those interviews. With this survey we intend to validate some statements by asking you if you agree or do not agree with them regarding your experience in the hospital environment.

Your answers are completely anonymized. We request your honesty by answering what you experience over what you know is the most correct procedure.

The survey takes approximately 7 minutes to fill. Click do proceed.

PERSONAL INFORMATION

1. Country of origin
2. Gender
3. Date of birth
4. Profession
    a. Speciality
5. For how long have your worked in the operating room?
6. Have you ever worked in other hospital areas? Where?
7. In which city are you currently working?

The answer options for the following questions were either “True/False” or a likert frequency scale between “Never” and “Always”.

EQUIPMENT

8. The position of nurses and anaesthetists inside the operating room differs according to surgical specialty
9. The position of nurses and anaesthetists inside the operating room differs according to the hospital they work in
10. Hospitals tend to maintain the same equipment distribution in all their rooms

11. The A/C sound disturbs me
12. I interact with equipment from different manufacturers daily/weekly
13. I interact with vital signs monitors from different manufacturers daily/weekly
14. When I interact with monitors from different manufacturers daily/weekly I notice sound different even though they communicate the same events and monitored parameters
15. Having equipment from a single manufacturer would be an advantage to my work in terms of interaction (visual and auditory interface)

REACTION TO AN ALARM

16. When I hear an alarm, I know what the event or problem is even without looking at the monitor
17. When I hear an alarm, my first reaction is to look at the monitor
18. The fact that the same alarm coming from devices from different manufacturers sounds differently has a significant impact on my work activities
19. I have difficulties in associating the alarm sound to the event or problem it signals
20. I have difficulties in associating the alarm sound to the priority of the event or problem it signals
21. Alarms have a significant impact on my work activities
22. I can predict or identify the source of a false alarm (eg.: interference with electrical equipment, sensor fell)
23. Noise is part of my daily professional lives
24. The main sources of noise are:
    a. Conversations
    b. Instrumentation
    c. Music/Radio/TV's
    d. Personal devices (smartphones, pagers)
    e. Auditory alarms from medical devices
25. Whenever an alarm sounds during surgery, the most common action is to silence it for a given amount of time (eg.: 2-3 minutes)
26. Whenever an alarm sounds during surgery, the most common action is to change the thresholds of the alarm

PARAMETERS

27. For my work, monitoring alarms are the most important
28. For my work, laparoscopic alarms are the most important
29. The parameters which are most relevant to my work are…
    a. Pulse oximetry
    b. Blood Pressure
    c. Heart Rate
    d. Respiratory Rate
    e. Temperature
    f. CO_2
    g. ECMO
    h. Invasive Ventilation
    i. Peak Pressure
    j. Infusion Pumps

30. Without looking at the monitor I can distinguish between, for instance, blood pressure and respiratory rate
31. Even though some surgeries are noisier than others, I always hear the auditory alarms
32. The three surgeries I consider the noisiest are…
33. I believe having distinct alarm sounds for the most important parameters would not improve my work activities
34. Think of the pulse oximeter monitoring sound. Would you consider it a well-designed alarm?
35. If you could choose one parameter to have its alarm sound redesigned, which one would it be?

PATIENTS

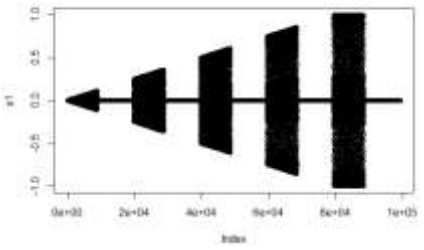
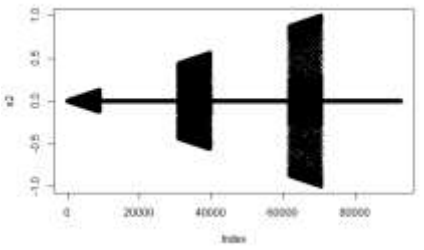
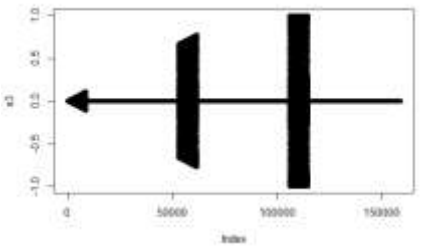
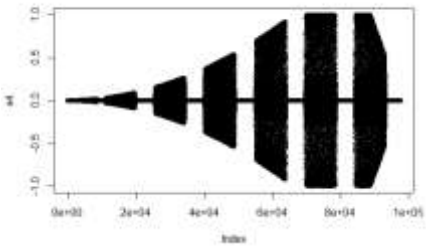
36. Monitoring sounds generate anxiety on patients during surgery when they are awake
37. Monitoring sounds generate anxiety on patients in the recovery rooms
38. I have seen patients or family members getting scared with sound alarms which were signalling trivial information

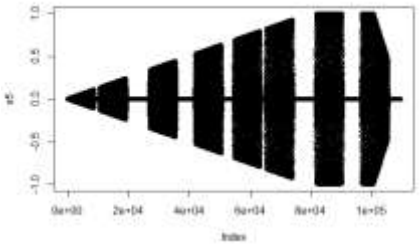
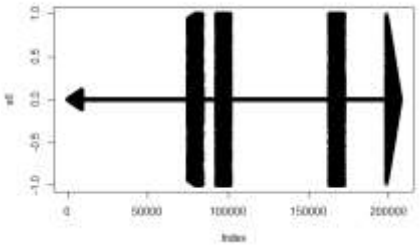
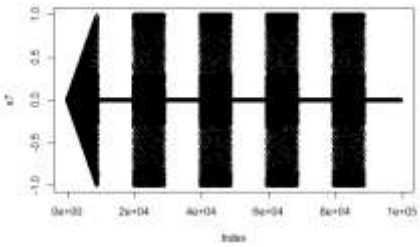
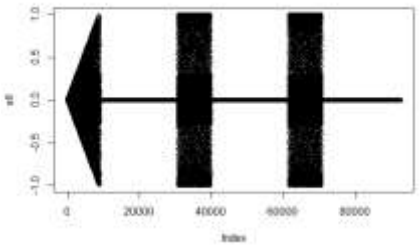
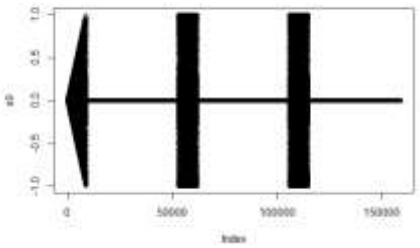
RECOVERY ROOM

