# Industrial Cobot Sound Design Study

Audio Design Principles for Industrial Human-Robot Interaction


Master Thesis Marvin Resing

DPM2061


A thesis submitted to the faculty of Engineering Technology at the University of Twente, for the degree of Master of Sciences, in the Industrial Design Engineering program, for the Human Technology Relations sub-track.

Augsburg / Enschede

28.11.2023


UNIVERSITY OF TWENTE.

Professor:
Daniel Saakes

2nd Assessor:
Sebastian Thiede

Company Supervisors:
Anja Riedelsheimer
Nadine Reißner

KUKA


# Abstract

Marvin Resing: Industrial Cobot Sound Design Study - Audio Design Principles for Industrial Human-Robot Interaction

(Under the Direction of Daniel Saakes)

In the field of industrial robotics, the new class of collaborative robots, or cobots, has emerged, which is intended to allow for closer interactions. With their improved safety features, intuitive user-centered interfaces, and adaptability to flexible use-cases, developments in cobot technology could redefine interaction principles not only for collaborative work across various industries but also our general interactions in the emerging field of robots for various applications. In this paper, the current requirements caused by these new developments will be examined, particularly regarding communication and interaction through audio with cobot users in industrial settings. From the use of naturally occurring sounds and how they are used in these environments already, insights can be gained, regarding the possibilities of audio. Also, with some insights into the underlying principles and characteristics of sound generation, transmission and perception, conclusions can be drawn on how to employ it effectively. Together with a view on requirements and needs being introduced from the interaction with these new collaborative systems, their varied user group, and the challenging environmental conditions, a broad perspective on possibilities for audio augmentation can be gained. Considering sonification and mapping methods for representing information through sound effectively, as well as the relevant principles and scenarios found for designing audio solutions, concrete prototype sounds are created. These are then used to evaluate, validate, and improve the underlying principles accordingly. A user study in which users and experts can react to the prototype signals investigates the perception of the sounds. A semi-structured interview is included in this, to gain deeper, uninfluenced insights into their understanding, preferences, and the reasons for those. This feedback is then used for a final iteration of the underlying principles. These final guidelines were related to factors like representing and using the natural characteristics of sound, musical principles, sonification techniques and further factors influencing the user experience. Also, a need to match the auditive communication to other modalities was established. The final principles aim to guide the development of audio for current and future cobot interactions.

# Industrial Cobot Sound Design Study

## Table of Contents

# Glossary

| Term | Definition |
| --- | --- |
| Auditory Channel | Refers to the human capacity to hear, process, and interpret sounds, as opposed to for example the visual channel. It includes the entire auditory system, from the ear to the perception in the brain. |
| Auralize | To represent something in audible form. Similar to how visualize is used for making something perceivable to the eye, creating sound or auditory representations of events, actions, or processes that are not inherently associated with sound to make them audible. Auralization is not limited to data representation like sonification; it will be used in this thesis as the general broader concept of using sound to convey something in various contexts. |
| Cobot | Collaborative robots, commonly known as cobots, are a type of industrial robot designed to work safely and interactively alongside human workers. They allow for simpler human-robot collaboration, through beginner-friendly programming methods, and enable tasks that require both human dexterity and robot precision or force. They are differentiated from traditional industrial robots mainly by their inherent safety features like force limiting, which allow them to work alongside humans without requiring any safety barrier. |
| Handguiding (Mode) | A way of interacting with the robot employed on a variety of industrial robots, and especially cobots. It entails moving the arm using your hands to bring it into position, and then saving the motion or position to use in further programming. It simplifies this process, and makes it more intuitive compared to manually adding coordinates. |
| HCI | Human Computer Interaction |
| HMI | Human Machine Interaction |
| HRC | Human Robot Collaboration |
| Sonify | Similar to Auralize it means the representation of data in the form of sound. However, in this thesis it will be used in a bit stricter way, relating to the academic concepts of sonification techniques. |
| Sound-concept | The overall audible design of something like a product. Mostly designed sounds in many cases but will also contain any other perceivable sounds related to it like consequential, natural sounds generated by its components or materials. |
| Soundscape | Refers to the overall acoustic environment of a particular space or setting, as it can be heard. It includes all kinds of sounds from various natural or artificial sources and can also be affected by the environment through effects like echo. |

