Utrecht University

Master's Thesis HCI-6245293

Human Computer Interaction

Towards a Better Understanding of Auditory Feedback in Warehouse Management Systems to Improve Usability

By
Lode Dams - 6245293
l.dams@students.uu.nl

Coordinators:
First Supervisor: Frans Wiering
Utrecht University - f.wiering@uu.nl

Second Supervisor: Christof van Nimwegen
Utrecht University - c.vannimwegen@uu.nl

External Supervisor: Rick Soons
Rentman - rick@rentman.nl

January 19, 2024

Universiteit Utrecht

Preface

Presented here is the Master’s thesis titled "Towards a Better Understanding of Auditory Feedback in Warehouse Management Systems to Improve Usability." This thesis serves as the concluding piece for the Master’s program in Human-Computer Interaction at Utrecht University. I dedicated significant effort from May 2023 to January 2024 to accomplish this work.

After a well-deserved break in Central America, I embarked on the search for a thesis topic in February. Following numerous emails, phone calls, and meetings, it became apparent that my current employer was the optimal choice for my thesis. They offered considerable flexibility in choosing research topics, and I was confident in receiving excellent support. Given my strong interest in sounds and sound design, influenced by a Sound & Music Technology course during my Master’s program, I identified this as a fitting subject that allowed me to blend creativity with design. Rentman graciously supported my research efforts and provided insights into their product development process.

Another influential factor in my decision to pursue a sound-related topic was the availability of a supervisor. Frans Wiering expressed his interest in supervising a topic related to sound, providing the decisive push for me to choose this subject. With his support and expertise, I knew I was in capable hands. I am grateful to him for the constructive feedback sessions, in-person meetings at my office, his organized approach, and great communication. Our collaboration was enjoyable, as it felt like he regarded me as an equal, fostering discussions on a pleasant and mutually respectful level.

I would also like to express my gratitude to Rick Soons for providing me with the opportunity at Rentman and for expressing trust in my process. It felt like I could openly discuss anything with him and turn to him for any problems I encountered. He inspired and ensured that the pertinent business aspects of the thesis were adequately addressed. I hope to continue this collaboration with him in the form of colleagues in the near future.

Lastly, I would like to thank my family and closest friends. Even though they may not realize it, they have been the consistent driving force that kept me motivated over the past few months. I cherished every moment spent with them, although they might have found it less enjoyable when I delved into the intricacies of my thesis issues. They provided the energy I needed to stay on track and were always available for advice or a listening ear when I needed it most. I am grateful and proud to call them my family and close friends. I cannot thank them enough for all the wonderful moments we have spent together.

I hope you have an enjoyable reading experience.

Lode Dams
19th of January, 2024

Contents

Abstract

This study investigates the impact of enhanced auditory feedback on the usability of equipment scanning on handheld devices in Warehouse Management Systems (WMSs). Today, warehouses deal with labor shortages and optimization challenges. There is a growing recognition of the significance of incorporating multimodality to craft immersive user experiences (UX) and to provide more natural and robust interaction, which is important for WMSs due to the pursuit of low error rates and high efficiency. While visual feedback is commonly used in WMSs, audio feedback during barcode and RFID scanning is often overlooked and limited to beeps. Both auditory icons and earcons have proven to positively impact usability issues in other domains, but there is still a gap in the literature regarding usability impact in the warehousing domain. Our research includes a threefold of qualitative and quantitative studies consecutively exploring, verifying, and validating auditory feedback for WMSs. Potential improvements are explored, of which the sound design will then be verified. During the validation step, four conditions are quantitatively evaluated on task completion time, number of errors, perceived workload, annoyance, and subjective rating. Enhanced auditory feedback has lower perceived workload, is less frustrating and less annoying compared to conventional auditory feedback. RFID scanning is more efficient and effective, while barcode/QR code scanning is less mentally demanding and has a higher SUS score. The results of this study contribute to a better understanding of how usability in warehouses could be improved, which in turn impacts the ongoing optimization challenges and solves potential labor shortages because of improved UX. Results could be extended to other domains where scan processes are used, like in retail or transport.

Keywords

Auditory icons, earcons, auditory feedback, warehouse management system, RFID, barcode, QR code, usability

1 Introduction

In today's digital landscape, user interfaces (UIs) play a crucial role in enabling seamless interactions between humans and technology. One of the most successful improvements to the user-computer interface made is the inclusion of icons – graphical symbols that visually represent information in the computer display. Icons can present a great deal of information concisely [13]. While traditionally the majority of displays have been designed for the human visual system, the importance of multimodality, in specific audio, in creating immersive and engaging user experiences (UX) is nowadays recognized as we are moving towards multimodal user interfaces [20, 61, 89]. Overly dense visual displays can lead to cognitive overload, negatively impacting the user's decision-making performance, which in turn explains the exploration of alternate modalities to convey information [13].

Moreover, we continuously interact with our environment using our five senses. Past studies have shown that information spread over more than one modality helps to reduce the user's cognitive load, is perceived as more usable and enjoyable, and helps support the working memory [2]. More specifically, audio has proven to give valuable feedback to users' actions, carry information, provide information beyond the field of view, enhance visual representation, strengthen the emotion that UI creates, and immerse users in the environment [61]. Therefore, audio seems like a solid additional modality to the visual display and should not be neglected.

Audio in UI occurs in three different types: auditory feedback, music, and voice [47]. While the latter two are part of a product's landscape, this study will solely focus on auditory feedback. Although many variations exist, roughly three types of auditory feedback are distinguished. Firstly, *auditory icons* are familiar sounds based on experiences in the real world [47]. These are brief sounds that represent objects, functions, and actions. They are the auditory equivalent of visual icons, like the sound of paper crumbled up when an item is moved to your trash can [29, 49]. Secondly, *earcons* are abstract, synthetic, and musical tones or sound patterns, having a metaphorical relation with the object (the referent) they represent [21, 29, 47]. Blattner et al. [13] define earcons as "audio messages that are used in the computer/user interface to provide information to the user about some computer object, operation or action". An example of an earcon includes the initiation of Apple's Siri.

Both auditory icons and earcons are types of nonspeech auditory feedback. *Spearcons*, on the other hand, are spoken words or phrases sped up until they may no longer be recognized as speech. They represent an action or object in the software, just like auditory icons and earcons. Despite being easy to learn, efficient, effective, and easy to create, this research will solely focus on nonspeech auditory feedback. One reason for this is that spearcons are often deemed more irritating and disturbing compared to auditory icons and earcons, rendering them a less optimal choice for most UIs [29, 73, 89, 98]. Furthermore, spearcons should be considered with other speech-based auditory feedback methods as well, like spindexes and normal speech [89]. These methods are not in the scope of this study. Figure 1 shows the exact focus of this study.

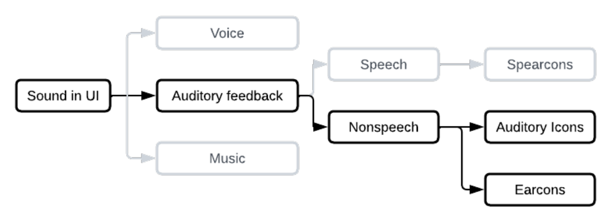

Figure 1: Scope of this study.

Having been neglected for a long time, the use of audio in UI is now receiving more attention in e.g. the video game industry, retail, and automotive domain [39, 73, 77, 86]. However, despite advancements in audio technology and lots of research, many UIs still fail to fully leverage the potential of sound, leading to UX issues like frustration, incomprehension, and annoyance [36]. Unlike the visual Gestalt principles, there is no clear heuristic guidance for sound when designing interfaces [75, 89]. Many domains still need extensive research to fully leverage the potential of auditory feedback.

This also extends to the warehousing domain with Warehouse Management Systems (WMSs), which have a dominant role in modern supply chain management by linking the material flows between the supplier and customer [84]. While visual feedback during scan processes in mobile WMS applications is typically provided through a visual display, audio feedback is often overlooked or limited to a beep by the scanner. This becomes a significant problem in e.g. stock control systems involving complex visual displays and where the volume of information to be communicated increases, or location tracking tasks with very limited auditory feedback [65, 87].

Additionally, warehouse managers have to deal with labor shortages. Both attracting qualified employees and retaining them is a major problem, resulting in increased training costs and operations slowing down [46]. An aging warehouse workforce proves that the new generation of workers is not willing to put the warehouse working conditions the way their predecessors did. To offset this issue, new technologies should be introduced. Mobile WMS could help with this and should include intuitive, functional, enjoyable, and ease-to-use interfaces [46, 57]. Proven to be efficient and immersive, enhanced auditory feedback in the form of auditory icons and earcons during scan processes should be considered in these WMSs [61].

Only little research has been done on auditory feedback in WMSs and current research is still lacking regarding a WMS's nonspeech auditory feedback. The introduction of technology in WMSs effectively changed how many employees conducted their jobs. Nevertheless, the effectiveness that many technological innovations offer in WMSs has rarely been scientifically evaluated [11]. In comparison to all technological advancements in warehouses, that contrast is noticeably skewed. One of these technological advancements is Radio Frequency Identification (RFID) for warehouse inventory management systems which should be thoroughly evaluated to get a clear overview of its usability [95].

Despite the fact that RFID and barcode scanning are widely used in warehouses, a good auditory solution for scanning RFID tags is still absent and auditory feedback for barcode scanning has, to our knowledge, never been evaluated. As there is a gap in scientifically testing and evaluating technological possibilities in WMSs, this study aims to contribute to a better understanding of auditory feedback in WMSs. More specifically, this study will focus on the usability of auditory feedback during scan processes in a warehouse, including handheld RFID and barcode scanning. The following research question will be answered:

Research question: How does enhanced auditory feedback impact the usability of equipment scanning on mobile devices in Warehouse Management Systems?

By answering this research question, we not only contribute to the need for scientifically testing and evaluating technological solutions of WMSs, but we also gain knowledge about the usability of auditory feedback about the relatively new RFID technology and already existing barcode technology implemented in warehouses. If a user-friendly, easy-to-use, enjoyable, and efficient interface with auditory feedback can be created, we might see workforce retention and process optimization as long-term usability results.

The research question is answered by performing three different studies, divided into an exploration, verification, and validation phase. First, feedback regarding the auditory feedback of the current scan processes is gathered. Next, different sounds will be created which will be evaluated in the verification phase of the study. Participants will evaluate the sounds by giving their opinions on perception and interpretation. After this feedback is gathered, the auditory feedback will have a final iteration to change the sounds accordingly. The final study will conduct a quantitative experiment where participants test both the conventional and enhanced auditory feedback for barcode and RFID scanning. Based on these results, conclusions can be drawn and finally, the research question can be answered.

2 Related work

A literature review is performed to research the state-of-the-art regarding auditory feedback in both general UI applications and warehousing applications. It addresses guidelines, advantages, and pitfalls of auditory feedback and divides this more specifically into auditory icons and earcons. It compares the two types and also touches upon evaluation methods to understand how auditory feedback is assessed. This is important as it allows us to gain inspiration on how we should conduct our own research. On top of that, the literature review addresses the current shortcomings of auditory feedback in warehousing solutions. Altogether, the related work section defines the scope and sketches current gaps in knowledge, leading to our research question that contributes to the existing body of knowledge regarding auditory feedback in warehousing solutions.

The literature review conducted started with a basic search including the keywords "auditory feedback", "earcons", and "auditory icons" to gain a general understanding of the field. Some highly cited publications were thoroughly analyzed, resulting in more important keywords that were included: "usability", "UI", "UX", "design", and "multimodality". Abstracts and conclusions of publications were read to determine their usefulness and relevance. Among these publications, highly relevant papers were identified on which backward and forward citation search was performed to get a clear and comprehensive overview of the state-of-the-art. After analyzing and reporting on auditory feedback in general, the search was expanded with the keywords "Warehouse Management System (WMS)", "RFID scanning", "Barcode scanning", and "AIDC". Relevant publications with these keywords were used to create a clear overview of the current warehouse challenges and how scan technologies are now being used.

Publications before the studies by Sumikawa [94] and Gaver [42] about earcons and auditory icons respectively, were excluded from the search as these studies are known to have introduced these concepts. The review was limited to studies in English. Google Scholar has been used as a search engine, both in normal and incognito mode.

First, the role of auditory feedback will be discussed, with a specific focus on its use in UIs. Next, auditory icons and earcons are discussed. Their key characteristics like meaningfulness, learnability, identification, musical characteristics, and user preferences are stated. Applications are mentioned to discuss the current use of these concepts. Hereafter, a short comparison will be made between auditory icons and earcons, after which we dive into the evaluation methods. Finally, WMSs and the auditory feedback in this field will be discussed in section 2.6.

2.1 The role of auditory feedback in UI

There have been huge strides in the domain of auditory feedback in wearables and hand-held devices such as smartwatches, tablets and mobile phones [2, 89]. Other than being used as the typical alerts or notifications of events (e.g., incoming calls, messages, or feedback in games), nonspeech sounds are also used as simple aids in tasks, such as audio-tactile feedback when typing on touch interfaces, swiping a page in an e-reader, going over the speed limit detected by a mobile GPS, etc. The role of auditory feedback is becoming increasingly important for conveying information, enhancing UX, and improving accessibility.

2.1.1 Conveying information

Sight is considered the most valued sense by most people while hearing is ranked second, which partially explains the success of Graphical User Interfaces (GUIs) [13, 17, 34]. Although proven to be successful, due to limited screen space there is a high chance to get overly dense displays which in turn leads to cognitive overload, negatively impacting the user's decision-making performance [13, 89]. Another downside is that visual information is not transient as obsolete or irrelevant information remains visually available [17]. While we should not diminish the importance of visual feedback in UIs, one could argue we have overly prioritized this modality. Past studies have shown that information spread over multiple modalities helps to minimize users' cognitive load and increases task performance [2, 17]. Displaying information on the auditory channel is useful because it is still an under-utilized channel in UI and transient in nature [17]. It enables users to receive immediate and informative feedback, without solely relying on visual elements [21]. Although less researched, audio is therefore a promising addition to the visual channel in UI.

Aside from relieving the visual channel, audio has more advantages for information transfer. Audio is useful when the user cannot view a display, such as while driving or walking in a crowded area [31]. Compared to the foveal area of the retina which subtends an angle of only two degrees around the point of fixation, sound can be heard from 360 degrees [21, 61]. No concentration has to be paid to the output device, providing greater perception flexibility overall [17]. However, while the transient nature of audio can assist with dense visual displays, it also presents a drawback as audio cannot always be replayed, resulting in potential information loss [1, 17]. Figure 2 by Gaver [43] briefly summarizes important differences between time and space for sound and vision when conveying information in UI. A simple way to contrast the auditory channel with the visual channel is the fact that sound exists in time and over space, while vision exists over time and in space.

Our ears are precise instruments making them useful to convey information [36]. Human sensations produced by sound, researched in psycho-acoustics, are well-developed and can be used to increase situation awareness [32, 73]. Compared to visual stimuli, auditory stimuli evoke a faster reaction time [14]. Yet, humans have approximately ten times as many cortical neurons devoted to vision as there are to hearing. Perceptual judgments made with the eyes are

	TIME	SPACE
SOUND	**Sound exists <u>in</u> time.** • Good for display of changing events. • Available for a limited time.	**Sound exists <u>over</u> space.** • Need not face source. • A limited number of messages can be displayed at once.
VISION	**Visual objects exist <u>over</u> time.** • Good for display of static objects. • Can be sampled over time.	**Visual objects exist <u>in</u> space.** • Must face source. • Messages can be spatially distributed.

Figure 2: Differences between sound and vision by Gaver [43].

usually more precise than those made with the ears. When comparing the visual and auditory representational dimensions of length and pitch respectively, it becomes evident that the percentage of pitch change needed to perceive a difference is twice that of the percentage of visual line length change required [76]. This perception difference is important to take into account when one designs an interface or transfers knowledge from the visual to the auditory channel.

Despite a well-developed auditory channel, psycho-acoustic considerations, such as the influence of context on sound perception and individual differences in music cognition, are crucial [24, 76]. Musicians, with heightened pitch discrimination, demonstrate better acuity in pitch and time. On top of that, psycho-acoustics have told us interrelationships among sound characteristics exist, like loudness affecting timbre [14, 76]. Such variability in music cognition contributes to variability in auditory feedback comprehension and therefore the conveyed information. Being aware of those psycho-acoustics helps sound design to be effective and accurate.

Concluding, the under-utilized auditory channel can play a significant role by providing immediate and informative feedback, reducing cognitive load, and offering advantages in situations where visual attention is limited. While the perception of sound is context-dependent, individually different, and entails interrelationships among sound characteristics, the auditory modality offers valuable opportunities to improve user interfaces and optimize information transfer. By using the strengths of both visual and auditory channels, designers can create interfaces with increased usability.

2.1.2 Enhancing UX

There is a growing recognition of the significance of incorporating multimodality, particularly audio, to craft immersive UX in the evolving landscape of multimodal user interfaces [20, 61, 89]. As human communication is multimodal, multimodality is assumed to provide a more natural and robust interaction than unimodal systems. Consequently, incorporating multimodality into a system should enhance usability and UX [99]. Besides the positive effects of multimodality on cognitive load, multimodal interfaces are also perceived to be more usable and enjoyable, where audio is believed to increase stimulation and lead to pleasurable experiences [2].

Current mobile devices can generate high-quality polyphonic sounds, immersing users and enhancing emotional engagement by incorporating realistic sounds for UI objects [61]. Additionally, it is important to consider that melody-based feedback, rather than simple beeps, improves effectiveness, efficiency, and user satisfaction with the UI. Furthermore, low-intensity sounds can enhance the positive perception of the UI [79]. Research has demonstrated that combining audio with graphics in UIs significantly enhances usability by leveraging our natural ability to process information across multiple senses. By facilitating multimodal interfaces, seamless and intuitive communication between the device and the user is created, fostering an enhanced UX [20].

Absar and Guastavino [2] published general guidelines for designing audio to enhance UX. These guidelines include: sounds should be aesthetically pleasing; sounds should be short; sounds should all be aesthetically homogeneous to the interface, although each have to be different to each other to convey appropriate information; the total number of sounds should be kept to a minimum; multiple sounds overlaying each other should be avoided as much as possible; and commonly occurring events should be designed with less obtrusive sounds. A major challenge with audio in mobile phone UIs is the varying contexts in which they are used. Individuals often feel embarrassed or disturbed by the sounds emitted from their devices in public spaces, leading them to opt for turning off the device's volume. One of the reasons to turn the volume down is the fact that users do not think the sound is informative [61]. Following guidelines makes sure a designer knows how to implement audio in UI and can help to detract as little from the UX in different contexts.

Although some guidelines exist, Frauenberger et al. [36] claim that designing auditory feedback in UI is poorly understood as audio as an integrative part of UI has been widely neglected in the past. However, both Frauenberger et al. [36] and Korhonen et al. [61] say that the importance of audio in UI is increasing. The main UX concern is the fact that users quickly link auditory feedback to annoyance: a feeling of displeasure associated with audio known or believed by an individual or a group to be adversely affecting them [20]. Besides audio being uninformative, excessive intensity is also a reason for perceived annoyance [20, 61]. To avoid annoyance, audio hardware must have good quality; audio should be kept to a low level, slightly louder than background music; parameters other than intensity should be manipulated; and audio must be kept within a narrow intensity to allow users to pick the best volume level [20].

Concluding, there is an increasing recognition of the importance of incorporating multimodality, particularly audio, in the design of UIs to create engaging UX. Current mobile devices provide high-quality audio capabilities, allowing for the use of aesthetically pleasing sounds that enhance immersion and emotion within the UI. The design of auditory feedback requires careful consideration to avoid potential negative impacts on the UX, such as annoyance and disturbance. Following guidelines can help create a positive auditory experience and is crucial for user acceptance. While designing auditory feedback in UIs has been overlooked in the past, the increasing recognition of the significance of audio highlights the need for well-designed auditory feedback to enhance the overall UX. Future research should focus on mitigating perceived annoyance and improving user satisfaction by considering factors such as informativeness and aesthetics.

2.1.3 Improving accessibility

Nowadays, devices are seen as integrated environments, accessible by anyone, anytime, and anywhere. This highlights the importance of high-quality user interfaces, accessible and usable by a diverse population with different abilities, skills, requirements, and preferences, in many different contexts [93]. Nielsen's "visibility of the system status" heuristic focuses on such accessibility. A design should always keep users informed about what is going on, through appropriate feedback within a reasonable amount of time [78]. It is important to provide feedback on actions initiated by the user, as the feedback serves as acknowledgment for the user. Not only visual feedback but also auditory feedback like the boot sound of an Apple Mac is therefore crucial to inform the user about actions [8].

Audio provides information that vision cannot [45]. Gaver [45] captures the essence of sound in UI : "*Sound is a powerful medium for conveying information. Sound complements vision, [...]. Sound can reveal patterns in data, give feedback about user actions or allow monitoring of system processes. [...] Moreover, when combined with speech, sound can be used to make computers accessible without any need for vision or visual displays at all.*" Audio can thus provide additional information, complement vision, and contribute to a more accessible interface. Especially when viewing this in the scope of the amount of information that can be distributed and displayed with current technologies, the importance of audio for accessibility becomes apparent. Multimodal interfaces simplify information and guide user attention, which in turn positively influences decision-making, response times, and general performance [83].

With the rise of GUIs, visually impaired users struggle more with digital interaction [100]. In older command line interfaces, screen readers could read text aloud, providing visually impaired users with the same information as sighted users. However, modern interfaces with the information presented pictorially pose a challenge for screen readers [18]. Auditory feedback, in the form of auditory icons and earcons, provides an alternative way of conveying information for user interfaces and is accessible to visually impaired users [40]. Auditory feedback, especially auditory icons, using environmental, familiar sounds is received well by visually impaired users [100]. Utilizing auditory feedback not only helps to show the visibility of the system status, but also assures a wider range of users and enhances their overall experience, immersion, and satisfaction [40].

Contrary to visual feedback, auditory feedback is transient, meaning the information cannot be stored to display. However, this does allow auditory feedback to be attention-demanding, as it breaks in on the attention of the operator. Additionally, visually hidden aspects in the background of a certain process can be made perceptible with auditory feedback, like a warning when a certain threshold is reached [85]. By leveraging the attention-capturing nature of auditory feedback, multimodal interfaces thus can enhance overall accessibility and mitigate the negative impact of information and cognitive overload.

In conclusion, accessibility considerations play a crucial role in ensuring digital interfaces cater to the needs of all users in various contexts. The rise of GUIs has presented challenges for visually impaired people and shifted the focus to mainly vision as feedback mechanism. While vision is often considered the primary sense, sound offers unique advantages in improving accessibility that vision alone cannot provide. Utilizing auditory feedback not only enhances the visibility of system status, it also helps visually impaired users by conveying information, contributes to overall user experience, immersion, and satisfaction, and reduces information overload.

2.2 Auditory icons

Auditory icons are everyday sounds meant to convey information about events in the computer by analogy with everyday events [42, 43]. Auditory icons are familiar sounds based on experiences in the real world [47]. Gaver [42] claims that using sound in the form of auditory icons is an appealing idea, as it is based on the way people listen to the everyday world. Auditory icons represent multidimensional, organized information in an intuitive way, providing information that visual displays do not and thus extending the consistency of a model world on a computer. Auditory icons and visual icons together create a more encompassing world for the user. In the next section, we will discuss auditory icons in detail, by highlighting their key characteristics and applications.

2.2.1 Key characteristics

When designing both auditory icons and earcons, many factors may affect their usability. We have divided these factors into five categories: 1) *Meaningfulness*: the relation between the auditory feedback and referent, 2) *Learnability*: the extent to which the auditory feedback can be easily learned, 3) *Identification*: the extent to which auditory feedback can be easily perceived and separated from other cues, 4) *Musical characteristics*: the characteristics belonging to well-designed auditory feedback, and 5) *User preferences*: the opinions of users on the auditory feedback. We will now discuss these characteristics for auditory icons.

Meaningfulness

Auditory icons have a clear and recognizable association with the action or object they represent (the referent). Three levels of physical equivalence exist between the auditory icon and the referent [42].

- Symbolic: the relation between the auditory icon and the referent is based on social convention, like sirens for an approaching ambulance.
- Iconic or Nomic: the relation between the auditory icon and the referent is based on the physical source of the referent, like a wood-hitting sound to represent wooden objects.
- Metaphorical: the relation between the auditory icon and the referent is not completely arbitrary, yet also not fully dependent on physical causation. It is based on some similarities between the sound and referent, but not as strongly as the iconic level, like breaking glass to indicate something fragile is present.

These mappings are not as mutually exclusive as they seem and some may fall between these categories. For instance, a weak metaphorical mapping becomes increasingly symbolic. Similarly, nomic mappings depend on models of the source events that produce the sound. If these sounds are becoming more approximate, the sounds map more to a metaphorical connection [42]. The meaningfulness is important to understand, as it, in turn, influences the learnability of the auditory icon.

Learnability

In general, auditory icons have great learnability [24]. Learnability is determined by *articulatory directness*, which represents the degree to which a sound corresponds to the referent [54]. Nomic mappings have the most articulatory directness, as it is based on the physical source of the referent. Metaphorical mappings have in turn more articulatory directness than symbolic mappings, as the latter is only based on social conventions not necessarily mapping to interaction with the referent. More articulatory directness also means improved learnability, meaning the more a sound depends on its meaning, the easier it is to be learned. This makes nomic mappings the easiest to learn, followed by metaphorical mappings, which are in turn easier to learn than symbolic mappings. However, once a mapping has been learned, articulatory directness does not affect performance anymore [42].

A sound designer should always strive to minimize the learning curve for understanding auditory icons to promote efficient and intuitive interaction. Users' experience with previous auditory icons is a factor influencing learnability. Previous experience can hamper learnability, making inexperienced learners quickly adapt [24].

Identification

Cabral and Remijn [24] state that the usefulness of auditory icons diminishes when the sound is not identifiable. The identification of an auditory icon is context-dependent and is vulnerable to masking by other (environmental) sounds, meaning it is possible that the sound blends in with the other nearby sounds. Masking can be avoided by diversifying auditory icons by time-varying frequency, amplitude, and timbre. Auditory icons should be distinctive and should not produce multiple semantic interpretations in order to keep swift identification on a high level.

A study by Mynatt [74] assessed the identifiability of auditory icons. Subjects were asked to describe the sounds as best as they could. 64 everyday sounds were recorded and played for 83 students in a classroom. Only 15% of the auditory icons had high rates (>80%) of correct identification. A reason for this low percentage might be that the sounds were context-independent, not focused on a specific part in a UI. The majority of sounds had high partial or alternative scores, meaning the sound is partially identified (like an object with the same affordance) or a sound had a consistent alternative answer. 10% of the sounds had overall low rates of identification. The study also stressed that there was a clear distinction between identifying a sound as an object and identifying a sound as an action. We learn that despite auditory icons benefit from great learnability, the initial identification might not be of great precision in such studies.

Another study by Belz et al. [10] compared conventional warning signal recognition (tonal, nonverbal sounds) with auditory icons. Upon completing a driving simulator, participants were asked to identify the meaning of each warning signal presented. While only half of the participants were able to correctly identify the meaning of the conventional auditory signals, 96% of the participants correctly identified the auditory icons. Somewhat similar results were found in an intensive care setting study by McNeer et al. [71]. Warnings based on auditory icons were easy to identify and learn, while conventional warnings were perceived as having higher fatigue and task load. Both studies highlight the importance of identification of auditory icons, as guessing or not knowing a warning based on audio could have catastrophic implications.

Musical characteristics

Musical characteristics of an auditory icon should focus on accurate identification [24]. Factors contributing to identification are sound duration, intensity, quality, and frequency range. Auditory icons should have a duration between 400ms and 2000ms to enable swift identification, with a frequency range of 300-3000Hz given that the human hearing is most acute within this range. However, auditory icons may be perceived differently according to the listening environment and the listeners themselves as researched in psycho-acoustics.

Creating complex auditory icons representing real-world events is challenging. Unlike simple computer-generated sounds (like earcons), auditory icons involve both real-world recording and software manipulation, offering more realism but facing drawbacks like technology-induced coloration, shaping difficulties with music-oriented software, and limited real-time modification [44]. Because sound creation of auditory icons is more difficult compared to earcons, it is more difficult to describe detailed musical characteristics. The most crucial aspect while designing auditory icons is to make sure that the listener can hear the sound clearly and understand any changes in specific parts of the sound. The focus is on the perception of the message rather than the content itself [97].

User preferences

How the user responds emotionally to an auditory icon is important [74]. Auditory icons enhance the interaction between a computer and the user, making an interface more intuitive, efficient, and enjoyable [24]. Besides, research has shown that users prefer auditory icons with strong relations to their referents, meaning they prefer nomic mapping where the sound of the icon is directly related to the source, having a high level of articulatory directness [24].

Even though many studies have found potential performance advantages for auditory icons compared to abstract sounds, auditory icons have not been widely employed for human-computer interaction. A reason for this might be the lack of user acceptance for auditory icons [55]. In the study by Belz et al. [10] about auditory icons used for vehicle warning systems, users' acceptance of and subjective response to auditory icons is considered. Although auditory icon identification was successful, half of the participants were skeptical about them, indicating they did not like the warning signal. The context, expected response, and importance of the sound of an auditory icon itself will all play a part in listener acceptability of these warnings [55]. We conclude that it is important to not only research the efficiency and effectiveness of auditory icons but also user satisfaction as this is a critical usability factor too.

Lessons learned

All in all, auditory icons have three different mappings. Their articulatory directness determines the learnability, with nomic mappings being the easiest to learn. Identification of different auditory icons is important, but can sometimes still be difficult due to problems like masking or general confusion. It is expected that identification is higher when the context in which the sound is played is known. Auditory icons should be short and within the frequency range where human hearing is most acute. Despite auditory icons evoke intuitive, efficient, and enjoyable interaction in some studies, they are not widely employed due to low user acceptance. User satisfaction is a very important factor in determining success.

2.2.2 Applications

Although not widely employed yet, auditory displays equipped with both auditory icons and earcons are becoming increasingly prominent in a great variety of applications, like home appliances, computers, smartphones, and automotive, aviation, medical, financial, and military applications [89]. We highlight several domains in which auditory icons are currently used. In the automotive domain, auditory icons are used to inform drivers about the condition of the vehicle, prevent drivers' misbehaviors such as falling asleep behind the wheel or forgetting to lock the door, and improve situation awareness by reducing brake response time [10, 24]. The study by Belz et al. [10] use tire skidding and a horn honk to represent impending collisions from different angles. Despite the fact that their study highlights the skepticism about auditory icons not being preferred by participants, one could discuss that the pros of this study outweigh the cons due to the positive impact on response time and learnability.

Although the importance of sound is well-known in video games, there is a growing interest in using auditory icons as means of providing the player with additional information [77]. Video games with auditory icons give the user an idea as to what generated the sound, whether the sound was important enough to warrant further attention, and if required how the user might take action. Auditory icons providing multidimensional information are promising, as they convey many different attributes of their source [77]. An example includes the sound of a bullet nearly missing its target to detect the position of an enemy.

Airplane pilots have to deal with a great amount of information, including a visually demanding interface. A study by Perry et al. [81] investigated the efficacy of auditory icons as warning signals in an aviation context. Metaphorical mapping was used, like couching and a car failing to start, to indicate a high level of carbon monoxide and low fuel levels respectively. Significantly fewer training trials were required to learn auditory icon warnings compared to conventional warnings, and accuracy in the test phase was higher for auditory icons. Conventional warnings elicited slow reaction times and poor accuracy. These findings can help pilots to learn the interface quicker and make fewer mistakes.

Studies on hospitals and medical equipment have reported similar results regarding learnability too [24]. Nevertheless, it is important to note that in instances where auditory icons are undesired, they have been found to contribute to errors during the execution of secondary tasks. Lastly, auditory icons also have proven to positively impact notifications on mobile devices. Learnability, memorability, and intuitiveness are better when auditory icons are used for notifications [24].

We conclude that learnability is the main positive driver of auditory icons. It is worth considering changing from conventional feedback to auditory icon feedback when it takes a long time to understand an interface or many different warnings exist. However, aesthetically it might not always be the best option. Evaluating user satisfaction is crucial in applications where quick reactions and quick learnability are not necessarily important.