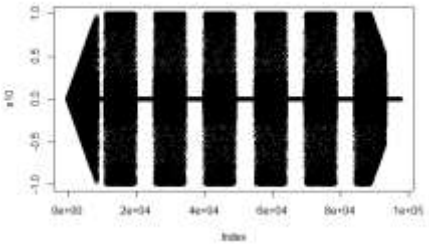
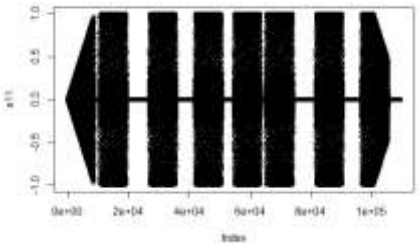
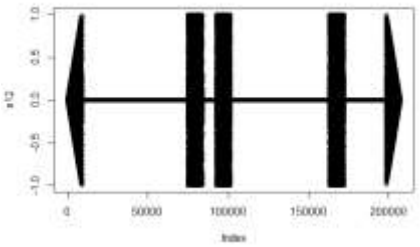
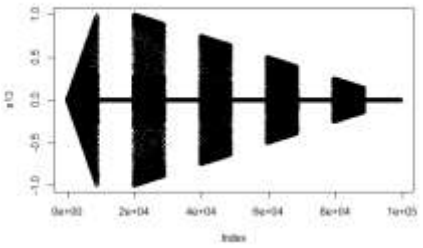
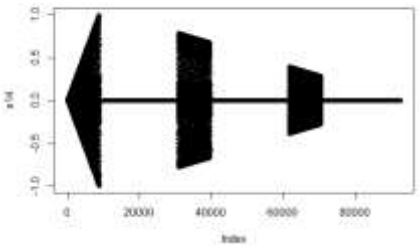
39. In the recovery room sound alarms are more important than in the operating room
40. The recovery room in my hospital has … beds
41. When the recovery room has many patients and one alarm sounds in the room, I have difficulty in identifying from which machine it is coming from
42. The vital signs monitors in the recovery room have visual aids (eg: lights) that signal when an alarm is triggered

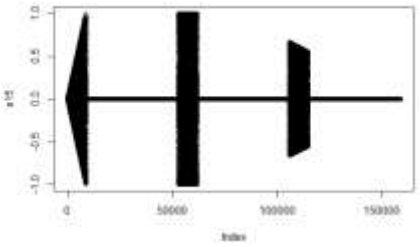
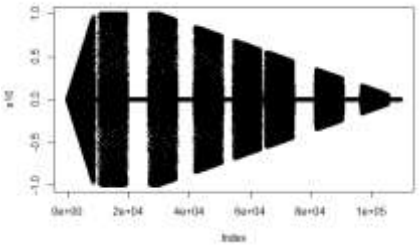
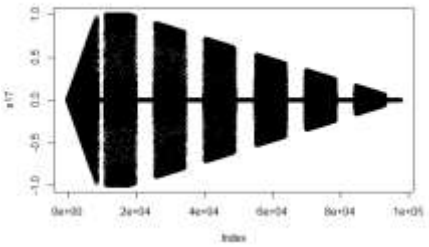
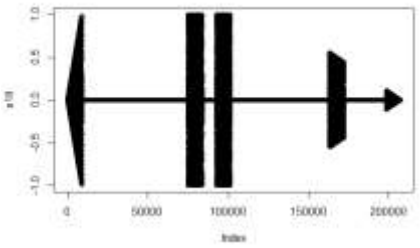
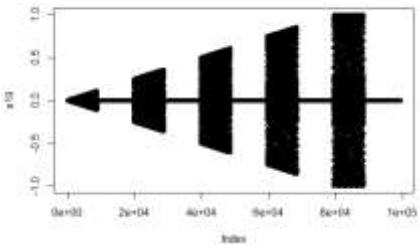
APPENDIX 4 - STIMULI USED IN SUBJECTIVE PERCEPTION OF AUDITORY WARNING SIGNALS

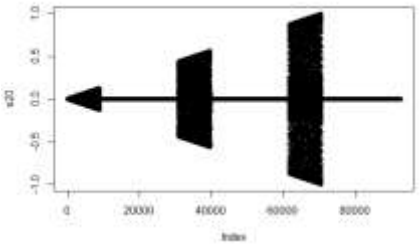
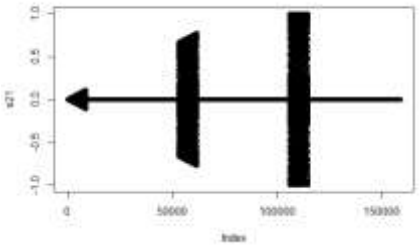
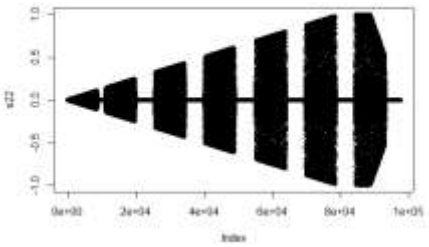
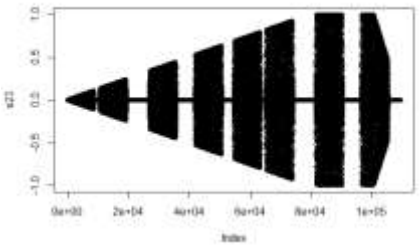
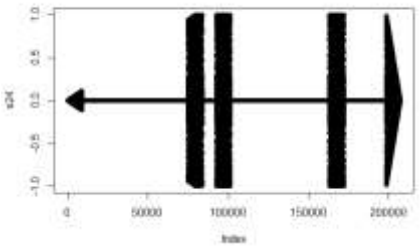
Sounds available in the following link:
https://figshare.com/articles/media/Using_semantic_differential_scales_to_assess_the_subjective_perception_of_aufditory_warning_signals/14905200

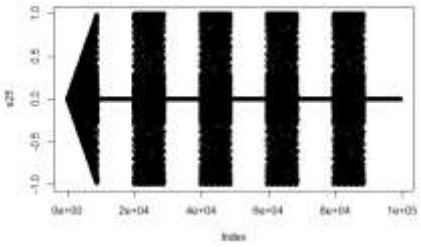
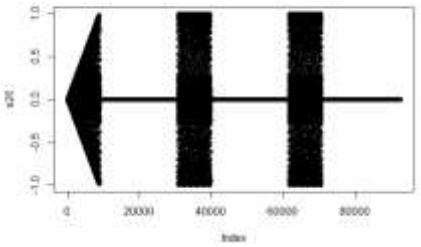
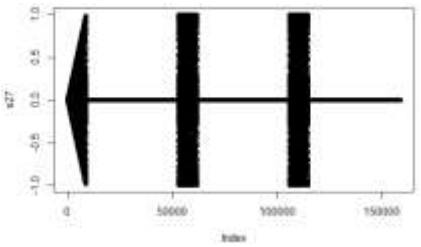
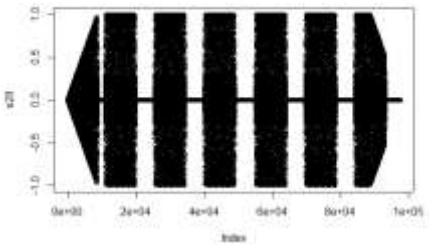
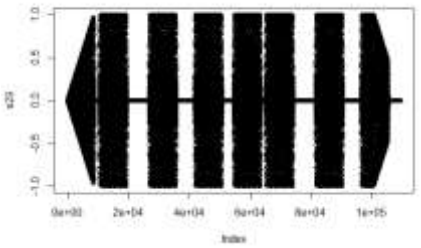
	Frequency	**Speed**	**Rythm**	**Onset**	
S1	2500	x4	Regular	Slow onset	
S2	2500	x2	Regular	Slow onset	
S3	2500	x1	Regular	Slow onset	
S4	2500	x4	Synchopated	Slow onset	