# 1 Introduction

Collaborative robots, also referred to as 'cobots', are designed to work more closely alongside humans in a shared workspace. Unlike in traditional industrial robotics, they are intended to interact directly with human workers, which puts additional importance on the quality of these interactions. Ensuring effective and intuitive communication between humans and cobots is not just a matter of efficiency but also of safety. With cobots being designed to be easily and flexibly operatable, even for untrained users, new perspectives on the interaction and communication also become more relevant, to be intuitively understood.

## 1.1 The Current State of Human Robot Collaboration

Cobots represent a relatively recent advancement in the history of automation and industrial robotics, offering new ways of employing robotics. While many established robotics manufacturers as well as new competitors are actively engaged in the development these collaborative systems, the trajectory of the field is still slightly unclear. While some workflows and usability related features have already been successfully implemented, the widespread adoption of cobots in the industry is still slowly ongoing, and the full potential and promises of collaborative industrial work are still not realized (Zaatari et al., 2019; Michaelis et al., 2020).

Interaction improvements coming from these developments out of the necessity to make the close collaboration safe for users, might partially also find their way into other industrial robotics applications, and the wider field of Human Robot Interaction (HRI). This field is still in its earlier stages and thus malleable. With improved safety features, intuitive user centered communication and interfaces, and natural adaptability to flexible situations, developments in the field of cobots could redefine interaction principles not only for collaborative work across various industries but also our general interactions in the emerging field of robotics for various applications.

After already being an established actor in the development and production of industrial robots, KUKA entered the cobot market with the LBR iiwa ("intelligent industrial work assistant"), their first industrially applicable, HRC capable robot. This model allowed flexible use for a range of different locations and applications. The LBR iisy followed as their second iteration of cobots, and together with the new robot operating system iiQKA, the company already made important steps towards a user-friendly setup and programming experience.

While traditional robots used to be kept away from human access behind large fences, to comply with the relevant safety regulations, the development of cobots aims to reshape this approach, allowing humans to work closely together with them. This can be especially advantageous to small and medium-sized enterprises (SMEs), as cobots offer increased flexibility and simplified setup in terms of required space, as well as necessary employee expertise regarding installation, safety, and programming. These advantages could make cobots a viable option even when space or robotics knowledge is limited, so making them be available and suited to the needs of these new customer groups is one of the main development goals.

The existing body of research has delved into the general application of cobots in the industry, as well as technical aspects of cobots, like their safety features, but have particularly pointed out the need for considering interactions and communication in these contexts (Zaatari et al., 2019). In this regard, especially with a focus on sound, Alexanderson (2004) looked at how existing sounds are used in the industrial context, and even the sounds a cobot naturally makes and how they influence the perception of it were researched

(Tennent et al., 2017). And while generally auditive communication is also studied in the wider HCI field, including for example methods on sonification (Barrass, 1997; Henkelmann, 2007; Stienstra et al., 2011), the application of such methods in this context, fit to support the new interaction scenarios with cobots still opens a range of needed research (Bergman et al., 2019; Alexanderson, 2004).

In the frame of this thesis, the changed requirements caused by these new developments will be examined, particularly regarding audio communication and interaction with cobots. Further, the available options, especially regarding the possibly more elaborate and interactive scenarios in the future, will be looked at, to ensure a future-proofed audio communication concept for this field.

For interfacing with technology, the focus of the research has mainly been on visual methods, as also notes by Macaulay & Crerar (1998), and we have grown accustomed to using especially screens to transfer information. Especially with our ability to read and our general reliance on sight in life, this might not be surprising. High densities of information can easily be presented on a display especially in textual form, or indications can be given with something as simple as an LED. However, the other senses are of course not negligible, and are an important part of our natural experience. This can also be seen in the study of how the rich soundscape of natural noises in the workplace is already being used, allowing for many useful conclusions about relevant processes (Alexanderson, 2004).