2.3 Earcons

Earcons, as introduced by Sumikawa [94], are non-verbal audio cues used in UI to provide information and feedback to the user about some computer object, operation, or interaction. Unlike auditory icons with direct analogies, earcons are abstract, synthetic, and musical tones with a symbolic mapping between the sound and the referent [29]. Examples include a Mac startup sound, a notification sound, and the sound you hear when pressing a key on your touchscreen.

Blattner et al. [13] discuss the four types of earcon structures: one-element, compound, hierarchical, and transformational earcons.

One-element earcons

One-element earcons consist of single-pitch earcons and single-motive earcons. Single-pitch earcons are used to transmit only a single bit of information, like saving, clicking on, or opening a file. These earcons have attributes pitch, duration, and dynamics. They cannot be decomposed further to obtain more information [3, 13]. Single-motive earcons, on the other hand, have a brief succession of pitches arranged to produce a rhythmic and tonal pattern sufficiently distinct to allow it to function as an individual, recognizable entity. The attributes included are rhythm, pitch, timbre, register, and dynamics. They represent common computer entities such as error messages, system information, windows, and files [13].

Compound earcons

Compound earcons, also combined earcons, are formed by placing two or more one-element earcons in succession. If a single-pitch earcon is created for the icon "file" and a single-pitch earcon is created for the action "open", then a compound earcon can be "open file" by placing the two single-pitch earcons after each other as can be seen in figure 3 [89]. Repeating audio elements gives the same advantages as repeating visual elements: ease of construction, set expansion, and ease of identification and retention [13].

Hierarchical earcons

Hierarchical earcons are constructed around "grammar", where each earcon is a branch of a tree and each branch receives all the properties of the branches above it in the tree [3]. Modifications could be made with rhythm, pitch, timbre, register, and dynamics. Although the number of earcons that may be learned is much greater compared to auditory icons, hierarchical earcons are useful for systems with a large number of earcons present as the message can become very sophisticated [13]. Figure 4 shows how such hierarchical earcons could be made.

Figure 3: Creation of compound earcons by Roginska [89].

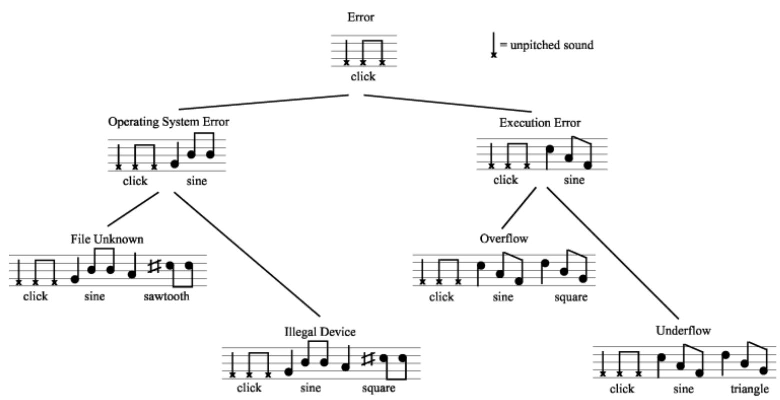

Figure 4: Creation of hierarchical earcons by Blattner et al. [13].

Transformational earcons

Transformational earcons are constructed around "grammar" too, but only the rules by which earcons are formed need to be learned. Transformed earcons are modified in simple ways that clearly retain perceptual equivalences so that the contour, also musical shape, is not changed too much. Changes in timbre, dynamics, and register are no problem, but pitch changes change the contour of the earcon and should be administered with care [13]. Brewster et al. [21] discuss that differences between transformational earcons can only be heard by skilled musicians and are therefore not useful. Figure 5 shows how such transformation earcons could be made.

2.3.1 Key characteristics

Just like auditory icons, earcons have many factors that affect usability.

Meaningfulness

Auditory icons are not suited for situations where there is no intuitive sound to represent the referent [69]. Many items in a UI have no clear iconic representation. For these cases, earcons can yield an effective solution, such as three notes diminishing loudness and pitch to represent file deletion. Most earcons have solely symbolic mappings with the information they represent and are thus flexible [29, 41]. Therefore, earcons should help users maintain awareness of their actions and current location in the UI for actions that do

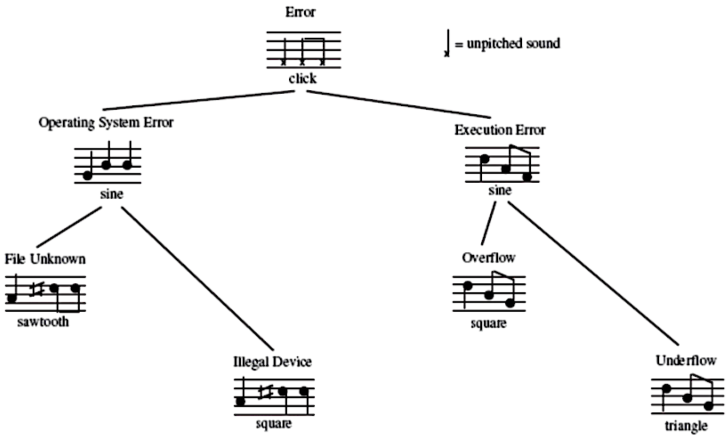

Figure 5: Creation of transformational earcons by Blattner et al. [13].

not have a clear iconic representation [73, 98].

Learnability

Contrary to auditory icons, earcons lack a meaningful relationship with their referent and thus lack articulatory directness. Users will have to learn and memorize these relationships from scratch, compared to the relatively easily learned auditory icons [41]. During the learning process of an earcon, the listener codes an internal representation of an earcon in the brain, which includes musical characteristics. Listeners use internal representations to compare known earcons with new incoming earcons. If the earcon is sufficiently similar to an existing internal representation, the listener recognizes the earcon. If it is not similar, it can be considered a new earcon. This mental structure for storing audio information facilitates memory recall [13]. Based on this information, we infer that hierarchical earcons demonstrate the best level of learnability among the four types of earcons, because each node inherits the properties of the earcon above [29].

Blattner et al. [13] address the importance of space complexity for earcons. Space complexity measures the amount of memory required to retain an earcon. The design goal for earcons is to minimize space complexity, as it is easier for users to recall and identify the earcon while conveying maximal information content. Each note of an earcon can be seen as a five-tuple of rhythm, pitch, timbre, dynamics, and register. An earcon with n notes has therefore space complexity 5n, or $O(n)$. In family earcons, like transformational or hierarchical, motives share relationships, reducing space complexity by introducing slight variations compared to different individual sounds. Compound earcons are useful when it comes to reducing space complexity too, as they use common audio elements to represent similar features of different computer entities. A user has less than a full set to remember, as we can conclude from figure 3. Only the earcons for create, destroy, file, and folder must be remembered, as from these the user could easily recognize and identify earcons like "create file" and "destroy folder". This approach aims to create earcons that are easy to learn and remember while conveying essential information.

Earcons use the Western tonal scales, as tonal sequences appear to be easier for Western listeners. Both the pitches and intervals of atonal melodies are much harder to learn. The intervals of the Western tonal scales are embedded in the minds of even untrained listeners, positively impacting the learnability compared to atonal melodies [13].

Identification

Brewster et al. [21] discuss that timbre is a very important factor to identify and recognize earcons. Just like auditory icons, masking can be a problem. To avoid masking and increase identification, timbres with multiple harmonics can be used. Additionally, a greater difference in pitches and registers allows for easier differentiation between earcons. In general, small and subtle changes between earcons are unlikely to be noticed by anyone but skilled musicians. Therefore, to clearly identify different earcons, musical characteristics must be sufficiently distinct.

Musical characteristics

Because earcons are synthesized sounds, more accurate musical characteristics can be determined compared to auditory icons. Blattner et al. [13] set up guidelines for earcon creation, which are improved by Brewster et al. [21] and Dingler et al. [29].

***Timbre*)** Brewster et al. [21] argued that musical timbres are more effective than simple tones like sine waves and square waves. Therefore, synthesized musical instrument timbres should be used. Where possible use timbres with multiple harmonics, as it helps perception and avoids masking. Use timbres that are easy to tell apart, like "brass" and "organ" instead of "brass1" and "brass2".

***Pitch*)** It is hard for users to distinguish two earcons solely differing in pitch. Therefore, one should not use pitch on its own unless there are very big differences between those used [21, 29]. Besides, it is difficult to distinguish very low and very high pitches, which should not be taken into account because of that reason [13]. Suggested ranges for pitch are Min.: 125Hz – 150Hz (an octave below middle C), and Max.: 5kHz (four octaves above middle C) [21, 29]. This guideline is somewhat different to the 300-3000Hz range from auditory icons.

***Register*)** The eight octaves in our Western system could be divided into low, medium and high register. As with pitch, it is better to combine changes in register with other sound dimensions or at least use large differences of two or three octaves in order to achieve good rates of recognition [29]. If used solely, the difference of three or more octaves gives good rates of recognition [21]. Besides, earcons constructed with pitch sequences contained in the low, medium, or high registers are easily differentiated, and therefore, easily perceived [13].

***Rhythm*)** As discussed in Patterson [80], sounds using similar rhythm are very likely to be confused. Therefore, to create effective earcons, it is important to make the rhythms of separate earcons as different as possible [29]. To accomplish this, one can vary the number of notes in each rhythm; however, to avoid overlooking small note lengths, a minimum duration of 0.125s is essential [21, 29]. Lastly, to make an earcon sound like a complete rhythmic unit, the first note should be played slightly louder and the last note should be slightly longer [51].

***Intensity*)** Perception of loudness differs from person to person and should thus be carefully designed. If a sound is too loud it may become annoying and too quiet it may be lost. Earcons should be kept in a close range of intensity, so no sounds get lost when the user adapts the volume. Intensity in the range of Min.: 10dB above background sound threshold and Max.: 20dB above background sound threshold is suggested [21, 29]. Always let the user decide the volume to reduce annoyance by sound pollution [18].

***Duration*)** Earcons should be kept as short as possible so they can keep up with the interaction in the interface. Blattner et al. [13] say earcons composed of three different note values are easiest to remember. Based on that reasoning, Engeln et al. [33] advice to not use more than 4 notes in an earcon to keep the learning effect small. However, other studies have proven that earcons with up to six notes played in one second are usable [21]. Monsaingeon et al. [73] used earcons of 900ms in his study. We can conclude there is no one golden rule for the duration of earcons, but the general sentiment is to keep earcons short and concise.

User preferences

In the study by Garzonis et al. [41], earcons not only scored lower in subjective preference compared to auditory icons, but they also seemed to cause strong negative feelings and frustration. This can be explained by the inefficient learning process. The same study suggests that earcon design needs to both meet literature-based requirements in terms of structure and distinguishability and involve users in order to produce aesthetically pleasant sounds. It is interesting that the same earcons received very positive feedback in another experiment by Garzonis et al. [41], where their validity was established. Besides, Brewster et al. [21] claims that previous work has proven that, if designed carefully, earcons are useful and not annoying.

Contrary to Garzonis et al. [41]'s first experiment, a study by Amer and Johnson [6] showed better results for earcons. Although their first experiment showed that participants more quickly learned the relationships between computing events and auditory icons, their second experiment showed that participants preferred to hear earcons rather than auditory icons. On top of that, they indicated that auditory icons would be more irritating after repeated hearings. Taken together, these results show that the more effective mode of communication is less preferred by users. There is not a clear distinction between the preferred mode of auditory feedback.

Lessons learned

All in all, four different types of earcons exist. While all of them have low articulatory directness, hierarchical earcons and compound earcons have fairly good learnability because of a three structure and reduced space complexity respectively. Musical characteristics to identify earcons depend on things like timbre, pitch, register, rhythm, and intensity. Although earcons evoke negative responses at first because of their learnability, other experiments have proven that once known, earcons are preferred over auditory icons.

2.3.2 Applications

The automotive domain benefits from earcons in many ways. Monsaingeon et al. [73] performed a study with earcons to indicate the hierarchy of automation modes while driving (manual, longitudinal, and longitudinal with lateral). The results indicated that the earcons were efficiently perceived and provoked a small decrement to a visual task, meaning the driver no longer has to look at his display as often to see which automation mode is active. Earcons thus effectively conveyed information, making sure participants had to rely less on their visual channel.

A study by Reynolds-McIlnay and Morrin [86] discussed the importance of retail transaction auditory confirmation (RTAC): providing earcons during the purchase transactions. Purchase transactions happen in a visual and auditory complex environment, having a distracting nature as one could potentially interact with other individuals. This situation could negatively impact shopper attention by increasing cognitive load and decreasing trust. RTAC enhances trust by utilizing earcons that are associated with purchase transactions, such as the beep heard when scanning an item at the checkout. These earcons offer confirmation to users that the technology has successfully registered their actions, minimizing any uncertainty during the transaction process. Additionally, this helps alleviate the cognitive load caused by distracting retail environments. Lastly, earcons also help with reducing transaction scanning errors. This proves the great usability increase that earcons provide.

Like auditory icons, earcons are used in games too. Earcons have the advantage of being without context and thus can represent any event or interaction in the interface. They also tend to be more precise than auditory icons. However, the disadvantage is that earcons have no intuitive knowledge to draw on when interpreting the sound. They have to be learned. Earcons in video games are used to provide detailed information to players in the form of warning signals, like potential ambushes, damage levels, and team orders [77].

Earcons are used in interfaces where visual feedback is not possible, like in telephone-based interfaces (TBIs). In these systems, earcons provide navigational cues and represent menu hierarchy. With TBIs, like voicemail systems, auditory navigation is important to make the user aware of the current options. Earcons were able to represent a 27-node hierarchy and showed to be a robust method of presenting navigation information [19].

We conclude earcons have many relevant applications. Although the learning process is a major drawback and might be a reason to pick auditory icons over earcons, earcons have proven to convey information effectively, decrease cognitive load, increase trust, and are able to provide navigational cues.

2.4 Comparison of auditory icons and earcons

Auditory icons and earcons appeared in the research literature almost simultaneously, in 1987 and 1985 respectively. Because they were both proposed by researchers from different fields; Denise Sumikawa as a cognitive psychologist for auditory icons and William Gaver as a computer scientist for earcons; they were initially considered separately for the most part [27]. More recently, several studies have combined the two and compared dimensions such as learnability, memorability, and user preference. However, due to the large number of application domains and evaluation methods, such results are often contradictory and lack generality [27].

Plenty of studies suggest that auditory icons can be easier to learn [24, 41, 98], and that user reaction times to auditory icons can be shorter than to earcons [41, 75], while other studies have demonstrated that earcons can be more pleasant in certain cases [6]. On top of that, Brewster [19] claims that users' retention of sounds depends heavily on the individual sound (not the sound type) as well as the learning method used. It is certain that auditory icons and earcons evoke different kinds of cognitive capabilities. For auditory icons, this means they are easier associated with iconic entities, while earcons are used where no reference to a physical entity is available [27].

Some studies try to compare auditory icons with earcons. However, these studies depend on the design of the sounds meant to represent auditory icons and earcons. If a well-designed set of auditory icons is compared with a poorly-designed set of earcons, it is easy to suppose that auditory icons are proven to be better. The generalizability of these results could be suspect at best [45]. Besides, the ratings of mappings in most studies do not look at long-term memorability. Similarly, ratings in laboratory settings may not relate well to ratings after long-term exposure in a work setting. Concluding, the results of comparison studies must be taken with some skepticism [45].

Several researchers have argued that this strict discrimination between auditory icons and earcons may be limiting for real-life applications, as no clear qualitative dominant superiority has been found after all studies performed. They suggest that both types should be used together, while others also highlight that auditory icons and earcons are theoretical extremes along a continuum of semi-abstract nonspeech sounds [27]. Gaver [45] highlights the fact that there has always been an implicit rivalry between auditory icons and earcons. However, he also discusses that the two approaches may turn out to complement one another and differences are not as great as the theories would suggest. Take for example parameterizing auditory icons along acoustic dimensions like the pitch to indicate the size of a real-world object in UI. When this is done, the acoustic dimensions used to parameterize auditory icons are not much different from those used to build hierarchical earcons.

While research on auditory icons and earcons has provided insights into their characteristics and applications, the conflicting results and limitations of comparative studies suggest that a more nuanced approach is necessary. We infer that auditory icons and earcons should not strictly be seen as two separate entities. Instead, we should focus more on what kind of individual sound, instead of sound type, must be used and researched.

2.5 Evaluation of auditory icons and earcons

To decide upon the usability of auditory icons and earcons, we should take a look at how previous studies have evaluated auditory feedback. Nees and Liebman [75] performed a meta-analysis of brief audio alerts in human-machine interfaces. They divide auditory feedback into five evaluation categories: accuracy (43 studies), reaction time (50 studies), subjective ratings (36 studies), workload (13 studies), and dual-task interference (11 studies). We will shortly discuss each of the five categories to gain insight into how evaluations are performed in those included studies.

2.5.1 Accuracy

Accuracy is measured in different ways, depending on the goal of the study. Many times, like in the highly cited paper of Brewster et al. [21] or the study by Bonebright and Nees [15], accuracy is determined by the percentage of correctly perceived auditory icons or earcons. First, participants listen to a set of auditory feedback and view visual icons (the training phase), after which they need to map the auditory feedback to the visual icons again (the recognition phase). This determines accuracy in terms of memorability. Garzonis et al. [41] is an exception, as no training phase with the actual auditory feedback was used because they wanted to measure the intuitiveness of the associations. Fake sounds were used to familiarize the participant with the process. We conclude that accuracy of auditory feedback is tested either in combination with visual icons, or is compared with other auditory feedback, to determine memorability. Mostly, a training phase is preceded.

2.5.2 Reaction time

Reaction time is often studied by measuring the time between hearing an auditory cue and selecting a visual cue. Like McKeown and Isherwood [70], the procedure often starts with a demonstration and a training phase. A computer is used to play the auditory cue as soon as the participant clicks on the screen, while at the same time visual icons are shown. The reaction time is the time between the start of the auditory cue and the selected icon. Graham [50] tests reaction time in a driving simulation. The study evaluates the differences in brake response time between auditory icons and conventional warnings. At the same time, it looks at inappropriate reactions, therefore measuring accuracy. Another study used fake sounds in the training phase, as intuitiveness was part of the research [41]. Like with McKeown and Isherwood [70], a sound was played while nine visual icons are shown. During the experiment users had to map the intuitive sound to the correct visual icon as fast as possible. In order to address possible learning effects, the order of presentation of the sounds was pseudo-randomized, ensuring that the same type of sound was not played in two consecutive trials. Most reaction time studies present in the meta-analysis by Nees and Liebman [75] involved the automotive domain, which makes sense as this requires quick reactions from the users to guarantee safety. Reaction time for other domains, like generic UI sounds to improve UX, is not as important and therefore seems not much covered in literature.

2.5.3 Subjective ratings

Subjective ratings about auditory feedback can be captured in many different ways, like with questionnaires or semi-structured interviews. Absar and Guastavino [2] used five participants to discuss audio in a system and evaluated if changes had to be made regarding length, loudness, or other parameters. Before the actual experiment started, a 20-minute training session demonstrated the use of the system. A combined method of think-aloud, followed up with a semi-structured interview gave insights to verify the auditory feedback. The interviewer asked subjective questions about impression, ease-of-use, and clarity. This study design, preceded by three panels of end users collaboratively and iteratively designing nonspeech sounds, ensured that the sounds designed are not based on designers' personal or ad hoc choices and instead exploit the creativity of participatory sound design. It serves as a great framework when sounds have to be made from sketch.

A short questionnaire could also be used to capture subjective ratings, like in many other studies [15, 41, 88]. One way of testing participants' preferences and general usability is with the industry standard system usability scale (SUS) questionnaire, used in the study by Alseid and Rigas [4]. Other studies focus more on an understanding of the user's perception of the system across the four dimensions usefulness, ease of use, ease of learning, and satisfaction [90]. In that case, the 30-item USE questionnaire created by Lund [67] is more useful. However, if a less obtrusive questionnaire is needed, single ease question (SEQ) could be used to ask about the task difficulty. It is designed to interfere as little as possible with the flow of using the system, and as a post-task questionnaire it allows you to compare which parts of the interface are perceived as most problematic. It is recommended to use standard questionnaires over homegrown ones since the former demonstrate their validity, reliability, and sensitivity [64].

Landry et al. [62] uses a somewhat different approach to capture subjective ratings. Each participant was presented with 30 auditory cues in total. Participants were allowed to replay the cue as many times as they chose, and filled out seven-point Likert scales for seven dimensions deemed relevant. These dimensions were discriminability, meaning, urgency, natural response, annoyance, startle, and overall appropriateness.

Concluding, subjective ratings are measured in many different ways, including SUS, USE, SEQ, and homegrown questionnaires. Each questionnaire has its own pros and cons, like interference, specificity, and length. Besides questionnaires, other approaches like think-aloud and semi-structured interviews are used to gain more in-depth knowledge and create user-centered sound design.

2.5.4 Workload

Another frequently used and widely accepted post-task questionnaire is NASA-TLX (Task Load Index), containing 6 questions on a 21-point scale. Each question addresses one dimension of the perceived workload: mental demand, physical demand, time pressure, perceived success with the task, overall effort level, and frustration level. The scale allows scores to be distributed between 0 and 100 in increments of 5. The downside is that NASA-TLX is a relatively complex questionnaire that needs to be answered after every key task, and so will add a lot of time to the overall test process [64]. To reduce time spent on the NASA-TLX, researchers use the Raw TLX (RTLX): an unweighted combination that skips pairwise comparison of each dimension. It assumes each dimension is equally important and simply averages or adds the six ratings of the dimensions to get an overall workload score [52].

Studies about estimating workload for auditory feedback all use (variations) of NASA-TLX. Examples include studies by Brewster and Crease [20], Finlayson and Mellish [35], and Šabić et al. [90]. Brewster and Crease [20] included an additional seventh factor: annoyance. They ran a statistical analysis on both the overall workload and individual dimensions to see whether different audio conditions varied significantly.

We infer that NASA-TLX can be very useful in estimating workload during auditory feedback testing. It is well-known and widely accepted, enabling researchers to make quantitative comparisons between different conditions.

2.5.5 Dual-task interference

Dual-task interference refers to studies where performance is measured on another (non-auditory) task in the presence of auditory feedback. When users engage in dual-task scenarios, it might lead to interference due to limited attentional resources. The cognitive load associated with processing and interpreting the auditory cues can compete with the cognitive resources needed for the primary tasks. As confirmed by Nees and Liebman [75], there indeed is parity among auditory feedback for workload and dual-task interference.

One example of such dual-task interference is the study by Gable et al. [37]. Participants performed a search task accompanied by sound through a list of 150 songs on a cell phone while performing lane change tasks in a driving simulator. For data analysis, eye-tracking data, driving performance, cognitive workload, user preferences, time to find a song, and errors made were collected.

Dual-task interference studies are specifically interesting for our current study, as we try to evaluate the impact of auditory feedback during scan processes. Although sounds should still be aesthetically pleasing, more focus is going towards how sounds might help users during their primary task. In our case, we could collect different kinds of data too, like completion time, accuracy, and subjective ratings, to estimate how such auditory feedback influences the primary task of scanning equipment items.

2.6 Audio for Mobile Warehouse Management Systems

A WMS is an important aspect of a supply chain network as it aims to control the movement and storage of materials within a warehouse and process the associated transactions, including shipping, receiving, put-away, and picking. It is the interface used to manage processes, people, and equipment on the operational level [56, 84]. Both efficiency and effectiveness are crucial factors for WMSs, as companies always strive for minimizing warehousing costs and increasing throughput rates [84].

WMSs often use Auto ID Data Capture (AIDC) technology such as barcode scanners, mobile computers, wireless LANs, and RFID to efficiently monitor the flow of products [84]. Within a warehouse, pickers receive instructions from a warehouse or project manager and pack projects with materials equipped with barcodes or RFID tags. These pickers usually have mobile devices with integrated scanners to process the packed items with their barcode or RFID tag, see figure 7. Auditory feedback in the form of beeps are used to confirm a scan is performed, while other auditory feedback is not frequently used. As soon as a barcode or RFID tag is read, the material is packed and processed.

Now we have a thorough understanding about auditory feedback with its pitfalls and potentials, we are able to focus on how it might aid WMSs with their current challenges. In next sections, we will shortly touch upon understanding the barcode and RFID technologies, and address why research about auditory feedback for WMSs is crucial.

2.6.1 Pick-by-scan technology

Warehouses utilize a wide range of technologies, with ongoing development and advancements in this field. In the context of the order-picking process, a plethora of solutions have emerged. The conventional approach involved pick-by-paper, where a list on a sheet of paper was used to manually mark checkboxes. More recently, new technologies have introduced alternative methods. Pick-by-voice and pick-by-vision use voice commands and head-mounted displays respectively to navigate users through the warehouse and improve the order-picking efficiency [11, 58, 82]. Nevertheless, most order-picking processes nowadays rely on AIDC in the form of barcodes and RFID, making pick-by-scan one of the preferred order-picking methods [25, 95].

Auto ID Data Capture - Barcodes

The traditional AIDC technology is barcode technology, consisting of a series of bars and spaces arranged regularly effectively representing a distinct value [53]. A scanner emits a beam of light that reads the barcode by detecting the contrasting bars and spaces. As soon as the barcode is read, the scanner plays auditory feedback (a beep) to confirm it has read the barcode. By utilizing the scanner in combination with the WMS's application, the barcode read can be connected to the material, enabling swift identification of the scanned item [53, 56, 95]. For order-picking, this means that after the beep the material is packed and processed.

Barcode scanning offers an immense advancement made over normal text labels because the staff no longer has to manually enter data into the system, avoiding human errors [53, 95]. However, barcode scanning is still a meticulous task requiring pickers to be near items while focusing their scanning device on the barcode [38].

Auto ID Data Capture - RFID

Radio Frequency Identification (RFID) is a newer material identification technology compared to barcodes. RFID transmits the identity of a material, in the form of a unique identification number, using electromagnetic fields. RFID consists of three major components: the application which commands an instruction to the reader to start scanning and look for tags, the optical character reader (transceiver) which transmits the instruction through an antenna to and from the transceiver, and tags which are located at the materials that are required to be identified [25, 95]. Like barcodes, the identification number is connected to a material in the WMS database. Once the reader captures the unique identification number from the tag, a beep is played to confirm something is scanned, and the identification number is transmitted to the application for final processing [95]. Figure 6 summarizes this process and explains the data flow from the application to the tag and back. Figure 7 shows the RFID gun, having both the device running the application and the transceiver.

RFID tags are mainly defined into active, passive, and semi-passive tags. Passive tags play a vital role in WMSs because of their small size, low power consumption, low cost, robustness, and little interference [95]. Passive tags do not require a power source, as the energy is transferred from the reader to the tag. Semi-passive and active tags require built-in batteries, are larger in size, and are more difficult to handle [30]. This study assumes passive tags are used as we focus on the warehousing domain. See figure 8 for a RFID tag.

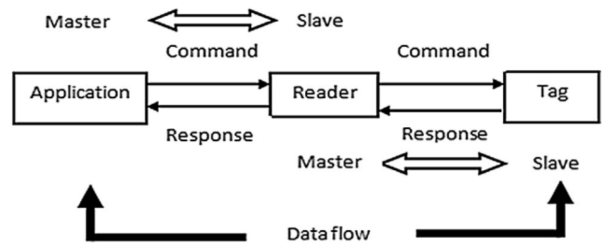

Figure 6: The components of RFID system, by Tejesh and Neeraja [95].

(a) Top view. **(b) Side view.**

Figure 7: The RFID gun, having both the application and transceiver. It can also scan barcodes.

Figure 8: A passive RFID tag.

RFID is a simple technique for indoor localization, where GPS and other (satellite) technologies lack precision or fail entirely. It outperforms many technologies based on accuracy, processing time, hardware architecture, and cost development [25]. Bluetooth for example, limits itself to a maximum confined number of 7 slaves, and cost increase if more range is required [95]. Ultrasonic systems serve as alternative, calculating the distance of an object by using echoes. However, the cost of implementation is very high, the system may cause health effects, and both the transmitter and receiver must have a line of sight [95].

RFID could be seen as a replacement for barcode technology. Although the barcode technology is low-priced, compact, and has low power consumption, it still needs a direct line of sight and is susceptible to light sources. RFID tags have more data capacity storage and are not dependent on undamaged labels. On top of that, multiple RFID tags can be scanned at once, while barcodes can only be scanned consecutively [95].

2.6.2 General challenges in the warehouse

Several challenges exist in current warehouses. Firstly, minimizing warehousing costs is an ongoing process in today's highly competitive global business environment [84]. Despite the initial investment required, AIDC systems, particularly RFID, offer significant benefits by minimizing warehouse time and enhancing overall efficiency, eventually reducing costs [95].

Secondly, warehouse managers have to deal with labor shortages as the work is physically demanding, repetitive, and the working environment is austere and stressful. To attract and retain qualified employees is a major problem, resulting in more training expenses and operations slowing down. Besides, the workforce is aging, showing that the new generation is not yet willing to join the warehousing industry [46]. Annual estimates indicate that between 20% and 75% of all warehouse workers will leave their job within one year of the hire date, and the cost of replacing each employee is estimated to be in the thousands of dollars [7]. Not monetary incentives, but job security, company size, and experience turned out to be important factors to increase job satisfaction and reduce employee turnover [72]. Technology can be used to increase job satisfaction and reduce turnover [68]. It is also claimed by Jacobs [57] and Generix Group [46] that technology in the form of mobile WMSs could be introduced to reduce turnover. Mobile WMSs should include intuitive, functional, enjoyable, and easy-to-use interfaces. Proven to be efficient and immersive, improved auditory feedback for WMSs, AIDC in specific, should be considered to solve labor shortages [61].

Thirdly, the effectiveness that many technological innovations offer for performance gains has rarely been scientifically evaluated for warehousing solutions [11]. Despite the recognized importance of RFID in the knowledge-based economy, its research applications in warehousing operations have been relatively scarce compared to domains such as retail, healthcare, and logistics [66]. Publications on RFID in warehousing have primarily focused on RFID case-based reasoning for managing different warehouse operations, investment in improving inventory accuracy, RFID adoption challenges, implementation issues, and cost-benefit evaluations [66]. Notably, these studies have predominantly emphasized the "identification" aspect, while research encompassing the full usability scope of RFID, including efficiency, effectiveness, and user satisfaction, has been limited [66]. However, especially with the labor shortages in mind, getting a thorough understanding of the usability of RFID, or AIDC technologies in general, is important to make sure job satisfaction stays on a satisfactory level.

2.6.3 Auditory feedback challenge for AIDC in the warehouse

There is a notable absence of publications investigating the impact of nonspeech auditory feedback in the pick-by-scan process using AIDC technologies. In the current practice, users heavily rely on the familiar beep sound as confirmation of a scan, without receiving additional information. While research has shown that immediate feedback, such as auditory icons or earcons, can enhance performance and reduce errors in warehouse settings, only a few studies have explored auditory feedback during scan processes [11]. For example, a study by Beckham et al. [9] examined the effects of combining different feedback modes (auditory, auditory-visual, auditory-tactile, and auditory-visual-tactile) on operator performance during scan tasks on a handheld device. The results indicated that while the combination of auditory-visual-tactile feedback produced the fastest task completion time, there was no significant improvement in operator performance across all conditions.

The importance of improved auditory feedback is evident in other domains as well. In the healthcare sector, bedside barcode scanning systems rely on auditory beeps for medication verification. However, the use of identical beeps for correct and incorrect scans can lead to confusion among nurses, who may mistakenly assume that the correct medication has been scanned [26]. The retail domain showed that auditory confirmation with a beep enhanced trust, reduced cognitive load, and positively impacted customer satisfaction [86]. Drawing from these findings, the application of more comprehensive auditory feedback, beyond a simple beep, during the pick-by-scan process in the warehouse domain could potentially improve usability.