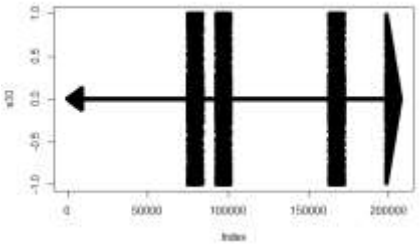
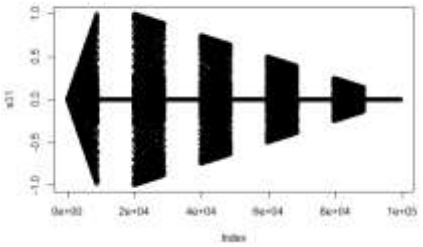
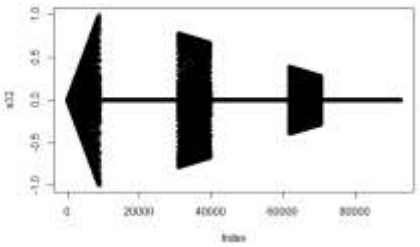
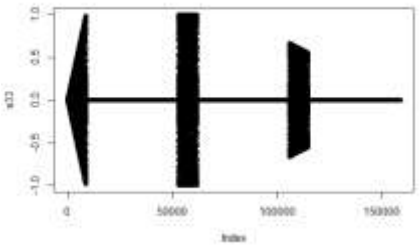
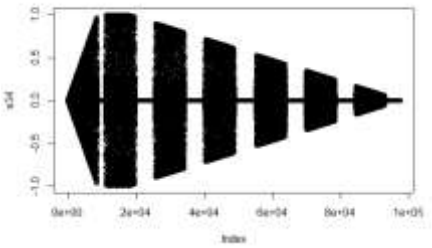
S5	2500	x2	Synchopated	Slow onset	
S6	2500	x1	Synchopated	Slow onset	
S7	2500	x4	Regular	Regular onset	
S8	2500	x2	Regular	Regular onset	
S9	2500	x1	Regular	Regular onset	

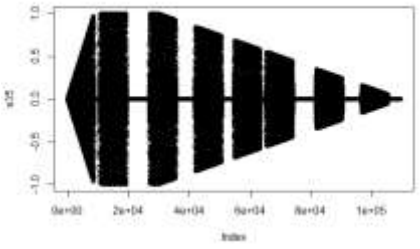
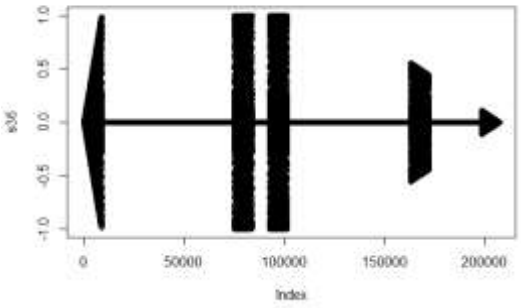
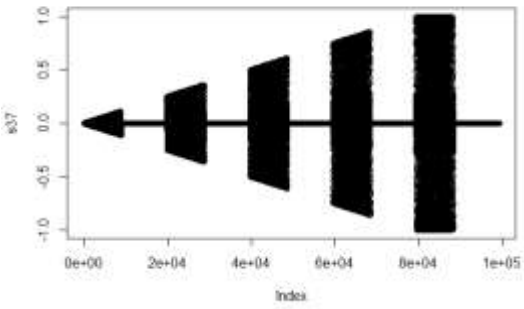
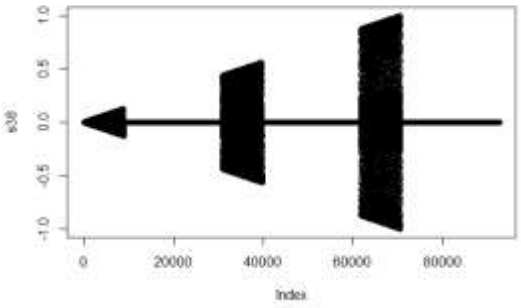
S10	2500	x4	Synchopated	Regular onset	
S11	2500	x2	Synchopated	Regular onset	
S12	2500	x1	Synchopated	Regular onset	
S13	2500	x4	Regular	Slow offset	
S14	2500	x2	Regular	Slow offset	

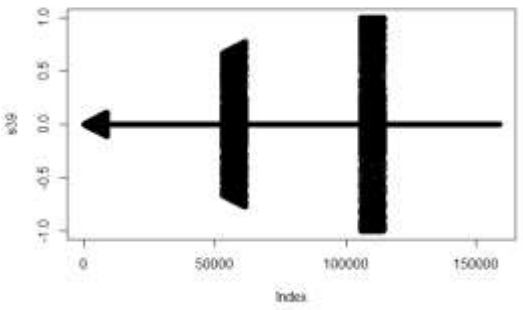
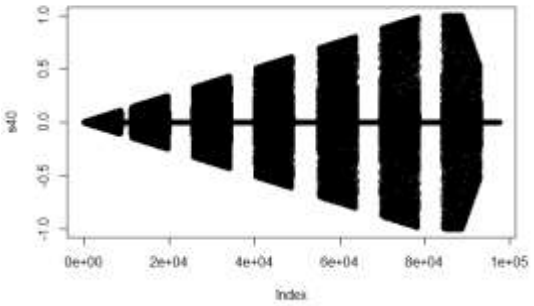
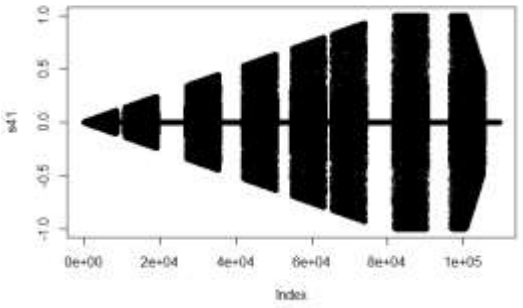
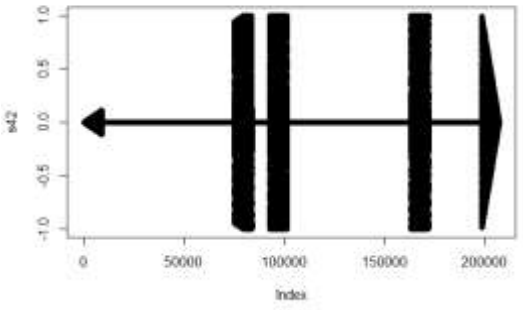
S15	2500	x1	Regular	Slow offset	
S16	2500	x4	Synchopated	Slow offset	
S17	2500	x2	Synchopated	Slow offset	
S18	2500	x1	Synchopated	Slow offset	
S19	1500	x4	Regular	Slow onset	