In the light of making cobots easier to use and understand for a wide range of users, this might offer valuable support. Not only does this open up another communication channel in addition to the visual one, but audio also offers some unique advantages, which will be further explored and applied in this thesis.

While interaction with cobots and other technologies generally has been studied, and also the effective design of auditive communication for general contexts has seen extensive research, the intersection of this seems still less explored. This might also be the case as new technological developments and the system of cobots have only recently allowed for, or necessitated such more elaborate means of communication.

With the generally more prevalent and closer collaboration in these systems, the need for effective communication becomes more apparent, and the workers side of this interaction, especially in terms of perception, interpretation, and safety implications, needs further research. Addressing this gap can be a crucial part for the safe and efficient integration of cobots into workspaces. As industries continue to adopt cobots, understanding the nuances and principles of audio communication will be of growing importance in ensuring smooth human-cobot collaborations.

## 1.2 Purpose/ Research Question

The soundscape in industry settings has been shown to be quite information rich to operators that grew accustomed to it. Furthermore, using the auditory channel for certain signals can be much less demanding in terms of attention than for example visuals, especially text. While there is already a large amount of noise in such settings, the signals appropriate to the personal situation can often still be identified. Some phenomenological analysis of subtle naturally occurring sounds and their meaning to workers has already been done (Alexanderson, 2004), however implementing the principles into designed signals is not as straightforward as copying them. Firstly, appropriate situations must be identified in which they can deliver meaningful information, and secondly a recognizable, representative sound design needs to be chosen. While sounds that also occur naturally in the setting can have the advantage of being easily understandable and relatable to the field, the artificial version likely will not come close in richness initially. The option of freely synthesizing sounds, however, can be used to encode additional information, as well as giving a natural distinction between the sources, and still being noticeable in the big and complex soundscape of an industrial setting.

For these reasons, and in coordination with the development directions of KUKA, a number of goals for this thesis are made out:

## Goals:

- Worker understanding for safety
- Enhanced worker understanding of the cobot for confident and efficient work approach
- Natural collaborative user experience
- Fast learning/onboarding for inexperienced users
- Fast & easy error Handling
- Problem avoidance (predictive)
- Feeling and understanding of progress
- Customizability to specific working situation, while keeping safe recognizability

Additionally, to further guide the research and development activities, considering the knowledge gathered so far, some preliminary research questions and hypotheses were formulated:

## Research Questions

- How can the auditory channel be used effectively to improve understanding of the robot's internal processes?
- What situations in a typical industrial setting (using collaborative robotics) can be suitable to augmentation with auditive signals?
- How would an effective sound design for these applications and cobots in general be executed, to allow for…
    - meaningful, information-rich signals,
    - that are still able to blend in without distraction,
    - But are recognizable at the right moments to the right people,

…and are understandable and relatable to the cause and meaning of it, even when not learned explicitly?

## Research Hypotheses

H1: Using sound signals of different levels can allow for a richer communication of messages to the worker without information overload.

H2: Using distinguishable groups of sound, that are related to the category of the message conveyed can allow a bigger amount of easily recognizable signals for novice users aswell as more acquainted experts.

H3: The understanding of the robots' internal processes, as provided through these audio signals, leads to:

- higher safety
- Improved efficiency
- Improved satisfaction through a sense of progress and control

…during collaborative task execution

The study stands out in its focus on the cognitive aspects, especially of intentionally designed audio communication in collaborative industrial robotics, which still poses a gap in current research. By exploring the fundamental principles of sound and sonification, and the experiential perspective of workers using such solutions for the communication with cobots, this research aims to provide insights that will support the development of successful, user-friendly audio interactions with cobots in the workplace. By delivering not only some concrete concept sound solutions, but also analyzing the underlying working principles, this research aims to guide cobot manufacturers in designing more intuitive and safer audio communication systems for cobots, enhancing the efficiency and safety of human-cobot collaborations.

