Jun 25th, 9:00 AM

Functional and sensible: Patient monitoring alarm tones designed with those who hear them

Avery Sen
Sen Sound, United States of America

Yoko Sen
Sen Sound, United States of America

Matt Barile
Sen Sound, United States of America

Sage Palmedo
Sen Sound, United States of America

Andreas Walden
Philips Medical Systems, Germany

See next page for additional authors

Citation

Sen, A., Sen, Y., Barile, M., Palmedo, S., Walden, A., and Antunes, V.V. (2022) Functional and sensible: Patient monitoring alarm tones designed with those who hear them, in Lockton, D., Lenzi, S., Hekkert, P., Oak, A., Sádaba, J., Lloyd, P. (eds.), *DRS2022: Bilbao*, 25 June - 3 July, Bilbao, Spain. https://doi.org/10.21606/drs.2022.684

Authors

Avery Sen, Yoko Sen, Matt Barile, Sage Palmedo, Andreas Walden, and Vitor Vicente Antunes

DRS 2022
Bilbao 25th June — 3rd July

Functional and sensible: Patient monitoring alarm tones designed with those who hear them

Avery Sen[a,*], Yoko Sen[a], Matthew Barile[a], Sage Palmedo[a], Andreas Walden[b], Vitor Vicente Antunes[b]

[a]Sen Sound, USA
[b]Philips Medical System, Germany

*corresponding e-mail: avery@sensound.space

doi.org/10.21606/drs.2022.684

Abstract: This is a case study in participatory design of alarm tones for the Philips IntelliVue patient monitoring system. Through interviews and workshops, we asked clinicians and other stakeholders what mattered to them as we designed new tones. We distilled responses into criteria with which to evaluate new tone options that we created by adjusting the tones' pitch, timbre, and other parameters. In surveys, participants compared these options using the criteria distilled from interviews. The results were: 1) new tones that stakeholders judged to be improvements over the originals, and 2) criteria for evaluating future tones, based on "functionality" (i.e., their ability to be heard, understood, and prompt response) as well as "sensibility" (i.e., avoidance of unintended consequences: annoyance, fatigue, patient distress). We found that we could engage stakeholders meaningfully in the definition and design of "better" tones. We also found it possible to make tones that are both functional and more sensible.

Keywords: sound design; clinical alarms; participatory design; medical devices

1. Introduction

> "The patient monitor considers me a sleeping opponent who has to be beaten to wake up. I wish him to be a colleague, with the same interest as myself to do the best for the patient. The monitor helps sense physical states that I can't see myself, so we are a team. As a colleague, we have to cooperate, not be opponents."
> – Anaesthesiologist

These are the words one clinician used to describe a patient monitor, regarding its alarm tones. This was the answer to a question that we asked dozens of clinicians: "if the patient monitor were a person, who would it be to you – based on the way it sounds?" Others said it was a dictator, a drill sergeant, an ignored boss, a mother-in-law, or a toddler. When we asked who they wished it could be, many envisioned a supportive person such as a friend, a coach, a colleague, a monk, or a mom when you are sick. The answers revealed how clinicians perceive their relationship to patient monitoring alarms tones – how it is and how it could be.

To redesign these tones, we spent time among those who hear them. Here, we detail a participatory design project to improve the alarm tones of a widely used patient monitoring system: the Philips IntelliVue. As Özcan, Birdja, and Edworthy (2018) have argued, holistic and collaborative approaches are needed in the design of alarm tones that include a wider range of stakeholders in healthcare. In that spirit, this project engaged clinicians to co-create new alarm tones, as well as new criteria on which to evaluate them. Through the process, we established that medical device companies can improve device sounds through discourse with those who use them.

This project was a transdisciplinary endeavour, integrating the expertise of designers, social scientists, artists, and engineers (many who had been patients or family caregivers) working with nurses, doctors, and other stakeholders as advisors and participants – all centered on creating better alarm tones and a more dignified experience for clinicians and patients. Our objectives were to 1) learn from clinicians what it means for alarms to sound "better," and then 2) redesign them accordingly. It is important to note that this study did not focus on alarm management or the architectural acoustics; it focused on the tones themselves. The products of our work were digital audio files to be played by patient monitoring systems under established alarming conditions of low, medium, and high priority.

2. Background

The unintended consequences of noise in hospital environments, particularly from alarms, are well documented. Noise has negative impacts on patients, who may suffer from stress, lack of sleep, and alarm fatigue and associated health risks (Topf, 2000; Basner, 2011; Shivers et al., 2013; Basner et al., 2014; Sakallaris et al., 2015; Sen & Sen, 2020). It affects the health, perception, cognition, and learning of hospitalized infants and children (Brown, 2009; Wachman & Lahav, 2011; Erickson & Newman, 2017; Smith et al., 2018). Noisy alarms also have a significant impact on clinicians. Alarm fatigue, stress, and burnout are common, particularly for nurses (Topf & Dillon, 1988; Topf, 2000; Morrison et al., 2003; Ryherd et al., 2008; J. P. Keller et al., 2011; Watson et al., 2015; Cho et al., 2016). Alarm sound can also hinder communication and cognition in the operating room (Murthy et al., 1995; Hasfeldt et al., 2010; S. Keller et al., 2016). Risks to clinicians, in turn, become additional risks to patient safety.

Efforts to address alarm noise tend to focus on alarm management (J. P. Keller et al., 2011; Drew et al., 2014), rather than the design of the tones. Studies of the design of alarm tones often focus on functional attributes, i.e., the performance they induce in clinicians. Such attributes include: audibility (C. L. Bennett et al., 2015; Hasanain et al., 2017; Bolton et al., 2019); identifiability (J. R. Edworthy et al., 2018; McNeer et al., 2018; C. Bennett et al., 2019); localizability (J. Edworthy et al., 2017, 2018); and learnability (Phansalkar et al., 2010; Gillard & Schutz, 2016; McDougall & Edworthy, 2018; McDougall et al., 2020).

Only recently have studies been directed to the design of alarm tones as a way of addressing the consequences of noise in hospital environments. In a study of alarm volume in the operating room, Schlesinger et al (2018) found that – even though conventional wisdom is that for alarms to be heard, "louder is better" – task performance was maintained even when alarm volume was noticeably lower than background sound levels, and it may be safe to decrease volumes in operational settings. Özcan et al (2018) note that tones themselves contribute to adverse experiences of alarms, that medical devices continue to "beep" even though digital technology allows nearly any sound to be used. They offer the *CareTunes* prototype as demonstration that pleasant yet informative tones are possible. Industry guidelines, too, are becoming more specific about both performance requirements for alarms tones and detrimental side-effects. For example:

> "Candidate sounds should be conspicuous, distinctive, and reasonably pleasant… Attention-getting ability, distinctiveness, clear communication of the desired information (source, urgency, and meaning), and freedom from annoyance and aversion" (Association for the Advancement of Medical Instrumentation, 2018).

Thus, it seems useful to distinguish between two types of alarm tone design criteria: functional and sensible. *Functional* means that alarm tones have direct performance benefits (realized by patients and hospitals), for which there are well-established criteria, e.g., audibility and identifiability. *Sensible* means that alarm tones avoid unintended consequences, e.g., fatigue and stress (costs incurred by clinicians and patients) and may also yield indirect performance benefits for patients and hospitals, for example: less fatigue leading to fewer missed alarms. We can think of functionality and sensibility as axes for an production possibilities curve (see **Figure 1** in section **9. Figures and Tables**). "Conventional wisdom," as Schlesinger et al (2018) put it, holds that sensibility must come at the expense of functionality and vice versa; we must move along the utility curve.

But perhaps – through unconventional wisdom – we can create ways to push the curve out to get more sensible tones without sacrificing functionality, or even get more of both. Consideration for sensibility may actually improve functionality, as it does (for example) in the design of the driving experience of automobiles and cutting experience of cooking knives (Norman, 2002). In the medical field, experience design in some hospitals has focused on cuisine: better hospital food made by professional chefs to improve patient nutrition (Schiffman, 2018). As such, we wanted to understand how alarms are heard by those who hear them regularly and what is important to them.

3. Case study

The subject of our project was a widely used patient monitoring system: the Philips IntelliVue. Patient monitors provide real-time physiological information regarding patients' vital signs to clinicians, but are audible to others. Alarms are sounded at the bedside, as well as at the central station, on mobile caregiver apps, and patient worn telemetry systems. In ICU settings, patient monitors are the source of up to 82% of alarm tones (Cho et al., 2016).

There are over one million IntelliVue monitors in use worldwide and they are used in a variety of care settings, such as the emergency department, intensive care unit, operating and recovery rooms as well as on the general ward. Because of its ubiquity, improving the sounds of patient monitoring could be an important step to improving entire hospital soundscapes. The three IntelliVue alarm tones that were the focus of this project (high, medium, and low priority) are detailed in **Table 1**.

We examined the original alarms of the IntelliVue and alternatives through interviews, workshops, and surveys primarily with direct users of the patient monitoring system (e.g. nurses and anaesthesiologists) as well as indirect users (e.g. patients, family members, and hospital staff, who also hear and respond to alarms), interspersed with two rounds of sound design. From this, we developed a set of criteria for improving alarms – requirements grounded in concerns voiced by those on the frontline. We developed and refined new alarm concepts based on what clinicians and other stakeholders had to say about existing alarms, then decided among and further refined those concepts based on pairwise comparisons using the criteria developed. The final product was a set of new alarm tones for patient monitoring and a set of new criteria with which to assess them. The tones were intended to be as functional as, and more sensible than, current alarms.

The tension to be resolved in this project concerned the degree and speed of change. At the start of this project, we knew (and through the course of this project confirmed) that patient monitor alarm tones – while essential – can have unintended detrimental effects. As such, we must explore ways to improve them. On the other hand, Philips's patient monitoring sounds have not changed in over 40 years. Clinicians around the world have grown accustomed to those sounds and depend on them to provide care. To change these sounds too drastically or quickly, without their input, would be not only be unsafe, but also lack respect for their lived experiences. Our resolution of this tension was to keep changes incremental, to base any changes on input from clinical stakeholders, and to use the understanding we develop as the basis for more radical changes in future projects, having established that change is viable.[1]

As such, this project had two aims:

- Patient monitoring sounds are thoughtfully re-examined and incrementally re-designed to reflect the needs of those who have to hear them.
- Trust is built among stakeholders through co-creation, non-disruptive change, and mutual learning to prepare us for more transformative changes to follow.