Furthermore, the challenge of providing effective auditory feedback during scan processes becomes even more apparent in the context of RFID scanning on handheld devices. The current practice typically involves playing a beep for each received tag or unique tag, resulting in repetitive sounds when multiple tags are scanned simultaneously (see video here). Moreover, the nature of RFID scanning, which involves moving the device around to capture signals from different angles, limits the continuous availability of visual feedback. A study in the retail domain by Lee et al. [65] revealed that participants heavily relied on auditory feedback (beeps) during an RFID-based location tracking task, using visual feedback only as a secondary cue to confirm the accuracy of the auditory feedback. Participants reported difficulties in continuously monitoring the visual display due to item height and therefore difficult visibility angles. The study also highlighted the irritation caused by the fast and repetitive beeping sound, which disrupts the shopping experience of customers. Suggestions to address these issues included providing easy access to mute or volume control, using headsets, and incorporating varying tones or sounds to convey gradual changes in the signal.

We can conclude that there is a limited scientific evaluation of new technological innovations in the warehousing domain, a scarcity of AIDC (specifically RFID) research focusing on usability, and a lack of research on enhanced auditory feedback in pick-by-scan processes. Considering these factors, additional research about the inclusion of more informative auditory feedback in AIDC technologies during the pick-by-scan process to enhance usability is helpful, eventually possibly aiding with general warehouse challenges like labor shortages and improved efficiency.

2.7 Conclusion

In conclusion, audio has emerged as a valuable tool for conveying information, improving user experience, and increasing accessibility across different digital applications. Audio can be implemented to alleviate the heavily used visual channel, contribute to effective and efficient multimodal interaction, and widen the user range by using alternate communication channels. The existing, but limited, guidelines should be administered to reduce the negative impact of annoyance and optimally benefit from the potential audio has to offer in UIs.

Research makes a generic distinction between two types of auditory feedback: auditory icons and earcons. Earcons have proven to be effective in conveying information and are able to reduce cognitive load. Auditory icons, on the other hand, have great learnability. Although studies have shown that auditory icons evoke better reaction time and better learnability compared to conventional sounds, it should be emphasized that user preference is an important factor to consider too. In most situations, auditory icons and auditory feedback in general should be aesthetically pleasing to positively impact UX.

The perception of auditory icons and earcons as rival concepts is shifting among researchers. It is now increasingly recognized that they should be viewed as complementary elements of sound design. By integrating both concepts effectively, designers can create optimal sound designs that cater to the specific needs of their applications. Many different methods exist to evaluate the impact of these sounds. Most researches focus on factors like accuracy, reaction time, subjective rating, workload, and dual-task interference. These evaluation methods serve as an inspiration and help us structure our methodology.

In modern warehouses, AIDC in the form of barcode and RFID scanning is used. Although the two are widely accepted and used, the latter is rarely covered in contemporary warehouse domain-related research publications. Additionally, auditory feedback for both barcode and RFID scanning is still relatively untouched. Current applications mostly rely on a confirmatory beep after a scan is performed, which is even found to be annoying in the RFID case. The motion involved in RFID scanning with handheld devices limits continuous visual feedback and is an example highlighting the need for alternative auditory feedback. Little research is done to investigate how elaborate auditory feedback might help during AIDC warehouse processes to increase usability, despite evidence showing that immediate feedback improves performance and reduces errors in warehouse settings. Lessons learned from other domains, such as retail and healthcare, indicate the potential benefits of auditory confirmation and the importance of customizable and non-intrusive auditory feedback.

Hence, this research endeavors to investigate the usability of auditory feedback in AIDC technologies specifically within the warehouse domain. If an enhanced auditory profile in an application turns out to increase usability, this might positively impact efficiency, effectiveness, and user satisfaction in warehouses. This, in turn, might contribute to cost reduction and reduced employee turnover. The results could be generalized if enhanced auditory feedback during scan processes is shown to increase usability. By improving the auditory feedback during scan processes, users in domains such as healthcare and retail may also derive benefits from this approach, like error reduction or trust increase.

3 Methodology

The purpose of this study was to evaluate auditory feedback in the pick-by-scan process of a warehouse while using a WMS. It aimed to assess the usability of handheld devices for both barcode and RFID scanning by analyzing the impact of enhanced auditory feedback. More specifically, it made an effort to answer the following research question:

Research question: How does enhanced auditory feedback impact the usability of equipment scanning on mobile devices in Warehouse Management Systems?

A mixed-methods research approach was employed, combining qualitative data about users' opinions and perceptions, and quantitative data about users' performance to provide a comprehensive understanding of the impact of auditory feedback on usability in WMSs.

The study was divided into three consecutive steps and provided answers to seven sub-questions in total. The three-step structure was designed as the results of the first study influenced the content of the second study, which, in turn, influenced the content of the third study. The three studies were divided into an exploration phase, a verification phase, and a validation phase. Study 1 determined which actions during pick-by-scan would require auditory feedback and what kind of auditory feedback was expected. Study 2 verified the sounds created for these actions from study 1 to check whether the sounds themselves did not negatively impact the experiment. Lastly, study 3 was an experiment that tested the usability of auditory feedback by researching the effectiveness, efficiency, perceived workload, annoyance, and subjective rating with four different conditions. All three studies were preceded by a pilot study. The remainder of this chapter discusses the studies one by one, including motivations, participant recruitment, materials, procedures, and data analysis. The chapter closes with a short note on Rentman, the company aiding this research by providing the required resources like participants and the mobile application. That information is useful for the common ground before we dive into the results of chapter 4.

3.1 Study 1 - Exploration

The primary objective of the first study was to determine the appropriate way in which enhanced auditory feedback should support scan processes on mobile devices in WMSs compared to the current situation. The following sub-question was addressed:

SQ1: *In what way should enhanced auditory feedback support different actions of the scan processes on mobile devices in Warehouse Management Systems compared to the current situation?*

3.1.1 Motivation

To answer this sub-question, a qualitative research approach using semi-structured interviews was employed. The study specifically focused on understanding the key characteristics and requirements of auditory feedback during barcode and RFID scan processes on mobile devices. User preferences were taken into account for the different scan processes. The results informed the design and implementation of enhanced auditory feedback systems that effectively support the scan processes and hopefully enhance the overall usability of WMS on mobile devices.

Semi-structured interviews were chosen as a research method, as they allowed an in-depth understanding of participants' experiences with technology and their hopes for future technology [12]. They were chosen over focus groups and surveys, as they bring more nuances and allow participants to elaborate on their thoughts more compared to the other two methods. On top of that, surveys would have limited flexibility and creativity when answering a question, while this was deemed important during this phase of the research. Participants should feel free to mention any ideas or opinions they have.

3.1.2 Participants

Participants for the study were selected using purposive sampling to ensure representation from individuals with relevant experience and expertise in warehouse management and mobile device usage in a WMS context. It was acknowledged that bias could occur depending on the participants recruited. A sample from Rentman's user population was chosen and carefully recruited via their network. In total, 7 participants (100% male) were recruited in 5 interviews, differing in roles from warehouse manager to order picker. Most participants were Dutch.

3.1.3 Materials

The interview guide consisted of open-ended questions designed to gather insights into participants' preferences regarding auditory feedback during the scan processes of order picking on mobile scan devices in Warehouse Management Systems (WMS). The focus was to explore specific aspects of auditory feedback, including types of sound preferred, timing and frequency of the sound, and perceived benefits and challenges associated with auditory feedback in the WMS context. The interview guide is in Appendix A.

Participants were requested to have a Zebra barcode and RFID scanner equipped with the Rentman mobile application with them to aid them in understanding and discussing the specific scan processes during the interview. The interviews were conducted online via Microsoft Teams. Audio was recorded to ensure accurate capture of participants' responses. The audio recordings were transcribed verbatim, preserving the participants' spoken language for subsequent analysis.

3.1.4 Procedure

Before the start of the interview, the informed consent was shared. The online interviews were conducted in a quiet and comfortable environment, allowing participants to freely express their opinions and experiences. It was scheduled for approximately 30 minutes.

The interview started with a short introduction to highlight the aim of this research and put the participant at ease. The audio recording was then started. Next, the interview guide from Appendix A was used. The interview guide served as a flexible framework, guiding the conversation while also allowing for the exploration of emerging themes. It started with some general questions, followed by questions about the current auditory feedback, and ended with questions regarding enhanced auditory feedback. After the questions were asked, the interviewee was free to ask any remaining questions or highlight previous statements made. Finally, the interviewer concluded the interview by giving a brief summary.

3.1.5 Data analysis

The recordings of the interviews were transcribed with the help of OpenAI's Whisper AI[1], and were analyzed using Nvivo. Thematic analysis was used, as it entails searching across a data set to identify, analyze, and report repeated patterns in a flexible manner. As the study was looking for patterns and shared experiences, thematic analysis was considered the most suitable. It echoes the steps of the Grounded Theory method, as that also relies on coding and searching data sets for themes as part of its process. However, the goal of this study was to understand more than just the description and categorization, without developing a theory, which is the main idea behind the grounded theory method [60].

The six steps of thematic analysis were followed, as introduced by Braun and Clarke [16] and described in Kiger and Varpio [60]. First, we familiarized ourselves with the data, after which we inductively generated initial codings. Next, themes were searched. These themes were derived from the coded data by analyzing, combining, and comparing codes. The themes identified were closely linked to the original data. After this step, the themes were reviewed. Coded data in each theme was reviewed to verify relevance. The individual themes were also reviewed in relation to the entire data set to verify their fit. Defining and naming themes followed. It included answering the question of why a theme was important for the current sub-question that needed to be answered. The final step involved writing up the final analysis and description of the findings. These findings told us about current issues with barcode / QR code and RFID scanning in the warehouse process and in what way enhanced auditory feedback should support these scan processes to improve usability. From the final results of this study, we were able to conclude which actions to apply auditory feedback to and how the feedback should support the user. With the help of Apple's Garageband, three sound sets were created to verify during the second study. More information on that in Chapter 6.

3.2 Study 2 - Verification

The primary objective of this study was to investigate how users perceive and interpret the enhanced auditory feedback created after processing the feedback from the first study. In total, three different sound sets were created, each with eight different sounds mapping to eight actions in the Rentman mobile app. Both the sound sets in general and the individual sounds had to be evaluated to ensure the auditory feedback was useful, aesthetically pleasing, and well-distinguishable. The following sub-question was addressed:

SQ2: *How do users perceive and interpret the newly created enhanced auditory feedback during the different scan processes of a Warehouse Management System?*

3.2.1 Motivation

To answer this sub-question, a qualitative contextual inquiry method was used, consisting of an observational think-aloud with a semi-structured interview. This study specifically focused on verifying and evaluating the three sound sets that resulted from study 1, following a similar structure as Absar and Guastavino [2] mentioned in section 2.5.3. However, different from that study, the sound design was done by the researcher himself. It was important to know whether the newly created feedback was perceived and interpreted well, to ensure it would not negatively impact the results of the third study. The study's findings provided valuable insights into the translation process from the initial interviews to the actual auditory feedback and verified that the auditory feedback reflected what the participants had in mind during the first study.

Contextual inquiry refers to any approach that involves interleaving observations and interviews within the work setting [12]. Contextual inquiry is especially well-suited for understanding users'

[1]Code can be found here, and is a slightly altered version from Dwarkesh Patel.

interactions with complex systems and in-depth processes, as well as the point of view of expert users [91]. It is an ethnographic study but also involves interleaving observation with focused, situated interview questions concerning the current action and the role of the technology during that action [12]. One of the grounding principles includes partnership, meaning that the user and researcher are partners in the process of understanding the work. Both participant and researcher have the freedom to guide the conversation toward the relevant topics and considerations [91].

Using a think-aloud as an observational method allowed us to focus on the interaction with an interface and to identify the strengths and limitations of that interface. Think-aloud provides an understanding of why users perform certain actions and what their attitude is during these actions [12]. Think-aloud provides rich real-time insights, as participants verbalize their thoughts and interpretations during prototype interaction. As it was part of a contextual inquiry, the researcher was allowed to interrupt the process and ask questions. Semi-structured interviews complemented the think-aloud by providing an opportunity to explore participants' perceptions and interpretations in a more structured and focused manner, both during and after the think-aloud. This allowed participants to highlight and explain actions they had or emotions they felt in retrospect. It gave a more nuanced insight and could clarify any uncertainties during the think-aloud process. On top of that, the semi-structured interview allowed us to evaluate and compare the individual sounds of the three different sets.

Other methods were considered too but were evaluated as less suitable. Guerilla testing could be useful to estimate the aesthetics of the sounds. However, as sounds are perceived differently depending on the context, it was important to test the sounds in the right context with the right users reflecting the population [24]. A diary study could have been used to evaluate the experiences of participants for a longer period [12]. Long-term experience can be important, as the users might hear the auditory feedback frequently, leading to e.g. annoyance or frustration. This could not be properly evaluated during the contextual inquiry. However, longitudinal studies bring difficulties like participant recruitment and ongoing motivation [12]. Besides, we were mostly interested in verifying whether we had understood the feedback of the first study and translated that correctly to the auditory feedback. This could be done with contextual inquiry, without spending too much time and burdening the participant unnecessarily. Lastly, surveys were considered but not used as it was expected that non-musically trained participants would find it difficult to reflect on audio. Having a more open-ended and free form of data gathering allowed us to create more valuable insights.

3.2.2 Participants

In total, 9 participants were recruited, ranging between 21-54 years of age (M = 30.0, SD = 11.0) and partially overlapping with the participants from the first study. New participants were also recruited via Rentman's network and were Rentman users. All participants were male with job titles like Owner, Warehouse Manager, Order Picker, and Operations Manager.

3.2.3 Materials

The contextual inquiry consisted of a 4-step protocol, with both questions for during the think-aloud and semi-structured interview, which can be seen in Appendix B. The questions of the think-aloud method were focused on understanding the user and asking for clarification about certain actions. They served as help when the participant fell silent. The questions of the semi-structured interview focused more on the actual interpretation and perception of the sounds if they did not mention that during the think-aloud already.

A Zebra barcode and RFID scanner equipped with the Rentman mobile app with three different sound sets of enhanced auditory feedback were used during the contextual inquiry. A separate app build had to be made and sound design was done by the researcher. The scanner was used to perform predefined tasks for every sound set. 30 items differing in weight and size equipped with a barcode and RFID tag were used. A smartphone was used to record audio, and a laptop was used to set up tasks for the app.

3.2.4 Procedure

Before data collection, participants were reminded about their already signed informed consent and their rights. They could discontinue the study at any time. New participants had to sign a new informed consent. The study was scheduled for approximately 45 minutes.

The contextual inquiry was performed in the user's working environment. The study was conducted in a warehouse, where background noises were present to come as close to the actual context. The 30 items were placed in a small area (approximately $2m^2$) in this warehouse, which the user then needed to scan to process.

The 4-part session structure by Salazar [91] was followed. Part 1 was called the primer, and focused on easing the participant into the session. The researcher introduced himself, indicated what he would like to achieve, and discussed confidentiality. Hereafter, the transition started, which was phase 2. An explicit transition into the contextual interview was made by explaining what would happen during the rest of the session. The researcher let the participant know he would interrupt him to ask questions, but that the participant could communicate when not to interrupt. Participants were instructed to verbalize their thoughts and perceptions as they interacted with the scan processes accompanied by enhanced auditory feedback.

Phase 3 was the actual contextual think-aloud, where the user performed two tasks per sound set (6 tasks in total). The audio recording was now started. Task 1 was QR code scanning, task 2 was RFID scanning. The tasks were in broad terms the same, as they all processed 30 equipment items from one status to the other status by scanning the items. During each task, the participant thought aloud and commented on the process, while the researcher took notes and interrupted when he wanted more elaborate feedback. The participant instantly gave feedback on the sound he heard. After both tasks had been performed, the sound set was shortly evaluated.

After the first sound set was evaluated, the same steps were followed for the second and third sound sets. The order of the sound sets was randomized. After all tasks with all sound sets had been performed, the study transferred to phase 4, the wrap-up. The three different sound sets were compared with each other in a semi-structured interview, and individual sounds for certain actions could be compared. They delved deeper into what made certain sounds better than others and compiled a preferred sound set based on all sounds.

3.2.5 Data analysis

The recordings of the think-aloud and semi-structured interviews were transcribed with the same Whisper AI code from study 1, and analyzed using Nvivo. Unlike the emergent coding strategy of the first study, this study used a priori coding where codes were deductively generated. As this study was about evaluating the new enhanced auditory feedback of the different actions, the code structure could relatively easily be generated. A division was made between General feedback, QR code sounds, RFID sounds, and Error/Pop-up sounds for every sound set. Every individual sound had a node with positive, negative, and chosen in the final set to get an indication of the popularity of the specific sound. Figure 9 shows the branches and node structure. The general feedback branch was expanded with emergent coding depending on the feedback given during the study. Perhaps new actions that could use auditory feedback or general usability issues were mentioned.

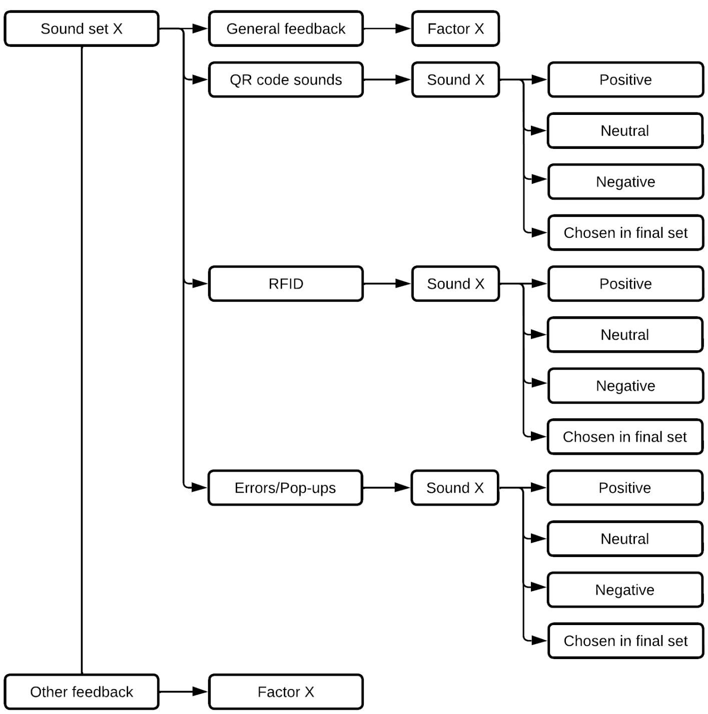

Figure 9: A priori code structure used to evaluate the interpretation and perception of the auditory feedback.

The results of this study provided valuable insights into how users perceived and interpreted the new auditory feedback used in scan processes in a warehouse. For the third study, a final sound set was picked, consisting of a combination of the individual sounds from the three sound sets. Choices for the final sounds were based on the number of favorite picks by participants and positive and negative comments. If a sound was chosen frequently as the final sound, it should be a good indicator of its success. Besides, many positive or negative comments indicate a much-discussed sound and were therefore important indicators. As also happened during a study by Brewster et al. [21], a final redesign of the sounds was performed based on the feedback of this study. Those redesign choices can be found in chapter 7. The findings contributed to the design and improvement of auditory feedback in WMS, ultimately enhancing its usability. With the feedback gathered during this study, we were able to make adjustments that benefit usability and aesthetics. After the findings were processed, the last study could begin with the finalized sounds in one ultimate sound set.

3.3 Experiment 3 - Validation

The primary objective of this study was to investigate the impact of enhanced auditory feedback on the usability of the pick-by-scan process of a WMS. It compared the conventional feedback, consisting solely of beeps, with the newly designed enhanced auditory feedback in four different conditions: 1) BC - barcode scanning with conventional sounds, 2) BA - barcode scanning with enhanced auditory feedback, 3) RC - RFID scanning with conventional sounds, and 4) RA - RFID scanning with enhanced auditory feedback. The following sub-questions were answered:

SQ3: *How do conventional sounds and enhanced auditory feedback affect usability in terms of efficiency during the scan process in a Warehouse Management System?*

$$H_0: \mu_{BC} = \mu_{BA} = \mu_{RC} = \mu_{RA}$$

H_1: The means are not all equal

SQ4: *How do conventional sounds and enhanced auditory feedback affect usability in terms of effectiveness during the scan process in a Warehouse Management System?*

$$H_0: \mu_{BC} = \mu_{BA} = \mu_{RC} = \mu_{RA}$$

H_1: The means are not all equal

SQ5: *How do conventional sounds and enhanced auditory feedback affect usability in terms of cognitive workload during the scan process in a Warehouse Management System?*

$$H_0: \mu_{BC} = \mu_{BA} = \mu_{RC} = \mu_{RA}$$

H_1: The means are not all equal

SQ6: *How do conventional sounds and enhanced auditory feedback affect usability in terms of annoyance during the scan process in a Warehouse Management System?*

$$H_0: \mu_{BC} = \mu_{BA} = \mu_{RC} = \mu_{RA}$$

H_1: The means are not all equal

SQ7: *How do conventional sounds and enhanced auditory feedback affect the usability in terms of subjective rating of the scan process in a Warehouse Management System?*

$$H_0: \mu_{BC} = \mu_{BA} = \mu_{RC} = \mu_{RA}$$

H_1: The means are not all equal

3.3.1 Motivation

To answer these sub-questions, a quantitative experimental research approach was employed. As four conditions were tested, a within-subject design was used to limit the required number of participants to detect statistically significant differences. On top of that, a within-subject study has more statistical power with fewer participants compared to between-subject studies. However, fatigue and learning effects are important factors for within-subject designs that had to be taken into account [22]. The experiment was a 2x2 factorial design, with scan type and sound design as independent variables, as can be seen in table 1. The five dependent variables were 1) task completion time, 2) number of errors made during the scanning task, 3) perceived workload, 4) annoyance, and 5) subjective rating.

Table 1: Factorial design set-up for experiment 3

	Conventional feedback	Auditory feedback
Barcode scanning	Dependent variable	Dependent variable
RFID scanning	Dependent variable	Dependent variable

3.3.2 Participants

Participants for the experiment were selected using a non-probability convenience sampling procedure. Recruitment was done via Rentman and university. To have statistically reliable results with a confidence level of 95%, $n > 40$ had to be recruited [23]. The sample included 41 adults, of which 27 were male (66%) and 14 were female (35%), with a mean age of 29.7 years (SD = 6.7). Most participants were Rentman employees, 4 students participated.

3.3.3 Materials

As stated by Laubheimer [64], it was best to use standard questionnaires to demonstrate validity, reliability, and sensitivity. Therefore, the NASA-RTLX survey was used to quantitatively measure perceived workload. Like the study by Brewster and Crease [20], annoyance was added to this questionnaire but was analyzed separately. The slightly adjusted NASA-RTLX can be found in Appendix C. For the subjective rating, the SUS survey was used, as this is an industry-standard scale to determine usability. The questions used for SUS can be found in Appendix D.

The experiment used the same 30 items with barcodes and RFID tags as study 2. The Zebra scanner to scan these codes was the same as in the previous two studies. To measure the time, a stopwatch on a mobile phone was used. Errors made during the scan process were noted on paper, while the researcher observed the participant. The SUS questionnaire was filled in on the researcher's computer, while NASA-RTLX was offered on paper due to digital visualization problems on a large scale. These questionnaires were not translated, as the participants in this sample were fluent in English.

3.3.4 Procedure

The experiment was conducted in a quiet and comfortable environment where participants could walk around to simulate the pick-by-scan process. The experiment started with sharing and signing the informed consent. Next, a detailed introduction was given to explain the experiment's purpose and procedures. Participants were given a brief overview of the WMS interface, scanning tasks, and auditory feedback. They had an opportunity to familiarize themselves with the scanning device by scanning both barcodes and RFID tags. The actual study began as soon as the researcher felt the participants understood the scanning device and the task they needed to perform.

Participants underwent the scanning tasks in four different conditions, which were counterbalanced to minimize fatigue and learning effects. The four conditions tested were:

(1) Barcode scanning with Conventional feedback (BC)
(2) Barcode scanning with improved Auditory feedback (BA)
(3) RFID scanning with Conventional feedback (RC)
(4) RFID scanning with improved Auditory feedback (RA)

The researcher counted down and started the timer to begin the task. The participant had to scan 30 equipment items as if they were order pickers and move them from location A to location B. Error messages were included to improve realism and test auditory feedback for them. As soon as the participants thought they were ready, they said "stop," after which the time was stopped and noted. The time and errors made were noted, and the participant was asked to fill in the SUS and NASA-RTLX questionnaires. These steps were repeated for every condition.

After completing the four conditions, participants had the opportunity to ask questions, provide feedback, and briefly share their experiences.

3.3.5 Data analysis

The four conditions used during the experiment all contained a combination of two independent variables. For every dependent variable, a Shapiro-Wilk test has been performed to decide the distribution of the data. If the p-value was below .05, the data was not normally distributed. If Levene's test for equality of variances had a p-value of < .05, there was no equal variance between the four groups. If only one statistical assumption was not met, we decided to use a parametric two-way repeated measures ANOVA. In case both statistical assumptions were not met, two Wilcoxon signed rank tests were used (one on each independent variable). With a confidence level of 95%, we have used a p-value of .05 to determine statistical significance.

Because the study included a 2x2 factorial design, there was also a possibility for interaction effects. Interaction effects can be found when the effect of one independent variable on the dependent variable depends on the other independent variable. The statistical significance of the interaction effects was decided with the same two-way repeated measures ANOVA and also included checking the visualizations. Because there were only two conditions for each independent variable, no post-hoc tests had to be performed.

The results of the significance tests were analyzed and reported clearly and concisely, including any significant main effects and interaction effects. They provided insights into the impact of auditory feedback on the efficiency, effectiveness, cognitive workload, annoyance, and subjective rating in WMSs. These results are the final results we needed to answer the main research question.

3.4 Rentman

Rentman is the company aiding this research. Their resources, like hardware, software, and developers, are used during this research.

Rentman is a SaaS company built for the rental-based events and media production industry and has a WMS solution for its users. Order pickers use Rentman's mobile application to process their projects both on-site at the customer and in the warehouse. It is important to get to know the workflow of the order pickers to better understand the scope of our research.

The standard workflow starts by a customer who confirms a rental project. A day before the start of the project, or on the day itself, the project is packed by an order picker. An order picker walks around the warehouse with a mobile scanner to pack equipment items scheduled on that project. When the order picker arrives at an equipment item, either a barcode/QR code[2] on the shelve/equipment item itself or an RFID tag is scanned with the scanner. Once all equipment items have been placed in the holding area of the warehouse, the project status goes from confirmed to packed. When the project must be loaded for transport, an additional check with RFID is performed so the order picker is sure all equipment items will be on location. After the festivity is over, the equipment items are picked up and transported back to the warehouse. Back in the warehouse, the equipment items are processed again by scanning with RFID or barcodes. Any lost equipment items or defects are noted and processed accordingly.

Some adaptations to this workflow exist. Some users manually count equipment items and perform RFID scanning afterward, while some integrate RFID scanning immediately during the picking process. Others rely on RFID scanning as an additional check, following barcode scanning, before loading items in a truck for transport. Some users try to RFID scan the entire project at once at the holding area, while there is also a possibility to scan separate cases one by one and load these to reduce errors.

4 Study 1: Exploration - Results

In this section, we present the most important results of our first study. Transcripts can be found in the .zip folder, in the folder "Study 1 - Data analysis". This section discusses in what way enhanced auditory feedback should support different actions of the scan processes on mobile devices in WMSs compared to the current situation and highlights the most important quotes to support these findings. After the first five steps of thematic analysis had been followed, the results were organized into three distinct thematic areas that emerged during our analysis. The first theme "Technology Sentiment" delves into the intricate details of how warehouse personnel engage with the technology, particularly how they are using both QR code and RFID technology. We examine the circumstances under which QR code or RFID tag scanning methods are preferred, as well as the prevailing attitudes toward RFID implementation in warehouse settings to be able to gain an overview of how users use the Rentman application.

Moving on to the second theme, "Satisfaction with Current Auditory Feedback" we delve into the users' general sentiment regarding auditory feedback during current scanning activities. This theme presents an overview of the significance of auditory feedback in user interactions and the level of trust associated with these sounds. Furthermore, we address frustration, confusion, and challenges encountered by users, with a specific focus on evaluating the auditory feedback provided during QR code and RFID tag scanning, including the handling of error messages.

Finally, the theme "Potential Improvements" explores the scope for enhancing the existing auditory feedback. Notably, we observe that users are partially content with the current situation; nevertheless, their expectations for future improvements and new features are also covered. Additionally, we discuss design recommendations to further optimize the auditory feedback system. Remarkably, it becomes evident that RFID-related auditory cues hold substantial potential for improvement, while QR code scanning remains relatively unaltered.

Table 2 at the end of this chapter summarizes the main feedback discussed in a concise manner. The column priority is based on an estimated combination of the number of occurrences, the expressed urgency, feasibility, and the expected impact on usability.

4.1 Technology sentiment

All participants use QR codes on their equipment items to process them within a project. They were happy with the QR code system and did not have any major issues with how this technology works. Most participants were not completely finished with their transition to RFID, as its functionality is still relatively new to the rental industry and the Rentman application. Participants use a combination of QR codes and RFID technology right now and QR codes will support the RFID technology in the future. Quote 4.1 discusses the current RFID state.

Quote 4.1
Participant 2: *"We are currently heavily engaged in RFID implementation; however, due to the busy season, we have not yet managed to affix an RFID tag to everything. [...] Nearly all devices are now equipped with RFID, and I intend for my team to depart from a festival in the near future with the assurance that we have all our belongings with us, thanks to the RFID labeling on everything."*

Although most of the participants have not finished their RFID labeling entirely, they are confident that RFID will be of great help in the near future. Some participants already experience fewer errors in their packing processes or benefit from major time savings as stated in quote 4.2.

Quote 4.2
Participant 4: *"While order picking may take a bit more time, we experience time savings when returning equipment items to the warehouse. Typically, it takes about 45 minutes to return a cable case. However, with RFID you can quickly check if the case is complete as you go along, allowing you to return it to the shelf immediately. You can finish each crate in just fifteen minutes. "*

Overall, participants expressed a predominantly positive sentiment toward the Rentman mobile application and its scan processes. They appreciated the continuous development of the application, which steadily improved its functionality. However, some users

[2]The words barcode and QR code are used interchangeably throughout this thesis, as the technology works the same but some warehouses use barcodes, while others use QR codes. They refer to the same concept.

had a "love-hate" relationship with the application due to occasional bugs, causing significant time loss. While the app covers crucial processes, some key functionalities are absent, hindering full warehouse process optimization.

The Rentman mobile application is mostly used in warehouses. Participants indicated that there are plenty of noises in the warehouse, with compressors running and other environmental noises disturbing users during their work with the application. On top of that, there are users with hearing disabilities due to the many exposures to loud noises while working in the events industry. Lastly, warehouse managers have pointed out not all individuals employed as order pickers exhibit the highest level of expertise, highlighting the importance of an easy-to-use and easy-to-learn application.

Concluding, both QR codes and RFID are currently used in the warehouse processes. QR codes are considered reliable, while RFID is still relatively new and offers room for improvement. Rentman's application enjoys sufficient usability, although there is still room for improvement on some parts to fully optimize the different workflows. The application is used in noisy environments and should be easy to use and easy to learn.

4.2 Satisfaction with current auditory feedback

First, the importance of auditory feedback in general in warehouses is discussed, after which we will focus more on auditory feedback for QR code and RFID scanning. Lastly, auditory feedback during error handling will be discussed.

4.2.1 Importance of auditory feedback

Participant responses consistently underscored the essential nature of auditory feedback within Rentman. Initially, Rentman did not include auditory feedback during their QR code scanning, which made the app unusable for some users due to the constant need for visual verification of successful scans, see quotes 4.3 and 4.4.

Quote 4.3
Participant 2: *"When we initially started using Rentman, there was no sound included. After a week, we immediately realized that it wasn't effective. We simply need to have sound, otherwise, you constantly have to check your screen to see if something has been scanned correctly. [...] Sounds are important. In fact, they are truly very important."*

Quote 4.4
Participant 7: *"If you hear something, there is no need to look at it."*

Simplicity and clarity were identified as key attributes of effective auditory feedback. Participants expressed a preference for straightforward sounds that offered immediate, unambiguous indications of scan success or failure. Such sounds were deemed particularly efficacious within the dynamic and often noisy warehouse environment. Sounds were deemed attention-grabbing, improving response time to visual actions on the application. Although audio is very important during the scan processes, visual feedback will always be considered primary as it will actually tell the users what happened and whether the equipment items have actually been processed correctly. Participants said the auditory feedback is supplementary to their visual system when working with Rentman.