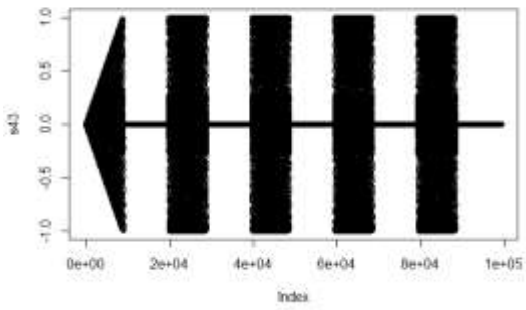
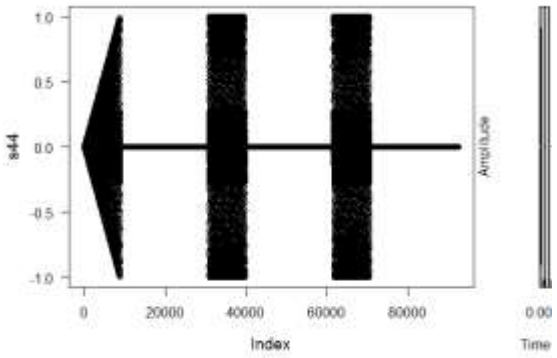
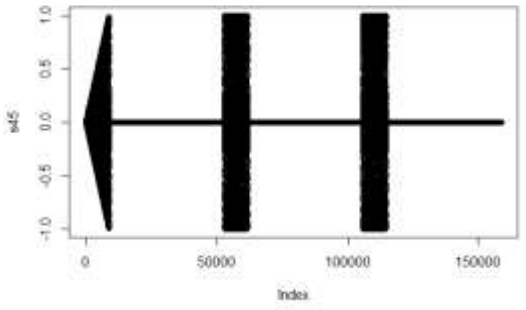
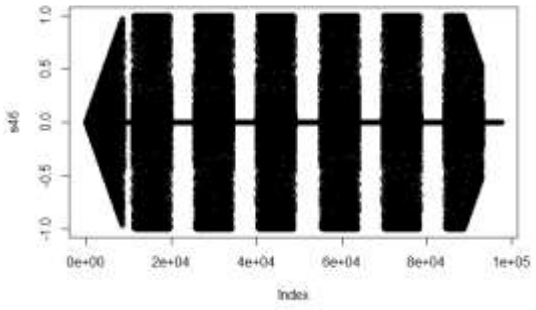
S20	1500	x2	Regular	Slow onset	
S21	1500	x1	Regular	Slow onset	
S22	1500	x4	Synchopated	Slow onset	
S23	1500	x2	Synchopated	Slow onset	
S24	1500	x1	Synchopated	Slow onset	

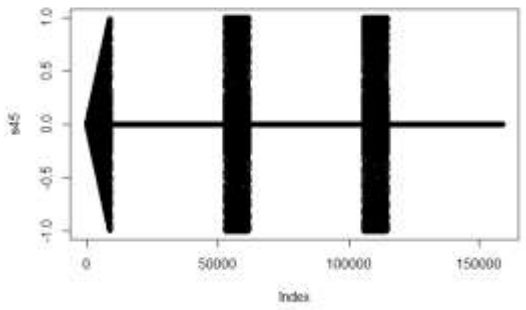
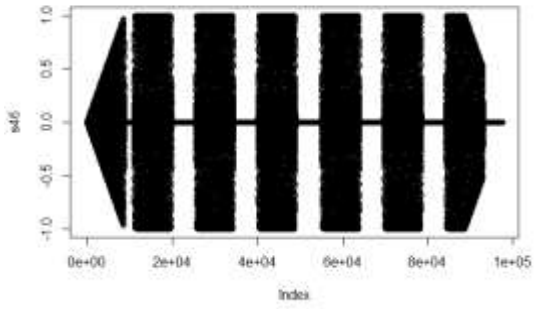
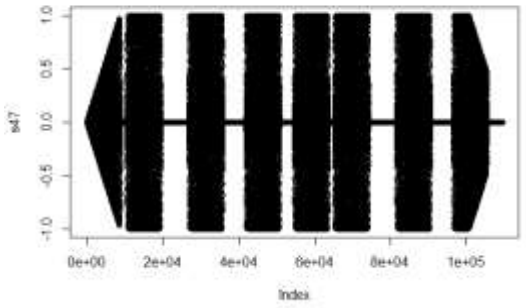
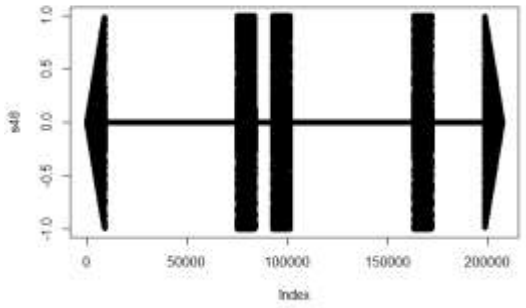
S25	1500	x4	Regular	Regular onset	
S26	1500	x2	Regular	Regular onset	
S27	1500	x1	Regular	Regular onset	
S28	1500	x4	Synchopated	Regular onset	
S29	1500	x2	Synchopated	Regular onset	