The key ingredient for any innovation, including but not limited to patient monitoring alarm tones, is trust (Clegg et al., 2002; Dovey, 2009; Sol et al., 2013). Changes in technology can

[1] For this reason, we did not create alarms that were radical departures from the originals. We aimed for tones that were distinctly recognizable as a Philips patient monitoring system. To ensure safety and performance, Philips will continue to test the efficacy of these tones and ensure compliance with applicable regulations prior to product release. IEC standard tones will also continue to be included as options in the IntelliVue.

be unfamiliar and frightening, and they can shift power dynamics within communities. A responsible innovation process helps stakeholders to feel safe. It assures clinicians that they will recognize and act upon new alarm tones. It supplements, but does not replace, engineers' and scientists' objective expertise with clinicians' subjective experience. Taking it slow (spreading changes out over multiple years and multiple projects), as well as making the entire process as transparent as possible (through exposure in venues such as this), is essential.

4. Methodology

This project employed a participatory approach to the design of patient monitoring alarm tones. Engagements with clinicians and other healthcare workers recognized their subjective experiences and latent desires to drive systematic change. As recommended by Özcan et al (2018), our research methods were ethnographic in nature (Coffey & Atkinson, 1996; Creswell, 2007), embedded within an iterative human-centered design process (IDEO, 2021; Liedtka et al., 2013). As such, the purpose was to empathize and share semantics with those who use alarming devices to make sense of what "better" means to them (Denzau & North, 1994; Weick, 1995).

Across technology areas, non-expert stakeholders have a great deal to contribute. They excel in articulating ethical concerns and normative judgments that experts – operating from a stance of objectivity – often do not address (Sclove, 2010). Moreover, non-expert stakeholders are entitled to shape the innovations that will impact their lives (Weller et al., 2021). For patient monitoring, clinicians can help direct the creation of alarm tones that are the tools of their trade. Our process was structured for them to provide such direction.

Research and design occurred in 2 rounds over 6 months. In each round of engagement, we diverged through open-ended questions in individual interviews and group workshops followed by qualitative analysis, then converged through close-ended questions in an online survey followed by quantitative analysis. All the while, sound design also diverged and converged per feedback, eliminating certain options and creating or refining others. Here, we describe research and sound design methods separately, but it should be noted that they occurred together in each round. The research and design process is illustrated and described further in **Figure 2.**

4.1 Research methods

Research used both qualitative and quantitative methods. We first conducted 39 semi-structured interviews with several communities of stakeholders and 5 semi-structured workshops with more than 100 participants in 12 countries. Participants were mostly nurses and anaesthesiologists, who are the primary, active listeners of the patient monitoring alarm tones, as well as other healthcare workers, such as ICU physicians, a hospital chaplain, a unit secretary, and some patients and family members. We asked about experiences with alarms and assessments of current and alternative tone designs, which we played in sessions.

Participants were encouraged to talk about their perceptions and preferences. The purpose of interviews and workshops was to understand how stakeholders thought about the patient monitor alarm tones and what was important to them about those tones.

Most interview and workshop participants were clinical specialists employed by Philips, and introductions were made and workshops organized by the Philips team. In addition, other interviewees were part of the authors' professional network, resulting from prior work with US hospital systems such as Johns Hopkins and Northwell Health. 5 virtual workshops were conducted with participants in Germany, Netherlands, Switzerland, and United Kingdom (02 Feb 2021); Canada and United States (on 05 February 2021); Japan (also on 05 February 2021); Columbia and Brazil (on 26 February 2021); and Australia, New Zealand, and Korea (on 09 March 2021). However, tangible input used for coding and synthesis was only available for 3 workshops (participants in Germany, Netherlands, Switzerland, United Kingdom, Canada, United States, and Japan). Detailed demographic data are provided in **Table 2**.

In each engagement we took verbatim notes, excerpts of which were then coded and clustered (Saldaña, 2015). Data – in the form of excerpts – were analyzed using Miro, a virtual whiteboard wherein excerpts existed on "sticky notes" with "tags" to apply codes. Coding was an abductive process (Coffey & Atkinson, 1996) of comparing empirical data with existing knowledge. Some codes were pre-set, given the literature reviewed, particularly industry standards for alarm tones (Association for the Advancement of Medical Instrumentation, 2018; International Electrotechnical Commission, 2020); other codes emerged through the process of reading, coding, and organizing. Two of the authors (Avery Sen and Sage Palmedo) performed the coding. This yielded over 7500 excerpts, tagged by one or more of 38 codes:25 initial codes to specify which question was being answered, and 13 for important alarm tone factors. A list of these latter codes is provided in **Table 3**.

Coded excerpts were then clustered into typological models by linking similar or related factors together (see **Figure 3**). Developing models was a process of defining and organizing clusters visually as a concept map (Bryson et al., 2004; Novak & Cañas, 2008). Clusters were formed organically and spatially. Separate models were made to capture: important design factors for each alarm priority; factors related to learning and changing alarm tones; audio parameters participants noticed; as well as personification of the patient monitor; and worldview about the harshness of alarms. Criteria for "better" alarms were derived from the first models: important design factors for each alarm priority. These criteria then served as indicators by which alternative tones were compared and assessed in surveys with respect to each other (via pairwise comparisons).

Surveys were conducted using the Qualtrics platform, which allows audio files to be played within the survey. Survey takers were recruited from interviewees and workshop participants, as well as from audiences at virtual conference presentations about the project, and who were informed about the problem of alarm noise as well as our approach to a solution. In the first survey (n=98), multiple options were presented for each alarm (6 for low, 5 for medium, and 5 for high, plus 9 combinations of low, medium, and/or high) across 18 comparisons. The options in each pair were carefully selected to isolate changes in single parameters, e.g., pitch or timbre (See **Figure 4**). Candidate tones were embedded in a simulated, stereo recording of hospital ambience. Participants used sliders to indicate relative preference between options A and B on each of the eight criteria (on a scale of -10 for A to +10 for B), then a discrete all-things-considered judgement of either A, B, or no preference. Per the results of the first survey, options for new tones were refined and narrowed down to one each for high, medium, and low priority. The second survey (n=25) employed the same structure, but the options for each tone were limited to two: the original versus a final option for the new tone. The differences between this final set of new tones and the originals are detailed in **Table 4**.

Participants in the first survey were invited from the list of interviewees and attendees at workshops. Participants in the second survey were invited as part of a presentation about the project at two separate venues: an internal hospital conference at Children's' National Hospital in Washington, DC, and the 2021 American Association of Critical-Care Nurses National Teaching Institute convention. Surveys were conducted online with sound files embedded in the survey. While we could not control the listening conditions of participants, survey instructions told them, "headphones are encouraged. If using headphones, please start with the volume half way to make sure it is not too loud. Feel free to use this audio clip to adjust the volume." An audio clip was provided that represented the volume range of alarm tones to follow. Questions were ordered to prevent hearing fatigue and retraumatization by alternating among gentler and harsher tones. **Figure 5** shows the template for all survey comparison questions.

4.2 Sound design methods

The sound designers for the project were co-authors Yoko Sen and Matthew Barile, whose qualifications are described briefly in the author information at the end of this paper. They used production software such as Ableton Live and Max/MSP. We started with the original, proprietary Philips alarm tones and created alternatives per analysis of 1) the capabilities of the audio hardware, as determined by measures of frequency response and total harmonic distortion and 2) stylistic and branding preferences and opportunities for improvement identified by Philips Research and Development. For example, original low and high priority alarms were 8-bit audio files. As a starting point, 16-bit versions were developed to eliminate quantization noise, leaving all other alarm parameters the same.

Design of alternative tones considered not only stated preferences on the audio parameters of each sound, but also the possible cognitive effects of variations across parameters. For example, regarding the new low priority tone, the main difference is its amplitude envelope. Compared to the original low priority tone, the new tone uses a "rounder," non-instantaneous attack with a longer loudness sustain (to maintain audibility) followed by an extensive decay. This design decision was driven by clinicians' wish for "less technical," "less square," and "rounder, more natural" sounds, especially for the most commonly heard, low priority alarm. The resulting sound is more percussive than flat, as defined by MAPLE Lab (2018), which is beneficial as there can be a 60% increase in the memorability of sequence-object associations when using percussive as compared to flat tones.

Sound design also had to take into consideration the acoustic properties of IntelliVue hardware and of the human ear. For example, regarding the high priority, the original is a 1s naturally decaying (percussive) tone consisting of a fundamental frequency just below 1kHz and its 3rd harmonic. The 3rd harmonic was 2dB louder than the fundamental as measured using the WAV file (versus through speaker, with casing and room effects). The frequency response of the IntelliVue speaker also amplifies the harmonic by 4-5dB. Amplification is compounded in the ear canal, which increases sound pressure of frequencies between 2600 to 3000Hz by 14-18dB (Silva et al., 2014). Speaker and ear effects combine to increase the 3rd harmonic by 18-23dB, making perceptual loudness 4 times greater than intended.

In fact, many clinicians wished for high priority alarms to be "less shrill," "less grating," "less penetrating," "less harsh" – referring to timbre. Several clinicians noted that this alarm sounds when a patient is "coding" (having cardiac arrest, possibly passing away) and how this tone stays in their head, "goes home with you, you can hear in your sleep." While many acknowledged that the annoyance of the high priority tone prompts action, several expressed concerns for patients, particularly infants and children, those with autism, PTSD, or delirium, and women in labor. They wanted tones that are "softer but still be able to alert you enough," and to be "warned without [being] jolted." The new high priority tone used in this comparison compensates for the perceptual loudness increase by reducing this harmonic (keeping other parameters the same) to create a more balanced timbre.