A conflict arises when it comes to the trust in the auditory feedback. Multiple participants said they trusted the auditory feedback from the Rentman application. However, it is not always considered reliable. The system occasionally failed to accurately track scans, particularly in the context of large projects with multiple subprojects. The application might omit audio feedback after a series of rapid QR code scans, leading to discrepancies in item counting during order picking., see quote 4.5.

Quote 4.5
Participant 4: *"I count along with the sounds in my head. There was this case where I counted eight cables (and sounds) in my head, the system said I scanned nine, and I needed ten cables in total. These are three different numbers, so I had to empty the shelf to start counting again. I counted eight cables in total, while the application said I processed nine, which was the result of one sound not playing."*

Users trust the sounds in the Rentman application, but because of before mentioned issues, they are not always able to fully rely on auditory feedback, explaining their primary focus on the visual system. Most participants have confidence in the audio feedback during the warehouse processes but occasionally notice that the audio feedback does not work properly, which can damage their trust. On top of that, the malfunctioning of auditory feedback can lead to additional and time-consuming steps in the warehouse, leading to issues like frustration and confusion.

4.2.2 QR code scanning

The general sentiment about QR code scanning was positive amongst most participants. Currently, two sounds exist during QR code scanning. One confirmation sound at 440Hz and one error sound at 300Hz, the latter will be discussed in more detail in a separate section. The two sounds are one-element single-pitch earcons with a duration of 100ms and 400ms respectively, and both have a triangular waveform. A reason participants like the sounds is because of the clarity and ease of perception. The simplicity of these two sounds is praised as can be seen in quote 4.6, highlighting the significance of unambiguous auditory feedback. Most participants found the two sounds distinct enough and easily distinguishable from background noises. There was one participant for whom the error sound was not intense enough to perceive through his earbuds. While the ability to distinguish the two sounds and separate them from background noises was deemed crucial by most participants, opinions regarding aesthetics were more varied. The majority appreciated the sound for its functional utility, although one participant found the high-pitched confirmation sound to be irritating, see quotes 4.6 and 4.7.

Quote 4.6
Participant 1: *"The short confirmation like now is a good, clear, and nice sound."*

Quote 4.7
Participant 5: *"When I work alone during a quiet morning, I hear the beep echoing through the warehouse. [...] I would prefer a different sound, something a bit friendlier and less shrill. It could use some reduction in the high frequencies."*

Due to the unambiguous auditory feedback, participants experienced very few confusions. Instances of confusion arose due to a malfunctioning Rentman application. The absence or delay of sound emerged as the most common source of confusion, indicating a negative user experience with the application. It disrupts participants' workflows, forcing them to look at the screen.

Although the auditory feedback during QR code scanning is very limited, the derived information from this sound was deemed crucial to perform order picking efficiently. For mobile handheld scanners, counting the sounds when scanning multiple of the same equipment item is important to keep track of the packed amount. When scanning only a few QR codes, the confirmation beep signifies a successful scan, eliminating the need for users to consult the screen for information, therefore increasing operational efficiency, see quote 4.8.

Quote 4.8
Participant 7: *"With QR code scanning, I aim for the QR code and scan the equipment item. I hear a beep, knowing I am done with scanning. I do not have to look at my screen on my handheld scanner anymore, because I heard the beep. In this case, the beep is very important to us."*

4.2.3 RFID scanning

Currently, the RFID scanner beeps for every tag it reads. So even when only one tag is present, the scanner will repetitively beep as long as the RFID tag is in the proximity of the scanner and the scanner is sending out signals. This beep, a one-element single-pitch earcon, is the only auditory feedback during the RFID scan process. The general sentiment about the auditory feedback of RFID scanning was negative, mainly caused by the repetitive beeping of the RFID scanner. Many participants mentioned the distractions and confusion arising from the repetitive auditory feedback. Like in quote 4.9, participants emphasized that when scanning numerous items, the continuous beeping does not help at all and could lead to sensory overload.

Quote 4.9
Participant 1: *"When I am scanning with RFID, I do not want to hear 500 beeps. Those beeps confuse me."*

Besides confusion, quotes 4.10 and 4.11 highlight that users are not able to derive useful information from the beep. The scanner emits a sound, indicating it is actively scanning the tags in the environment. However, it does not convey substantial additional information. Some participants even indicated they were annoyed by the repetitive beeping, which indicates a bad user experience.

Quote 4.10
Participant 7: *"The sound itself doesn't offer any functionality. Whether it emits a beep or remains silent is inconsequential because I cannot derive any information from it."*

Quote 4.11
Participant 1: *"You brain cannot keep track of how many scans he has done by listening to the beeps. So I guess that makes the sound a bit redundant?"*

Visual feedback is considered more important than auditory feedback in this process because of above-mentioned reasons and the nature of RFID scanning. With RFID, the user is less in control about what is being scanned compared to QR code scanning. Users need to verify the equipment items they have scanned, which is done on a visual display. Participants mentioned they must commonly rely on the visual screen to verify the completion of scans, suggesting that the auditory feedback played a secondary or even irrelevant role in their workflows, like stated in quote 4.12.

Quote 4.12
Participant 7: *"When I am scanning with RFID, I have to look at my screen as I do not have any control on what is actually being scanned. Auditory feedback is therefore less important I think. "*

4.2.4 Error messages

Error messages in the Rentman mobile application occur in many different scenarios. They can be roughly divided into two categories: 1) error messages where no direct action is possible, like an equipment item not in stock, not active, or in repair, see figure 10; and 2) error messages where direct action is possible, like scheduling an additional equipment item on a project, see figure 11.

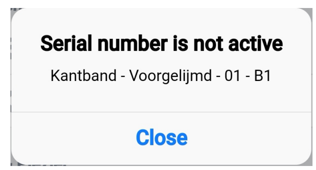

Figure 10: Error without direction action. The only option is close.

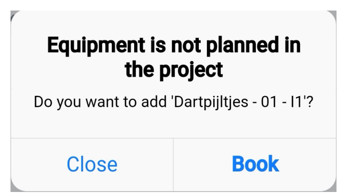

Figure 11: Error with direct action possible. The error can be immediately solved by booking an additional equipment item or closing the pop-up.

All error messages in Rentman have the same sound. The results of the study revealed a consistent preference among the participants for simple and clear auditory feedback of error messages. Almost none of the participants expressed a desire for increased differentiation in the sounds of error messages. The participants were concerned that additional sound variations could lead to confusion and decreased effectiveness in distinguishing between different error states, as stated in quotes 4.13 and 4.14.

Quote 4.13
Participant 5: *"There is currently a single error message sound that draws my attention to the screen. It would be enjoyable to instantly recognize the screen's content through the sound. However, in such a scenario, you would encounter a multitude of distinct sounds,*

creating a cacophony. The purpose of the sound is to capture your attention, a task it presently accomplishes effectively and satisfactorily. Incorporating additional variations in error sounds would not yield any time-saving or efficiency-enhancing benefits."

Quote 4.14

Participant 1: *"I don't desire additional sounds for various error messages. I'm more old-school. An error is simply an error, and one should strive to comprehend the root cause. Clarity is paramount when it comes to identifying errors. I'm skeptical about the potential benefits of employing different samples for distinct error messages."*

The participants shared the opinion that the auditory feedback of an error message is solely meant to be attention-grabbing, see quote 4.15. The importance of maintaining a visual component for error message resolution was emphasized. The power is in its simplicity, as all users (even new ones) currently know how to differentiate between the confirmation sound and the error sound, making the application easy to work with.

Quote 4.15

Participant 1: *"A mistake is a mistake; following an error, it's essential to focus on resolving it, typically by reading the screen to discern the issue. Was the error due to my own actions, a system glitch, or the account manager's input? It's akin to honking a car's horn – it prompts you to cease your current activity and direct your attention to the screen."*

The sound was perceived as clear and easy to separate from the confirmation sound by most of the participants, indicating no change was required. It is perceived as loud, clear, and clunky. Only one participant, working with earbuds, highlighted difficulties with perceiving the lower-pitched error message sound.

Concluding, participants generally expressed satisfaction with the current use of QR codes for equipment item processing. The simplicity and clarity of the auditory feedback during both a confirmation scan and an error message were appreciated. Differentiating error sounds for various error types was not considered necessary. While auditory feedback played a crucial role in aiding users during QR code scanning, its relevance was diminished during RFID scanning. RFID scanning received mixed reviews. Participants found the continuous beeping distracting, and the feedback was considered redundant as it did not provide additional information. Users frequently had to rely on the visual screen for verification, making auditory feedback less relevant in this context. It has become clear that participants valued simplicity, clarity, and ease of perception in auditory feedback within Rentman. These findings underscore the importance of considering user preferences and practicality when designing auditory feedback for Rentman.

4.3 Potential improvements

It is worth noting that participants found it challenging to articulate detailed improvements in auditory feedback. They recognized the potential for improvements but had difficulty specifying exactly how these could be achieved. A common sentiment was that finding the ideal auditory feedback for the Rentman application might be a trial-and-error process. They recognized the importance of auditory feedback but were cautious about introducing too many sounds. They were concerned that an excess of auditory feedback could lead to confusion or annoyance, making the application less user-friendly. The participants' hesitancy to propose specific improvements in auditory feedback during QR code scanning and error messages indicates that the current auditory feedback is effective and meets their expectations. However, it also highlights the challenge of fine-tuning auditory feedback in such a way that it enhances user experience without introducing unnecessary complexity or confusion. Quote 4.16 summarizes the sentiment.

Quote 4.16

Participant 1: *"In the warehousing industry, people generally prefer stability and minimal change, unless there's a need for correction. Here, individuals tend to adopt a structured and consistent approach. When operations are running smoothly, there's no need for alterations. Therefore, a brief confirmation sound, similar to the current one, is appreciated for its clarity and pleasantness. The error messages are also straightforward. However, for certain elements like RFID, there might be room for some variation. In such cases, practicality should be a key consideration."*

We will discuss the mentioned improvements for QR code scanning, error messages, RFID scanning, and general improvements.

4.3.1 QR code & error message improvements

QR code scanning is expected not to change much. If QR and error message sounds change, there should be a distinct difference between them. A participant also suggested reducing high frequencies for a more pleasant QR code confirmation sound.

Another minor improvement could be added when scanning multiple equipment items of the same product. It might be a good idea to introduce additional auditory feedback once the last item of that product is scanned, signaling the completion of packing that equipment item. An example: six fridges have been scheduled on a project. After you have scanned and processed the sixth fridge, an additional sound could be played to indicate you have finished packing the fridges.

A participant tentatively suggested that distinguishing auditory feedback between error messages for which no direct action can be taken and those for which direct action can be taken could be beneficial. Nevertheless, it was also noted that you still need to check your screen to precisely identify the type of error message displayed, making such differentiation potentially unnecessary.

Lastly, the delay of auditory feedback was considered important. Almost instant auditory feedback is expected, because waiting two seconds on a sound after you see the light of the scanner turn off will result in confusion.

4.3.2 RFID improvements

The participants' feedback and suggestions regarding RFID can be categorized into several key areas: starting sound, sound during scanning, end sound, and sounds for new features.

Firstly, participants acknowledge the need for a start sound upon entering the RFID module, serving as confirmation and signaling the beginning of the RFID process. It would reassure users about the scanner's functionality, especially when occasional double-clicking was required for activation. The start sound should only play if the scanner is working correctly, see quote 4.17.

Quote 4.17
Participant 5: *"I would greatly appreciate it if the startup sound would indeed play upon receiving feedback from the reader confirming that it has successfully scanned its first tag."*

Participants found the repetitive beeping for every tag read during scanning to be unnecessary and annoying. While some initially expressed a desire to keep this auditory feedback to confirm ongoing scanning, upon further consideration, they suggested improvements. Most participants suggested that the scanner should emit a sound only when it reads a new and unique tag. This change would reduce the redundancy of auditory feedback and make the scanning process less disruptive. Additionally, there was a suggestion for a distinct sound when scanning an equipment item that is not part of the project. This unique sound would alert users that they are scanning an item that is not scheduled for the project, helping to prevent errors and increase efficiency. However, only one participant came up with this idea, indicating that it is not of great importance. Besides, it is expected that additional confusion will arise when more than one tone is used for reading equipment item tags, negatively impacting usability.

The concept of an end sound was discussed as a means to signify the completion of the scanning process. Participants acknowledged that defining an appropriate end sound was more complex than a start sound or sound during scanning. Several ideas were proposed. Firstly, a cluster sound was suggested, indicating a group of equipment items had been covered with a single sound, as stated in quote 4.18. This would minimize repetitive beeping and create a more fluid user experience.

Quote 4.18
Participant 4: *"Utilize a single "prominent" sound, such as a checkmark sound or something similar, to indicate that you have successfully reserved multiple equipment items simultaneously. [...] It should be more intense than a normal QR code sound because you are scanning more items to the next status."*

Somewhat ironically, an end sound for participants can also just be silence. Quote 4.19 explains the fact that if the scanner is programmed to beep only for unique tags and it goes silent, users would interpret this as an indicator that no new tags are being detected. Silence could serve as a cue that the scanning process is complete.

Quote 4.19
Participant 5: *"The scanner remains silent for a moment as it doesn't detect anything new while I continue to hold down the trigger. No more beeps? Alright, I believe I've processed everything. I don't count along with all the beeps as I can't keep up due to the speed."*

Participants cautioned against using a celebratory sound as an end sound, as it might not necessarily mean that all expected items have been scanned. Instead, it should convey the message that the expected scanning process is done, but there could be additional equipment items to scan. It happens regularly that more equipment items are present in the holding area than those scheduled on the project in the software.

Lastly, participants also provided suggestions for auditory feedback related to potential new features in RFID scanning, as this functionality is still under continuous development. In scenarios where equipment items are divided into different areas, such as lighting and audio sections in a holding area, participants proposed a feature that would allow users to scan specific areas selectively. The application could then play a sound after scanning all items in the selected area, improving organization and efficiency. Another mentioned feature was equipment item localization. RFID technology has the potential to help locate lost equipment items. Participants suggested a feature where users could enter the serial number of an item and then scan while walking around the warehouse. If the scanner detected proximity to the specified item, it would emit beeping sounds or increase the frequency of beeps to guide the user toward the equipment item, effectively functioning as a locating tool.

4.3.3 Other improvements

Currently, after booking all equipment items from one location to another location, there is a pop-up stating whether the user would like to change the project status to the next one, see figure 12. This pop-up is currently accompanied by the same auditory feedback as the error messages. One participant mentioned the fact that a separate sound for this pop-up could be a nice addition. It is a heads-up for the user, but not really an error message so improved sound design would make sense.

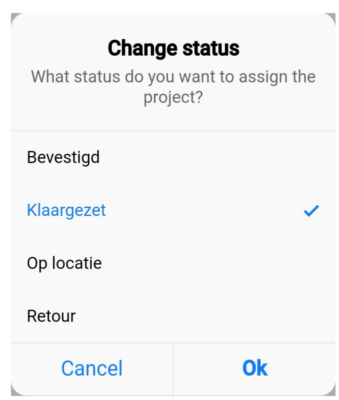

Figure 12: Project pop-up with direction action possible.

Participants preferred earcons over auditory icons, mostly they found earcons to be better perceivable in the warehouse as compressors are running and people are talking. The sound should definitely not disappear in the background, so participants think abstract sounds are easier to separate from other sounds. Furthermore, it was expected that earcons would be easier to perceive as both positive and negative compared to auditory icons. This distinction is one of the most important parts of the sound design in the application, therefore important to consider.

Finally, opinions were divided about whether QR code scanning and RFID scanning should have different scan sounds. One participant mentioned he wanted to have two different sounds because he performs two different actions. He did not want to get confused with his neighbor's scanning who is using another scanning method, see quote 4.20. However, another participant mentioned a distinction was not necessary, because of the same reason. The action performed is different, so you already know what you are doing, therefore making it unnecessary to use two different sounds, as can be seen in quote 4.21.

Quote 4.20
Participant 5: *"I expect distinct sounds because I'm performing a different action; I'm scanning something in a different manner. When someone scans QR codes nearby, I want to be able to distinguish those sounds from my RFID sounds I'm hearing."*

Quote 4.21
Participant 7: *"It involves a distinct action. With QR codes, you target and scan, whereas with RFID, you move the entire device around. A QR code isn't scannable unless you aim for it, so you're already aware of what you're doing. I don't believe using two different sounds would make a significant difference."*

4.3.4 Summary

While participants expressed overall satisfaction with the current auditory feedback for QR code scanning and error messages, they offered some subtle suggestions for refinement. Participants mentioned bigger changes for the RFID scanning, differing from starting sounds to cluster sounds, ultimately reducing the repetitive beeping. New ideas accompanied by auditory feedback were also discussed.

In summary, participants acknowledged the potential for auditory feedback improvements in Rentman but recognized the need for a cautious and iterative approach. Their feedback highlighted the delicate balance between enhancing usability and avoiding excessive complexity or annoyance of the auditory feedback. The results provide valuable insights for us to fine-tune auditory feedback in such systems, with a focus on user-centric design.

Table 2: Feedback summary of study 1. NF stands for new feature.

Category	Description	Priority
1. General	Concern of excessive feedback	High
2. General	Earcons over auditory icons	Medium
3. General	Unclear necessity distinction QR/RFID sound	Medium
4. QR	If multiple QR sounds, big difference in perception	Medium
5. QR	Reduce high frequency of confirmation beep	Medium
6. QR	Other sound after scanning final item of same sort	Low
7. QR	Minimal auditory delay	Medium
8. Error	Direct / no direct action difference	Low
9. RFID	No repetitive beeping	High
10. RFID	Only a sound for a unique tag	High
11. RFID	Start sound	Medium
12. RFID	Different sound for unscheduled equipment items	Low
13. RFID	Multiple options are available for an end sound	Medium
14. RFID NF	Select and scan an area to get completion sound	Low
15. RFID NF	Lost equipment item localisation sound	Medium

5 Study 1: Exploration - Discussion

The previous section has provided valuable insights into the use of auditory feedback for the Rentman application. The study highlighted the positive impact of auditory feedback in enhancing user efficiency and reducing errors in equipment item processing. It also emphasized the need for tailored auditory feedback for different scanning technologies, taking into account user preferences and the specific characteristics of each technology. The findings underscore the importance of simplicity and ease of perception in auditory feedback. This section puts the results into perspective and delves more into answering SQ1: "In what way should enhanced auditory feedback support different actions of the scan process on mobile devices in Warehouse Management Systems compared to the current situation." Table 2 is used for references to the users' needs. Guidelines are set up for the sound design, which will be validated in the second study. Table 3 at the end of this chapter summarizes the newly created guidelines. Chapter 6 will discuss the three different sound sets in more detail.

5.1 General guidelines

Simplicity, ease of use, ease of learning, and ease of perception have proven to be the key drivers that determine how sound design should be done in WMSs. These aspects are all connected to point 1 of table 2, which is the concern of excessive feedback resulting in potential confusion and decreased effectiveness. The results of this study reflect the guidelines created by Absar and Guastavino [2] mentioned in section 2.1.2. These guidelines state, among other things, that auditory feedback should be short; the total number of sounds should be kept to a minimum; and commonly occurring events should be designed with less obtrusive sounds. Guideline 1 will therefore be aligned with Absar and Guastavino [2] guideline, which is to *"Use a minimalist approach"*. The mentioned key drivers benefit from this approach and it prevents excessive feedback.

While one might anticipate auditory icons to perform better in terms of the learnability, users have expressed a preference for earcons because they find them easier to distinguish from background sounds, see point 2 in table 2. There are several additional reasons, apart from masking, that justify choosing earcons over auditory icons. One such reason is the necessity to keep the number of earcons in the system to a minimum, as we aim for a minimalist approach to maintain simplicity. While earcons do require learning from scratch due to their lack of articulatory directness, having a limited number of them enhances learnability and, consequently, distinguishability [41]. Moreover, in Rentman and similar WMSs, many actions lack a clear iconic representation, making earcons a more practical choice for implementation. Their symbolic mappings offer flexibility and are therefore better suited for this particular scenario [41**?**]. Lastly, as already discussed in the user preferences section 2.2.1, it is not only important to consider efficiency and effectiveness, but also user satisfaction with auditory feedback as user acceptance could be a deal breaker for the actual use of auditory feedback [55]. Therefore, guideline 2 will be: *"Prefer earcons unless strong evidence suggests using auditory icons"*. Earcons are more suitable compared to auditory icons when no elaborate auditory feedback is required because of masking prevention, offer more flexibility in an abstract system, and are preferred by the users.

The intensity of the earcons is another point of discussion. Research has shown that excessive intensity can lead to annoyance, while low-intensity sounds can positively influence the UI's overall perception [21, 79]. Nevertheless, both the scan sound and error sound in the Rentman application were praised for their clarity. Having a clear, clunky, and shrill sound allows users to distinguish it from the background noise in the warehouse, demonstrating a preference for functionality over aesthetics. Point 5 of table 2 was mentioned by only one participant, but matches the guideline by Absar and Guastavino [2] to eventually make the sounds aesthetically pleasing to reduce end-user fatigue. When redesigning the sounds in the Rentman application it is essential to administer guideline 3: *"Carefully weigh the trade-off between aesthetics and functionality"*. However, it is important to recognize that these two aspects do not necessarily have to be mutually exclusive, although finding the right balance might be a trial-and-error process. For example, a QR scan sound, which is heard hundreds of times a day, should be functional, unambiguous, and perceivable, yet not annoying.

While stated in the QR category, point 4 of table 2 is important to consider for the entire UI. Rentman users have mentioned the fact they like sounds to be well-distinguishable as this aids in ease of perception and ease of learning. Therefore, guideline 4 will be: *"Create sounds that are easy to tell apart."* Although our ears are very precise instruments, psycho-acoustics should be taken into account for this guideline [36]. Perception of sound is context-dependent and individual perception differences also exist [24, 76]. Non-musicians, like most WMS users, have a higher threshold for pitch discrimination compared to skilled musicians, which could lead to the variability of auditory feedback comprehension [76]. The musical characteristics stated in section 2.3.1 could help us create sounds that are easy to learn and recognize by most, if not all, users.

Melody-based feedback, as opposed to single-pitch earcons, improves effectiveness, efficiency, and user satisfaction with the UI [79]. However, when considering a scanning perspective, this may not hold true. Scanning, whether it is QR or RFID, involves rapid and repetitive actions. As mentioned in section 2.3.1, earcons should keep up with the interaction of the interface, which is also relevant for point 7 in table 2, emphasizing the importance of minimal auditory feedback delay. A melody-based sound takes longer to play due to its multiple notes, which could potentially lead to a delay if repetitive scans are performed. Furthermore, many other scanning devices found in settings like grocery or clothing stores employ a single-pitch earcon for scan sounds. Changing to a melody-based sound could disrupt the user's established mental model of a scan. On top of that commonly occurring events should be designed with less obtrusive sounds [2]. In conclusion, guideline 5 will be: *Use single-pitch earcons for commonly occurring events like a QR code and RFID scan.*

Although it is not clear whether there must be a distinction between sounds between an RFID scan and a QR code scan, see point 3 table 2, it is interesting to see how different sounds will be evaluated in the second study. It seems that transformational earcons are well-suited for this purpose. Brewster et al. [21] mentioned differences between transformational earcons can only be heard by skilled musicians. However, by using distinct timbres it may become easier to distinguish between them. Developing a transformational earcon could enable users to swiftly recognize a scan event and further categorize it as either a QR or RFID scan, as the musical shape of the earcon is roughly the same, but has different characteristics. Guideline 6 includes: *Use transformational earcons to differ between scan methods.* The second study will teach us whether this guideline will hold true or not.

5.2 QR code guidelines

The trust issue, mentioned in section 4.2.1, mainly occurred during QR code scanning as the scanner failed to count along with repetitive QR code scans. Such issues are important, as users lose count and make mistakes during the scan processes. The issue is fixed by solving a bug in the `createOscillator()` method of the `AudioContext` interface. No guideline is created for this improvement.

Point 6 of table 2 could pose a conflict with the concern of excessive feedback. It is therefore difficult to determine whether another sound should be added when you scan the last item of that sort. However, due to technical limitations in the Rentman application, it is not possible to develop such feedback and is therefore not implemented nor considered.

There are no QR code specific guidelines, as it is only about the sound of a confirmatory scan. The general guidelines mentioned in the previous section are considered enough to create valuable QR code scan sounds.

5.3 Error message guidelines

It was observed that error message differentiation would lead to confusion and distractions, as discussed in section 4.2.4. The primary function of an error message is to capture the user's attention, allowing them to subsequently read the accompanying visual error message. Therefore, introducing distinct auditory feedback for error messages may not be necessary.

However, one participant suggested that it might be worth considering a distinction between error messages that require immediate action (choosing between two or more options) and those that do not (only possible to accept the error), as stated at point 8 in table 2. Differentiating a bit could make sense, as there is a significant difference between the two errors. One is blocking the workflow as the error cannot be resolved right away, while the other could be solved by selecting the "Book" option, meaning it is not inherently "bad" (see figure 11). This, in combination with the hesitancy of participants to propose improvements, has led us to implement two distinct earcons for these error messages. A small error message hierarchy, like in figure 4, adds a bit more complexity but might contribute to a better understanding of what is going on and perhaps an increased user satisfaction. Guideline 7 is described as follows: *"Keep the auditory error message differentiation to a minimum.*

5.4 RFID guidelines

Auditory feedback during RFID scanning fails in every aspect of the discussed role of auditory feedback in UI in section 2.1. Firstly, it fails to convey information, see quotes 4.10 and 4.12. While past studies have shown that information spread over multiple modalities helps to minimize users' cognitive load and increase task performance, quotes 4.9 and 4.11 contradict these findings, highlighting a poor sound design for RFID [2, 18]. The repetitive beeping from point 9 in table 2 is the cause of this problem. In combination with point 10 of table 2, guideline 8 will be: *"Reduce repetitive RFID beeping by only playing a sound for every unique tag received.* A reduced sound transmission will increase the perceived information, as no sound will eventually tell users that all tags in the proximity have been received.

Another useful addition for conveying information is the implementation of a start and end sound (point 11 and 13 in table 2. Guideline 9 will be: *"Convey useful additional auditory information about the start and end process of RFID scanning.* Users are in that case better aware of when the RFID process starts and whether or not their RFID scan was successful. Chapter 6 will discuss the exact implementation of those sounds.

One could argue to implement point 12 to convey additional information as well. However, it was expected that creating different sounds for receiving RFID tags would lead to confusion. Such confusion could potentially cause users to pause their scanning prematurely, leading to time losses and decreased efficiency.

RFID is also failing in its second role of auditory feedback. RFID, with the associated auditory feedback, falls short of realizing the potential promised by multimodal interaction. Multimodal interaction is able to craft immersive UX and provide a more natural and robust interaction than unimodal systems [20, 61, 89, 99]. They should lead to pleasurable experiences [2]. However, RFID auditory feedback appears to be failing in its role to enhance the UX. Feedback from users suggests that they find the auditory cues generated during RFID scanning to be not just ineffective, but annoying too. This highlights a significant challenge in the design of RFID systems, where users have started to associate auditory feedback with irritation and dissatisfaction, thus undermining the broader UX goals that multimodal interaction seeks to achieve. It is expected that guidelines 8 and 9 will already help with solving the UX problem of RFID.

Finally, auditory feedback should enhance accessibility by providing a clearer indication of the system's current status, allowing users to understand its ongoing operations, as noted in previous studies [8, 78]. Currently, users are only aware that the system, in this case the scanner, is reading tags in its proximity by the repetitive beeping. However, crucial details and additional information remain concealed from users. Guideline 8 will cause the scanner to be silent after all tags in the proximity have been scanned. This is in conflict with Nielsen's heuristic to give a clear indication of the current status of the system [78]. To let users know they are still scanning but are not receiving new tags, a background sound is added when the RFID trigger is pressed. This will result in guideline 10: *"Provide a clear auditory system status during RFID scanning"*. This is important, as users will not look at their devices when scanning with RFID.

5.5 Remaining improvements

Upon successful booking of all equipment items from one status to another, users are prompted to modify the project status through a pop-up, see figure 12. The pop-up currently shares the same auditory signal as error messages in Rentman. The pop-up represents a slightly different message compared to the message where direct action is possible, making it prudent to expand the auditory hierarchy for error notifications with this project status action. Guideline 7 is still in place, only extending the error messages with one more sound. This enhanced earcon hierarchy aims to provide users with more nuanced auditory feedback, facilitating a clearer understanding of their interactions that require visual attention within the system.

Although points 14 and 15 in table 2 could be considered useful improvements, they fall out of the scope of current research as this would require not only new sound designs but also entire new feature development in the software. Future research should definitely focus on lost equipment item localisation. It would be interesting to see how Rentman users view this process, as they currently think RFID scanning is mainly a visual process. However, other RFID-based location tracking tasks, like in the retail industry, have proven to be very auditory-dependent, while the visual cues were only secondary [65]. Selecting and scanning an area instead of an entire project could also be interesting research, as it increases the feasibility of non-celebratory sounds because of project chunking.

In conclusion, our findings underscore the significance of simplicity, ease of use, learnability, and perceptibility in sound design within WMSs, mostly aligning with established guidelines. User preferences, masking, and flexibility in representing actions led us to prioritize the use of earcons over auditory icons. Balancing aesthetics and functionality emerged as a critical consideration, as users favored functional sounds, while guidelines mentioned auditory feedback should be aesthetically pleasing. In QR code scanning we resolved issues related to sound malfunctioning during repetitive scans. Error messages have been slightly differentiated to let users know whether an action is possible or not. In RFID scanning, a more significant overhaul of the sound design was required as it did not meet any of the three roles of auditory feedback. The repetitive beeping is reduced and a start and end sound are added to convey additional information. Lastly, a continuous background sound has been added to better represent the system's status. Table 3 sums up the guidelines.

6 Sound design

With the results from study 1 and the guidelines that emerged from that data analysis, three sound sets have been created. 8 sounds per sound set have been created for 8 actions in the Rentman application. Those 8 actions are listed in table 4. In total, 3 sound sets have been

Table 3: Sound design guidelines from study 1.

Category	Guideline
1. General	Use a minimalist approach
2. General	Prefer earcons unless strong evidence suggests using auditory icons
3. General	Carefully weigh the trade-off between aesthetics and functionality
4. General	Create sounds that are easy to tell apart
5. General	Use single-pitch earcons for commonly occurring events
6. General	Use transformational earcons to differ between scan methods
7. Error	Keep the auditory error message differentiation to a minimum
8. RFID	Reduce repetitive RFID beeping by only playing a sound for every unique tag received
9. RFID	Convey useful additional auditory information about the start and end process of RFID scanning
10. RFID	Provide a clear auditory system status during RFID scanning

created, comprising a total of 24 sounds. The sets are called "First Things First" (FTF), "Beeps and Bells" (BNB), and lastly "Old 'n Gold" (ONG). This chapter will briefly highlight sound design decisions made during the process. All figures in this chapter are in treble clef.

Sounds created are made and recorded with the help of Apple's GarageBand, which can be found in the .zip folder "Study 2 - Sound design". Different sound packs have been installed to be able to have a wide variety of synthesizers and other instruments available. Picking the right instruments for the sound design was a trial-and-error process, by clicking through the thousands of options and perceiving the timbre. After the first round, useful instruments were noted based on their fit with Rentman's character and the industry they are in. From that set, a more accurate selection was made by comparing the different sounds with each other. This resulted in a subset of instruments that were expected useful for the sounds. The instruments used were altered by applying changes to the controls, equalizers, and plug-ins to give the right sound.

6.1 QR code and RFID tag earcons

The current QR code and RFID tag sound when scanning an equipment item is a single-pitch 440Hz earcon of 100ms, as outlined in section 4.2.2. Since no complaints were received regarding the tone's duration, all three sound sets maintain a 100ms scan action. For FTF it was decided to change the pitch from 440Hz (A4) to 524Hz (C4 for RFID, C5 for QR code[3]) because of aesthetic reasons with the chosen synthesizer. BNB and ONG retain the same 440Hz frequency. Guideline 6 is implemented to differentiate between QR code and RFID tag scans by modifying the timbre for these distinct scanning methods. While both sounds are encountered frequently throughout the day, the RFID tag sound was intentionally designed to be slightly softer than the QR code sound. The RFID sound will be played in quick succession due to the possibility of multiple tag receptions per second, in contrast to the relatively slower QR code scanning process. This adjustment is expected to reduce annoyance when hearing the sound in rapid succession. Additionally, guideline 8 is applied, transitioning from beeping for all received tags to beeping solely for uniquely received tags. Both scan methods refrain from utilizing echo, delay, or reverb effects, adhering to a direct and short sound without decorations [48]. The different sounds can be heard here: QR code scan sounds and RFID tag scan sounds.