## 1.3 Thesis Structure

The thesis focusses on a holistic analysis of the possibilities of cobot audio communication in the industrial context. Bringing together the underlying principles of the use of audio, and the needs emerging from the new system and context, it aims to craft a natural user experience for cobot operators. Considering the current academic research state, as well as a grounded understanding of the challenges in real application, it aims to offer well-fitting solutions making best use of the possibilities, and capabilities of humans and cobots alike.

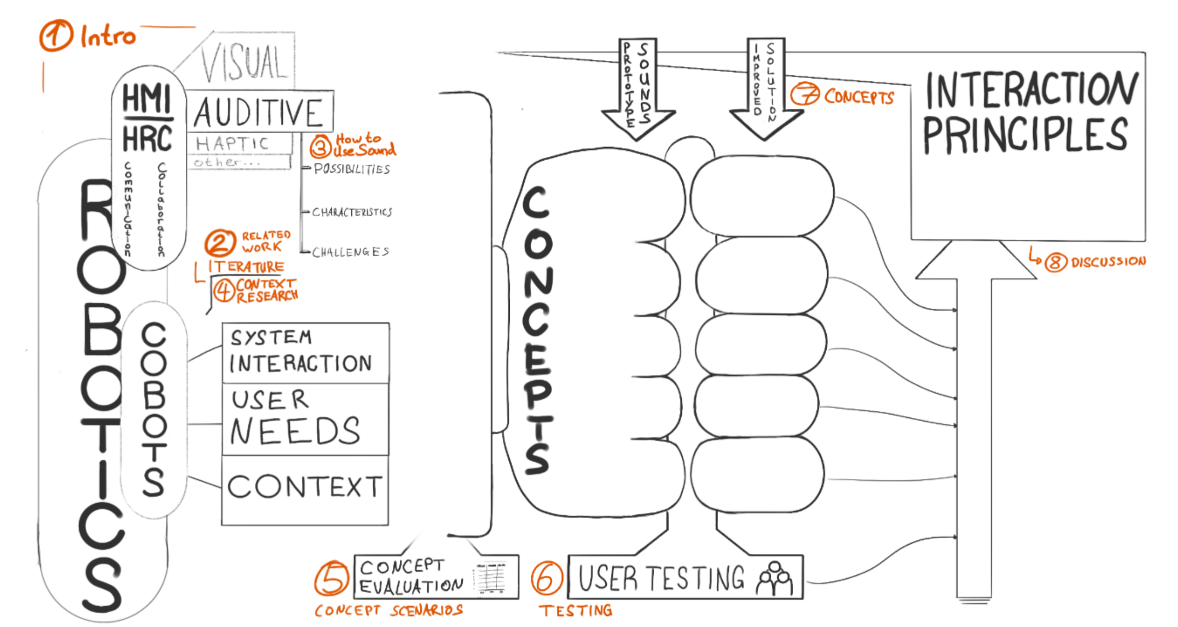



*Figure 1: Thesis Structure Design Process Flowchart*

The first chapter serves as an **Introduction** to the topic of cobots, their emergence, significance, and the context of their interactions in the industrial sector. By highlighting the challenges and opportunities within Human-Robot Interaction (HRI), especially in the context of cobots, it introduces some of the research directions followed in the chapters to come. It first identifies the research gap at the intersection of audio communication design and sonification, with collaborative industrial robots.

In the second chapter, **Related Work** is considered to gain an overview over the current cobot landscape as recognized in the literature. Existing research will be evaluated, covering foundational works on sound in the industrial context, and the needs during collaborative interactions, that aid the basic understanding of the research field. It also points out first possible audio guidelines and further research directions regarding cobots, their interfaces, and auditive communication.