Interpreting participant feedback is a nuanced tradecraft. Whenever possible, we rely upon common interpretations of colloquial language (e.g. "this sounds too square" translates to an amplitude envelope with a short attack and release). When this is not possible, we may ask participants (in interviews or workshops, and with open comment spaces on surveys) for elaboration, for onomatopoeia, or to mimic sound, or refer to analogous sounds with comparable qualities, or other means. In our interpretations, we focused on the elements of greatest consistency amongst participants.

5. Findings

In interviews and workshops, we asked participants to talk about their experiences with alarm tones, including but not limited to patient monitoring. These engagements revealed a mixed relationship that clinicians have with alarms: they are necessary, yet fatiguing, but they can be made less so. When asked about times they had to learn the sounds of a new device, they reported that learning happens by association; connecting sounds to what they mean requires repetition. One nurse, who specializes in training other nurses in the use of medical devices, observed that "if you like it, you quickly adopt it, but if you dislike it, it will take longer." By and large, participants said that learning happens on-the-job, however dedicated training would be desirable.

For most of each session, participants listened and reacted to the current patient monitor alarm tones, as well as – in later sessions – tone concepts we created per feedback in earlier sessions. For each sound, we asked "what do you like, wish, or wonder? What might you change or not change? What are the benefits/risks of making changes?" Responses to these prompts were clustered, ultimately forming the following ten criteria, the top eight of which were used in surveys.[2] The criteria are as follows. The first five address functionality and the latter five sensibility*:*

1. More easily heard and attention-getting (stands out above the background)
2. More distinct from other sounds (not mistaken for something else)
3. Sounds more like what it means (low, medium, or high priority)
4. Better stimulates a prompt response (not ignored or unattended)
5. Easier to isolate (among possible locations if could be coming from) *
6. Less startling or aggravating (beyond what is necessary to get attention)
7. Less fatiguing to hear over time (everyday, possibly for many years)
8. Less distressing for patients to hear (either short or long term)
9. Not distracting from important conversations (with clinicians or family) *
10. More pleasant to me (given my personal taste)

We also took note of which audio parameters participants used to distinguish among alarms. When distinctions were made, it was mostly in terms of pitch (perception of aggregate frequencies in the sound) and interval (time between beeps), less often in terms of amplitude envelope (volume change over time), timbre (texture of the source "instrument"), or dyad use (two overlapping notes). While not always using the jargon of sound designers, interviewees were able to use colloquial language to describe and compare sounds in terms of pitch (8 participants, e.g. "pitch"), interval (7 participants, e.g. "tempo"), amplitude envelope (1 participant, e.g. "softer edge") and timbre (2 participants, e.g. "tone"). For example, one participant said, "high pitched & louder... and the tone too, the actual tone of the sound," which we interpreted as a distinguishing timbre from pitch and volume.

To conclude sessions, we asked, "if the patient monitor were a person, who would it be? Who do you wish it were?" In general, participants imagined the monitor as an authoritarian but preferred it to be a partner. This also revealed contrasting perspectives. Some felt alarms should be harsh and there must be a trade-off between functionality and sensibility. They were afraid of human error and said that people often need to be forced to do what is needed, and even pleasant alarms can be fatiguing. Others felt that alarms could be gentler, that there need not be a trade-off, that clinicians and patients alike are victims of aggressive

[2] In the interest of reducing the survey length and participant time commitment, two criteria (indicated by the asterisk * in the list above) were not used in the surveys. These were the two that, in interviews, participants indicated were least important.

alarms, that more or stronger alarms are not the answer, and that even harsh alarms can be ignored.

The first survey (n=98) tested for preferences across alternatives for the three alarm priorities on parameters of interval, amplitude envelope, pitch, timbre, and dyad use. For low priority options, comparisons revealed an overall preference for the original pitch and timbre, but no preference between intervals of 4 or 6 seconds. For medium priority, there were mixed preferences for more acoustic timbres, and no preference between the original pitch and a higher pitch. For high priority, there were very clear preferences for an interval of 1 rather than 2 seconds, and timbres that did not contain 4 harmonics, as recommended by the IEC for greater audibility (International Electrotechnical Commission, 2020).

In terms of preferences among criteria, for low priority options, there were few trade-offs made between those for functionality (e.g., “easily heard, gets attention”) and those for sensibility (e.g., “not fatiguing to hear often”); preferences on each criterion reflected their overall preference. For example, when comparing low priority alarms that differed in amplitude envelope, most participants preferred a new option over the original, and their preferences on each of the 8 criteria reflected that overall preference (see **Figure 6**). In contrast, there were clear trade-offs made for medium and high priority options; preferences on functionality versus sensibility criteria did not always reflect overall preferences. For example, when comparing high priority alarms that differed in interval, most participants had an overall preference for 1s over 2s. Those who preferred 1s overall also preferred 1s on functionality criteria, but preferred 2s on sensibility criteria (See **Figure 7**).

The second survey (n=25) compared the final option for a new alarm set with the original set. The changes made from original to new are detailed in Table 1. It revealed overall preferences for the new versus original sounds by wide margins: low priority (23:1), medium priority (16:4), high priority (18:5), excluding responses of no preference. Participants preferred all three on sensibility criteria, but there were mixed results on functionality criteria: the new high priority was much preferred, the new medium was somewhat preferred, and the new low was only preferred on a single criterion (“sounds like what it means”).

All three of the final redesigned tones were rated much higher on every dimension than the original sounds, with few exceptions. On all three, sensibility ratings were high, and higher than functionality ratings. For low-priority, ratings did not suggest a clear preference for the functionality of new over original sounds (despite other studies showing benefits of percussive tones over flat tones in sequence-object associations). For medium-priority, ratings suggested a clear but moderate preference for the functionality of new over original tones. And for high-priority, ratings suggested the strongest preference for the functionality of new over original tones. It was surprising that the high-priority alarm, with the most incremental change from the original (of the three alarms), was considered to have the greatest improvement in functionality.

The final question on the survey asked participants to prioritize the criteria for improving alarms – in rank order on the first survey (n=98) and on a scale of 1-10 on the second survey (n=25). When rank-ordering, clinicians put functionality criteria (in general) above sensibility criteria. However, they also ranked one sensibility criterion equally important as some functionality criteria. "Easily heard, gets attention" and "stimulates a prompt response" were tied for top priority. "Not fatiguing to hear often" was in the 2nd tier with "distinct from other sounds" and "sounds like what it means." Tier 3 had "not overly startling or aggravating" and "not distressing to patients," and "reasonably pleasant was the lowest priority. Scale of 1-10 results yielded only two tiers; the top priorities (9 of 10) were "sounds like what it means," "not overly startling or aggravating," and "not fatiguing to hear often," while all other criteria were in the second tier (8 of 10). Both ways of prioritizing revealed that some sensibility criteria were as important to participants as some functionality criteria. When a trade-off was forced with rank order, functionality criteria are generally more important; when not forced, all criteria were similarly important (averaging at either an 8 or 9 of 10).

6. Discussion

The most recent international standards for medical devices state that "the two main requirements of the new auditory alarm tones are that they are audible and recognizable. Almost everything else is a matter of taste and preference" (International Electrotechnical Commission, 2020 p.92). Our research suggests that this understanding may benefit from more nuanced consideration of sensibility. Fatigue, stress, and delirium are serious unintended consequences beyond "taste and preference." Moreover, "taste and preference" might also improve performance.

Whether addressed by setting broad standards or through the design for particular tones, our findings show that alarm tone sensibility – reducing clinician fatigue over the long term, shock and aggravation in the moment, both consequences for patients, and general pleasantness – is important to those who use alarms to care for patients. Sounds that are too loud would be an obvious workplace hazard. The findings of this study indicate that this may be true for not-so-obvious parameters, such as the number of harmonics, or amplitude envelope. Short-term gains in functionality (i.e. a more prompt response) can bring with them long-term losses for both clinicians and patients (i.e. alarm fatigue).

This project served as a proof-of-concept in other ways, as well. It showed that criteria for medical alarm tone design – for functionality, sensibility, and perhaps other factors yet to emerge – can be developed empirically through deliberate engagement with those who use these sounds as tools of their trade, followed by qualitative data analysis to provide structure to free-form responses. Criteria thusly developed are meaningful to clinicians and can then be used in more close-ended engagements (e.g., surveys) as the basis upon which to empirically assess alarm tone options. Perhaps most importantly, this study suggests that functionality need not be sacrificed for sensibility, at least for low and medium priority alarms, which are heard most often. For high priority alarms, the trade-offs become more apparent. Clinicians tend to feel that more important alarms can be harsh, but that less important alarms need not be.

This project also demonstrated how alarm tones can be effectively analyzed and designed in terms of changes to a small number of discrete parameters. Clinicians can perceive changes in these parameters, name them, isolate changes in them, express how changes to parameters elicit in themselves specific cognitive and physiological reactions, and they have reasonable ideas about how to change some and not others. Clinicians need not be passive recipients of the alarms they hear; they can be active, insightful partners in the creation of alarms that better meet their needs, if we – the designers and engineers of alarming devices – invite them. What more might we learn about how to improve the sound of alarms by creating shared vocabularies, experimenting with new media, and creating new venues for conversation with those we design for?

The findings and conclusions presented here are the product of the particular populations we engaged. Participants were clinical specialists within Philips, plus some who work for hospitals; mostly nurses and anaesthesiologists; largely those in English-speaking countries. Future studies of this kind might yield additional findings or interpretations by engaging clinicians in a wider range of contexts. Indeed, the qualities of alarm tone “sensibility” may well vary from culture to culture.

Other limitations concern the survey results. While the first survey had 98 respondents, the second had only 25 – low from a statistical perspective. While that survey offers positive initial results with respect to the new tones, more data from additional usability testing would help to increase confidence. Further, as noted above, all survey takers were knowledgeable about this project and its goals, and so there may be some selection bias. On the other hand, this may indicate the importance of helping clinicians be aware of the full range of unintended consequences of alarm tones – for themselves and for patients – and the importance of training on new alarms for their adoption and use. Future studies might test the efficacy of alarms and education in combination, rather than alarms alone.