[3]Note labeling differs because different sound packages use different frequencies for middle C. For example, some synthesizer keyboards (61 keys) use C3 - 131Hz as middle C, which is MIDI note #48, while on piano keyboards (88 keys) C4 - 262Hz is considered middle C, which is MIDI note #60. This results in a different notation, but the same frequency. All note notations in this chapter include the frequency to avoid confusion, so alignment with the GarageBand file notation is maintained.

Table 4: Earcons created with a short explanation for the Rentman application.

Earcon	Explanation
1. QR code scan	Confirmation upon scanning QR code
2. RFID tag scan	Confirmation upon scanning RFID tag
3. Error no action	Blocking error (see figure 10)
4. Error action	Non-blocking error, direct action possible (see figure 11)
5. Project pop-up	Upon booking the last equipment item of the project the app asks to change project status (see figure 12)
6. RFID start	First time RFID trigger is pressed to open the module
7. RFID cluster	If all RFID tags are successfully booked to the next status
8. RFID sweep	Background sound to let users know the device is scanning

6.2 Errors and project pop-up

As stated in section 5.3 and 5.5, error messages and the project pop-up earcons (points 3-5 in table 4) have been structured hierarchically in their design. This choice aligns with guideline 7, aiming to minimize auditory differentiation in error messages. The hierarchical approach was anticipated to facilitate users' recognition of sound similarities and to improve learnability, thus adhering to guideline 1 for minimalist auditory feedback. The hierarchy varies across each sound set, so we will briefly outline the decisions made. The same instruments are employed within each hierarchy for consistency. The different sounds can be heard here: Error no action, Error action, and Project pop-up.

6.2.1 FTF hierarchy

Figure 13 shows the hierarchy of FTF. In its creation, our primary emphasis was on establishing a hierarchy with three levels. The hierarchy starts with two identical notes for errors where no action is possible. For errors with action possible, a perfect fourth interval is introduced to evoke a neutral response [59]. Given the possibility of immediate resolution of these errors, we anticipated that a neutral interval would be fitting. The project pop-up constitutes a fusion of a transformational and hierarchical earcon, transitioning from an F to a G, accompanied by the addition of a C. Nevertheless, we anticipate that the transition from F to G is unlikely to significantly affect the musical contour, likely only perceivable to professional musicians, in line with Brewster et al. [21]. This set employs a combination of two synthesizers with different timbres, with the first covering the ranges C6 to C7 (523Hz and 1047Hz) and the second covering C2-C4 (65Hz, 131Hz, 262Hz). This diverse timbral combination mitigates masking. A little bit of decay and reverb is added to give the sound more weight. The notes used are short, but approximately as long as the minimum length of 0.125s [21].

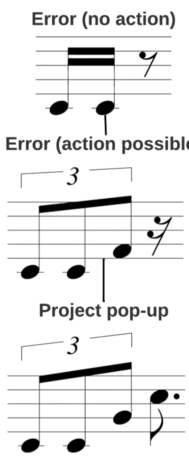

Figure 13: Hierarchy of (error) pop-ups - FTF.

6.2.2 BNB hierarchy

Figure 14 shows the hierarchy of BNB, which exhibits a flat structure. All earcons originate from the same foundational note, which is not used by itself, but differ in their final note. This design prioritized the musical motion of the earcon, having variations between descending, constant, and ascending sequences. The descending motion was chosen to convey a more negative impression, symbolizing a blocking error, while an ascending perfect fifth interval was employed to evoke stability and cheerfulness [59]. High pitches (C5 - 1047Hz) are rendered via a marimba and an electric keyboard, while the lower pitches (C2 - 131Hz and C3 - 262Hz) are generated using a synthesizer and another electric keyboard. The use of distinct timbres, once again, aids in mitigating masking. Note duration in this variant is slightly longer than those in the FTF set and notes are played with a staccato touch.

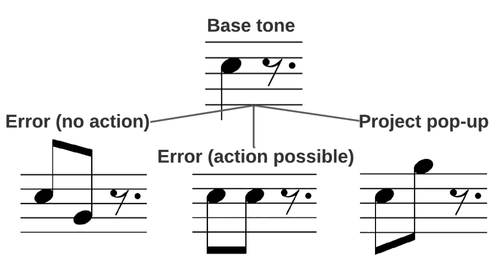

Figure 14: Hierarchy of (error) pop-ups - BNB.

6.2.3 ONG hierarchy

Figure 15 shows the hierarchy of ONG with a specific emphasis on mirroring the current error message sound in Rentman, as most participants liked the existing error sound. The "Error (no action)" retains the same sound (D4 - 293Hz, 400ms) with an identical synthesized timbre. The hierarchical level below the default error sound maintains a 400ms duration but is now divided into two 200ms notes. Unlike the BNB variant, this version does not have a staccato touch, resulting in a slightly longer perception. The "Error (action possible)" employs a major third interval to convey a friendly and hopeful feeling [59]. The "Project pop-up" employs an octave interval to give a feeling of completeness, signifying the conclusion of a project phase [28]. Given its focus on resembling the current error sound, this set exclusively employs one synthesizer.

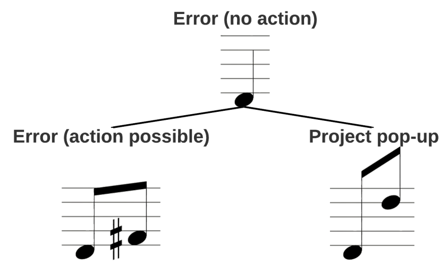

Figure 15: Hierarchy of (error) pop-ups - ONG.

6.3 RFID start

Figure 16 shows the RFID start sounds for the three sound sets, which are activated when the scanner trigger is pressed to initiate the RFID module. These sounds indicate the start of the RFID scanning process, allowing users to hold the trigger to receive any tags nearby.

During sound design, the softness of the RFID tag sound was copied to this sound. While each sound employs a different synthesizer, their timbre remains consistent. FTF represents the most musically intricate RFID start sound, driven more by musical aesthetics than functional considerations. It is anticipated that this earcon may be the least favored, potentially conflicting with guideline 3. BNB uses an octave interval. An octave interval can not only evoke a feeling of completeness but also a feeling of lightheartedness, conveying a cheerful state with minimal concerns, which suits this action [28]. ONG employs the same synthesizer as the ONG RFID tag sound, aiming to create a major chord with the final note's pitch aligning with the RFID tag sound. A major chord, comprising a major third and perfect fifth interval, conveys feelings of joy and happiness, hopefully aligning with the sentiment of embarking on a new scanning process [59]. The different sounds can be heard here: RFID start sounds.

6.4 RFID cluster

Figure 17 shows the RFID cluster sounds for the three sound sets. The cluster sound addresses potential improvement point 13 of table 2 and aligns with guideline 8 in table 3. Currently, after successfully completing an RFID scan of multiple equipment items,

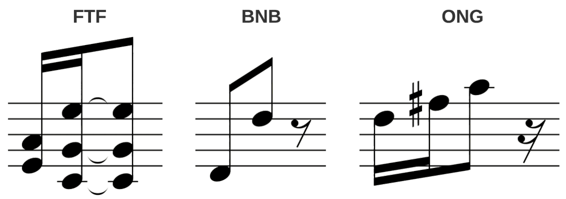

Figure 16: The different RFID start sounds

each item receives individual status updates with accompanying beeps, resulting in a repetitive beeping sequence that conflicts with guidelines 1 and 8.

The newly designed cluster sound aims to enhance the current situation. Instead of marking each item individually with multiple beeps, a single sound will be used to alert the user to the successful booking of all items of that scan. The cluster sound carries a more pronounced quality compared to a single QR/RFID scan sound, signifying the simultaneous successful booking of multiple equipment items. FTF and ONG share a similar structure, utilizing the RFID tag tone as a base and elaborating the earcon from there. The same intervals are employed, with FTF playing them simultaneously and ONG sequentially. Both use decay to impart a larger and slower feel compared to a single RFID scan, with FTF also adding reverb for increased real-life presence [48]. Additionally, FTF introduces a second synthesizer to amplify the sound's weight. BNB, on the other hand, differs by employing a bell-like sound to alert the user, distinguishing itself from the other sounds in the set and adhering to guideline 4. Decay is used in BNB for similar reasons as in FTF and ONG. The different sounds can be heard here: RFID cluster sounds.

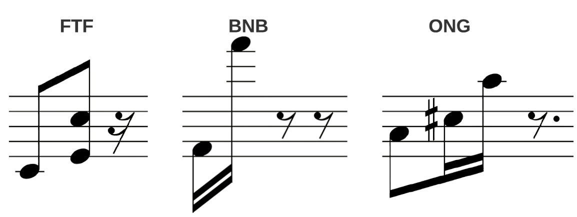

Figure 17: The different RFID cluster sounds

6.5 RFID sweep

Currently, no auditory feedback is given when the RFID trigger is pressed. The RFID sweep sound refers to the new continuous background sound during RFID scanning, audible as long as the trigger is held down to indicate ongoing scanner activity. As previously mentioned, this aids in conveying the scanner's active state even when no new tags are being received. The sweep sound is an excellent example of how guideline 3 can effectively be implemented in two ways. This action enables us to design a sound that is both aesthetically pleasing and discernible in the background while also fulfilling the purpose of indicating the system's status.

All three earcons are designed with a single or simultaneous dual note and incorporate echo. FTF simultaneously uses C4 and C5 (523Hz and 1047Hz), replaying the sound twice within the echo. The synthesizer used differs from that of the RFID tag sound, ensuring clear differentiation between the sweep sound and the tag sound. BNB utilizes the same synthesizer as FTF but with slightly adjusted echo settings. Lastly, ONG employs the same synthesizer as the RFID tag sound in that set but features less echo compared to FTF and BNB and has a different pitch (F#4 - 370Hz). The different sounds can be heard here: RFID sweep sound.

6.6 Summary

The three different sound sets have been carefully designed with the guidelines and user feedback of the first study in mind. An effort is made to ensure that all sounds within a set fit together well. During the sound design, we ensured that the error messages and project pop-ups were significantly different from the other earcons, as those must be attention-grabbing and require visual action. The second study allows us to not only verify that the new earcons are well-designed, well-perceived, and accepted by the user but also verify that the actions that have gained additional auditory feedback are in line with expectations. Sounds 6-8 of table 4 are new to their connected actions and also the error message/pop-up differentiation is different from the current version. The second study will tell us where we can still improve and which sounds per sound set are preferred the most.

7 Study 2: Verification - Results

In this chapter, we present the results of the second phase of the study, as outlined in section 3.2. Transcripts can be found in the .zip folder "Study 2 - Data analysis". Participant comments pertaining to each sound, as detailed in table 4, have been categorized into positive, negative, and neutral sentiments. A high score on (one of) those categories signifies it is much discussed. The term "Chosen in final set" denotes a participant's selection of a specific sound as their favorite among the three options (FTF, BNB, and ONG) for a given action. Per sound, a participant was allowed to only choose one as the final earcon.

The data analysis revealed that none of the three sound sets demonstrated superiority over the others. Each sound set exhibited its advantages and drawbacks concerning individual sounds. Figure 18 provides a concise overview of the sentiment associated with the three distinct sound sets. Each participant could mention multiple positive, negative, and neutral comments per sound and therefore sound set, but was only allowed to pick one of the three sounds per action. In total, one participant could have 8 favorite sounds. Although ONG may be perceived as the most "successful" sound set overall, not every sound within that set was the most frequently chosen. A more detailed breakdown of individual sounds is presented in figure 19.

The remainder of this chapter will address each sound individually, following the sequence outlined in table 4. This allows us to address our second research question (SQ2): "How do users perceive and interpret the newly created enhanced auditory feedback during the different scan processes of a Warehouse Management System?" Each subsection follows the same structure. It highlights the chosen sound including necessary changes and compares all sound results afterwards. Table 5 summarizes the sounds chosen with their adjustments for the final study.

During the sound selection process, the "Chosen" column of figure 19 was most important, leaving out repetitive mentions by a participant. If there was no clear winner, careful considerations were made by looking at the number of positive and negative comments. The final sound set can be listened to here: Final Sound Set, with the GarageBand file in the .zip folder "Study 3 - Sound design". The chapter closes with general feedback and a summary.

Table 5: Final sound set with changes from second study.

Earcon	Set	Adjustments
1. QR code scan	ONG	Changed timbre, less cutoff, and altered EQ
2. RFID tag scan	FTF	Increase pitch from C4 (523Hz) to A4 (880Hz), added double octave
3. Error no action	ONG	-
4. Error action	ONG	-
5. Project pop-up	FTF	From C-C-G-C to E-G-C and synthesizer changes
6. RFID start	ONG	Decreased decay, added RFID tag synth with sharp attack and short release
7. RFID cluster	FTF	Sharper attack, added the same synth as RFID start for coherence
8. RFID sweep	BNB	Slightly increase of volume

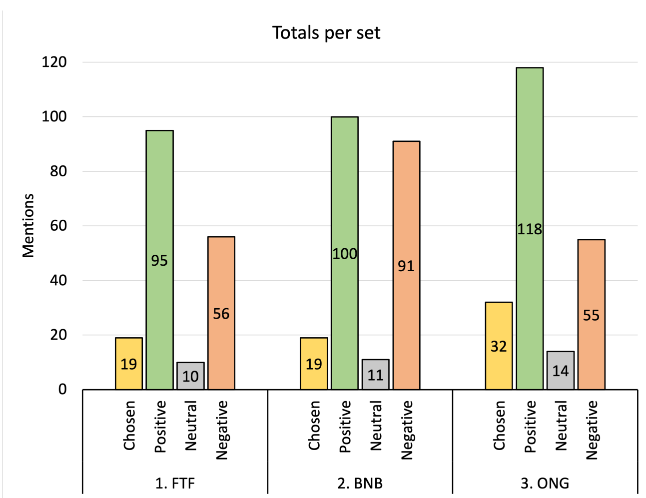

Figure 18: Sentiment per sound set. Each participant has multiple mentions per sound set.

7.1 Confirmation scan

Transformational earcons were employed to distinguish between QR code scans and RFID tag scans (see section 6.1 and figure 5), which was liked by the majority of participants. Participant 1 mentioned the exact reason why transformational earcons have been designed for the confirmation scan, see quote 7.1. Another participant emphasized the fact you cannot count along with the fast RFID tags, so the tag sound should be there, but it does not have to be too obvious. Both justify the need for transformational earcons.

Quote 7.1
Participant 1: *"The existing QR code sample is lengthy. When played repeatedly in rapid succession with RFID, it leads to interruptions in the sample. [...] RFID should be a more pleasing, lighter, and delicate sound in contrast to QR code scanning. Although the pitch may be identical, QR should be more of a "boop" sound, while RFID should produce more of a "beep" sound."*

7.1.1 QR code scan

Final sound: ONG
ONG is chosen because it was chosen the most and has good clarity, but is not perceived as too loud, making sure the aesthetics and functionality are well balanced (guideline 3 from table 3). The sound is improved by slightly changing the timbre of the synthesizer. The cutoff is changed so more higher frequencies are let through. On top of that, the EQ is slightly changed to enhance the perception of those higher frequencies. This results in a less old-school sound, while not differing too much from the originally chosen sound. Lastly, the sound will remain a single-pitch earcon, administering guideline 5 of table 3.

Results comparison
When looking at figure 19, it becomes apparent that BNB is the most disliked QR code scan sound. The sound was perceived as annoying and participants connected it to a blocking error sound. A participant even looked at the screen after a successful scan, as he thought an error had occurred. The sound was too intriguing, intense, loud, and not aesthetically pleasing. FTF is a better alternative compared to BNB. Although the opinions were slightly positive, the negative feedback is worth mentioning as participant 8 noticed a delay in the sound. He felt the time was increased between the actual scan and the sound played. Another participant indicated this sound as the most annoying of the three QR sounds. ONG was chosen the most and did not receive many negative comments. It was praised for its clarity. The sound was perceived well and relaxing, not being too loud compared to the other QR code sounds. A small improvement would be to make the sound slightly less old-school, adapting it to the modern action the user is performing. One participant could not choose between FTF and ONG, giving a total of 10 "chosen", instead of 9.

7.1.2 RFID tag scan

Final sound: FTF
FTF is chosen because it was picked by 7 participants, indicating that this sound is most favorable. Besides, the clarity was good and was perceived as pleasant sound. The FTF RFID tag is transformed into a lighter and softer sound, moving more to the aesthetic side instead of the functionality side, which is good as users are unable to count along and hear this sound in quick succession. Feedback received highlighted the need to enhance the pitch of the FTF sound for improved perceptibility. Furthermore, maintaining a consistent

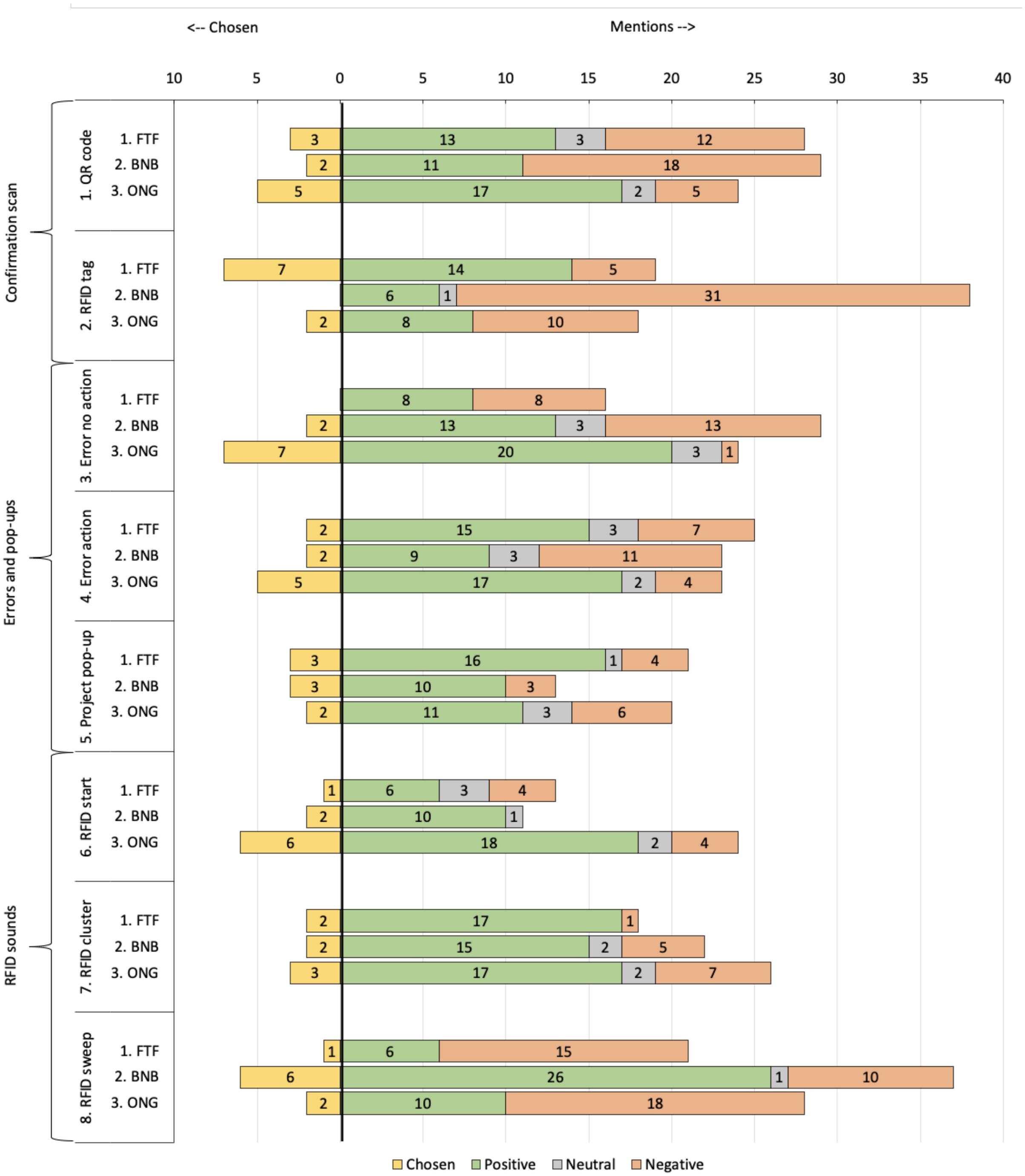

Figure 19: Sentiment per individual sound. Each participant could mention an arbitrary number of positive, negative, and neutral comments, but was only allowed to pick one favorite (chosen) sound per action.

pitch across different earcons within the transformational hierarchy was deemed essential to ensure a coherent musical contour, adhering to the principles of transformational earcons [21]. The initial pitch of the FTF RFID tag was 524Hz, while the pitch of ONG QR code was 440Hz. It was therefore decided that the FTF RFID tag would benefit from using a double octave. 440Hz, 880Hz, and 1760Hz are played simultaneously so the transformation hierarchy stays intact and clarity is better at lower volumes. This way, guideline 6 from table 3 is administered.

Results comparison
BNB is the least-liked sound for the RFID tag scan. Quote 7.2 highlights the concern:

Quote 7.2
Participant 9: *"In our warehouse, it is not very loud, but if forklifts are in operation and colleagues are throwing with cases... If you need a highly penetrating sound, this could be suitable. However, it is not suitable for our current environment."*

The sound was compared to a lightsaber, being too aggressive, extreme, and violent, eventually becoming unpleasant with continuous exposure. The sound clearly failed and should not be considered to be implemented. ONG was considered a gentle and pleasant sound by most participants. However, it was only chosen once by the participants, because it might become less distinct in noisier surroundings and it coincided with the RFID sweep, confusing the scan process. FTF's clarity was slightly better than ONG's, but still had a pleasant sound, making it suitable for continuous use. It was chosen by 7 participants, indicating that this sound is most favorable. A point of improvement was to slightly increase the pitch to increase perception with lower volumes and make it easily distinguishable from the RFID sweep sound.

7.2 Errors and pop-ups

To keep error message differentiation to a minimum (guideline 7 of table 3), a hierarchy was implemented to limit the number of sounds and increase learnability, like stated in chapter 6.2. All participants agreed with the project pop-up having another sound compared to the default error sound. There was a small majority of participants preferring an auditory distinction between "Error no action" and "Error action". Results were divided, as for some participants the latter is just an error, while other participants do not see this as mistake. This highlights the importance of guideline 7, to have only a small hierarchy to have a logical and easy to learn error message system. Quotes 7.3 and 7.4 highlight the different opinions.

Quote 7.3
Participant 2: *"The current situation is awful, that was just developed once. Having multiple notes is great, it makes it very clear. "*

Quote 7.4
Participant 5: *"My personal opinion is that it is not necessary to differentiate [...] It does not add anything to the pace you're solving things with. I have to learn a new sound while that might not be necessary, I think?"*

Participants did not explicitly mention anything about the hierarchy created for the error messages and project pop-up. It seemed like they evaluated each sound separately, not knowing there was a structure behind it. Based on figure 19 we can also derive that there was no favorite sound set for this part of the application. Because there was a small majority for error message differentiation, we decided to continue with that for the actions "Error no action" and "Error action".

7.2.1 Error no action

Final sound: ONG
ONG is chosen because it was picked by 7 participants, indicating that this sound is most favorable. The sound was able to convey the necessary information well. No adjustments have to be made.

Results comparison
The opinions about FTF were mildly positive, but negatively influenced by the sound "Error action" as they were too similar and therefore negatively perceived. The sound itself was clear and indicated the current error was blocking for some participants. However, four participants also mentioned the sound was too kind.

The opinions about BNB were divided. Participants thought the sound was aesthetically pleasing, and for some of them, it resembled an error message like they know in other systems. Nevertheless, many participants indicated the sound was too kind and not pressing enough, even suggesting swapping the BNB QR code with BNB error no action, as can be seen in quotes 7.5, 7.6, and 7.7. As discussed in section 5.3, the primary function of an error message should be to capture the user's attention. Both FTF and BNB failed to convey this message effectively.

Quote 7.5
Participant 8: *"I would almost say you should swap them. [...] Error no action sounds positive, while the QR code sound is quite an "ehhh" sound."*

Quote 7.6
Participant 1: *"If I were blind and I heard this at a pedestrian crossing, I would cross the road."*

Quote 7.7
Participant 12: *"I think the earcon sounds happy. I think I have scanned something now, so I would not use it as an error message sound."*

ONG on the other hand, was perceived very well. It was the same error as currently used. Participants felt it effectively conveyed the message that a blocking error had occurred, with some even indicating that they heard a cross sign. The duration of the sound (which is a single-pitch earcon) was seen as positive, allowing users to distinguish it from other notifications. It was clear that participants liked this error message the most.

7.2.2 Error action

Final sound: ONG
ONG is chosen because of four reasons. Firstly, it fits the hierarchy with the ONG "Error no action" sound. Secondly, the ONG sound has the same structural alignment as the phrase "to-do", see quote 7.8. Thirdly, it was voted for the most. Finally, the two error messages were easy to distinguish. No adjustments have to be made.

Results comparison
The results for this sound exhibited more variations compared to "Error no action". Similar justifications were provided for both the preference and aversion to a particular sound. This consistency in arguments can be attributed to the diversity of opinions among participants regarding whether the "Error action" inherently conveys negativity. Participants expressed a positive view of the FTF sound, describing it as pleasant and aesthetically pleasing. Despite being

well-received, some participants remarked that it lacked a distinctive error-related quality, felt somewhat confirmatory in nature, or simply did not fit the action.

On the other hand, the BNB sound was characterized as clear, but few other positive attributes were noted. Participants reiterated that the sound conveyed an overly positive tone, leaving them with a sense of finality. Both BNB and FTF have a too positive tone, which would in this scenario mean that users will not look at their device, while attention is required. Both options therefore miss the primary function of an error message, which is grabbing your attention. Additionally, both options got multiple complaints saying the two error message hierarchies were too similar to each other.

In contrast, the ONG sound was perceived as the most practical. It was easily distinguishable from the "Error no action" sound and found to be pleasant when encountered multiple times a day. An illustrative insight, as presented in quote 7.8, highlighted the structural alignment of the ONG sound with the phrase "to-do," mirroring the exact action required for the error pop-up. Notably, ONG was the most frequently selected sound and received the fewest negative comments, underscoring its suitability, particularly when used in conjunction with the "Error no action" sound.

Quote 7.8
Participant 4: *"The sound "tu-du" sounds almost like to-do."*

7.2.3 Project pop-up

Final sound: FTF
FTF is chosen as it received most positive comments (the "Chosen" category was almost equal for all sounds). FTF successfully conveyed a sense of accomplishment. The sound will be adjusted to make it shorter and less bouncy, by changing C-C-G-C (see figure 13 to E-G-C (see figure 20). The synthesizer for the lower pitches is changed to a different one so it fits the entire sound set better. A third synthesizer is added to give a slightly more celebratory effect. All synthesizers play the same notes. Lastly, it will be removed from the error message hierarchy tree, see results comparison section.

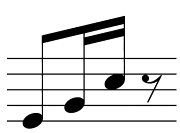

Figure 20: New project pop-up, based on FTF project pop-up.

Results comparison
8 votes were cast for the project pop-up as one participant was not completely satisfied with any of the sounds. There was no clear favorite for the project pop-up sound. It was mentioned that the design of this sound may be considered less significant, given that it is encountered only a few times each day, as stated in quote 7.9.

Quote 7.9
Participant 1: *"This sound occurs 6 times a day, so you [the designer] can be free to use any combination of positive notes."*

All participants provided positive feedback on FTF, generally appreciating its clarity, positivity, and the sense of accomplishment it conveyed, like stated in 7.10. Nevertheless, some participants mentioned that the sound was either too lengthy or too bouncy.

Quote 7.10
Participant 12: *"I actually enjoy this sound, it is nice. It is a beautiful sound. I now know I have scanned everything, it really is a happy sound."*

BNB also received positive mentions, albeit with moderately positive comments compared to FTF. A participant suggested reducing the level of distortion to gain a more positive character.

ONG's negative comments were the most dominant, as some participants perceived it as bearing too strong a resemblance to an error message. While its perception was good, it was noted it could benefit from an added celebratory element by some.

No clear "winner" emerged among these sounds, signaling the potential need for a redesign to address the current negative feedback associated with the project pop-up sound. The sounds were designed within the error message hierarchy as it is shown with a pop-up, like the other error messages, and used to have the same sound as the default error message. After evaluation, hierarchy exclusion is expected to be better, as it allows us to design the earcon freely, without hierarchy rules defined by the previous branches. Take for example the ONG hierarchy, as shown in figure 15. Participants linked the project pop-up sound with an error message, which can be explained by the fact that the first note of the project pop-up is used for "Error no action" too. With such a hierarchy, it is hard to bend a sound that starts with an error tone to a celebratory end sound. Therefore, deciding which actions to include and which not is of great importance to benefit from increased usability with such hierarchies.

7.3 RFID sounds

Section 5.4 mentioned that RFID scanning fails in the three roles of auditory feedback in UI. However, it has become clear that we are now moving in the right direction by making sure information is conveyed effectively, improving the UX, and increasing accessibility with enhanced auditory feedback. As will be discussed next in this section, guidelines 8, 9, and 10 of table 3 have been correctly implemented. As outlined by guideline 8 of table 3, the scanner was configured to emit a beep solely for unique tags detected, rather than beeping continuously for each tag scanned. The participants in our second study verified this adjustment as a significant enhancement, diminishing the overall annoyance experienced during the RFID scanning process.

7.3.1 RFID start

Final sound: ONG
ONG is chosen because it was picked by 6 participants and received most positive comments. It was straightforward and effective. The sound will be slightly redesigned to create less dullness and make it more snappy. This is achieved by using less decay on the same synthesizer, creating a faster and smaller feeling [48]. The synthesizer of the RFID tag sound is also added with the same start melody to create more coherence for the RFID sounds. This synthesizer has a sharp attack and a very short release so it feels energetic and snappy [48].

Results comparison
Adding a sound when starting the RFID module was regarded as a positive feature, as it aided participants in understanding their entry to the separate RFID module in the application. However, as stated in section 6.3, it was expected and now confirmed that FTF was the least liked sound for RFID start, being in conflict with guideline 3 and possibly guideline 1. Participants acknowledged its merit but did not express a clear and unanimous positive opinion. A few participants mentioned that the sound was harsh or unusual.

In contrast, almost all participants exhibited a highly favorable response to BNB. The sound was praised for its straightforwardness and effectiveness in signaling the start of a new RFID scan.

ONG also received positive feedback with similar arguments as BNB and was chosen most frequently to appear in the final set. Minor suggested improvements included increasing the sound's clarity, slightly elevating its pitch, or shortening its duration. Using either an octave (BNB) or a major chord (ONG) turned out to be a good choice for the RFID start process.

7.3.2 RFID cluster

Final sound: FTF
FTF is chosen because it received the fewest negative responses and got good positive feedback. Minor adjustments are made to make the sound less blurry. The synthesizer now has a sharper decay to better distinguish the different notes. On top of that, the same synthesizer of RFID start is also used to create more coherence in the sound set.

Results comparison
Only 7 votes were cast for the final set, as one participant refrained from selecting a favorite between FTF and ONG, and another participant expressed dissatisfaction with the complexity of all the sounds, as illustrated in quote 7.11.

Quote 7.11
Participant 1: *"It should not turn into a huge mishmash of all the sounds we will hear. It could also be as simple as... Just scanning, but repeated twice. "Tu-Tu". Like a double-click, indicating confirmation."*

FTF received positive feedback, as participants described it as providing a sense of confirmation and accomplishment. Only one participant remarked that the sound was a bit blurry.

BNB generally received positive feedback too, as it was described as a nice, short, and confirmatory sound, suitable for indicating successful actions. Minor suggestions for improvement included timing and volume adjustments to enhance its usability. The primary drawback of this sound was the last tone, perceived as somewhat irritating and high-pitched.