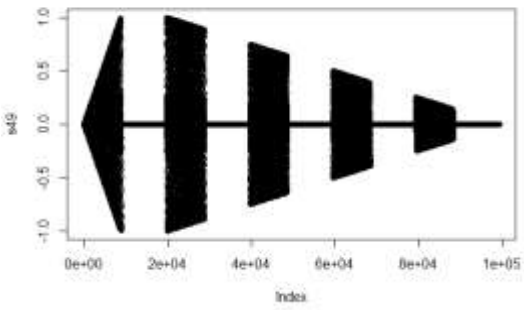
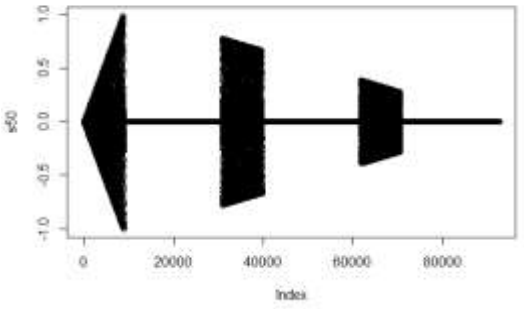
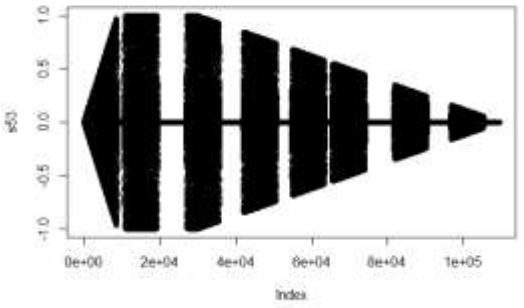
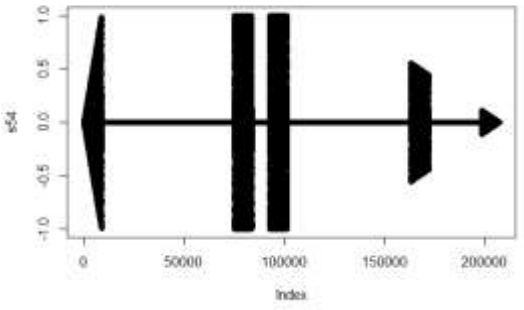
S30	1500	x1	Synchopated	Regular onset	
S31	1500	x4	Regular	Slow offset	
S32	1500	x2	Regular	Slow offset	
S33	1500	x1	Regular	Slow offset	
S34	1500	x4	Synchopated	Slow offset	

S35	1500	x2	Synchopated	Slow offset	
S36	1500	x1	Synchopated	Slow offset	
S37	500	x4	Regular	Slow onset	
S38	500	x2	Regular	Slow onset	

S39	500	x1	Regular	Slow onset	
S40	500	x4	Synchopated	Slow onset	
S41	500	x2	Synchopated	Slow onset	
S42	500	x1	Synchopated	Slow onset	

S43	500	x4	Regular	Regular onset	
S44	500	x2	Regular	Regular onset	
S45	500	x1	Regular	Regular onset	
S46	500	x4	Synchopated	Regular onset	

S47	500	x2	Synchopated	Regular onset	
S48	500	x1	Synchopated	Regular onset	
S49	500	x4	Regular	Slow offset	
S50	500	x2	Regular	Slow offset	

S51	500	x1	Regular	Slow offset	
S52	500	x4	Synchopated	Slow offset	
S53	500	x2	Synchopated	Slow offset	
S54	500	x1	Synchopated	Slow offset	

APPENDIX 5 – STIMULI USED IN CATEGORIZATION OF SOUNDS

The following table presents the psychoacoustical parameters measured and calculated for each sound. These include Harmonicity Noise Ratio, Spectral Centroid, Loudness, Roughness and Sharpness. All measurements were performed using Octave (Eaton et al., 2020), except Harmonicity Noise Ratio which was measured using Praat (Boersma & Weenink, 2019). Sounds available in the following link: https://figshare.com/articles/media/Categorization_of_sounds_for_he_design_of_clinical_auditory_alarms/14905113.

Name	Spec. Centroide (Hz)	Sharpeness (Acum)	Loudness (Sone)	Roughness (Asper)	Harmonicity (dB)
norm_powerdown	2508,1	1,8889	33,67	0,32297	1,23
robot_powerdown	4438,8	2,3999	30,04	0,6005	10,4
engine_quickfailure	422,4	1,5901	35,14	0,13035	2,32
engine_scifi	124,6	1,3517	44,68	0,42919	16,27
engine_doesntstart	2658,1	1,9417	34,32	0,26723	-3,09
norm_temperature	3516,1	2,5898	16,99	0,20554	21,96
steam	10063,4	2,7977	28,02	0,07002	-3,65
waterboil	350,1	1,2324	36,41	0,070024	3,06
frybacon	9198,1	2,7908	28,08	1,4363	-2,53
fryegg	5685,7	2,3394	31,71	2,0569	-1,97
norm_ventilation	7136,9	2,5624	29,58	0,025499	-5,9
AC	44,5	1,2635	45,07	0,14145	3,8
bellows	2088,4	1,9544	33,56	0,31835	-1,65
darthvader	463,52	1,5673	28,76	0,34056	6,9
breathein	5628,4	2,7621	23,85	0,017716	-4,42
mask	6000,6	2,5947	30,3	0,069992	1,39
breatheout	1629,8	2,2532	32,1	0,05057	5,5
baloon	3414	2,112	33,53	0,36474	-0,66
cuff_inflate	8740	3,0648	23,98	0,11121	-0,23
sound_koroktoff	157,86	0,6444	26,8	3,0213	0,43
koroktoffi	813,1	1,0941	28,05	2,1414	-3,53
pump_velcro	4523,9	2,2725	35,24	2,604	-0,04
pump_blood_pressure	1392,2	1,7624	35,37	0,79691	-2,15
norm_perfusion	1129	1,3486	20,4	1,6369	5,5
succion	4454	2,246	34,75	0,61227	-1,02
drops	297,16	0,6701	19,95	1,3861	12,98
splatter	2116,9	1,8749	37,01	0,52734	1,53
water_inside_glass	2024,8	1,7304	30,81	2,5141	4,71
constant_bubbles	809,36	1,342	21,61	1,4485	5,65
bubbling	696,49	1,2831	32,47	1,1825	2,63
water_plastic_bottle	2499,5	1,9383	31,45	0,5761	-4,65
norm_Oxygenation	244,1	0,6886	19,86	0,40943	23,62
aerosol	9197,5	2,9505	25,51	0,061076	-4,4
descompression	9027,5	2,7498	26,52	0,08596	-2,41
blow_tube	4059,2	2,7791	20,73	0,025937	11,8
gas	6721,4	2,4305	28,52	0,37011	-1,4
gas_champagne_bottle	1622,6	1,8254	34,08	0,56966	3,4

reverberant_water	1492,5	1,9711	23,59	0,26991	10,94
effervescent	1081,9	1,5611	19,58	0,57512	10,1
norm_drug_delivery	4768,6	2,6745	26,02	2,2261	0,56
spread_pills	7181,8	3,1992	22,97	1,812	0,27
take_pill	8847,5	2,7654	27,44	2,007	-4,03
norm_cardio	276,14	1,3063	40,37	2,1165	12,76
tictac	4682,2	2,6765	27,61	1,2878	-0,99
plastic_ruller	2497,8	1,9347	28,86	3,1068	1,01
clap	2050,6	1,8319	29,34	2,5792	-4,08
strings	168,53	0,69233	24,09	1,178	1,53
djambe_tribal	561,75	1,4471	36,05	1,3391	7,02
metal_sheet	1277,1	1,5176	26,15	0,394	-0,39
congas	271,7	0,77401	26,65	3,6239	31,05
snare_drum	1506,5	1,767	28,89	4,1541	1,88

APPENDIX 6 – STIMULI USED IN PERCEPTION OF URGENCY

Sounds can be heard on this link:
https://figshare.com/articles/media/Validation_of_clinical_auditory_alarms_with_users_Perception_of_Urgency/14905512.