The tones we designed were intended to be as functional as, and more sensible than, current alarms of the IntelliVue. Our initial findings indicate that this is likely the case, and ongoing usability testing may offer additional insights regarding how to make them more functional or more sensible prior to release. However, many of the unintended consequences that the sensibility criteria seek to address occur incrementally over time. Thus, whether they do reduce fatigue and stress, for example, must be the subject of longer term longitudinal studies.

Future research could address the interrelationship between functionality and sensibility. Many responses we received led us to wonder whether alarm sensibility, itself, could improve functionality. Is it possible that, as one interviewee suggested, if you like it, you learn it? Might it be, as other participants suggested, that clinicians “tune out” alarms they perceive to be too harsh? Could preference affect performance? Might it be that more aesthetically pleasing alarm tones work better (Norman, 2002)? If so, in what ways and under what circumstances?

This project purposefully kept changes incremental in the interest of keeping tones familiar. This limited how much the new tones could address sensibility criteria. However, as learning alarm tones occurs through association, it is possible that they could be made dramatically more sensible if more time could be dedicated to training. Future research could discover how much training is needed for sounds of increasing degrees of sensibility to induce equivalent performance. Clinicians are highly trained professionals, and perhaps the up-front costs of additional training on alarms would be worth the avoided daily costs of aggravation and fatigue that could accumulate over years.

Future research may also benefit from the using the sound design research process described here (and illustrated in Figure 2) as a template. While many aspects of the design process are tacit and unique to each designer (e.g. knowing how to talk with stakeholders, interpret their verbal and non-verbal input, and change a particular tone in response), other aspects are somewhat standardized. In particular, the iterative diverge-converge structure – wherein ideas and input increase, then decrease, in number and breadth over each iteration – is a staple of human-centered design and could be applied more widely in medical device sound design. Another aspect we would hope to see become more standardized is, as discussed above, purposely seeking out and including the voices of stakeholders beyond direct “users,” whose perspectives have not traditionally been represented, so that we can design with their values and institutional contexts in mind.

7. Conclusion

> “Currently the monitor is a very dominant person. I wish it would be more of a sensitive person, more of a partner that is supporting me. Commanding, empathetic, but not a jackass. Nice about stuff.”
> – ICU Nurse

This project provides insights about how to design alarm tones: criteria for sensibility are important to those who use alarms; such criteria can be developed empirically by engaging those people; such criteria and engagements can also be used to empirically assess alarm tone options; and it is possible to make alarms more sensible without making them less functional. Though our research, we also uncovered deeper conclusions about the culture of alarm development and how it fits into broader culture of healthcare, particularly regarding power dynamics. Per Özcan et al (2018), "People who have direct issues with alarms… have little authority to change alarms.”

Eisler (1987) defines culture as how relationships are structured and describes relationships on a continuum from dominance to partnership. In domination systems, relationships are hierarchical and driven by fear; in partnership systems, relationships are egalitarian, trust based, and involve shared decision making. Healthcare culture inclines toward domination, which impacts patient care (Oehlert, 2015). It is no coincidence that clinicians characterized patient monitoring alarms as a “dominant person.” The sounds that surround them embody the culture from which they emerged and hearing such sounds everyday may reinforce fear-based relationships. As such, designing the tones of devices to embody the characteristics of a trusted partner – kind, considerate, and respectful of human dignity during vulnerable times – is about more than improving devices; it is a way of transforming culture.

Acknowledgements: We would like to thank all of the clinicians, hospital team members and patients who participated in this design research without whom this project would not have been possible.

8. Bibliography

Association for the Advancement of Medical Instrumentation. (2018). ANSI/AAMI HE75:2009 (R2018) Human factors engineering—Design of medical devices (HE75:2009 (R2018)). American National Standards Institute Inc.

Basner, M. (2011, October 4). Noise: The Other Negative Effect of False Alarms. AAMI Alarms Summit, Herndon, VA.

Basner, M., Babisch, W., Davis, A., Brink, M., Clark, C., Janssen, S., & Stansfeld, S. (2014). Auditory and non-auditory effects of noise on health. The Lancet, 383(9925), 1325–1332.

Bennett, C., 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), 1050–1057.

Bennett, C. L., Dudaryk, R., Ayers, A. L., & McNeer, R. R. (2015). Simulating environmental and psychological acoustic factors of the operating room. The Journal of the Acoustical Society of America, 138(6), 3855–3863.

Bloomenthal, A. (2021, November 7). What the Production Possibility Frontier (PPF) Curve Shows. Investopedia. https://www.investopedia.com/terms/p/productionpossibilityfrontier.asp

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: The Journal of the Human Factors and Ergonomics Society, 001872081986291.

Brown, G. (2009). NICU Noise and the Preterm Infant. Neonatal Network, 28(3), 165–173.

Bryson, J. M., Ackermann, F., Eden, C., & Finn, C. B. (2004). Visible Thinking: Unlocking Causal Mapping for Practical Business Results (1st edition). Wiley.

Cho, O. M., Kim, H., Lee, Y. W., & Cho, I. (2016). Clinical Alarms in Intensive Care Units: Perceived Obstacles of Alarm Management and Alarm Fatigue in Nurses. Healthcare Informatics Research, 22(1), 46–53.

Clegg, C., Unsworth, K., Epitropaki, O., & Parker, G. (2002). Implicating trust in the innovation process. Journal of Occupational and Organizational Psychology, 75(4), 409–422.

Coffey, A., & Atkinson, P. (1996). Making sense of qualitative data: Complementary research strategies. Sage.

Creswell, J. W. (2007). Qualitative inquiry & research design: Choosing among five approaches. Sage Publications.

Denzau, A. T., & North, D. C. (1994). Shared Mental Models: Ideologies and Institutions. Kyklos, 47(1), 3–31.

Dovey, K. (2009). The role of trust in innovation. The Learning Organization, 16(4), 311–325.

Drew, B. J., Harris, P., Zègre-Hemsey, J. K., Mammone, T., Schindler, D., Salas-Boni, R., Bai, Y., Tinoco, A., Ding, Q., & Hu, X. (2014). Insights into the Problem of Alarm Fatigue with Physiologic Monitor Devices: A Comprehensive Observational Study of Consecutive Intensive Care Unit Patients. PLoS ONE, 9(10).

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: The Quarterly of Human Factors Applications, 26(4), 4–13.

Edworthy, J., 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: The Journal of the Human Factors and Ergonomics Society, 59(7), 1108–1127.

Edworthy, J., 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, 88–93.

Eisler, R. (1987). Chalice and the blade. Our history, our future. San Francisco, CA: Harper & Row.

Erickson, L. C., & Newman, R. S. (2017). Influences of background noise on infants and children. Current Directions in Psychological Science, 26(5), 451–457.

Gillard, J., & Schutz, M. (2016). Composing alarms: Considering the musical aspects of auditory alarm design. Neurocase, 22(6), 566–576.

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, 500–514.

Hasfeldt, D., Laerkner, E., & Birkelund, R. (2010). Noise in the Operating Room—What Do We Know? A Review of the Literature. Journal of PeriAnesthesia Nursing, 25(6), 380–386.

IDEO. (2021). Design Kit: Extremes and Mainstreams. https://www.designkit.org/methods/45

International Electrotechnical Commission. (2020). IEC 60601-1-8:2006+AMD1:2012+AMD2:2020 Medical electrical equipment. International Standards Organization.

Keller, J. P., Diefes, R., Graham, K., Meyers, M., & Pelczarski, K. (2011). Why Clinical Alarms Are a 'Top Ten' Hazard: How You Can Help Reduce the Risk. Biomedical Instrumentation & Technology, 45(s1), 17–23.

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, 59(12), 1541–1552.

Liedtka, J., King, A., & Bennett, K. (2013). Solving Problems with Design Thinking: Ten Stories of What Works. Columbia University Press.

MAPLE Lab. (2018). Amplitude Envelope. https://maplelab.net/overview/amplitude-envelope/

McDougall, S., & Edworthy, J. (2018, July 1). Soundscaping: Sound, meaning and vision in healthcare alarm systems. Proceedings of the 32nd International BCS Human Computer Interaction Conference.

McDougall, S., Edworthy, J., Sinimeri, D., Goodliffe, J., Bradley, D., & Foster, J. (2020). Searching for meaning in sound: Learning and interpreting alarm signals in visual environments. Journal of Experimental Psychology: Applied, 26(1), 89–107.

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, 129(1), 58–66.

Morrison, W. E., Haas, E. C., Shaffner, D. H., Garrett, E. S., & Fackler, J. C. (2003). Noise, stress, and annoyance in a pediatric intensive care unit. Critical Care Medicine, 31(1), 113–119.

Murthy, V. S., Malhotra, S. K., Bala, I., & Raghunathan, M. (1995). Detrimental effects of noise on anaesthetists. Canadian Journal of Anaesthesia = Journal Canadien D'anesthesie, 42(7), 608–611.

Norman, D. A. (2002). Emotion & Design: Attractive things work better. Interactions Magazine, 9(4), 36–42.

Novak, J. D., & Cañas, A. J. (2008). The theory underlying concept maps and how to construct and use them. Technical Report IHMC CmapTools 2006-01 Rev 01-2008.

Oehlert, J. K. (2015) "Themes in Health Care Culture: Application of Cultural Transformation Theory," Interdisciplinary Journal of Partnership Studies: Vol. 2: Iss. 1, Article 6.

Özcan, E., Birdja, D., & Edworthy, J. R. (2018). A Holistic and Collaborative Approach to Audible Alarm Design. Biomedical Instrumentation & Technology, 52(6), 422–432.

Phansalkar, S., Edworthy, J., Hellier, E., Seger, D. L., Schedlbauer, A., Avery, A. J., & Bates, D. W. (2010). A review of human factors principles for the design and implementation of medication safety alerts in clinical information systems. Journal of the American Medical Informatics Association, 17(5), 493–501.