ONG received the most negative feedback, as it was perceived as an outdated, too long, and hollow sound by some participants. The old sound does not align with modern expectations. Positive reactions were limited, mainly expressing moderate enthusiasm and confirming that the sound conveyed a sense of confirmation. Contrary to the comments, it was chosen most frequently, although be it with a 1 vote difference.

It is interesting to compare FTF with ONG for the RFID cluster sound. Both use the same intervals but employ different synthesizers and timing. Comments on FTF were very positive, while ONG was less liked with participants mentioning it should not be considered because it is outdated, long, and hollow. This highlights the importance of choosing the right timbre, note length, and timing for earcons as this massively influences the perception. Overall, the choice of the ideal cluster sound for RFID scanning depends on the specific user preferences and the desired user experience.

7.3.3 RFID sweep

Final sound: BNB
BNB is chosen because it was chosen the most and it was clear the sound performed better than its alternatives. The volume is slightly increased to increase the perception when present in noisy warehouses.

Results comparison
It was confirmed that a background sound during RFID scanning is beneficial to understanding the system's status. Although initially, some participants confused the sweeping sound with echo or scanning another kind of RFID tag, all participants agreed the background sound is a useful addition, see quotes 7.12 and 7.13.

Quote 7.12
Participant 2: *"This is very clear, something we do not have at the moment. Previously, you were uncertain about what was happening exactly, but now you know you are searching. [...] A repetitive sound is effective. [...] The longer the gaps between the sounds, the more uncertainty arises about whether you are scanning or not."*

Quote 7.13
Participant 9: *"I really like the sweep sound, as it provides a clear indication that the scanning process is happening. Currently, I often press the RFID button, only to discover that the device is not scanning, either due to a device malfunction or an incorrect button press. The sweep sound serves as confirmation that the scanning is in progress, allowing me to discern whether the issue lies with my equipment or tag placement, rather than the scanner itself."*

Upon reviewing figure 19 it becomes evident that BNB's background sound is preferred the most. The sound was considered subtle, friendly, and not too obvious, while it is still clear and conveys the message well. It was nicely balanced with a lower volume, allowing other RFID tag sounds to stand out when played simultaneously. However, it must be noted that participants 11 and 12 both voted for another sound. They had the nosiest warehouse and were not able to perceive BNB well.

In contrast, comments received on both FTF and ONG were predominantly negative, as they were considered nervous and chaotic, ultimately leading to annoyance. Besides, participants encountered difficulty in distinguishing between the RFID tag and the RFID sweep when played simultaneously. Notably, FTF and BNB have used the same synth with almost the same settings. It is important to notice that FTF has had many negative responses, highlighting the impact of subtle differences in sound design.

7.4 Other feedback

Besides the comments regarding the sounds themselves, participants provided further feedback on the auditory enhancements. Firstly, participants agreed with expanding the number of sounds in the application, despite the notable increase from 2 to 8 sounds. The expanded auditory feedback is considered beneficial, provided that the sounds remain brief and easily distinguishable.

Secondly, there was a prevailing sentiment among most participants concerning sound configuration within the application. Participants were asked about their interest in configuring the sounds themselves, allowing them to select a subset that suited their preferences. Quote 7.14 underscores the rationale behind sound configuration, while quote 7.15 underscores the substantial variation in personal sound preferences. Most participants favored the idea of sound configuration but noted it should be tailored to the company's preferences rather than individual users. Scanners are exchanged between users and users are operating in close proximity, so individual sound configuration would lead to confusion, emphasizing the importance of company-wide sound standardization.

Quote 7.14
Participant 9: *"From all sets I just heard, there are a couple of sounds appealing to me and some who are not. I think mixing and matching different sounds yourself is important to create the best user experience for all users, not just a few."*

Quote 7.15
Participant 2: *"I would just design it right once. This is an error message for everyone, full stop! But well, on many devices, you can change your sound settings. It would be funny if we could, but if it really helps... On the other hand, it is sound. What you like, I do not. What I think is nice, you think is unpleasant. That already starts the discussion."*

Furthermore, participants stressed the importance of coherence within a sound set if there would only be one final set. The balance of a sound set is important, as sounds should fit together, but must also be easy to distinguish. It can be hard to create sounds that are significantly different, easy to distinguish, but connected to each other in a way too. Distinctiveness should be achieved through pitch differences and differing the number of notes in a sound, rather than relying too heavily on volume or timbre differences. A participant suggested that error messages could benefit from a higher volume level to capture the user's attention. Additionally, a participant proposed increasing the pitch of all sounds to improve perception at lower volumes.

Moreover, three participants mentioned RFID tag sound differentiation. Currently, all RFID tags produce the same sound. RFID tags not present in a project appear in a separate list on the scanner, see figure 21. Participants suggested implementing two different RFID tag sounds: one for RFID tags present in the project and another for RFID tags not associated with the project. This differentiation would assist users in promptly identifying unscheduled equipment items, with the modified sound signaling its distinctiveness and capturing attention.

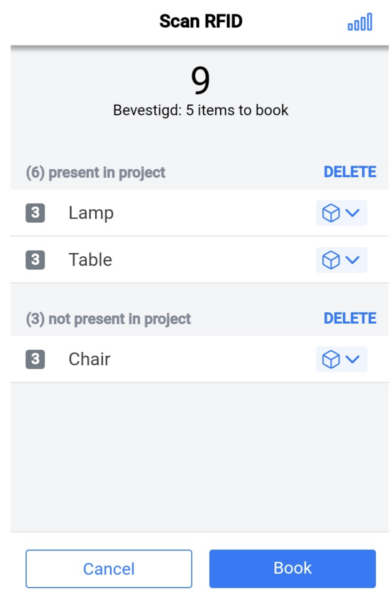

Figure 21: RFID table with (un)scheduled equipment items.

Lastly, participants were questioned about their expectations regarding the impact of the enhanced auditory feedback on UX. All participants agreed they would benefit from the enhanced auditory feedback with the selection of appropriate sounds. Clarity, learnability, and a user-centered approach were highlighted as crucial factors in achieving a positive UX. The enhanced auditory feedback contributes to professionalizing the application and gaining a better understanding of the actions taken within the application. Additionally, four participants agreed their efficiency would increase with the enhanced auditory feedback, as it would allow them to not look at the screen as often as they do now.

7.5 Summary

We have learned that transformational earcons could very well work to distinguish different scan methods from each other. If a sound is played in rapid succession, it would benefit from a light, short, and soft sound, while for QR code scanning more clarity is useful.

Error messages could benefit from a concise and comprehensible hierarchy. Although some error messages can be resolved immediately, they still require a negative character to grab the user's attention. A sound with a positive character does not necessarily direct the user's attention to the screen, leading to time losses and confusion during the initial use, and decreased UX in the long term. Besides, proper evaluation must be done to decide what errors or pop-ups are included in the hierarchy. Wrong decisions could lead to unfavorable outcomes where earcons are created with a misplaced character.

The RFID process has proven to benefit from limiting the repetitive beeping. Adding indicators to highlight the start, process, and end of the scan session results in better information transfer, improved UX, and increased accessibility. The sounds should be snappy to reflect the modern action the user is performing. While the start and end sound could be on a normal volume, the RFID sweep (background sound) can have a lower volume as it purely serves as a small confirmation the system is working correctly. However, it should still be perceived in noisy warehouses too.

An increase from 2 to 8 sounds is still acceptable, although all sounds should be useful and coherent with each other. Configuration on a company level could be a solution to satisfy the broadest range of users. In the future, it might be valuable to distinguish between correct and incorrect RFID tags during an RFID scan. Research should focus on determining whether the inclusion of an extra RFID tag will introduce additional noise and potentially lead to confusion during the scanning process, or whether it will expedite error detection, ultimately reducing the error margin during project preparations.

In conclusion, participants were satisfied with the direction the enhanced auditory feedback is going and have confirmed to expect a better UX and increased usability. There is no clear favorite sound set, as each sound set has its own pros and cons and sounds that stand out compared to the others. Different sounds were praised for different reasons, and even the most preferred sounds sometimes still have room for improvement. It has become clear that opinions about sound can have great variety and finding a solution that fits all users is difficult, but possible. With the newly adjusted sound set, we are confident we have built a solution fitting to most users in this field.

8 Study 3: Validation - Results

In this chapter, we present the results of the third and last phase of the study, as outlined in section 3.3. All data can be found in the .zip folder "Study 3 - Data analysis". It focuses on answering five usability sub-questions related to both the scan method and sound design. The results of each sub-question are discussed in separate sections. The abbreviations for the four evaluated conditions are as follows:

- BC: Barcode / QR code scanning with Conventional sounds
- BA: Barcode / QR code scanning with enhanced Auditory feedback
- RC: RFID scanning with Conventional sounds
- RA: RFID scanning with enhanced Auditory feedback

As the sample size is <50, Shapiro-Wilk tests have been performed for every dependent variable to determine the distribution of the data. Those results can be seen in table 7 in Appendix E. Besides, Levene's tests have been performed to determine the homogeneity of variance. Both tests will be briefly discussed at the start of each section to determine the significance test.

8.1 Efficiency

The sub-question for efficiency was as follows:

SQ3: *How do conventional sounds and enhanced auditory feedback affect usability in terms of efficiency during the scan process in a Warehouse Management System?*

$$H_0: \mu_{BC} = \mu_{BA} = \mu_{RC} = \mu_{RA}$$

H_1: The means are not all equal

Levene's test for equality of variances showed that the variances for the four conditions were equal ($F = 1.29$, $p = .280$). A Shapiro-Wilk test was performed for every condition and showed a normal distribution for all four conditions as can be seen in Appendix E table 7. The statistical assumptions for normality and homogeneity were therefore met, resulting in a parametric two-way repeated measures ANOVA as a significance test. Figure 22 shows more information about the distribution (with a Gaussian kernel density estimation), mean, and median per condition.

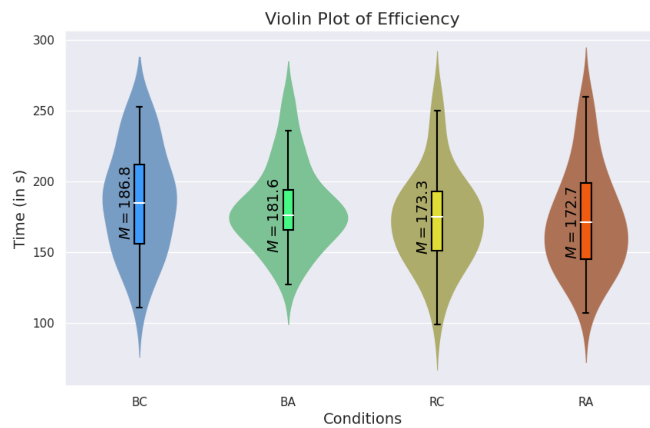

Figure 22: Violin plot efficiency, showing descriptive statistics and smoothed distribution shape. The white line inside the box plot shows the median, while the textual *M* on the side shows the mean value.

There was found one statistically significant effect at the .05 significance level, allowing us to reject the H0 of SQ3. The main effect for scan method yielded an $F(1, 37) = 15.78$, $p < .001$, indicating RFID scanning ($M = 173.0$, $SD = 35.4$) is significantly faster compared to QR scanning ($M = 186.8$, $SD = 33.4$). On average, RFID scanning was 7.3% faster compared to QR scanning.

No other significant differences were found. Although the current sounds ($M = 180.0$, $SD = 35.6$) was slightly slower compared to the enhanced auditory feedback ($M = 177.1$, $SD = 33.2$), it did not influence the efficiency significantly, $F = 0.76$, $p = .387$. Besides, no interaction effect was found between sound design and scan method for the efficiency, $F(1,37) = 0.36$, $p = .554$. We can conclude that only the scan method influences the efficiency, with RFID being more efficient compared to QR. Enhanced auditory feedback does not impact the work pace of the employee.

8.2 Effectiveness

The sub-question for effectiveness was as follows:

SQ4: *How do conventional sounds and enhanced auditory feedback affect usability in terms of effectiveness during the scan process in a Warehouse Management System?*

$$H_0: \mu_{BC} = \mu_{BA} = \mu_{RC} = \mu_{RA}$$

H_1: The means are not all equal

Levene's test for equality of variances showed that we have to reject the null hypothesis of equal variance between groups (F = 7.29, p < .001). On top of that, all Shapiro-Wilk tests performed also rejected the null hypotheses for normal distribution, with all data being right-skewed as can be seen in figure 32 in Appendix E. Two-way repeated measures ANOVA is known to be robust against violations of homogeneity of variance (if group sizes are equal) and normal distributions [5, 63]. However, both assumptions have been violated in this case, resulting in two non-parametric two-sided Wilcoxon signed rank tests to test statistical significance between QR code scanning and RFID scanning, and current auditory feedback and enhanced auditory feedback. Figure 23 shows more information per condition.

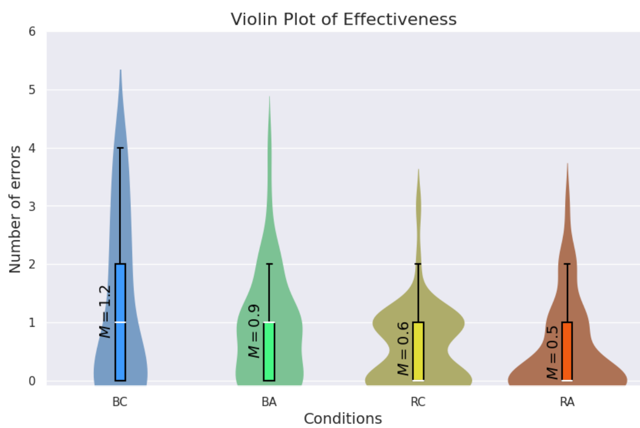

Figure 23: Violin plot effectiveness, showing descriptive statistics and smoothed distribution shape.

The Wilcoxon signed rank test for scan methods revealed that errors made were significantly lower during RFID scanning (M = 0.52, SD = 0.73) compared to QR code scanning (M = 1.02, SD = 1.18), T = 472.0, Z = -5.68, p = .004. RFID scanning was 51.0% more accurate compared to QR code scanning. We can therefore reject H0 of SQ4.

The Wilcoxon signed rank test for sound design revealed that errors made with the current sounds (M = 0.88, SD = 1.05) did not differ significantly from the enhanced auditory feedback (M = 0.67, SD = 0.86), T = 678.0, Z = -4.73, p = .224.

Lastly, when looking at the interaction plot in figure 24, we can conclude that no interaction was found between sound design and scan method for effectiveness, as the lines do not cross and are almost parallel to each other. We can conclude that only scan method influences the effectiveness, with RFID being more efficient compared to QR. Enhanced auditory feedback does not impact the errors made during the picking process.

8.3 Perceived workload (RTLX)

The sub-question for perceived workload was as follows:

SQ5: *How do conventional sounds and enhanced auditory feedback affect usability in terms of cognitive workload during the scan process in a Warehouse Management System?*

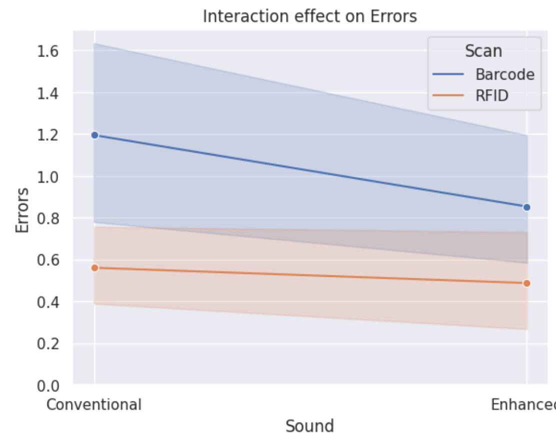

Figure 24: Interaction effect plot of the effectiveness. Error bars show the 95% confidence interval

$$H_0: \mu_{BC} = \mu_{BA} = \mu_{RC} = \mu_{RA}$$

H_1: The means are not all equal

The RTLX score was calculated by taking the average of the six NASA-TLX dimensions, see appendix C with annoyance excluded. Levene's test for equality of variances showed that the variances for the four conditions were equal (F = 1.30, p = .276). A Shapiro-Wilk test was performed for every condition and did not show evidence of non-normality for three out of four conditions. Only RA had a non-normal distribution (W = .94, p = .021). Based on this outcome, after a visual examination of the histograms (see figure 33 in Appendix E, and based on the robustness of ANOVA against violations of normality, we decided to use a two-way repeated measures ANOVA. Figure 25 shows more information per condition.

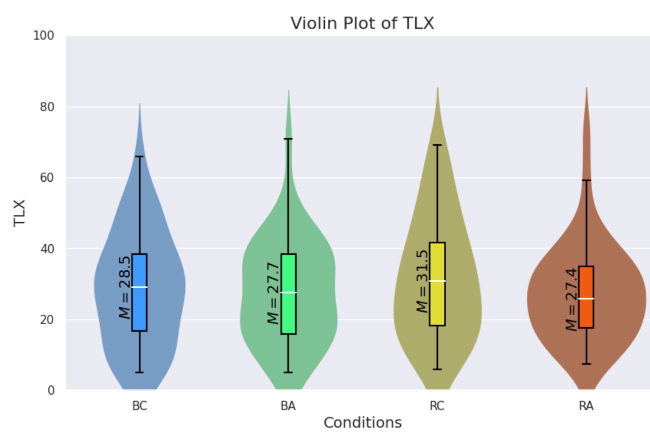

Figure 25: Violin plot perceived workload (RTLX), showing descriptive statistics and smoothed distribution shape.

There was found one statistically significant effect at the .05 significance level, allowing us to reject H0 for SQ5. The main effect for sound type yielded an $F(1, 37) = 4.37$, $p = .043$, indicating a significant difference between the current sounds ($M = 30.0$, $SD = 16.5$) and the enhanced auditory feedback ($M = 27.6$, $SD = 14.1$). The perceived workload of the enhanced auditory feedback was 2.4 points lower compared to the conventional sounds, meaning it is less demanding.

No statistically significant difference was found between the two scan methods. QR code scanning ($M = 28.1$, $SD = 15.2$) and RFID scanning ($M = 29.5$, $SD = 15.4$) had approximately the same perceived workload. Besides, although the lines in figure 26 are not parallel, there was no interaction effect found between sound design and scan method for the RTLX score, $F(1, 37) = 2.24$, $p = .141$.

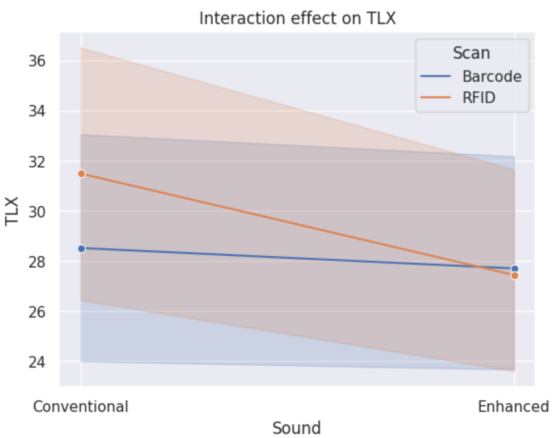

Figure 26: Interaction effect plot of the perceived workload. Error bars show the 95% confidence interval

8.3.1 Individual factors RTLX

Besides the calculated RTLX average, significance tests were also performed on the individual NASA-RTLX factors to get a more detailed insight. All six factors were non-normally distributed, see Appendix E table 8 for the scores. All of Levene's tests failed to reject the null hypothesis of equal variances between groups, so all groups have similar variances. As only one parametric assumption is violated, and two-way repeated measures ANOVA is known to be robust against non-normality for larger sample sizes, it was still decided to use this parametric significance test [5]. Figure 27 shows the mean and standard deviation per factor per condition.

Two statistically significant effects were found at the .05 significance level. There was a significant main effect of the scan method on mental demand, $F(1, 37) = 7.73$, $p = .008$. Barcode/QR code scanning ($M = 24.3$, $SD = 20.4$) was significantly less mentally demanding compared to RFID scanning ($M = 31.0$, $SD = 23.4$). Besides, there was a significant main effect of sound design on frustration, $F(1, 37) = 4.24$, $p = .046$. The enhanced auditory feedback ($M = 24.8$, $SD = 20.6$) was significantly less frustrating compared to the conventional sounds ($M = 29.5$, $SD = 22.4$). No other main effects or interaction effects were found for the remaining factors.

8.4 Annoyance

The sub-question for annoyance was as follows:

SQ6: *How do conventional sounds and enhanced auditory feedback affect the usability in terms of annoyance during the scan process in a Warehouse Management System?*

$$H_0: \mu_{BC} = \mu_{BA} = \mu_{RC} = \mu_{RA}$$

H_1: The means are not all equal

Levene's test for equality of variances showed that the variances for the four conditions were equal ($F = 0.59$, $p = .621$). A Shapiro-Wilk test was performed for every condition and showed evidence of non-normality for all conditions, with RA having the lowest score ($W = .88$, $p = .001$). This is confirmed when looking at the histograms in figure 34 in Appendix E. However, we decided to still use a two-way repeated measures ANOVA, as it is known to be relatively robust to violations of normality. Moreover, if all distributions exhibit a similar skewness pattern, such as being right-skewed in this instance, the impact of the violation is mitigated to a lesser extent [5]. On top of that, the Central Limit Theorem applies due to the large sample size, where the sampling distribution will always approximately follow a normal distribution [96]. Lastly, only one statistical assumption is violated, contrary to the effectiveness situation where the assumption of homogeneity is violated too. Figure 28 shows more information per condition.

No statistically significant effects were found at the .05 significance level. No main effect was found for the scan method with $F(1, 37) = 1.20$, $p = .281$, indicating no significant difference between QR code scanning ($M = 33.2$, $SD = 25.7$) and RFID scanning ($M = 29.2$, $SD = 23.3$). No main effect was found for the sound design $F(1, 37) = 3.79$, $p = .059$, indicating no significant difference between the conventional sounds ($M = 33.6$, $SD = 24.4$) and the enhanced auditory feedback ($M = 28.8$, $SD = 24.6$). No interaction effect was found between sound design and scan method for the annoyance, $F(1, 37) = 0.08$, $p = .772$, confirmed by figure 39 in Appendix E, showing two lines almost parallel to each other.

The p-value of the main effect of the sound design was close to the significance level, therefore an additional non-parametric Wilcoxon signed rank test was performed to determine whether there was a difference in the ranking of the conventional sounds and enhanced auditory feedback. Results of that analysis indicated that there was a significant difference between the two conditions, $T = 801.50$, $Z = -4.16$, $p = .023$, indicating that the enhanced auditory feedback is less annoying compared to the conventional sounds and rejecting H0.

8.5 Subjective user rating (SUS)

The sub-question for subjective user rating was as follows:

SQ7: *How do conventional sounds and enhanced auditory feedback affect usability in terms of subjective rating of the scan process in a Warehouse Management System?*

$$H_0: \mu_{BC} = \mu_{BA} = \mu_{RC} = \mu_{RA}$$

H_1: The means are not all equal

Levene's test for equality of variances showed that the variances for the four conditions were equal ($F = 1.15$, $p = .333$). A Shapiro-Wilk test was performed for every condition, rejecting all null

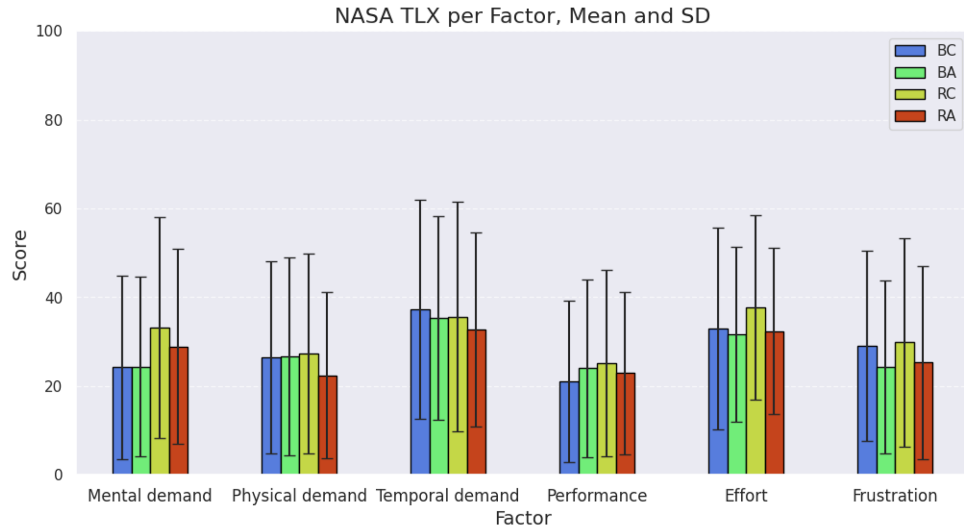

Figure 27: Mean and standard deviation per factor per condition for NASA-RTLX.

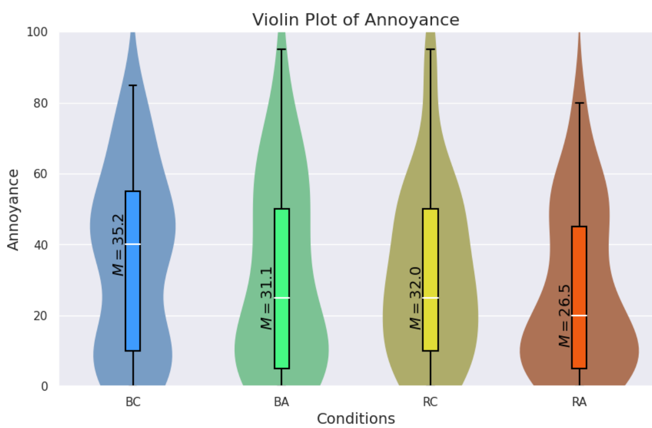

Figure 28: Violin plot annoyance, showing descriptive statistics and smoothed distribution shape.

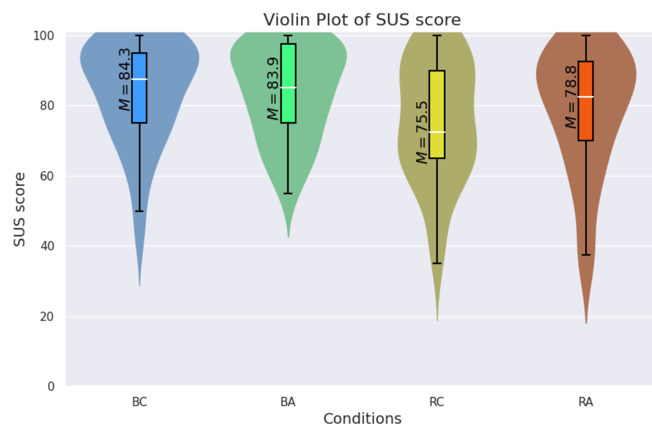

Figure 29: Violin plot subjective user rating (SUS), showing descriptive statistics and smoothed distribution shape.

hypotheses except for RC (W = .95, p = .101). The distributions in figure 35 in Appendix E show that most of the data are left-skewed. As previously stated in the annoyance section, ANOVA tests are relatively robust to violations of normality. Therefore, two-way repeated measures ANOVA was performed, but extra attention was paid to p-values close to the threshold. Figure 29 shows more information per condition.

There was found one statistically significant effect at the .05 significance level, allowing us to reject H0 for SQ7. QR scanning (M = 84.1, SD = 12.9) scored significantly higher on the SUS score compared to RFID scanning (M = 77.1, SD = 17.5), resulting in a main effect of $F(1, 37) = 10.92$, p = .002.

No main effect was found for sound design $F(1, 37) = 0.63$, p = .431. Therefore, no statistically significant difference was found between the conventional sounds (M = 79.9, SD = 15.8) and the enhanced auditory feedback (M = 81.3, SD = 15.5). There was also no interaction effect found between sound design and scan method for the SUS score, $F(1, 37) = 0.72$, p = .402.

8.6 Qualitative results

Participants were allowed to leave short notes of feedback after each condition. With both BC and RC, many participants mentioned the fact that the project pop-up did not have the sound they would expect, as it was the same as an error sound. There was a sentiment among the participants that BC was overly negative and discouraging because of the negative beeps. This was better during the BA condition, where the project pop-up got a more positive sound. Participants liked this, although someone mentioned the discrepancy was a bit odd compared to the other sounds in that set.

Participants mentioned they struggled with the repetitive beeping of the RC condition, not knowing whether they scanned everything or not. Opinions about the RA condition were divided. Although people knew better whether they were done or not, they also struggled with learning the new sounds. They did not know what was going on and only started to understand better how it worked during the actual experiment. For some of them, it was useful, creating a more intuitive scan process, while for others it increased the confusion and annoyance.

9 Study 3: Validation - Discussion

Table 6 summarizes the significant results of the previous section. These are used as references in this chapter so the results can be put into perspective and the implications of the findings can be discussed. The chapter will conclude with discussing the experiment limitations and external validity of the experiment.

9.1 Influence of scan method on usability

Two out of four results were in favor of RFID scanning. RFID scanning turns out to be faster compared to QR code scanning. This result was aligned with what was expected, as RFID does not need a direct line of sight and multiple RFID tags can be scanned at once. This means that RFID is the more efficient way of working and is able to speed up the warehouse processes by at least 7.3% (see Limitations section 9.3 for more information). These results are also in line with previous research telling that RFID enhances overall warehouse efficiency [95].

RFID turns out to be the more effective scan method too, as less errors were made during the packing process. This is a striking result, as one would think that scanning items one by one with QR code scanning gives a user more control about what has been scanned and what has not. However, many participants double scanned QR codes during the experiment, something which does not happen as frequently with RFID due to the nature of RFID scanning. The Limitations section 9.3 will discuss this in greater detail.

The other two significant results were in favor of QR code scanning. QR code scanning is less mentally demanding compared to RFID scanning. One of the reasons might be the fact that people are more used to barcode scanning in their daily life. Another reason might be that QR code scanning is seen as the simpler task, because you get immediate and very understandable feedback after every single scan that is performed. While using RFID, the user is less in control of what is actually scanned, making it more mentally demanding to understand if all the necessary items have been processed.

Lastly, QR code scanning has a higher SUS score compared to RFID scanning. Both scan methods are above 68, which means they score above average [92]. While RFID falls in the B category (68-80.3), QR code scanning is in A (>80.3), meaning the latter is in the highest usability segment. So although RFID turns out to be more effective and efficient, in terms of mental demand and usability QR code scanning is preferred which is probably because of its direct and simplistic nature.

9.2 Influence of sound design on usability

In terms of perceived workload, it has been proven that enhanced auditory feedback outperforms the conventional sounds for both QR code and RFID scanning. A lower perceived workload could potentially be beneficial for increased pleasure at work and less stress. With the enhanced auditory feedback primarily focused on improving the three roles of auditory feedback in UI (conveying information, enhancing UX, and improving accessibility), it is safe to say that it is useful to implement so users have a more enjoyable experience and experience less workload. Besides, it is in line with previous research showing that (useful) information spread over multiple modalities helps to minimize users' cognitive load [2].

Enhanced auditory feedback has also proven to be less frustrating and annoying compared to the conventional sounds. This shows that a thoughtful sound design could be beneficial for the UX. Definitely for systems where sounds are heard over a thousand times a day, this is an incredibly important finding. Having significantly less frustration and being significantly less annoyed increases the UX, keeping employees more in the flow, being potentially beneficial for long-term efficiency or even employee happiness.

These findings are in line with the research by Frauenberger et al. [36], stating that many UIs fail to fully leverage the potential of sound, leading to UX issues like frustration incomprehension, and annoyance. Rentman did not pay much attention to the auditory feedback in their application so far. However, we have proven that with more detailed and enhanced auditory feedback, UX issues like those stated above can be improved in the warehousing domain.

There were no significant results for enhanced auditory feedback on efficiency, effectiveness, and SUS. Although participants said they felt more confident performing a full RFID scan with the enhanced auditory feedback, time and error-wise this made no difference. However, there were also no results negatively impacting the enhanced auditory feedback. Like proven before, moving towards a better multimodal approach enhances usability [99].

9.3 Experiment limitations

While the current experiment contributes valuable insights to the understanding of both scan methods and sound design in the warehousing industry, it is essential to acknowledge several limitations that may impact the interpretation and generalization of the findings.