#	F1	SPEED	RHYTHM	INHARMONICITY	PULSE LENGTH
1	441 Hz	5		1 = f1*25/24 2 = f1*25/18 3 = f1*36/2 4 = f1*9/5 5 = f1*15/8	75ms + 25ms(IPI)
2	441 Hz	5		1 = f1*25/24 2 = f1*25/18 3 = f1*36/2 4 = f1*9/5 5 = f1*15/8	75ms + 25ms(IPI)
3	295 Hz	3		1 = f1*5/4 2= f1*4/3 3= f1*3/2	150ms + 50ms(IPI)
4	295 Hz	3		1 = f1*5/4 2= f1*4/3 3= f1*3/2	150ms + 50ms(IPI)
5	441 Hz	6		1 = f1*25/24 2 = f1*25/18 3= f1*36/2 4= f1*9/5 5 = f1*15/8	75ms + 25ms(IPI)
6	441 Hz	6		1 = f1*25/24 2 = f1*25/18 3= f1*36/2 4= f1*9/5 5 = f1*15/8	75ms + 25ms(IPI)
7	295 Hz	4		1 = f1*5/4 2= f1*4/3 3= f1*3/2	150ms + 50ms(IPI)
8	295 Hz	4		1 = f1*5/4 2= f1*4/3 3= f1*3/2	150ms + 50ms(IPI)
9	718 Hz	5		1 = f1*25/24 2 = f1*25/18 3= f1*36/2 4= f1*9/5 5 = f1*15/8	75ms + 25ms(IPI)
10	718 Hz	5		1 = f1*25/24 2 = f1*25/18 3= f1*36/2 4= f1*9/5 5 = f1*15/8	75ms + 25ms(IPI)
11	479 Hz	3		1 = f1*5/4 2= f1*4/3 3= f1*3/2	150ms + 50ms(IPI)

12	479 HZ	3		1 = F1*5/4 2= F1*4/3 3= F1*3/2	150MS + 50MS(IPI)
13	718 Hz	6		1 = f1*25/24 2 = f1*25/18 3= f1*36/2 4= f1*9/5 5 = f1*15/8	75ms + 25ms(IPI)
14	718 Hz	6		1 = f1*25/24 2 = f1*25/18 3= f1*36/2 4= f1*9/5 5 = f1*15/8	75ms + 25ms(IPI)
15	479 Hz	4		1 = f1*5/4 2= f1*4/3 3= f1*3/2	150ms + 50ms(IPI)
16	479 Hz	4		1 = f1*5/4 2= f1*4/3 3= f1*3/2	150ms + 50ms(IPI)
17	441 Hz	5		1 = f1*25/24 2 = f1*25/18 3 = f1*36/2	75ms + 25ms(IPI)
18	441 Hz	5		1 = f1*25/24 2 = f1*25/18 3 = f1*36/2	75ms + 25ms(IPI)
19	295 Hz	3		1 = f1*5/4 2 = f1*4/3	150ms + 50ms(IPI)
20	295 Hz	3		1 = f1*5/4 2 = f1*4/3	150ms + 50ms(IPI)
21	441 Hz	6		1 = f1*25/24 2 = f1*25/18 3 = f1*36/2	75ms + 25ms(IPI)
22	441 Hz	6		1 = f1*25/24 2 = f1*25/18 3 = f1*36/2	75ms + 25ms(IPI)
23	295 Hz	4		1 = f1*5/4 2 = f1*4/3	150ms + 50ms(IPI)
24	295 Hz	4		1 = f1*5/4 2 = f1*4/3	150ms + 50ms(IPI)
25	718 Hz	5		1 = f1*25/24 2 = f1*25/18 3 = f1*36/2	75ms + 25ms(IPI)
26	718 Hz	5		1 = f1*25/24 2 = f1*25/18 3 = f1*36/2	75ms + 25ms(IPI)
27	479 Hz	3		1 = f1*5/4 2 = f1*4/3	150ms + 50ms(IPI)
28	479 Hz	3		1 = f1*5/4 2 = f1*4/3	150ms + 50ms(IPI)
29	718 Hz	6		1 = f1*25/24 2 = f1*25/18 3 = f1*36/2	75ms + 25ms(IPI)
30	718 Hz	6		1 = f1*25/24 2 = f1*25/18 3 = f1*36/2	75ms + 25ms(IPI)
31	479 Hz	4		1 = f1*5/4 2 = f1*4/3	150ms + 50ms(IPI)
32	479 Hz	4		1 = f1*5/4 2 = f1*4/3	150ms + 50ms(IPI)

APPENDIX 7 – FINAL LIBRARY OF CLINICAL AUDITORY ALARMS

All previous work has led to the creation of a set of 48 alarms (8 x Category, 3 x Priority, 2 x Tendency). All alarms were designed by Rui Marques and developed in Cycling74 Max Msp (*Cycling '74*, n.d.).

- The final set of alarms without tendency indicators can be heard here [25] (https://figshare.com/account/projects/117336/articles/16574996) .
- The final set of alarms with tendency indicators can be heard here[26] (https://figshare.com/account/projects/117336/articles/16575002).

Priority Pointers

The priority of the icons was manipulated using three levels of pitch, speed, pulse length, inharmonicity, repetition and rhythm. Regarding the harmonic constitution of the signal, the used synthesis model used was additive synthesis with sinusoidal waves, with a total of 7 whole harmonics and a variation of 3 to 5 inharmonics depending on the values that k takes in the inharmonic variable and depending on the priority. The values of the inharmonics are: f1*5/4 (major3nd), f1*4/3 (perfect4th), f1*3/2, (perfect5th), f1*5/3 (major6th), f1*15/8 (major7th). These intervals were chosen because they are considered the intervals with greater consonance in relation to the fundamental frequency.

The shape of the envelope of harmonics and inharmonics is dynamic as the sound progresses over time. The shape of the envelope of each pulse is constant over time with an onset of 20 ms and an offset of 20 ms.

- Priority pointers can be heard here [27] (https://figshare.com/articles/media/Final_priority_pointers_for_clinical_auditory_alarms/16570683).