Ryherd, E. E., Waye, K. P., & Ljungkvist, L. (2008). Characterizing noise and perceived work environment in a neurological intensive care unit. The Journal of the Acoustical Society of America, 123(2), 747–756.

Sakallaris, B. R., Macallister, L., Voss, M., Smith, K., & Jonas, W. B. (2015). Optimal Healing Environments. Global Advances in Health and Medicine, 4(3), 40–45.

Saldaña, J. (2015). The Coding Manual for Qualitative Researchers (3rd edition). SAGE Publications Ltd.

Schiffman, R. (2018, September 20). Hospital Food You Can Get Excited About. The New York Times. https://www.nytimes.com/2018/09/20/well/eat/hospital-food-you-can-get-excited-about.html

Schlesinger, J. J., Baum Miller, S. H., Nash, K., Bruce, M., Ashmead, D., Shotwell, M. S., Edworthy, J. R., Wallace, M. T., & Weinger, M. B. (2018). Acoustic features of auditory medical alarms—An experimental study of alarm volume. The Journal of the Acoustical Society of America, 143(6), 3688–3697.

Sclove, R. (2010). Reinventing Technology Assessment: A 21st Century Model (STIP 01; Science and Technology Innovation Program). Woodrow Wilson Center for International Scholars.

Sen, Y., & Sen, A. (2020). The Future of Hospital Sound: Transforming Healthcare Through Sound Experience. Music and Medicine, 12(1), 5.

Shivers, J. P., Mackowiak, L., Anhalt, H., & Zisser, H. (2013). "Turn it Off!": Diabetes Device Alarm Fatigue Considerations for the Present and the Future. Journal of Diabetes Science and Technology, 7(3), 789–794.

Silva, A. P. R. da, Blasca, W. Q., Lauris, J. R. P., & Oliveira, J. R. M. de. (2014). Correlation between the characteristics of resonance and aging of the external ear. CoDAS, 26, 112–116.

Smith, S. W., Ortmann, A. J., & Clark, W. W. (2018). Noise in the Neonatal Intensive Care Unit: A New Approach to Examining Acoustic Events. Noise & Health, 20(95), 121–130.

Sol, J., Beers, P. J., & Wals, A. E. J. (2013). Social learning in regional innovation networks: Trust, commitment and reframing as emergent properties of interaction. Journal of Cleaner Production, 49, 35–43.

Topf, M. (2000). Hospital noise pollution: An environmental stress model to guide research and clinical interventions. Journal of Advanced Nursing, 31(3), 520–528.

Topf, M., & Dillon, E. (1988). Noise-induced stress as a predictor of burnout in critical care nurses. Heart & Lung: The Journal of Critical Care, 17(5), 567–574.

Wachman, E. M., & Lahav, A. (2011). The effects of noise on preterm infants in the NICU. Archives of Disease in Childhood - Fetal and Neonatal Edition, 96(4), F305–F309.

Watson, J., Kinstler, A., Vidonish, W. P., Wagner, M., Lin, L., Davis, K. G., Kotowski, S. E., & Daraiseh, N. M. (2015). Impact of Noise on Nurses in Pediatric Intensive Care Units. American Journal of Critical Care: An Official Publication, American Association of Critical-Care Nurses, 24(5), 377–384.

Weick, K. E. (1995). Sensemaking in organizations. Sage Publications.

Weller, N., Govani, M. S., & Farooque, M. (2021, May 19). Need Public Policy for Human Gene Editing, Heatwaves, or Asteroids? Try Thinking Like a Citizen. Issues in Science and Technology.

About the Authors:

Avery Sen, Co-founder and Research Director at Sen Sound, has 15 years of experience in R&D strategy and evaluation. He holds a PhD in Public Administration and MA in Science & Technology Policy from The George Washington University, and BA in Science & Technology Studies from Cornell University.

Yoko Sen, Founder of Sen Sound, award-winning electronic musician, and former patient. She was a citizen artist fellow at The Kennedy Center and artist-in-residence at Johns Hopkins Sibley Innovation Hub, Stanford Medicine X, and Kaiser Permanente, and a finalist for International Design Excellence Awards by IDSA.

Matthew Barile, Director of Sound Design at Sen Sound, has 11 years of experience as a composer, sound-designer, producer, and music educator. His design work includes Google Calendar notifications, Hangouts, Voice Search, as well as ringtones, app sounds and system notifications released on over 1 billion Android phones globally.

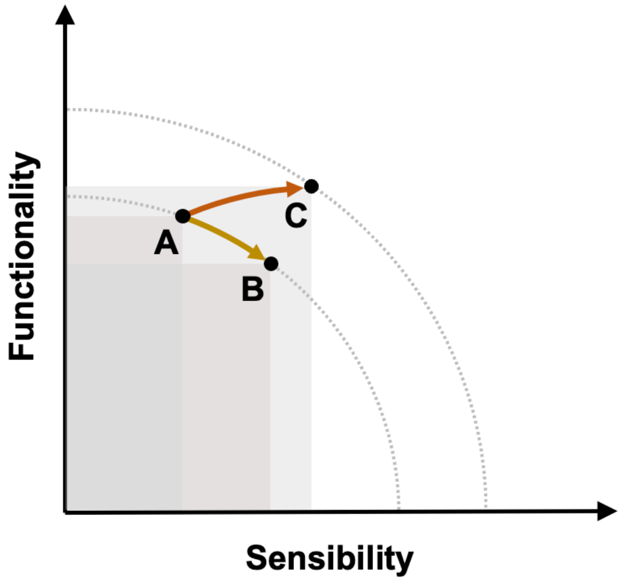
Sage Palmedo, a Health and Humanities Scholar pursuing an M.D. at Dartmouth Medical School, is a multi-disciplinary artist and scientist. She was a sound designer and researcher at Sen Sound and a researcher at the Murthy Lab at the Princeton Neuroscience Institute, studying the evolution of acoustic perception.

Andreas Walden, Usability Leader at Philips Medical Systems, works on human-centered design. With 15 years at Philips, he served as R&D engineer and user experience and usability expert. He holds an engineering degree in information technology from Berufsakademie Stuttgart and an IT systems degree from Xerox GmbH.

Vitor Vicente Antunes, Systems Architect at Philips Medical Systems, is an expert in alarming software and alarming. With 22 years at Philips, he has developed clinical applications for patient monitors. He holds a Dipl.Ing. (FH) in Cartography from Karlsruhe University of Applied Sciences.

9. Figures and Tables

How do we make tones better and what does "better" mean?



Functionality

- More easily heard and attention-getting (stands out above the background)
- More distinct from other sounds (not mistaken for something else)
- Sounds more like what it means (low, medium, or high priority)
- Better stimulates a prompt response (not ignored or unattended)
- Easier to isolate (among possible locations if could be coming from)

Sensibility

- Less startling or aggravating (beyond what is necessary to get attention)
- Less fatiguing to hear over time (everyday, possibly for many years)
- Less distressing for patients to hear (either short or long term)
- Not distracting from important conversations (with clinicians or family)
- More pleasant to me (given my personal taste)

Figure 1. What does it mean to sound "better?" Conventional wisdom assumes that, in order to make a tone more sensible, its functionality must be sacrificed. This is illustrated here by a shift from point A to B along a production possibilities curve (Bloomenthal 2021), representing a constrained trade off. However, our research suggests that the trade off need not be so constrained. Through careful design and training, it may be possible to move from a more constrained to a less constrained trade off space, enabling a shift from point A to point C.

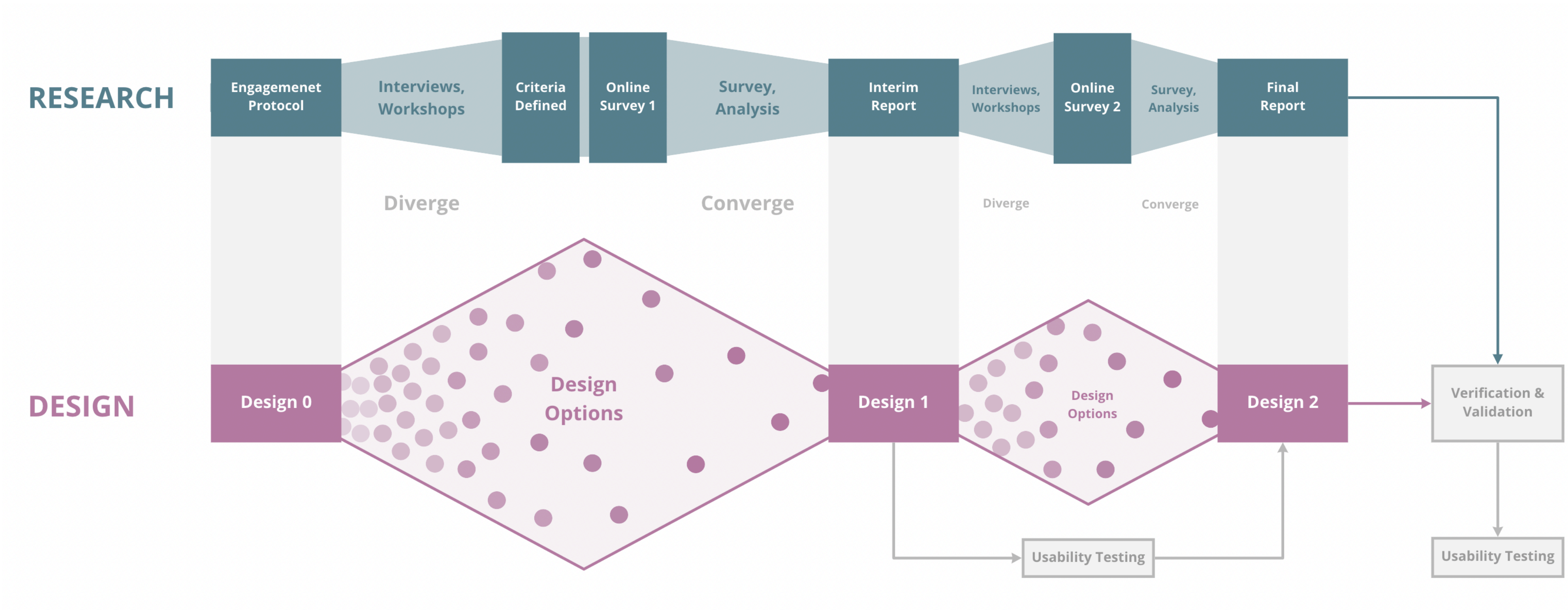


Figure 2. After conducting a literature review, a speaker hardware analysis, and establishing general constraints, the research and design process took a "double diamond" approach. Each diamond represents a design iteration. In the "divergent thinking" phase of each iteration, open-ended input was gathered and synthesized with participants though interviews with individuals and workshops with groups, and sound design concurrently responded to this input by exploring a growing number of alarm tone options. In the "convergent thinking" phase of each iteration, alarm tone options were narrowed down to include a limited number for inclusion in surveys, participants answered close-ended questions (largely pairwise comparisons), and survey data were analysed to inform the final prototype tone set for that iteration. "Design 0" was the initial prototype tone set that was used alongside the original tones for listen-and-react questions in the first set of engagements, "Design 1" was used this way in the second set of engagements, and "Design 2" was the final set that went into verification and validation processes at Philips.