Initially, the study sample comprised 35% females, a noteworthy proportion given the predominant male demographic in the industry. Previous interviews exclusively involved male participants, and the majority of warehouse personnel also consists of males. Furthermore, the participant pool primarily consisted of Rentman employees, who do not represent the end users. While an ideal scenario would involve exclusively end users, logistical considerations combined with experimental consistency dictated that the experiment could be conducted solely at Rentman's office. Performing the experiment at end user warehouses would introduce confounding factors such as background noises and setup variations, compromising the reliability and validity of the results.

As can be seen in figure 30, there was a learning effect during the experiment (see figure 41 in Appendix E for a more detailed version). On average, participants became faster after each condition, as they

Table 6: Significant results of experiment 3.

DV	IV	What	Comparison
Efficiency	Scan	RFID scanning is faster compared to QR code scanning	173s vs. 187s (7.3%)
Effectiveness	Scan	RFID scanning has fewer errors compared to QR code scanning	0.52 vs. 1.02 (51.0%)
Workload	Sound	Enhanced auditory feedback has lower perceived workload compared to conventional sounds	27.6 vs. 30.0 (8.0%)
Workload	Scan	QR code scanning requires less mental demand compared to RFID scanning	24.3 vs. 31.0 (21.6%)
Workload	Sound	Enhanced auditory feedback is less frustrating compared to conventional sounds	24.8 vs. 29.5 (15.9%)
Annoyance	Sound	Enhanced auditory feedback is less annoying compared to conventional sounds	28.8 vs. 33.6 (14.3%)
SUS	Scan	QR code scanning has higher usability score compared to RFID scanning	84.1 vs. 77.1 (9.1%)

learned where the QR code sticker was, got better at RFID scanning, and improved their efficiency by picking up items and moving them from A to B. This learning effect shows the importance that we carefully randomized the order in which each condition appeared. All conditions were executed first, second, third, and fourth approximately an equal number of times. However, it also indicates that the initial demo and tryout at the beginning of the experiment might have needed more extensive coverage and attention.

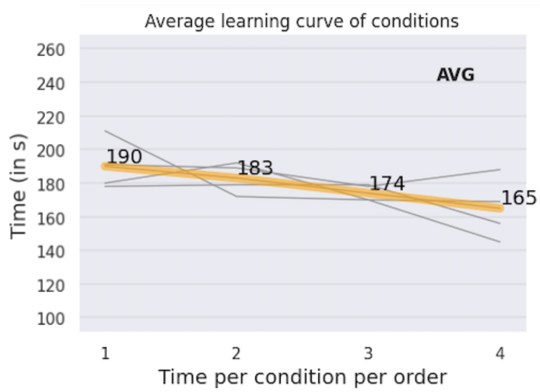

Figure 30: Line chart showing the learning effect over the different conditions during the experiment. Grey lines show the average times of the individual conditions, see figure 41 in Appendix E for details.

Because it is a lab experiment and not a field experiment or natural experiment, it negatively influences the validity of the results. Recreating a real warehouse setting with the corresponding tasks was difficult. Although we tried to come as close as possible with the core tasks of packing equipment items, a full scale experiment was not possible. In the warehousing industry packing volumes can be bigger, the work itself is more physically demanding, and other conditions are less controlled.

RFID has demonstrated to be faster compared to QR code scanning. However, it is anticipated that this disparity would be more pronounced in a real-world scenario. Consider for instance a flight case containing 50 small items. When all 50 items must be scanned, the conventional process entails extracting each item from the case individually for scanning. In contrast, RFID technology allows for scanning without the need to open the flight case, and it is envisaged that this streamlined process would yield a more substantial difference than the presently observed 7.3%.

In a real world context it is expected that RFID will exhibit an increased error rate, leading to a diminished effectiveness. Throughout the experiment, RFID scan errors manifested as instances of failing to scan one or multiple equipment items. Conversely, QR code scanning, involving a click for every equipment item, occasionally resulted in the inadvertent duplication of scans, recorded as errors. While such errors are circumvented in RFID scanning, the prospect of missing an equipment item due to a faulty RFID scan is deemed a more significant error in practical scenarios than the occurrence of a duplicate QR code scan, which ensures the verification of item packing. Despite the current substantial 51.0% advantage favoring RFID scanning, it is anticipated that these differences will diminish during actual order picking, potentially even tilting in favor of QR code scanning.

10 Conclusion

In this section, we will answer our final research question. A short recap on the sub-questions will be given before we delve into answering the main research question. This recap will serve as a contextual bridge, paving the way for a comprehensive examination and insightful answers to the overarching research problem.

10.1 Study 1 - Exploration

During the exploration phase, we focused on answering the following sub-question:

SQ1: *In what way should enhanced auditory feedback support different actions of the scan processes on mobile devices in Warehouse Management Systems compared to the current situation?*

Our study uncovered insights across three main thematic areas. Firstly, the "Current Workflows and Technology Sentiment" theme illuminated the diverse practices within warehouse operations, revealing a predominantly positive sentiment toward the Rentman mobile application. QR codes were seen as reliable, while RFID technology showcased room for improvement and optimization.

Secondly, the theme "Satisfaction with Current Auditory Feedback" emphasized the pivotal role of auditory feedback in user interactions. Participants underscored the importance of simplicity and clarity in auditory cues during QR code scanning, which significantly contributed to efficient order-picking processes. However, in the context of RFID scanning, the continuous beeping was met with negative feedback, highlighting the need for improvements in this area. Participants expressed a reliance on visual feedback as little information could be inferred from the RFID sounds.

Lastly, under the theme of "Potential Improvements" participants provided nuanced suggestions for refining auditory feedback in both QR code and RFID scanning processes. QR code scanning only got some minor enhancements, while RFID scanning advocated for a more streamlined auditory experience, including the introduction of a start sound, reducing repetitive beeping, and exploring innovative features like equipment item localization.

These three themes resulted in sound design guidelines, which are stated in table 3. Key lessons learned that answer SQ1 included creating a sound pack that is simple, easy to use, easy to learn, and easy to perceive. Different scan methods could benefit from different sounds as they require different actions. RFID scanning could benefit from a slightly more complex sound design as it requires more complex actions.

10.2 Study 2 - Verification

After the first study, three different sound packs were created which were evaluated by participants. Every sound set included eight sounds, see table 4. With those sounds, divided into the categories "Confirmation Scan", "Errors and Pop-ups", and "RFID sounds", the following sub-question was answered:

SQ2: *How do users perceive and interpret the newly created enhanced auditory feedback during the different scan processes of a Warehouse Management System?*

In general, there was no single sound set showing superiority over the others. Each set had its advantages and drawbacks. Although ONG was perceived as the most "successful" overall, not every sound within that set was the most frequently chosen. Take for example the confirmation scan. For QR code, ONG was chosen as the final option, while for RFID scanning FTF was picked most frequently. Minor adjustments were made to increase consistency and clarity for both sounds.

A slight majority of the participants preferred error message differentiation, making us decide to implement a small error message hierarchy. ONG was chosen for both error sounds, as it was praised for its clarity, attention-grabbing, and participants agreed that it effectively communicated a feeling of error. The project pop-up sound, however, did not have a clear favorite, and redesigning it was required.

Participants found the adjustment where a beep was emitted only for unique RFID tags to be a significant enhancement, reducing overall annoyance during the process. BNB was favored for RFID sweep due to its subtle and friendly characteristics. For RFID start, ONG was chosen despite some negative feedback, leading to slight redesign adjustments. The RFID cluster sound was more divided, with FTF ultimately chosen for further development.

Participants showed interest in increasing the number of sounds in the application, provided they remained short and distinguishable. There was also a preference for sound configuration tailored to company preferences rather than individual users. Participants emphasized the importance of coherence within a sound set and suggested RFID tag sound differentiation based on whether an equipment item is planned on a project or not.

The findings suggest that users generally appreciated the enhanced auditory feedback, anticipating improvements in UX and efficiency. However, there was no unanimous preference for a specific sound set, highlighting the subjective nature of user preferences. The study showed valuable insights into the strengths and weaknesses of different sounds, informing the researchers' decisions for the final sound set. Adjustments were made based on participant feedback, aiming to strike a balance between clarity, distinction, and user satisfaction in the auditory feedback design.

10.3 Study 3 - Validation

During the last study, we focused on answering 5 sub-questions about efficiency, effectiveness, perceived workload, annoyance, and subjective rating, all related to the usability of auditory feedback and scan methods in warehouse management systems. The findings emphasize the importance of considering both technological and user-centered aspects in designing warehouse systems.

No interaction effects between the independent variables were observed, indicating that the combined influence of the two variables did not yield a statistically significant effect. Concerning the scan method, RFID scanning demonstrated both a statistically significant increase in speed and accuracy compared to QR code scanning. While the expectation was that RFID would exhibit greater speed, the revelation of it being the more accurate condition was unexpected. It is anticipated that in real-world scenarios, RFID scanning would further amplify its speed advantage, albeit with a decrease in accuracy. Conversely, QR code scanning, characterized by lower mental demand, garnered a higher SUS score. This preference could be attributed to the widespread familiarity of users with barcode scanning in their daily routines.

Concerning the sound design, all significant results were in favor of the enhanced auditory feedback. Users perceived a lower workload compared to the conventional sounds. On top of that, users were less annoyed and less frustrated with the enhanced auditory feedback. There were no significant results for efficiency, effectiveness, and subjective user rating.

10.4 Final conclusion

During the three studies of this research, the conclusion for the main research question gradually unfolded. The research question was:

Research question: How does enhanced auditory feedback impact the usability of equipment scanning on mobile devices in Warehouse Management Systems?

The amalgamation of these studies provides a comprehensive perspective on the multifaceted aspects of usability, encompassing both scan methods (QR code and RFID) and sound designs (conventional and enhanced auditory feedback). The first study, a preliminary exploration, laid the foundation for subsequent investigations into the nuanced aspects of usability. Building upon the insights gained from study 1, the second study introduced a comparative analysis of sound designs across both QR code and RFID scanning methods. The results highlighted the impact of sound design on UX, indicating that the created guidelines from the first study could improve usability and verified that some of the created sounds were liked by the participants. The third study synthesized the findings from the previous studies and expanded the investigation to understand the combined impact of scan methods and sound designs on usability.

The synthesis of these studies underscores the intricate interplay between scan methods and sound designs in shaping the best usability within WMS. RFID scanning emerged as a frontrunner in terms of efficiency and accuracy, while QR code scanning demonstrated user-centric advantages. Enhanced auditory feedback proved influential in mitigating negative emotional responses and enhancing UX.

The cumulative findings advocate for a nuanced approach to usability in WMS, recognizing the diverse preferences and priorities of users. If efficiency and accuracy are most important, RFID is the recommended scanning method. If subjective user rating and mental demand are your priority, one should use QR code scanning. However, likely, there will always be an interplay as most users will use both methods in their warehouses with QR codes being used as back-up for RFID.

Lastly, as results have shown, enhanced auditory feedback does not have negative effects. The lower perceived workload, with less annoyance and frustration, could result in a more enjoyable working experience. Although it is difficult to design auditory principles similar to the well-known visual Gestalt principles, table 3 could serve as a high-level framework for optimizing usability via enhanced auditory feedback for WMSs. As can be seen by the results of the second study, there is no one-size-fits-all solution or golden standard, but iterating and verifying the created sound set is always necessary to ensure a user-centered design with maximum usability benefits. In conclusion, it has become clear that spending resources on getting an enhanced auditory feedback system is beneficial for usability.

10.5 Implications

As prior research stated, technology can be used to increase job satisfaction and reduce employee turnover, which is currently an important challenge in the warehousing industry [46, 68]. With our results, we can conclude it makes sense to spend resources on enhanced auditory feedback to increase the employee's emotional well-being, as it decreases perceived workload, annoyance, and frustration. This might lead to higher job satisfaction, by creating a working environment that is less austere and stressful. This implication has similarities with findings from a study by Beckham et al. [9], where they conclude that "any positive association [regarding multimodality] that can be made with a repetitive motion job should be of interest to management given the high rate of turnover in material- and package-handling positions."

Furthermore, our contribution extends to advancing research on the usability of emerging technologies in the warehouse industry. Typically, such aspects, definitely in combination with audio, are scarcely analyzed. However, our investigation takes a step in the right direction, providing a foundation for future studies and serving as inspiration for further research endeavors in this domain by filling a critical usability knowledge gap.

The integration of enhanced auditory feedback as a usability catalyst echoes the insights from Absar and Guastavino [2], affirming that a thoughtful sound design contributes to a more enjoyable user experience and aids in minimizing cognitive load. Our research extends this understanding to the specific context of WMSs, emphasizing the role of enhanced auditory feedback in reducing perceived workload, frustration, and annoyance. This aligns with the assertation by Wechsung and Naumann [99] that a multimodal approach, including auditory feedback, enhances usability.

While RFID may excel in effectiveness and efficiency, QR code scanning, with its higher SUS score and lower mental demand, reflects the user-centric aspect. Besides, enhanced auditory feedback is better for the emotional well-being of the users. Our research approach considers not only task completion metrics but also the users' subjective experiences, ensuring a more comprehensive evaluation of usability by also looking more at the human side of the usability spectrum. Definitely in the warehousing domain, where employee turnover is high, a more user-focused evaluation of usability is necessary to increase job satisfaction.

In conclusion, our research offers a framework for optimizing WMS usability. While previous research has barely covered a full usability scope, our research serves as an inspiration to fully cover the usability aspect, as significant results have been found for all usability parts, including user satisfaction on multiple levels. This framework not only focuses on efficiency and effectiveness but also highlights the importance of testing user satisfaction, which in turn could positively influence ongoing warehouse challenges like job satisfaction and minimizing warehousing costs by reducing employee turnover. The three-step structure we followed for designing and testing auditory feedback has proven fruitful and we would recommend following a similar pattern for comparable studies. The integration of RFID efficiency, QR code user-friendliness, and enhanced auditory feedback emotional well-being provides a nuanced approach to addressing the diverse needs and preferences of users. Although there is not a one-size-fits-all solution, we were able to stress the importance of certain features and functionalities that could increase usability in the warehousing domain. Our implications bridge empirical findings with established literature, offering a robust foundation for refining WMS practices and enhancing the overall usability of mobile equipment scanning in warehouse environments.

10.6 Future work

Our current study has unveiled several options for future research. Next, potential directions for extending and building upon our current investigation, as well as areas where our study opens new possibilities for exploration will be discussed.

Our study focused solely on the influence of enhanced auditory feedback including visual feedback. Future research could explore the integration of haptic feedback, to assess how a holistic multimodal approach enhances usability. Studying three combined modalities could provide a comprehensive understanding of the most effective design principles for WMS interfaces. Although Beckham et al. [9] have performed research about multimodality including haptics for barcode scanning, they have not yet done any research including RFID scanning, which could be an interesting addition.

Secondly, conducting studies in real-world warehouse environments, besides an artificially controlled lab setting, is crucial. Future research should aim to validate our findings by implementing the proposed feedback in authentic warehouse scenarios. This would involve considering factors such as variable noise levels, different warehouse layouts, and the physical demands of the work environment. Researching different warehouses, from small-scale storage facilities to large distribution centers, could enhance the generalizability of our findings.

Our study captured only a very short usability timeframe. A longitudinal investigation could assess how users' experiences evolve over an extended period of system usage. It would allow us to see whether there is any novelty effect regarding both RFID scanning and enhanced auditory feedback, offering a more nuanced understanding of the long-term impact on usability.

Lastly, comparative analyses with other industries that employ similar scanning and feedback systems could be performed, providing valuable benchmarks. This approach could help identify best practices, lessons learned, and potential transferable insights for optimizing usability in not only the warehousing domain, but also domains like retail, transport, and healthcare.

As technology and user preferences evolve, our research underscores the importance of continuous exploration. It serves as a foundational step in the exploration of usability in WMSs. By addressing these potential future research directions, we can continue to advance the field and ensure that WMS interfaces are designed to maximize efficiency, minimize errors, and enhance overall user satisfaction in the dynamic and evolving landscape of warehouse operations.

References

[1] Rafa Absar and Catherine Guastavino. 2008. Usability of non-speech sounds in user interfaces. In *Proceedings of the International Conference on Auditory Display (ICAD'08)*. Georgia Institute of Technology, 1–8.

[2] Rafa Absar and Catherine Guastavino. 2015. The design and formative evaluation of nonspeech auditory feedback for an information system. *Journal of the Association for Information Science and Technology* 66, 8 (2015), 1696–1708.

[3] Aditi Agarwal. 2019. Earcon, the new icon. https://yujdesigns.medium.com/earcon-the-new-icon-3ce28c6b9d02

[4] Marwan Alseid and Dimitrios Rigas. 2011. The role of earcons and auditory icons in the usability of avatar-based e-learning interfaces. In *2011 Developments in E-systems Engineering*. IEEE, 276–281.

[5] AM Statistical Consulting. n.d.. Two-Way RM ANOVA - Statistical Consulting. https://www.amstatisticalconsulting.com/two-way-repeated-measures-anova/

[6] T S Amer and Todd L. Johnson. 2018. Earcons versus auditory icons in communicating computing events: learning and user preference. *International Journal of Technology and Human Interaction (IJTHI)* 14, 4 (2018), 95–109.

[7] Chad W. Autry and Patricia J. Daugherty. 2003. Warehouse operations employees: linking person-organization fit, job satisfaction, and coping responses. *Journal of Business Logistics* 24, 1 (2003), 171–197.

[8] Nick Babich. 2020. 10 Usability Heuristics and How To Apply Them To Product Design. https://uxplanet.org/10-usability-heuristics-and-how-to-apply-them-to-product-design-dd4a4a06d78c

[9] Elizabeth C. Beckham, Reuben F. Burch, Lesley Strawderman, Kari Babski-Reeves, Linkan Bian, and Katherine King. 2020. Feedback mode preference and performance improvement with a handheld scanning device. *International Journal of Services and Operations Management* 35, 3 (2020), 321–338.

[10] Steven M. Belz, Gary S. Robinson, and John G. Casali. 1999. A new class of auditory warning signals for complex systems: Auditory icons. *Human factors* 41, 4 (1999), 608–618.

[11] Samuel M. Berger and Timothy D. Ludwig. 2007. Reducing warehouse employee errors using voice-assisted technology that provided immediate feedback. *Journal of Organizational Behavior Management* 27, 1 (2007), 1–31.

[12] Ann Blandford, Dominic Furniss, and Stephann Makri. 2016. Qualitative HCI research: Going behind the scenes. *Synthesis lectures on human-centered informatics* 9, 1 (2016), 1–115.

[13] Meera M. Blattner, Denise A. Sumikawa, and Robert M. Greenberg. 1989. Earcons and icons: Their structure and common design principles. *Human–Computer Interaction* 4, 1 (1989), 11–44.

[14] Sara A. Bly. 1982. *Sound and computer information presentation*. University of California, Davis.

[15] Terri L. Bonebright and Michael A. Nees. 2007. Memory for auditory icons and earcons with localization cues. In *Proceedings of the International Conference on Auditory Display (ICAD'O7)*, Vol. 2007. Georgia Institute of Technology.

[16] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[17] Stephen A. Brewster. 1992. *Providing a model for the use of sound in user interfaces*. University of York, Department of Computer Science.

[18] Stephen A. Brewster. 1994. *Providing a structured method for integrating non-speech audio into human-computer interfaces*. Ph. D. Dissertation. University of York York, UK.

[19] Stephen A. Brewster. 1998. Using nonspeech sounds to provide navigation cues. *ACM Transactions on Computer-Human Interaction (TOCHI)* 5, 3 (1998), 224–259.

[20] Stephen A. Brewster and Murray G. Crease. 1999. Correcting menu usability problems with sound. *Behaviour & Information Technology* 18, 3 (1999), 165–177.

[21] Stephen A. Brewster, Peter C. Wright, and Alistair D.N. Edwards. 1993. An evaluation of earcons for use in auditory human-computer interfaces. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*. 222–227.

[22] Raluca Budiu. 2023. Between-Subjects vs. Within-Subjects Study Design. https://www.nngroup.com/articles/between-within-subjects/

[23] Raluca Budiu and Kate Moran. 2021. How Many Participants for Quantitative Usability Studies: A Summary of Sample-Size Recommendations. https://www.nngroup.com/articles/summary-quant-sample-sizes/

[24] João P. Cabral and Gerard B. Remijn. 2019. Auditory icons: Design and physical characteristics. *Applied ergonomics* 78 (2019), 224–239.

[25] Chun S. Choong, Ahmad F.A. Nasir, Anwar P.P. Abdul Majeed, Muhammad A. Zakaria, and Mohd A.M. Razman. 2021. Automatic identification and categorize zone of RFID reading in warehouse management system. In *Advances in Mechatronics, Manufacturing, and Mechanical Engineering: Selected articles from MUCET 2019*. Springer, 194–206.

[26] Michael R. Cohen and Judy L. Smetzer. 2011. Scanner Beep Only Means the Barcode Has Been Scanned; Barcode Scanning after Drug Administration Has Little Value; National Alert Network: Dosing Confusion with Colistimethate for Injection; Mix-ups between Risperidone and Ropinirole. *Hospital Pharmacy* 46, 10 (2011), 738–742.

[27] Ádám Csapó and György Wersényi. 2013. Overview of auditory representations in human-machine interfaces. *ACM Computing Surveys (CSUR)* 46, 2 (2013), 1–23.

[28] David Czuchta. 2021. How interval music theory can trick your brain. https://gravitysound.studio/blogs/news/how-interval-music-theory-can-trick-your-brain#:~:text=The%20perfect%20octave%20interval%20involves,completeness%20as%20well%20as%20lightheartedness.

[29] Tilman Dingler, Jeffrey Lindsay, Bruce N. Walker, et al. 2008. Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. In *Proceedings of the International Conference on Auditory Display (ICAD'08)*. 1–6.

[30] Konstantinos Domdouzis, Bimal Kumar, and Chimay Anumba. 2007. Radio-Frequency Identification (RFID) applications: A brief introduction. *Advanced Engineering Informatics* 21, 4 (2007), 350–355.

[31] Inger Ekman, Laura Ermi, Jussi Lahti, Jani Nummela, Petri Lankoski, and Frans Mäyrä. 2005. Designing sound for a pervasive mobile game. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*. 110–116.

[32] Mica R. Endsley. 1995. Toward a theory of situation awareness in dynamic systems. *Human factors* 37, 1 (1995), 32–64.

[33] Lars Engeln, Maximilian Hanke, Toni Auerswald, Felix Kallenbach, and Rainer Groh. 2019. The earconizer-a tool for constructing hierarchical earcons. *Mensch und Computer 2019-Workshopband* (2019).

[34] Jamie Enoch, Leanne McDonald, Lee Jones, Pete R. Jones, and David P. Crabb. 2019. Evaluating whether sight is the most valued sense. *JAMA ophthalmology* 137, 11 (2019), 1317–1320.

[35] J. L. Finlayson and Chris Mellish. 2005. The 'audioview' - providing a glance at Java source code. In *Proceedings of the International Conference on Auditory Display (ICAD'O5)*. Georgia Institute of Technology.

[36] Christopher Frauenberger, Tony Stockman, and Marie-Luce Bourguet. 2007. A survey on common practice in designing audio in the user interface. In *Proceedings of HCI 2007 The 21st British HCI Group Annual Conference University of Lancaster, UK 21*. 1–9.

[37] Thomas M. Gable, Bruce N. Walker, Haifa R. Moses, and Ramitha D. Chitloor. 2013. Advanced auditory cues on mobile phones help keep drivers' eyes on the road. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 66–73.

[38] Eboni S. Gaddis, Lesley Strawderman, Harish Chander, Brian K. Smith, Charles Freeman, Courtney Taylor, et al. 2022. The impact of using wearable devices on the operator during manual material handling tasks. *International Journal of Industrial Ergonomics* 89 (2022), 103294.

[39] Nick Gang, Srinath Sibi, Romain Michon, Brian Mok, Chris Chafe, and Wendy Ju. 2018. Don't Be alarmed: Sonifying autonomous vehicle perception to increase situation awareness. In *Proceedings of the 10th international conference on automotive user interfaces and interactive vehicular applications*. 237–246.

[40] Franco E. Garcia and Vânia P. de Almeida Neris. 2013. Design guidelines for audio games. In *Human-Computer Interaction. Applications and Services: 15th International Conference, HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013, Proceedings, Part II 15*. Springer, 229–238.

[41] Stavros Garzonis, Simon Jones, Tim Jay, and Eamonn O'Neill. 2009. Auditory icon and earcon mobile service notifications: intuitiveness, learnability, memorability and preference. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1513–1522.

[42] William W. Gaver. 1987. Auditory icons: Using sound in computer interfaces. *ACM SIGCHI Bulletin* 19, 1 (1987), 74.

[43] William W. Gaver. 1989. The SonicFinder: An interface that uses auditory icons. *Human–Computer Interaction* 4, 1 (1989), 67–94.

[44] William W. Gaver. 1993. Synthesizing auditory icons. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*. 228–235.

[45] William W. Gaver. 1997. Auditory interfaces. In *Handbook of human-computer interaction*. Elsevier, 1003–1041.

[46] Generix Group. 2023. Can a Warehouse Management System Really Help Managers Retain Employees in the Warehouse? https://www.generixgroup.com/en/blog/can-warehouse-management-system-really-help-managers-retain-employees-warehouse

[47] Google. n.d.. Material design - Applying Sound to UI. https://m2.material.io/design/sound/applying-sound-to-ui.html#sound-use-cases

[48] Google. n.d.. Material design - Sound attributes. https://m2.material.io/design/sound/sound-attributes.html#tonality

[49] Richard Gould. 2016. Auditory Icons. https://designingsound.org/2016/11/24/auditory-icons/

[50] Robert Graham. 1999. Use of auditory icons as emergency warnings: evaluation within a vehicle collision avoidance application. *Ergonomics* 42, 9 (1999), 1233–1248.

[51] Stephen Handel. 1989. *Listening: An introduction to the perception of auditory events*. Cambridge, Mass.: MIT Press.

[52] Sandra G. Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[53] Sun Hong-Ying. 2009. The application of barcode technology in logistics and warehouse management. In *2009 First International Workshop on Education Technology and Computer Science*, Vol. 3. IEEE, 732–735.

[54] Edwin L. Hutchins, James D. Hollan, and Donald A. Norman. 1985. Direct manipulation interfaces. *Human–computer interaction* 1, 4 (1985), 311–338.

[55] Sarah J. Isherwood and Denis McKeown. 2017. Semantic congruency of auditory warnings. *Ergonomics* 60, 7 (2017), 1014–1023.

[56] Nadya A. Istiqomah, Putri F. Sansabilla, Doddy Himawan, and Muhammad Rifni. 2020. The implementation of barcode on warehouse management system for warehouse efficiency. In *Journal of Physics: Conference Series*, Vol. 1573. IOP Publishing, 012038.

[57] David Jacobs. 2022. How modern WMS permits workforce retention by empowering workers. https://www.generixgroup.com/en/blog/hold-tight-manufacturing-knowledge-retain-and-empower-your-workforce

[58] Shaliza Jumahat, Manjit S. Sidhu, and Sharulhizam M. Shah. 2022. Pick-by-vision of Augmented Reality in Warehouse Picking Process Optimization–A Review. In *2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*. IEEE, 1–6.

[59] Cengiz Kaygusuz and Julian Zuluaga. 2018. Impact of intervals on the emotional effect in western music. *arXiv preprint arXiv:1812.04723* (2018).

[60] Michelle E. Kiger and Lara Varpio. 2020. Thematic analysis of qualitative data: AMEE Guide No. 131. *Medical teacher* 42, 8 (2020), 846–854.

[61] Hannu Korhonen, Jukka Holm, and Mikko Heikkinen. 2007. Utilizing sound effects in mobile user interface design. In *Human-Computer Interaction–INTERACT 2007: 11th IFIP TC 13 International Conference, Rio de Janeiro, Brazil, September 10-14, 2007, Proceedings, Part I 11*. Springer, 283–296.

[62] Steven Landry, Myounghoon Jeon, Pasi Lautala, and David Nelson. 2019. Design and assessment of in-vehicle auditory alerts for highway-rail grade crossings. *Transportation research part F: traffic psychology and behaviour* 62 (2019), 228–245.

[63] James Lani. 2021. The assumption of homogeneity of variance. https://www.statisticssolutions.com/the-assumption-of-homogeneity-of-variance/

[64] Page Laubheimer. 2018. Beyond the NPS: Measuring Perceived Usability with the SUS, NASA-TLX, and the Single Ease Question After Tasks and Usability Tests. https://www.nngroup.com/articles/measuring-perceived-usability/

[65] Young S. Lee, Santosh Basapur, Harry Zhang, Claudia Guerrero, and Noel Massey. 2010. Usability evaluation of beep-to-the-box. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*. 345–348.

[66] Ming K. Lim, Witold Bahr, and Stephen C.H. Leung. 2013. RFID in the warehouse: A literature analysis (1995–2010) of its applications, benefits, challenges and future trends. *International Journal of Production Economics* 145, 1 (2013), 409–430.

[67] Arnold M. Lund. 2001. Measuring usability with the use questionnaire. *Usability interface* 8, 2 (2001), 3–6.

[68] Hem Malhotra. 2023. Leveraging Technology to Reduce Employee Turnover. https://news.sap.com/australia/2023/02/14/leveraging-technology-to-reduce-employee-turnover/#:~:text=The%20use%20of%20technologies%20such,a%20more%20positive%20work%20environment.

[69] David K. McGookin and Stephen A. Brewster. 2004. Understanding concurrent earcons: Applying auditory scene analysis principles to concurrent earcon recognition. *ACM Transactions on Applied Perception (TAP)* 1, 2 (2004), 130–155.

[70] Denis McKeown and Sarah Isherwood. 2007. Mapping candidate within-vehicle auditory displays to their referents. *Human factors* 49, 3 (2007), 417–428.

[71] Richard R. McNeer, Danielle B. Horn, Christopher L. Bennett, Judy R. Edworthy, and Roman Dudaryk. 2018. Auditory icon alarms are more accurately and quickly identified than current standard melodic alarms in a simulated clinical setting. *Anesthesiology* 129, 1 (2018), 58–66.

[72] Hokey Min. 2007. Examining sources of warehouse employee turnover. *International Journal of Physical Distribution & Logistics Management* (2007).

[73] Noé Monsaingeon, Loïc Caroux, Sabine Langlois, and Céline Lemercier. 2023. Earcons to reduce mode confusions in partially automated vehicles: Development and application of an evaluation method. *International Journal of Human-Computer Studies* 176 (2023), 103044.

[74] Elizabeth D. Mynatt. 1994. Designing with auditory icons: how well do we identify auditory cues?. In *Conference Companion on human factors in computing systems*. 269–270.

[75] Michael A. Nees and Eliana Liebman. 2023. Auditory icons, earcons, spearcons, and speech: A systematic review and meta-analysis of brief audio alerts in human-machine interfaces. *Auditory Perception & Cognition* (2023), 1–30.

[76] John G. Neuhoff. 2019. Is sonification doomed to fail. In *Proceedings of the International Conference on Auditory Display (ICAD'19)*. Georgia Institute of Technology, 327–330.

[77] Patrick Ng and Keith Nesbitt. 2013. Informative sound design in video games. In *Proceedings of the 9th Australasian conference on interactive entertainment: matters of life and death*. 1–9.

[78] Jakob Nielsen. 2020. 10 Usability Heuristics for User Interface Design. https://www.nngroup.com/articles/ten-usability-heuristics/

[79] Minyong Park. 2011. Evaluation of earcons for portable digital electronic products considering perceptual user characteristics. *Human Factors and Ergonomics in Manufacturing & Service Industries* 21, 5 (2011), 464–473.

[80] Roy D. Patterson. 1982. *Guidelines for auditory warning systems on civil aircraft*. Civil Aviation Authority.

[81] Nathan C. Perry, Catherine J. Stevens, Mark W. Wiggins, and Clare E. Howell. 2007. Cough once for danger: Icons versus abstract warnings as informative alerts in civil aviation. *Human factors* 49, 6 (2007), 1061–1071.