[25] Vieira, Joana (2021): Final set of clinical alarms (without tendency indicators). figshare. Media. https://doi.org/10.6084/m9.figshare.16574996

[26] Vieira, Joana (2021): Final set of clinical alarms (with tendency indicators). figshare. Media. https://doi.org/10.6084/m9.figshare.16575002

[27] Vieira, Joana (2021): Final priority pointers for clinical auditory alarms. figshare. Media. https://doi.org/10.6084/m9.figshare.16570683.v1

PRIORITY	LOW	MEDIUM	HIGH
Pitch (k=150 Hz)	f1 = 220 Hz	f1 = 295 Hz	f1 = 441 Hz
Harmonic constitution of pulses	1 harmonic = f1*1 2 harmonic = f1*2 3 harmonic = f1*3 4 harmonic = f1*4 5 harmonic = f1*5 6 harmonic = f1*6 7 harmonic = f1*7	1 harmonic = f1*1 2 harmonic = f1*2 3 harmonic = f1*3 4 harmonic = f1*4 5 harmonic = f1*5 6 harmonic = f1*6 7 harmonic = f1*7	1 harmonic = f1*1 2 harmonic = f1*2 3 harmonic = f1*3 4 harmonic = f1*4 5 harmonic = f1*5 6 harmonic = f1*6 7 harmonic = f1*7
Speed (k=1.66)	2	3	5
Inharmonicity (k=1.58)	2	3	5
Inharmonic constitution of pulses	1inharmonic = f1*5/4 2inharmonic = f1*4/3	1inharmonic = f1*5/4 2inharmonic = f1*4/3 3inharmonic = f1*3/2	1inharmonic = f1*5/4 2inharmonic = f1*4/3 3inharmonic = f1*3/2 4inharmonic = f1*5/3 5inharmonic = f1*15/8
Repetition (k= 0.70)	1	1	2
Pulse length (k=75 ms)	300 ms (pulse) + 100 ms (IPI)	150 ms (pulse) + 50 ms (IPI)	75 ms (pulse) + 25 ms (IPI)
Rhytmic pattern			
Envelope harmonics (Morphing)	**1 Pulse** f1 = -7.8dB f2 = -13.8dB f3 = -21.3dB f4 = -27.3dB f5 = -27.3dB f6 = -27.3dB f7 = -22.9dB **2 Pulses** f1 = -10.8dB f2 = -19.1dB f3 = -19.9dB f4 = -16.1dB f5 = -15.4dB f6 = -16.3dB f7 = -22.8dB	**1 Pulse** f1 = -10.8dB f2 = -19.4dB f3 = -19.3dB f4 = -16.3dB f5 = -15.0dB f6 = -21.9dB f7 = -22.9dB **2 Pulses** f1 = -10.8dB f2 = -19.1dB f3 = -28.9dB f4 = -34.1dB f5 = -27.4dB f6 = -19.3dB f7 = -15.8dB **3 Pulses** f1 = -10.8dB f2 = -21.5dB f3 = -13.9dB f4 = -18.0dB f5 = -27.9dB f6 = -25.3dB f7 = -31.9dB	**1 Pulse** f1 = -7.8dB f2 = -13.8dB f3 = -21.3dB f4 = -27.3dB f5 = -27.3dB f6 = -27.3dB f7 = -22.9dB **2 Pulses** f1 = -7.8dB f2 = -13.8dB f3 = -21.3dB f4 = -27.3dB f5 = -39.3dB f6 = -39.3dB f7 = -27.9dB **3 Pulses** f1 = -7.8dB f2 = -13.8dB f3 = -21.3dB f4 = -27.3dB f5 = -25.3dB f6 = -25.3dB f7 = -28.9dB **4 Pulses** f1 = -7.8dB f2 = -13.8dB f3 = -21.3dB f4 = -27.3dB f5 = -45.3dB f6 = -48.3dB f7 = -28.9dB **5 Pulses**

			f1 = -7.8dB f2 = -13.8dB f3 = -21.3dB f4 = -27.3dB f5 = -28.3dB f6 = -48.3dB f7 = -28.9dB
Envelope inharmonics (Morphing)	**1 Pulse** 1 inharmonic = -7.7dB 2 inharmonic = -28.3dB **2 Pulse** 1 inharmonic = -28.6dB 2 inharmonic = -7.8dB	**1 Pulse** 1 inharmonic = -9.7dB 2 inharmonic = -26.3dB 3 inharmonic = -26.4dB **2 Pulse** 1 inharmonic = -32.6dB 2 inharmonic = -9.8dB 3 inharmonic = -28.1dB **3 Pulse** 1 inharmonic = -9.3dB 2 inharmonic = -28.5dB 3 inharmonic = -27.3dB	**1 Pulse** 1 inharmonic = -11.1dB 2 inharmonic = -29.3dB 3 inharmonic = -28.4dB 4 inharmonic = -26.3dB 5 inharmonic = -26.4dB **2 Pulse** 1 inharmonic = -29.1dB 2 inharmonic = -28.3dB 3 inharmonic = -9.4dB 4 inharmonic = -24.3dB 5 inharmonic = -24.4dB **3 Pulse** 1 inharmonic = -28.1dB 2 inharmonic = -29.3dB 3 inharmonic = -9.4dB 4 inharmonic = -26.3dB 5 inharmonic = -26.4dB **4 Pulse** 1 inharmonic = -28.1dB 2 inharmonic = -29.3dB 3 inharmonic = -28.4dB 4 inharmonic = -11.3dB 5 inharmonic = -26.4dB **5 Pulse** 1 inharmonic = -27.1dB 2 inharmonic = -29.3dB 3 inharmonic = -28.4dB 4 inharmonic = -26.3dB 5 inharmonic = -11.4dB
Total duration in miliseconds	762 ms	782 ms	572 ms

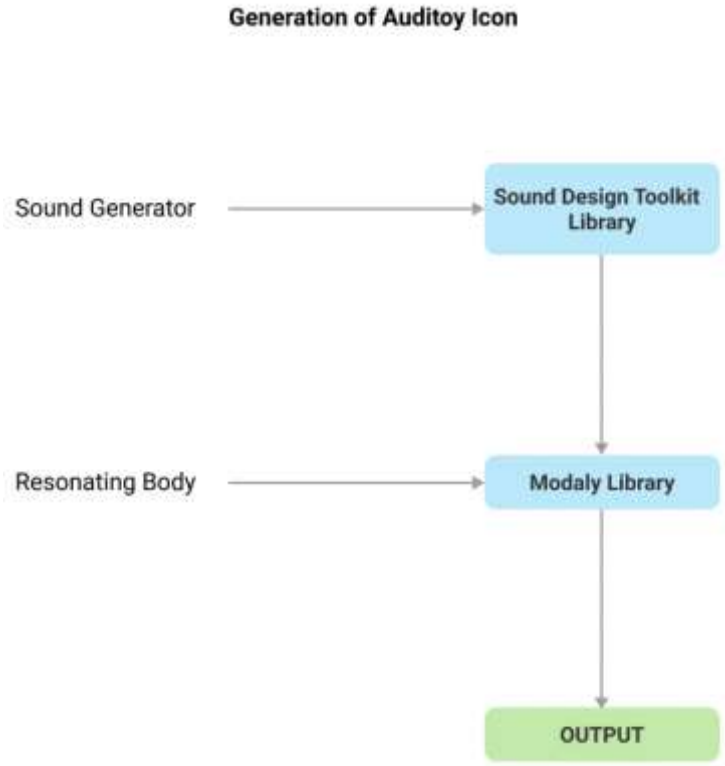
Auditory Icons

The Auditory Icons were designed using a sound synthesis algorithm library called "Sound Design Toolkit" developed by IRCAM, Genesis Acoustics, IUAV and KTH. The mathematical models used in this library simulate the physical characteristics of sound, such as mass and matter, distance, and acceleration. They are used to simulate ecological

sounds such as solid interactions (collision, friction, rolling), liquids (drops, water flow), gases (wind, air), and mechanical (combustion and electrical engines).