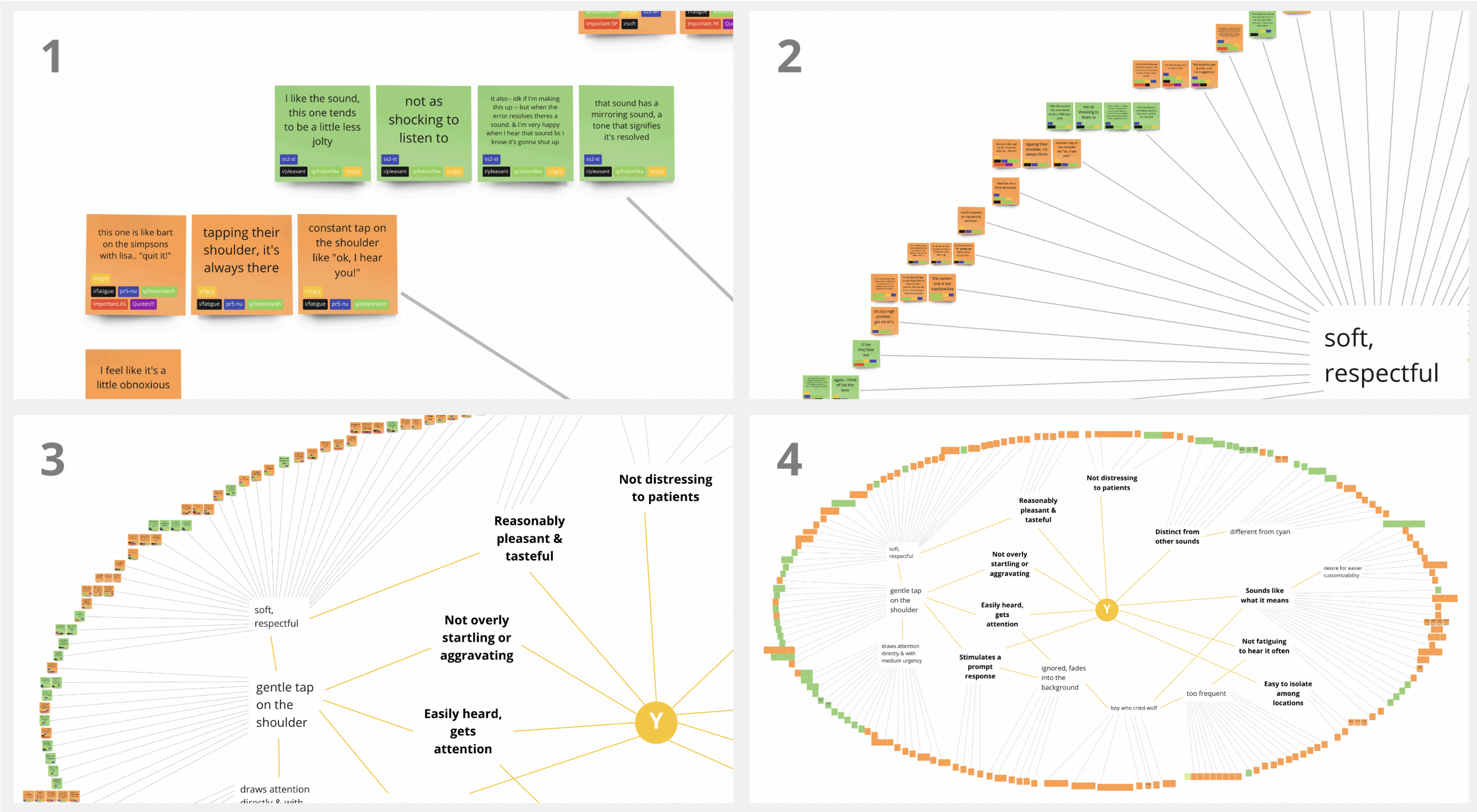


Figure 3. Example of coded excerpts and clustering for comments about what is important regarding the medium priority (yellow) alarm tone, zooming out from position 1 (individual excerpts) to position 4 (full typological model).

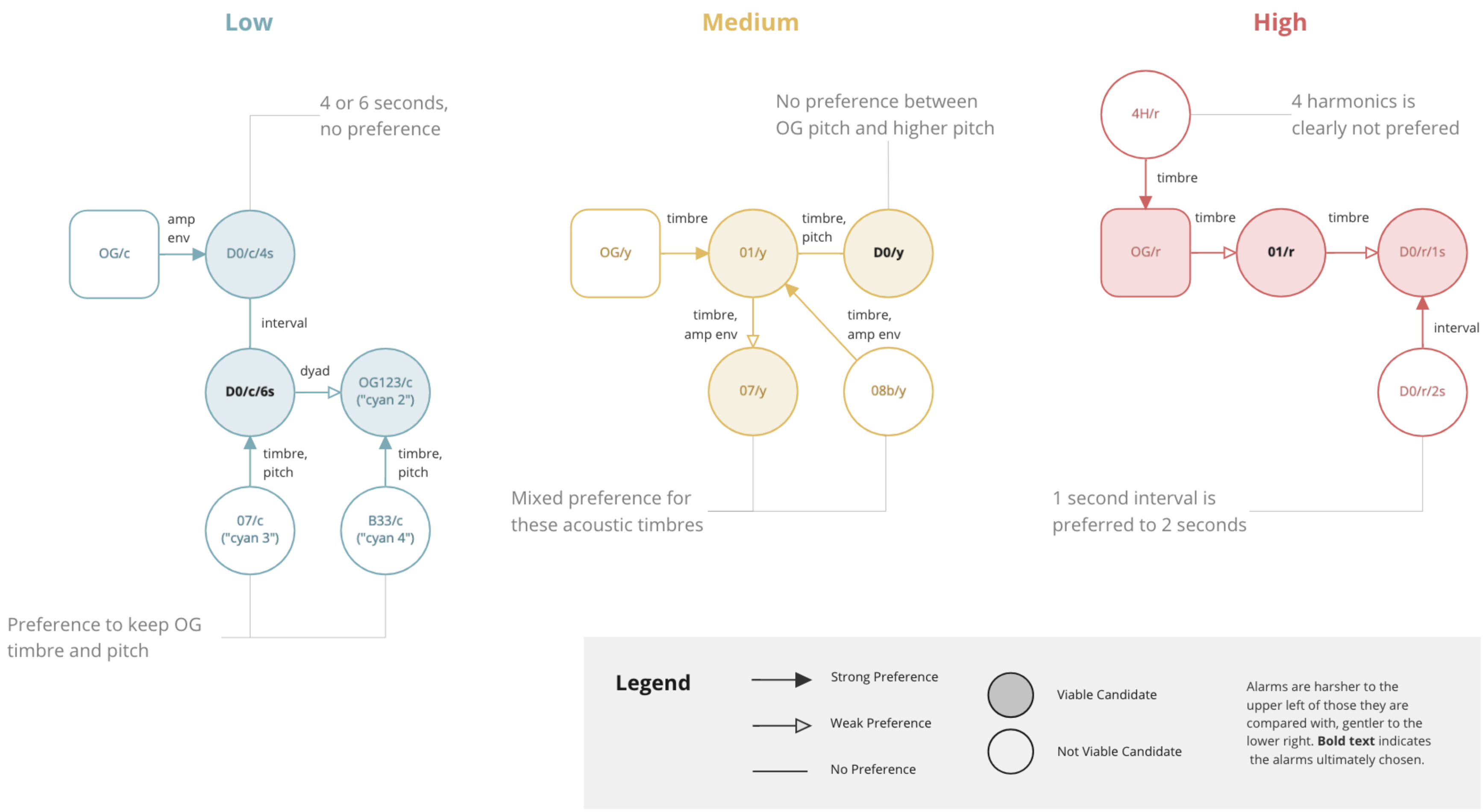


Figure 4. Comparisons and relative preferences among alarm tone options, per "All Things Considered" question in survey 1. Each question compared a specific difference on one or two parameters, such as pitch and timbre. Sound names start with their timbral "family" and include other parameter information as well. "OG" is the original sound family, while "D0," "01," "07," "08," "B33," "OG123," and "4H" are prototype alarm families created for this project. For each family, "/c" is a cyan (low priority) alarm tone, "/y" is a yellow (medium priority) alarm tone, and "/r" is the red (high priority) alarm tone. "/1s," "/2s," "/4s," and "/6s" represent the interval length for each tone.