[82] Nur K.U. Pratiwi, Sella Monica, Rully Yandri, and Doddy Himawan. 2020. The use of voice technology to enhance the picking activity in warehouse. *Advances in Transportation and Logistics Research* 3 (2020), 691–694.

[83] Matthew S. Prewett, Linda R. Elliott, Ashley G. Walvoord, and Michael D. Coovert. 2011. A meta-analysis of vibrotactile and visual information displays for improving task performance. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42, 1 (2011), 123–132.

[84] A. Ramaa, K.N. Subramanya, and T.M. Rangaswamy. 2012. Impact of warehouse management system in a supply chain. *International Journal of Computer Applications* 54, 1 (2012).

[85] Matthias Rauterberg. 2019. Different effects of auditory feedback in man-machine interfaces. In *Human factors in auditory warnings*. Routledge, 225–242.

[86] Ryann Reynolds-McIlnay and Maureen Morrin. 2019. Increasing shopper trust in retailer technological interfaces via auditory confirmation. *Journal of Retailing* 95, 4 (2019), 128–142.

[87] Dimitris Rigas and Ibrahim Bahadur. 2006. The use of audio in a stock control application. In *Proceedings of the 10th WSEAS international conference on Computers*. 1172–1177.

[88] Dimitrios Rigas and Antonio Ciuffreda. 2007. An empirical investigation of multimodal interfaces for browsing internet search results. In *Proceedings of the 7th WSEAS International Conference on Applied Informatics & Communication*. Citeseer, 194–199.

[89] Agnieszka Roginska. 2013. Auditory icons, earcons, and displays: Information and expression through sound. *The psychology of music in multimedia* 339 (2013).

[90] Edin Šabić, Daniel Henning, and Justin MacDonald. 2019. Adaptive Auditory Alerts for Smart In-Vehicle Interfaces. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 63. SAGE Publications Sage CA: Los Angeles, CA, 1545–1549.
[91] Kim Salazar. 2020. Contextual Inquiry: Inspire Design by Observing and Interviewing Users in Their Context. https://www.nngroup.com/articles/contextual-inquiry/#:~:text=Contextual%20inquiry%20is%20a%20type,of%20work%20practices%20and%20behaviors.
[92] Jeff Sauro, PhD. 2011. Measuring Usability with the System Usability Scale (SUS) – MeasuringU. https://measuringu.com/sus/
[93] Constantine Stephanidis. 2001. User interfaces for all: New perspectives into human-computer interaction. *User interfaces for all-concepts, methods, and tools* 1, 1 (2001), 3–17.
[94] Denise A. Sumikawa. 1985. *Guidelines for the integration of audio cues into computer user interfaces.* Technical Report. Lawrence Livermore National Lab., CA (USA).
[95] B.S.S. Tejesh and S. Neeraja. 2018. Warehouse inventory management system using IoT and open source framework. *Alexandria engineering journal* 57, 4 (2018), 3817–3823.
[96] Shaun Turney. 2023. Central Limit Theorem | Formula, Definition 038; Examples. https://www.scribbr.com/statistics/central-limit-theorem/#:~:text=The%20central%20limit%20theorem%20says,the%20mean%20will%20be%20normal.
[97] Bruce N. Walker and Gregory Kramer. 2004. Ecological psychoacoustics and auditory displays: Hearing, grouping, and meaning making. In *Ecological psychoacoustics*. Brill, 149–174.
[98] Bruce N. Walker, Jeffrey Lindsay, Amanda Nance, Yoko Nakano, Dianne K. Palladino, Tilman Dingler, and Myounghoon Jeon. 2013. Spearcons (speech-based earcons) improve navigation performance in advanced auditory menus. *Human Factors* 55, 1 (2013), 157–182.
[99] Ina Wechsung and Anja B. Naumann. 2009. Evaluating a multimodal remote control: The interplay between user experience and usability. In *2009 International Workshop on Quality of Multimedia Experience*. IEEE, 19–22.
[100] György Wersényi. 2009. Auditory representations of a graphical user interface for a better human-computer interaction. In *International Symposium on Computer Music Modeling and Retrieval*. Springer, 80–102.

A Appendix - Semi-structured interview - EN

Sub-question to be answered:

In what way should enhanced auditory feedback support scan processes on mobile devices in Warehouse Management Systems compared to current situation?

General questions

- Can you describe how you currently use the Rentman Mobile App in your warehouse processes?
- How does the scanning process differ between barcode scanning and RFID scanning?
- How is it currently clear to you that you are scanning with barcode and RFID?
- Can you describe your current experience with the app when using it for your warehouse processes?

Current auditory feedback

- What are your current experiences regarding auditory feedback during barcode scanning?
- What are your current experiences regarding auditory feedback during RFID scanning?
- How much do you rely on this auditory feedback?
- Can you describe any specific challenges or limitations you have encountered with the current auditory feedback during the scan processes in your warehouse?
- Can you identify any specific events where the current auditory feedback system has led to confusion or misunderstanding during scan processes?
- What do you think of the auditory feedback currently added to the error messages in Rentman?

Improved auditory feedback

- During which scan processes would you expect improved auditory feedback? Does this differ between barcode and RFID scanning?
- How do you think enhanced auditory feedback can improve your scan processes compared to the current situation?
  - Do you see any potential benefits regarding error reduction, task completion time, or overall user experience?
- What timing and frequency of auditory feedback do you think would be most effective in supporting the scan processes?
- In your opinion, what types of sounds would be most helpful in enhancing the barcode scan processes on mobile devices in a Warehouse Management System?
- In your opinion, what types of sounds would be most helpful in enhancing the RFID scan processes on mobile devices in a Warehouse Management System?
- How can the auditory feedback of error messages be improved? Should there be more distinction?
- Are there any concerns or potential drawbacks you anticipate with the integration of enhanced auditory feedback during scan processes on mobile devices?
- What benefits can you foresee with improved auditory feedback and do you see the added value of these?
- Can you think of any specific recommendations or suggestions for designing and implementing enhanced auditory feedback to optimize the scan processes on mobile devices in a Warehouse Management System?

Semi-structured interview - NL

Subvraag om te beantwoorden:

Op welke manier moet verbeterde auditieve feedback scanprocessen op mobiele apparaten in Warehouse Management Systemen ondersteunen in vergelijking met de huidige situatie?

Algemene vragen

- Kan je beschrijven hoe je momenteel werkt met de Rentman Mobiele App in jouw magazijn processen?
- Hoe verschilt het scanproces tussen barcode en RFID scannen?
- Hoe is het voor jou momenteel duidelijk dat je aan het scannen bent met barcode en RFID?
- Kan je jouw huidige ervaring omschrijven met de app als het aankomt op de processen in het magazijn?

Huidige auditieve feedback

- Wat zijn jouw huidige ervaringen met betrekking tot auditieve feedback tijdens het scannen met barcodes?
- Wat zijn jouw huidige ervaringen met betrekking tot auditieve feedback tijdens het scannen met RFID?
- In hoeverre vertrouw je op deze auditieve feedback?
- Kan je specifieke uitdagingen of beperkingen beschrijven die je bent tegengekomen met de huidige auditieve feedback tijdens de scanprocessen in het magazijn?
- Kan je bepaalde acties of situaties noemen waar de auditieve feedback tot verwarring of onbegrip heeft geleid?
- Wat vind je van de huidige auditieve feedback van foutmeldingen in Rentman?

Verbeterde auditieve feedback

- Wanneer zou je verbeterde auditieve feedback verwachten in het scanproces? Verschilt dit tussen barcode en RFID scannen?
- Hoe denk je dat verbeterde auditieve feedback jouw scanproces kan verbeteren ten opzichte van de huidige situatie?
  - Zie je enkele voordelen wat betreft minder fouten, sneller klaar zijn met je taak, of de algemene gebruikerservaring?
- Welke timing en frequentie van deze auditieve feedback denk je dat het meest effectief is tijdens het scanproces?
- Wat voor soort geluiden zouden het nuttigst zijn tijdens het scannen met de barcode op een mobiele scanner?
- Wat voor soort geluiden zouden het nuttigst zijn tijdens het scannen met RFID op een mobiele scanner?
- Hoe kan de auditieve feedback van foutmeldingen verbeterd worden denk je? Moet hier meer onderscheid in komen?
- Heb je momenteel bepaalde zorgen of nadelen die je kan voorzien met deze verbeterde auditieve feedback tijdens de scanprocessen?

- Welke voordelen kan je voorzien met verbeterde auditieve feedback en zie je hier de meerwaarde van in?
- Heb je specifieke aanraders/suggesties voor het designen en implementeren van de verbeterde auditieve feedback om het scanproces te optimaliseren?

B Appendix - Contextual inquiry - EN

Sub-question to be answered:

How do users perceive and interpret the earcons of the different actions required in the scan processes of a Warehouse Management System?

Part 1 - The Primer

- Introduction
- Express what hopefully will be achieved
- Discuss confidentiality
- Shortly mention the topic already

Part 2 - The Transition

- Explicitly state the transition from the introduction to the almost contextual interview.
- Explain what will happen and how interruptions work

Part 3 - The Contextual Interview

Useful questions for both during and after the think-aloud of the sound set to evaluate the sounds in that set:

- Why do you take this action?
- How did you interpret the auditory feedback during this task?
- Is it clear what must happen right now?
- How does this sound make you feel?
- Could you distinguish this sound from the previous one?
- Did the auditory feedback effectively guide you through the task? Why or why not?
- How would you describe the characteristics of the sound you heard?
- What is your opinion about this sound set in general?
- What did you like about this sound set?
- What did you dislike about this sound set?
- Were there any sounds that stood out to you? Both in a negative or positive way?
- Were there any sounds difficult to perceive?
- Were there sounds too similar to each other?
- Were there instances when you misunderstood or misinterpreted the auditory feedback?
- Based on your current experience, do you have any suggestions for improving certain sounds?

Part 4 - The Wrap-up

Now the final semi-structured interview is performed. Below are the questions used during this interview:

General feedback

- What did you think of the overall application?
- What did you think of the amount of new sounds? Too many, just right, or too few?
- Do you think the usability of the application has increased with this new auditory feedback?
- Would you expect to work more efficiently with this new auditory feedback?
- Would you prefer to configure these sounds yourself in the future?

QR scan sounds

- Which QR code scan sound did you like the most? Why?
- What do you find important about the sound of a QR code scan?
- Would you like to use one of these sounds permanently in the future?
- Do you have any improvements left for your preferred QR code scan sound?

RFID scan sounds

- What do you think of the four moments where sounds are implemented to RFID right now? So the start sound, sonar sound, tag sound, and cluster sound?
- Were these four moments of auditory feedback useful for you during the scan process?
- Would other actions/moments require additional auditory feedback too?
- Which RFID start sound did you prefer? Why?
- Which RFID tag sound did you prefer? Why?
- Which RFID sonar sound did you prefer? Why?
- Which RFID cluster sound did you prefer? Why?
- What do you find important about the sound during RFID scanning?
- Do you think your user experience got better with this new auditory feedback?
- Would you like to use these sounds permanently in the future?

Error/Pop-up sounds

- What is your opinion about the fact that there are now three different sounds for pop-ups?
- Was there a clear distinction between the different pop-ups?
- Which sound set of error messages/pop-ups did you prefer most? Why?
- Would you like to change things from that set?
- Did you feel that certain sounds drew your attention more than others?
- Was the distinction between the different notifications clear to you?
- Did you feel that after hearing the sound, you knew what was on your screen?
- Did these sounds improve your user experience?

Contextual onderzoek - NL

Subvraag om te beantwoorden:

Hoe zien en interpreteren gebruikers de verbeterde earcons en auditieve iconen van de verschillende acties die nodig zijn in de scanprocessen van een Warehouse Management Systeem

Deel 1 - De Inleiding

- Introductie
- Vertellen wat we hopelijk zullen behalen
- Bespreek vertrouwelijkheid
- Benoem kort het onderwerp

Deel 2 - De Overgang

- Benadruk duidelijk dat dit de overgang is van de introductie naar bijna het contextueel interview
- Leg uit wat er gaat gebeuren en hoe onderbrekingen werken

Deel 3 - Het Contextueel Interview

Een aantal vragen die nuttig kunnen zijn tijdens dit proces:

- Waarom neem je deze actie?
- Hoe interpreteerde je het geluid tijdens deze taak?
- Is het duidelijk wat er nu moet gebeuren?
- Wat voor gevoel krijg je bij dit geluid?
- Kan je dit geluid goed onderscheiden van het vorige geluid?
- Heeft de auditieve feedback je efficiënt door de taak begeleid? Waarom wel of waarom niet?
- Hoe zou je de kenmerken van de geluiden beschrijven die je zojuist hebt gehoord?
- Wat is jouw mening over deze geluidsset in algemene zin?
- Wat vond je goed aan deze geluidsset?
- Wat vond je minder goed aan deze geluidsset?
- Waren er geluiden die er voor jou uitsprongen? In zowel een positieve als negatieve zin.
- Waren er bepaalde geluiden moeilijk waar te nemen?
- Waren er geluiden die te veel op elkaar leken?
- Waren er gevallen waar je de auditieve feedback niet begreep of verkeerd interpreteerde?
- Gebaseerd op jouw huidige ervaring, heb je suggesties om bepaalde geluiden te verbeteren?

Deel 4 - De Afronding

Nu wordt het laatste gedeelte van het semi-structured interview gehouden. Hieronder staan de vragen van dit interview:

Algemene feedback

- Wat vind je van de applicatie in algemene zin?
- Wat vind je van de hoeveelheid nieuwe geluiden? Is dat te veel, goed, of te weinig?
- Denk je dat de bruikbaarheid van de applicatie is verbeterd met deze nieuwe auditieve feedback?
- Verwacht je efficiënter te kunnen werken met deze nieuwe audio feedback?
- Zou je het fijn vinden zelf de geluiden te kunnen configureren in de toekomst?

QR scan geluiden

- Welk QR code scan geluid vond je het fijnst? Waarom?
- Wat vind je belangrijk aan het geluid van een QR code scan?
- Zou je een van deze geluiden permanent willen gebruiken in de Rentman app?
- Heb je verbeterpunten voor het QR code scan geluid?

RFID Scannen

- Wat vind je van de 4 momenten waarop er nu RFID geluiden zijn? Dus de start, sonar, tag scan en het cluster geluid?
- Zijn deze 4 momenten van auditieve feedback nuttig geweest voor jou?
- Zou er nog op andere plekken auditieve feedback moeten komen?
- Welk RFID start geluid vond je het fijnst? Waarom?
- Welk RFID tag geluid vond je het fijnst? Waarom?
- Welk RFID sonar geluid vond je het fijnst? Waarom?
- Welk RFID cluster geluid vond je het fijnst? Waarom?
- Wat vind je belangrijk aan de geluiden tijdens RFID scannen?
- Vind je dat jouw gebruikerservaring is verbeterd met deze nieuwe feedback?
- Zou je deze geluiden permanent willen gebruiken in de Rentman app?

Error/Pop-up

- Wat vind je van het feit dat er drie verschillende meldingen zijn voor pop-ups?
- Was het onderscheid tussen de verschillende pop-ups voor jou duidelijk?
- Welke set van pop-ups/foutmeldingen vond je het fijnst? Waarom?
- Zou je dingen anders willen zien aan die set?
- Had je het gevoel dat bepaalde geluiden jouw aandacht meer trokken dan andere?
- Was het onderscheid tussen de verschillende meldingen voor jou duidelijk?
- Kreeg je het gevoel dat je na het horen van het geluid wist wat er op jouw scherm te zien was?
- Hebben deze geluiden jouw gebruikerservaring verbeterd?

C Appendix - NASA-TLX

Name:
Condition:

Mental Demand How mentally demanding was the task?

Very Low Very High

Physical Demand How physically demanding was the task?

Very Low Very High

Temporal Demand How hurried or rushed was the pace of the task?

Very Low Very High

Annoyance How annoyed were you?

Very Low Very High

Performance How successful were you in accomplishing what you were asked to do?

Very Low Very High

Effort How hard did you have to work to accomplish your level of performance?

Very Low Very High

Frustration How insecure, discouraged, irritated, and stressed were you?

Very Low Very High

D Appendix - SUS questionnaire

Instructions

For each of the following statements, please indicate your level of agreement by circling the appropriate number on the scale from 1 to 5, where 1 means "Strongly Disagree" and 5 means "Strongly Agree".

Questionnaire

(1) I think that I would like to use this application with auditory feedback frequently.
(2) I found the application with auditory feedback unnecessarily complex.
(3) I thought the application with auditory feedback was easy to use.
(4) I think that I would need the support of a technical person to be able to use this application with auditory feedback.
(5) I found the various functions in this application with auditory feedback were well integrated.
(6) I thought there was too much inconsistency in this application with auditory feedback.
(7) I would imagine that most people would learn to use this application with auditory feedback very quickly.
(8) I found the application with auditory feedback very cumbersome to use.
(9) I felt very confident using the application with auditory feedback.
(10) I needed to learn a lot of things before I could get going with this application with auditory feedback.

E Appendix - Results Data Distribution

E.1 Shapiro-Wilk

E.1.1 Shapiro-Wilk Dependent Variables

Table 7: Shapiro-Wilk for all dependent variables (DV) with all conditions.

DV ->	Efficiency		Effectiveness		TLX		Annoyance		SUS	
	Statistic	Sig.	Statistic	Sig.	Statistic	Sig.	Statistic	Sig.	Statistic	Sig.
BC	.976	.539	.789	<.001	.963	.199	.915	.005	.987	.001
BA	.951	.073	.798	<.001	.961	.173	.910	.003	.916	.005
RC	.978	.596	.710	<.001	.955	.107	.931	.015	.955	.101
RA	.962	.183	.668	<.001	.935	.021	.884	.001	.896	.001

E.1.2 Shapiro-Wilk TLX Factors

Table 8: Shapiro-Wilk for all TLX factors with all conditions.

TLX Factor	Mental demand		Physical demand		Temporal demand		Performance		Effort		Frustration	
	Statistic	Sig.	Statistic	Sig.	Statistic	Sig.	Statistic	Sig.	Statistic	Sig.	Statistic	Sig.
BC	.786	<.001	.835	<.001	.927	.012	.876	<.001	.921	.008	.902	.002
BA	.881	<.001	.883	<.001	.922	.008	.884	<.001	.920	.007	.878	<.001
RC	.899	.002	.861	<.001	.911	.004	.897	.001	.923	.009	.894	.001
RA	.907	.003	.857	<.001	.932	.017	.880	<.001	.942	.037	.865	<.001

E.2 Histograms

E.2.1 Efficiency

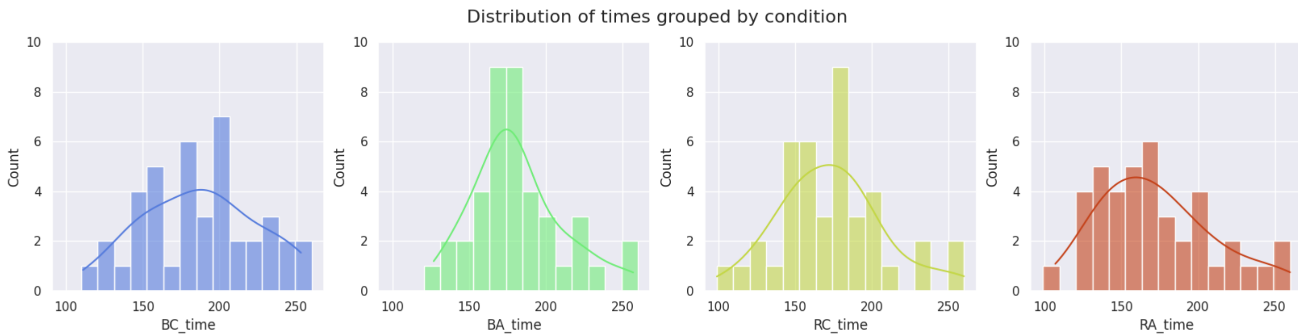

Figure 31: Histogram of the distribution of the efficiency data.

E.2.2 Effectiveness

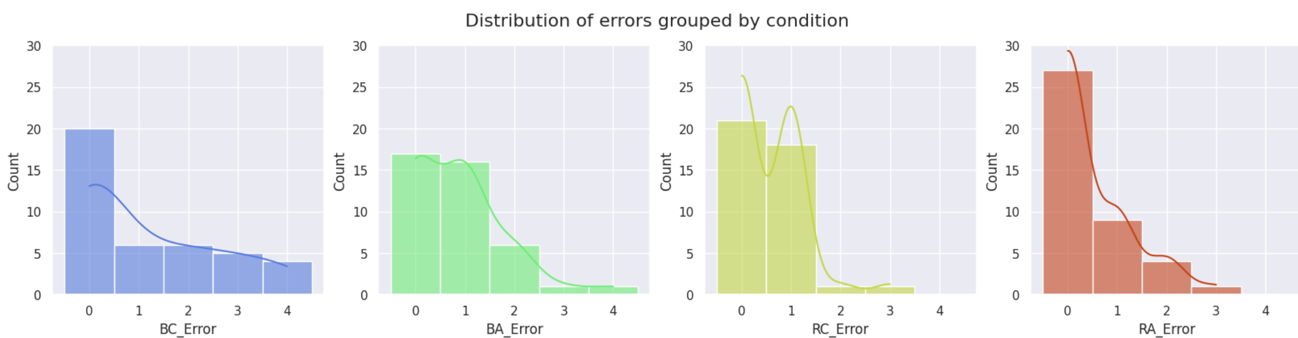

Figure 32: Histogram of the distribution of the effectiveness data.

E.2.3 Perceived Workload (TLX)

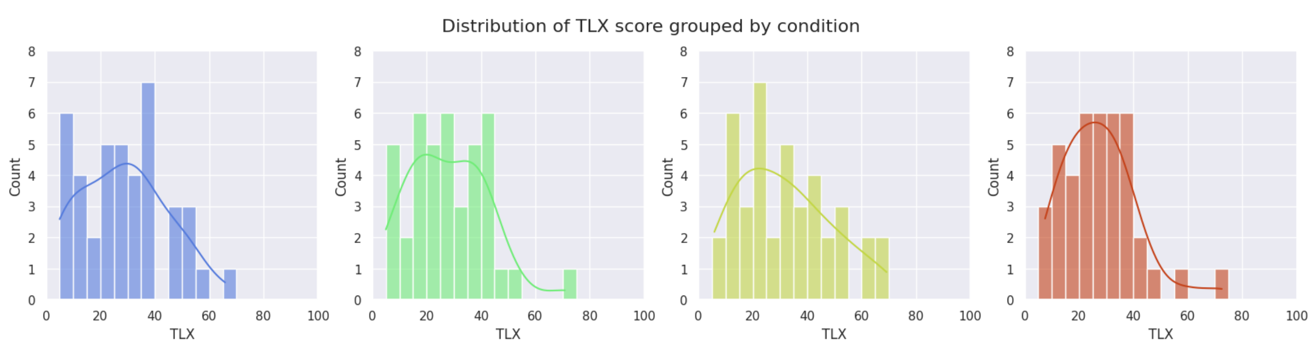

Figure 33: Histogram of the distribution of the TLX score data.

E.2.4 Annoyance

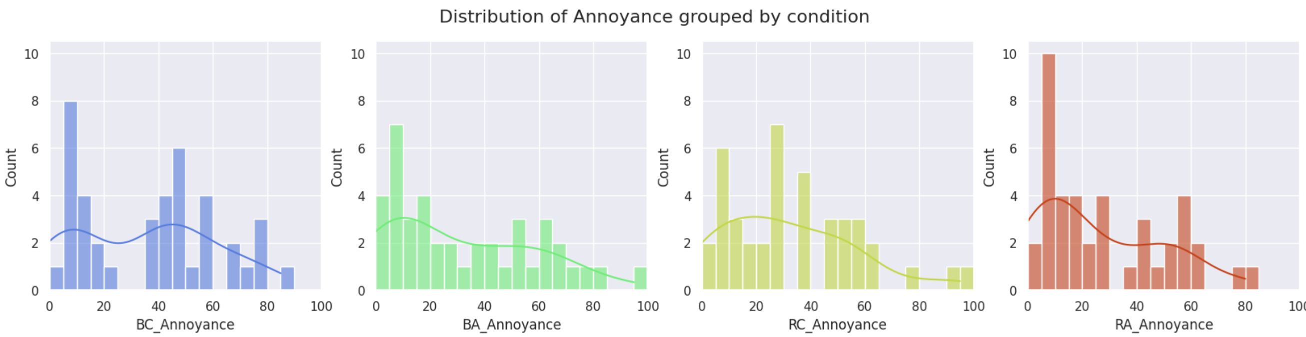

Figure 34: Histogram of the distribution of the annoyance data.

E.2.5 Subjective user rating (SUS)

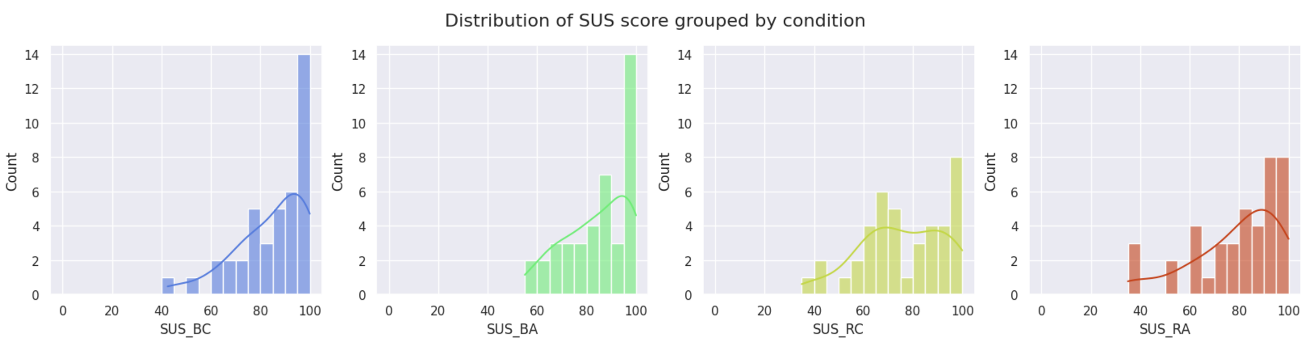

Figure 35: Histogram of the distribution of the SUS data.

E.3 Interaction plots

E.3.1 Efficiency

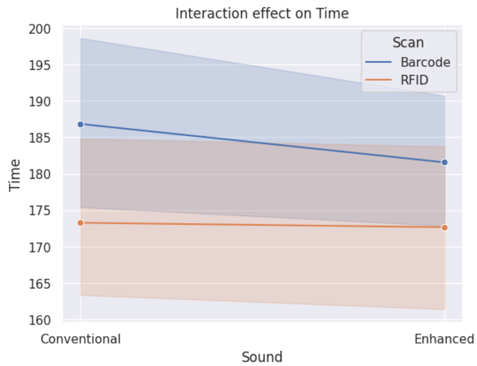

Figure 36: Interaction plot of efficiency.

E.3.2 Effectiveness

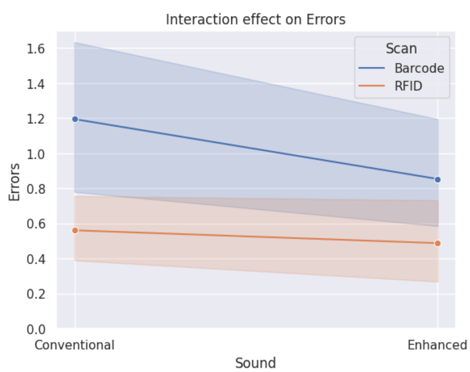

Figure 37: Interaction plot of effectiveness.

E.3.3 Perceived Workload (TLX)

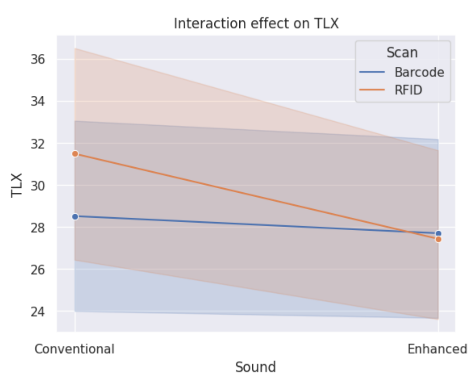

Figure 38: Interaction plot of TLX score.

E.3.4 Annoyance

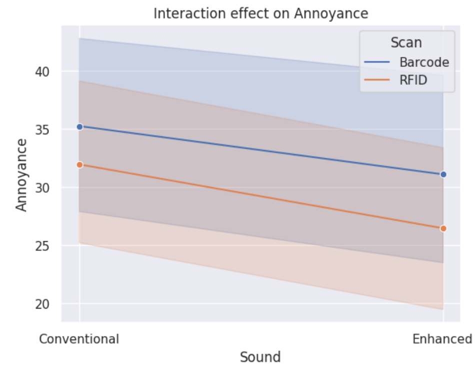

Figure 39: Interaction plot of annoyance.

E.3.5 Subjective user rating (SUS)

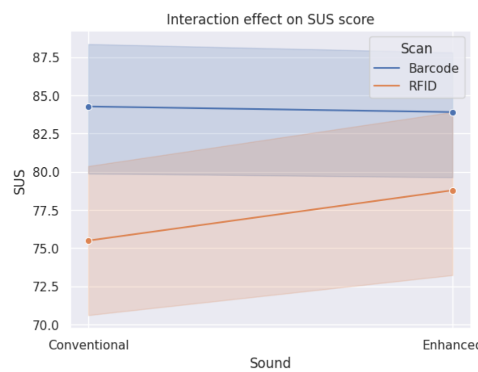

Figure 40: Interaction plot of SUS score.

E.4 Learning effect

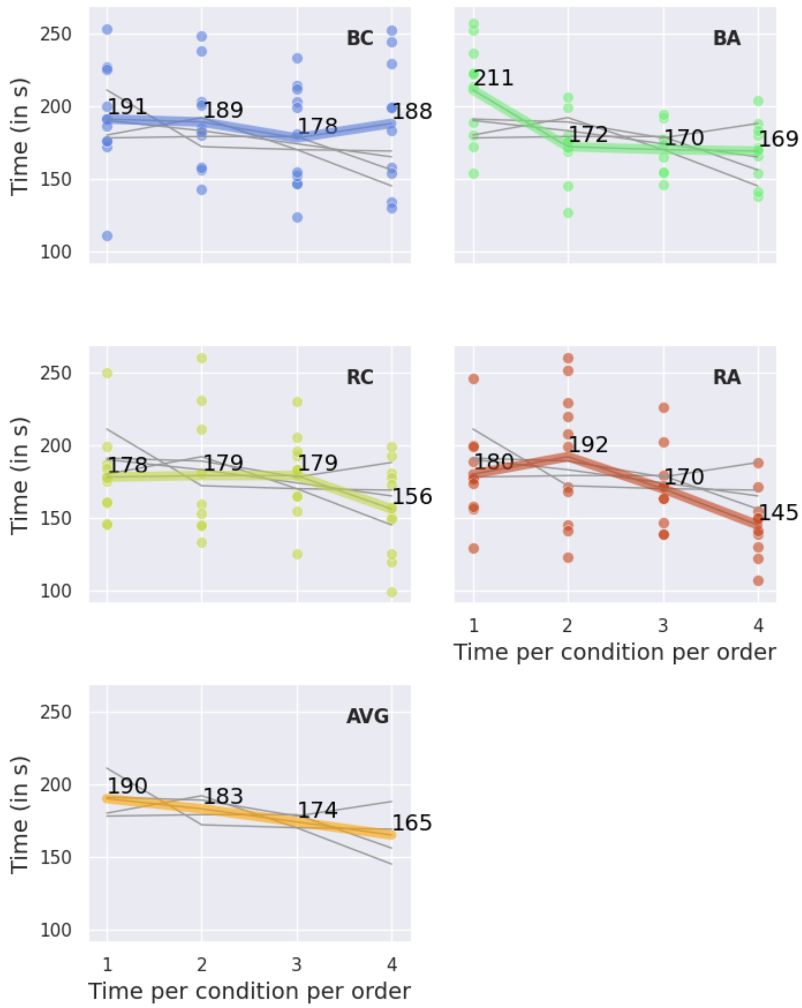

Figure 41: Learning effect during the experiment. Shows average time when a condition was performed first, second, third, or fourth in the experiment.