The algorithms of this library were implemented with three main considerations:

- Perceptual auditory relevance;
- Cartoonization. This technique simplifies and exaggerates the inherent physics to improve the computation efficiency and perceptual clarity;
- Parametric temporal control. This technique guarantees adequate, natural and expressive articulations of the sound processes.



Another library developed by IRCAM was used to harmonize the icons and thus transform these ecological sounds into more pleasant and musical sounds. Modalys includes mathematical models based on physics modelling. These models try to describe the physical characteristics of materials such as the shape and size of the resonating body, or the material's density.

Besides these two libraries, other algorithms based on filtering were used in the alarms (lowpass, high pass, bandpass and spectral filter).

The following table characterizes the final set of auditory icons. All icons have the duration of 2 seconds. The auditory icon **metaphor** describes a real, familiar sound used to represent the category. The **description** includes the icons' morphological composition.

Pitch refers to the harmonic with the greatest amplitude. **Harmonic constitution** describes the harmonization process. The harmonization of the icons was done by simulating a set of 4 tubes based on physical modelling, which were used as vibrating objects like resonators, obtaining tetrad chords (chords composed of four notes played simultaneously). The pitch of the pipes is constituted by 1 degree (Tronic), 3 degree (major third), 5 degree (perfect fifth) and 8 degree (an octave in relation to the tronic) thus forming a major chord. Harmonization assures the sounds are distinct from real sounds, and are more resistant to masking.

CATEGORY	METAPHOR	DESCRIPTION	PITCH	HARMONIC CONSTITUTION
Cardiovascular	Heartbeat	Time-phased rectangular and sinusoidal waves with frequencies between 150 Hz and 4000 Hz with different amplitudes, which are higher at low frequencies and more attenuated at high frequencies. Each pulse of the heartbeat is made up of two pulses separated by 200 ms. There are 4 heartbeats in a total of 8 pulses.	150.88 Hz	C2_E2_G2_C3
Blood Pressure	Blood pressure hand pump followed by heartbeat	Two pulses of white noise filtered to simulate the pressure pump being filled, followed by 3 pulses of heart beat. Each pulse of the heart beat consists of 2 pulses in a total of 6 pulses.	420.22 Hz	G2_B2_D3_G3
Drug Administration	Filling a glass of water	A glass of water being filled. The synthesis models used were physical modeling, amplitude and frequency modeling over time to give the sensation of a glass being filled.	614.20 Hz	E2_G#2_B2_E3
Oxygenation	Air bubbles bursting	Bubble events appear randomly with very short intervals. The synthesis model used was physical modeling.	290.84 Hz	A2_C#3_E3_A3

Artificial Perfusion	Liquis circulating inside a tube	Sequence of a strong liquid sound resonating inside a pipe. Sound synthesized through physical modeling.	423.62 Hz	F2_A2_C3_F3
Power down	A machine turning off	The synthesis model used is a physical model that represents fans and motors losing speed.	236.98 Hz	B2_D#3_F#3_B3
Temperature 1	Wood burning	White noise where frequency filtering by band is applied. To simulate the crackle of the wood an amplitude modeling model was applied, with close random values to obtain small pulses constituted by filtered noise. The synthesis model used was the subtractive synthesis.	1125.30 Hz	C3_E3_G3_C4
Temperature 2	Drop ofwater falling on hot plate + evaporation	White noise where frequency filtering by band is applied. The synthesis model used was the subtractive synthesis.	1125.30 Hz	C3_E3_G3_C4
Ventilation	Inhale and exhale sounds	White noise that undergoes a frequency filtering by band. It starts at a higher frequency and evolves to a lower frequency. The synthesis model used was subtractive synthesis.	743.26 Hz	D2_F#2_A2_D3

- Auditory icons without harmonization can be heard here [28] (https://figshare.com/articles/media/Auditory_icons/16570689).
- Auditory icons with harmonization can be heard here [29] (https://figshare.com/articles/media/Auditory_Icons_with_harmonization/16570686).

[28] Vieira, Joana (2021): Auditory icons. figshare. Media. https://doi.org/10.6084/m9.figshare.16570689.v1

[29] Vieira, Joana (2021): Auditory Icons with harmonization. figshare. Media. https://doi.org/10.6084/m9.figshare.16570686.v1

Alarm Tendency

The tendency or direction of an alarm intends to help the listener identify whether a particular alarm parameter is varying upwards or downwards. To do so, the pitch of one of the resonator bodies (tubes) is manipulated, so that it can be interpreted as an increase or decrease of the parameter.

The vibrating bodies used as resonators are a set of 4 tubes that obtain tetrad chords (chords composed of four simultaneous notes) with pitch variation over time in one of the tubes, keeping the pitch of the remaining 3 tubes static over time. The harmonic basis of the three pipes with static pitch over time has the following formation:

- pipe 1, first degree (Tonic)
- pipe 2, third degree (Third major)
- pipe 3, fifth degree (Perfect fifth)

The pitch variation of the tube that will communicate the tendency is presented in the following 3 situations:

1. If the tendency is increasing, the tonal variation of the tube takes values from the eighth degree in relation to the tonic of the chord until the 15^{th} degree. If the tendency is decreasing it goes from the 15th degree to the eighth.
2. If the tendency is increasing, the tonal variation of the tube takes values from the eighth degree in relation to the tonic of the chord until the 12^{th} degree. If the tendency is decreasing the reverse happens.
3. If the tendency is increasing the tonal variation of the tube takes values in the eighth degree in relation to the tonic of the chord, in the 10^{th}, 12^{th} and 14^{th} degrees. If the tendency is decreasing the reverse happens.

Tests should be performed to understand which pitch variation better communicates the perception of increasing and decreasing values.

APPENDIX 8 – GENERAL ORGANIZATION OF STUDIES, TIMEFRAME AND NO. OF PARTICIPANTS