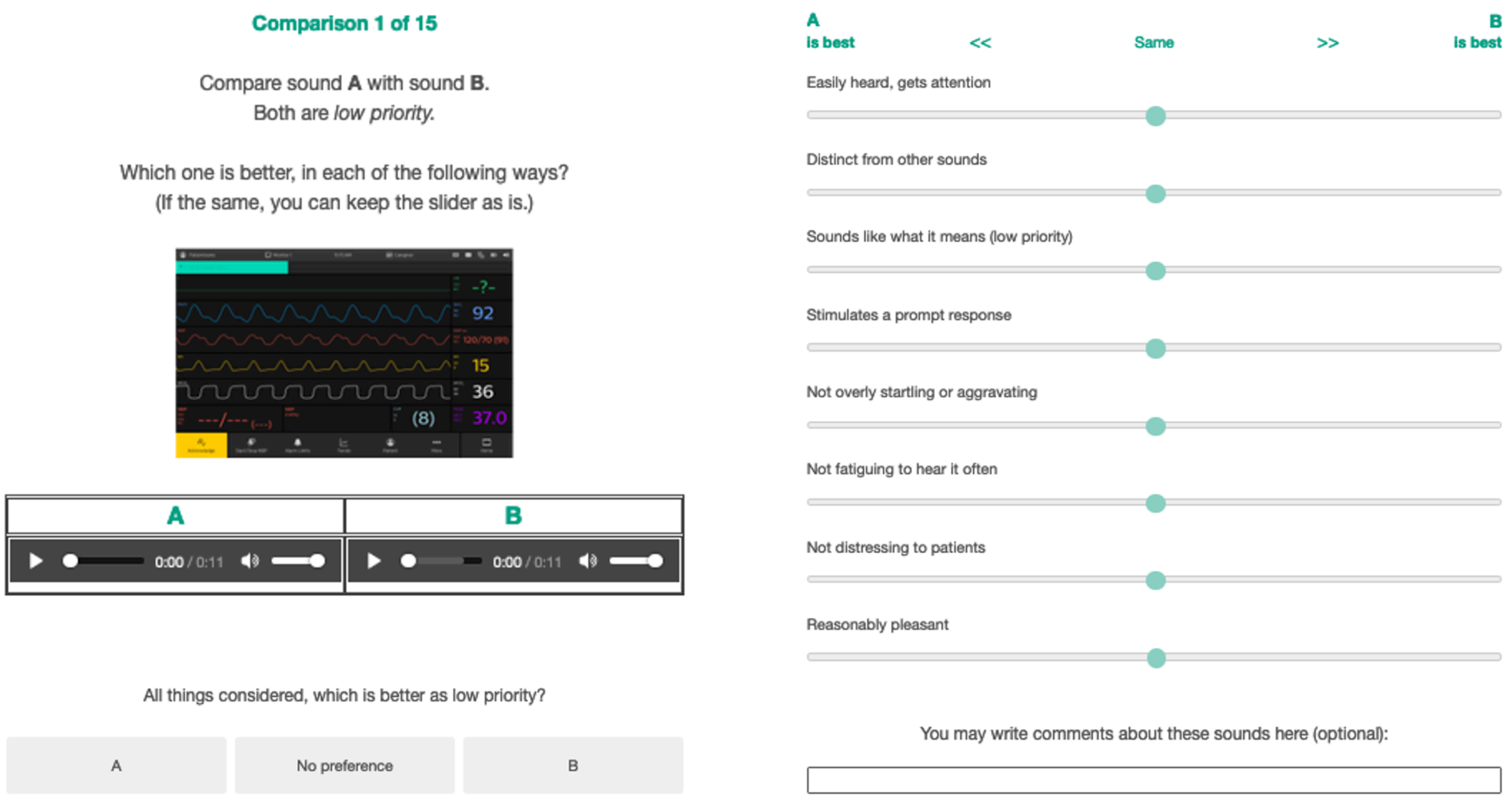


Figure 5. Example of question language and format for both surveys. Audio files were embedded in the survey page, with sound "A" on the left and "B" in the right. Actual sound designations were not displayed for the survey taker. Participants could move sliders in either direction to state preferences independently for each of 8 criteria. The position was captured on a 10 point scale in either direction (-10 to +10). Participants also were asked to provide an overall preference, A or B, and to write any comments they might have.

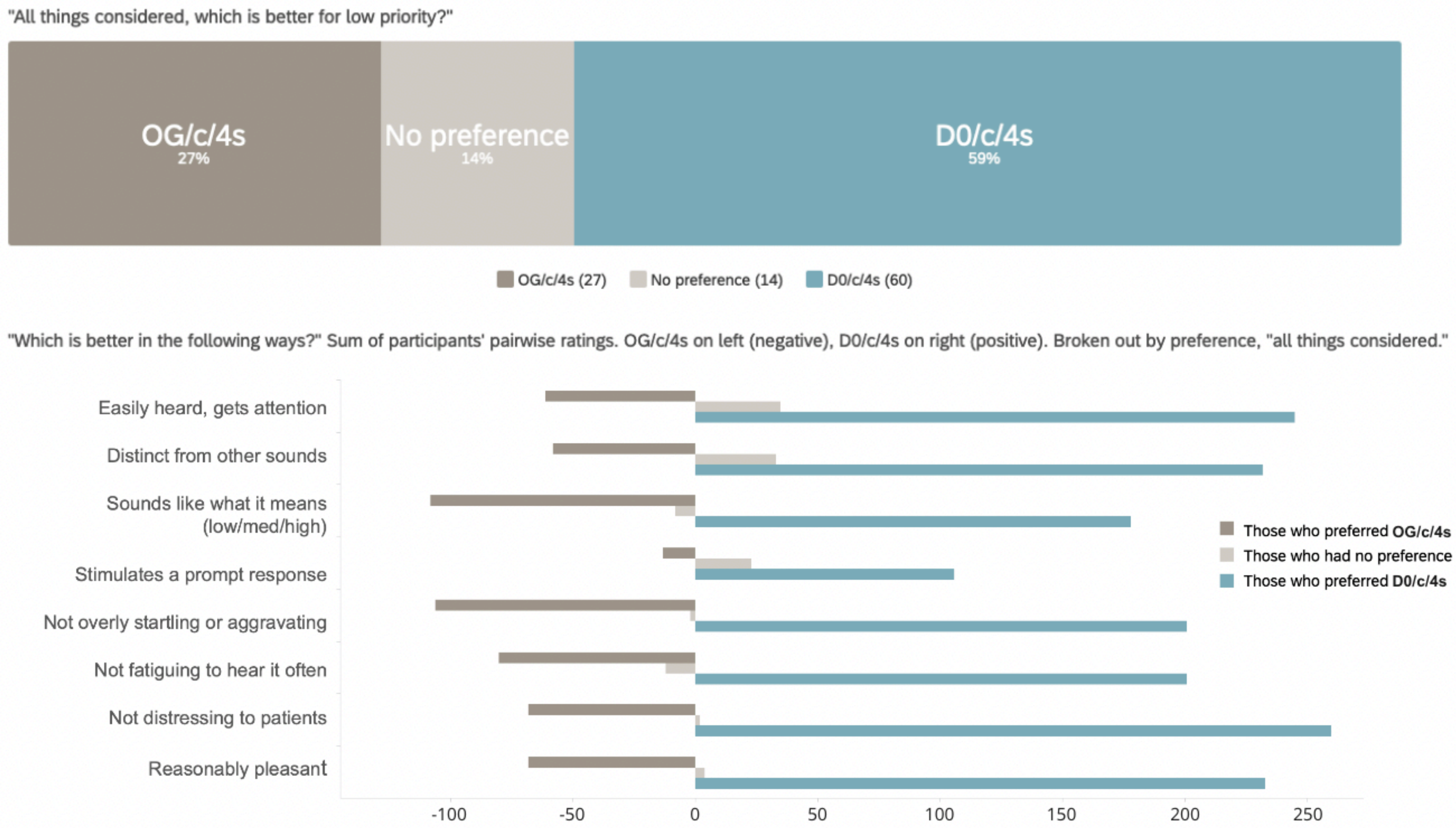


Figure 6. Participant preferences for two low priority alarms differing only in amplitude envelope, original ("OG") versus new ("D0"). For each criterion, preferences for the original extend to the left, and preferences for the new extend to the right.

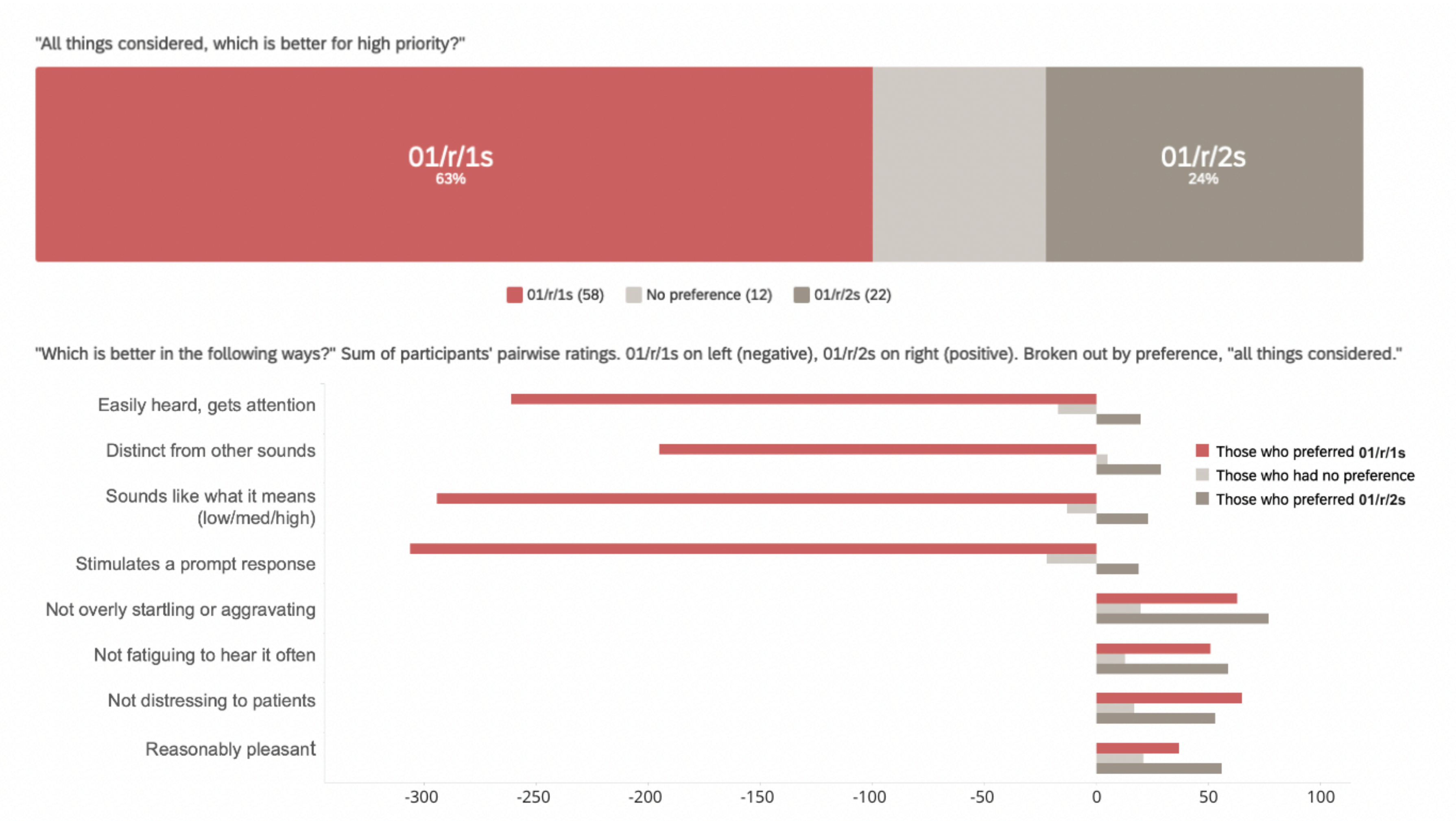


Figure 7. Participant preferences for two high priority alarms differing only in interval, 1 second versus 2 seconds. For each criterion, preferences for the 1 second version extend to the left, and preferences for the 2 second version extend to the right

Table 1. Alarm priorities and their meanings. (Definitions per International Electrotechnical Commission, 2020, pp.13-14)

Priority	Color	Definition	Example condition
Low	Cyan	Operator awareness is required and future action might be needed. Awareness implies the planning of future workflow is expected.	• Failure of an infusion pump for maintenance of intravenous fluids • Failure of an enteral feeding pump • Failure of a patient weighing system
Medium	Yellow	Prompt operator response is required. Prompt implies the re-planning of current workflow is expected.	• High or low blood pressure • Mild hypoxemia • High or low pCO2
High	Red	Immediate operator response is required. Immediate implies the interruption of current workflow is expected.	• Asystole • Ventricular fibrillation • Extreme hypoxemia

Table 2. Participants demographics in each engagement. All interviews and workshops were conducted virtually over Zoom. Dashes indicate no data available. (Note: numbers for workshops represent those who contributed input via online webform during the event, and which was formally synthesized. However, there were more who attended and whose voices were heard.)

	Interviews	**Workshops**	**Survey 1**	**Survey 2**
Responsibility				
I am/was directly responsible for attending to patient monitoring alarms	32	---	91	23
I have never been directly responsible for attending to patient monitoring alarms.	7	---	7	2
Experience with patient monitoring alarms				
Nurse	21	25	60	23
Physician	4	3	8	0
Anesthesiologist	4	1	15	0
Medical student or resident	1	0	1	1
Another healthcare worker (paramedic, chaplain, secretary)	3	5	7	1
Not a healthcare worker (patient, family caregiver, device engineer, device marketer)	6	6	7	0
Tenure in a clinical setting				
0-5 years	6	2	13	3
6-10 years	5	9	20	4
11-15 years	7	8	18	8
16-20 years	0	2	10	0
20+ years	11	11	32	9
unknown (but more than 0 years)	4	1	0	0
never	6	7	5	1
Location				
North America (Canada, United States)	32	24	---	---
Europe (Germany, Netherlands, Switzerland, United Kingdom)	6	12	---	---
Asia (Japan)	1	4	---	---
Total	**39**	**40**	**98**	**25**

Table 3. Codes used for developing criteria that guided sound design. Sources for codes derived from literature are provided, otherwise labelled as emergent.

Code	Meaning	Source	Criterion in Final Phrasing
i/hear	I can hear it, it stands out above the rest	ANSI/AAMI, 2018; IEC, 2020, C. L. Bennett et al., 2015; Hasanain et al., 2017; Bolton et al., 2019	More easily heard and attention-getting (stands out above the background)
i/attention	It gets my attention (or not), can't ignore (or can)	ANSI/AAMI, 2018; IEC, 2020	
i/means	I know what it means (or not)	ANSI/AAMI, 2018; IEC, 2020; Phansalkar et al., 2010; Gillard & Schutz, 2016; McDougall & Edworthy, 2018; McDougall et al., 2020	Sounds more like what it means (low, medium, or high priority)
i/confuse	It may (or may not) be confused with another sound	ANSI/AAMI, 2018; IEC, 2020; J. R. Edworthy et al., 2018; McNeer et al., 2018; C. Bennett et al., 2019	More distinct from other sounds (not mistaken for something else)
i/action	It makes me move, take action (or not)	Emergent	Better stimulates a prompt response (not ignored or unattended)
i/find	I can find it, isolate it among others (or not)	ANSI/AAMI, 2018; IEC, 2020; J. Edworthy et al., 2017, 2018	Easier to isolate (among possible locations if could be coming from)
i/soft	It is soft, gentle, round (or not), aggressive (or not)	Emergent	Less startling or aggravating (beyond what is necessary to get attention)
i/startle	It is startling in the moment (or not)	ANSI/AAMI, 2018	
i/fatigue	It is stressing or fatiguing over time (or not)	IEC, 2020; Topf & Dillon, 1988; Topf, 2000; Morrison et al., 2003; Ryherd et al., 2008; J. P. Keller et al., 2011; Watson et al., 2015; Cho et al., 2016	Less fatiguing to hear over time (everyday, possibly for many years)
i/patient	It is startling or fatiguing to the patient (or not)	Topf, 2000; Basner, 2011; Shivers et al., 2013; Basner et al., 2014; Sakallaris et al., 2015; Sen & Sen, 2020	Less distressing for patients to hear (either short or long term)
i/distract	It is distracting (or not), I can (or can't) hear what I need	Murthy et al., 1995; Hasfeldt et al., 2010; S. Keller et al., 2016	Not distracting from important conversations (with clinicians or family)
i/mech	It sounds technical, mechanical, clinical (or not)	Emergent	More pleasant to me (given my personal taste)
i/pleasant	It is pleasant, likeable (or not) vs annoying irritating (or not)	ANSI/AAMI; 2018	

Table 4. How and why alarms were changed

Alarm	Example Excerpts	Parameter	Change	Benefit of Change	Risk of Change
Low Priority	• *Maybe more present but less frequent* • *I would have it softer and less frequent* • *[I wish] a slower pace of the beat could achieve the same level of attention/alarm* • *I'm fine with the fact that this alarm is a little difficult to notice* • *I wish there was more pause between beeps* • *I wonder how it could be enriched to become less dull, less square, less strict* • *[I wish] it would sound a bit more smooth and organic. Not so mechanic* • *I wish they would be less technical, more human or more natural, like music* • *reestablish a connection that 'this is a person we care about'* • *I like it's pitch, it's in a good & useful range* • *[I like that it's] distinguishable from other higher priority alerts.*	Interval	Increased	Reduced fatigue, more distinguishable from yellow	Potential to not hear (mitigated by note length increase)
		Amplitude envelope	Attack, sustain, and release increased. Overall length increased	Rounder, less machine-like, and less mistakable for a QRS tone	Potential to be mistaken for yellow (mitigated by yellow pitch increase)
		Pitch	Unchanged	Familiar, immediately recognizable	None
		Timbre	Unchanged	Familiar, immediately recognizable	None
Medium Priority	• *Could the interval be longer?* • *It is the most common and constant alarm.... when mixing several monitors with the same sound is annoying* • *if it's moderate priority I would want something to cue me in about the urgency* • *We all tend to tune out what occurs repetitively*	Interval	Increased	Reduced fatigue, more distinguishable from cyan and red	Possibly mistaken as a lower priority (mitigated by even greater interval for cyan)
		Amplitude envelope	Sustain decreased, but overall length maintained.	More urgent (percussive) feel than cyan, with smoother tail	Potential not to hear (mitigated by pitch increase)

	• *My opinion is: change the INOP to the last "beep" sound (or maybe the yellow) and leave the red as is. the inop and yellow are too close in nature* • *If I had a wish, Philips monitors should have a kind of harmony in their tones... that is a major not a minor harmony* • *The INOP and yellow are too close in nature* • *It gets your attention but not in an abrasive way* • *Ideal world - still a sound that gains attention... but less harsh*	Pitch	Increased	More distinguishable from cyan	Potential for annoyance (mitigated by sustain decrease)
		Timbre	Harmonics removed	Less aggravating	Potential not to hear (mitigated by pitch increase)
High Priority	• *The red alarm is annoying, but it's supposed to be annoying* • *On my ears this sound is a little annoying, but I know this is a red alarm & only plays for life threatening event & I need to look at my patient* • *I'm hyper sensitive to it, which is good, but I wish there was some way where it wouldn't make me keep hearing it after the shift is over* • *I can't stand the tone of it either.. the beeping & it's so loud creates anxiety just from hearing it, by the way it resonates* • *It's like a bell ringing inside my head* • *I'm very sensitive for the high frequencies* • *Laboring patients who have preeclampsia can be affected and are potentially hypersensitive to seizures with this sound.* • *I wish those alarms were more pleasant*	Interval	Unchanged (even though most interviewees stated that an increase would be OK)	Maintain feeling of urgency	None
		Amplitude envelope	Mostly the same	Keep highest priority alarm most recognizable	None
		Pitch	Unchanged	Familiar, immediately recognizable	None
		Timbre	Harmonic attenuated	Less aggravating	May be slightly less audible, but difference is negligible