Transitioning from the literature review, in the third chapter, **Using the Medium of Sound**, the nuances of the auditory medium and its capabilities and limitations as a communication tool will be further investigated. This phase identifies why sound stands out as a medium, for enhancing interactions, and its possible advantageous use cases in the context of cobots. It builds on the existing body of research as analyzed before, by inducing some subjective observations and new directions this thesis aims to explore, with relation to the user experience.

The fourth chapter, titled **Contextual Research**, more thoroughly investigates the real-world settings, in which cobots could be employed, and the implications this could give for sound design. Through these contextual inquiries and research, the practical challenges, opportunities, needs, and dynamics faced in these environments will be identified.

In the fifth chapter, **Concept Scenarios** are created, based on insights from the prior contextual research, combined with the found needs and potentials in sound interaction scenarios. These scenarios more clearly describe and categorize the possible applications of auditory signals in HRI. This helps focusing the research on the most relevant areas, and to generate scenarios relevant for further exploration and testing, aiming to start implementation in the most relevant areas.

The sixth chapter, **Testing**, synthesizes all previous insights and knowledge into concrete prototype sounds, to allow for real world validation and iteration together with users and experts. The aim is to assess the viability and effectiveness of prototype sounds based on the identified scenarios, as well as general preferences of users. Laying the focus on open interviews, together with the blind prototype sound evaluations, deeper understanding can be gained from participants about relevant areas. The goal is not only to verify the concrete sound designs, but instead to use these as working examples for gaining a better understanding of the underlying principles and how to apply them effectively.

The seventh chapter, **Conclusion**, collects and further processes the insights gained from the testing phase and previous research. It summarizes the made improvements on the sound solutions, as found through user input, and finally presents the used and refined interaction principles of auditive communication between humans and cobots, as the result of this research.

The final chapter serves as a **Discussion**, giving retrospective remarks on the findings, their limitations, and openings for further research. It critically evaluates the results and mentions some implications for the general field of HRI, and the further implementations of the results. It again summarizes the contributions provided by this work, aswell as its limitations.

# 2 Related Work

In this Chapter, the current utilization of audio in the industrial context will be examined, as well as potential applications for additional audio. The focus lies on understanding the medium of sound, its characteristics, practical application, possibilities, and value in interactions. Further some findings relevant to this, from the general field of HRC will be discussed.

By reviewing existing literature and analyzing practical examples of the use of naturally occurring, as well as intentionally designed audio signals, insights can be gained, regarding the relevant parameters for designing the use of sound in collaborative robotics.

For this, firstly the reasons for using sound in this context will be examined, highlighting its potential for conveying rich information easily and enhancing understanding and communication naturally. Also, the requirements of the demanding industrial work environment will be evaluated in this regard.

Through this review of the current research literature of the relevant fields, a solid foundation for understanding the potentials and limitations of audio for collaborative robotics can be laid. These insights will form the basis for further exploration of audio communication and interaction principles with cobots, which forms the central focus of this thesis.

Especially, when considering the changing robotics landscape, auditive communication might gain relevance. Visuals and GUIs obviously have received the focus of attention in design research, and even kinematics have been employed to increase communication (Bergman et al., 2019; Mok et al., 2015). Features like handguiding/programming by demonstration can be considered haptic or tangible interactions, which are already used on typical cobots. A modality which hasn’t received much attention in this field yet however is audio feedback. With the evolving nature of robotics and the increasing need for effective communication between robots and humans this might become a helpful or even necessary addition, to create a well-rounded, natural user interaction. Especially, since the new generations of cobots should allow also relatively untrained workers to interact with the robot's capabilities, an easy-to-understand sound design can be an essential part in an effective interaction design.

This need for increased understanding and communication in the changing robotics landscape is also supported by a number of researchers (Bergman et al., 2019; Michaelis et al., 2020; Villani et al., 2018; Zaatari et al., 2019), to make use of the capabilities fully.

To be able to exploit the advantages of cobots in the best way possible, more than it is currently done, according to Michaelis et al. (2020), new approaches "including designing for human-augmentation, and more intuitive and capable worker interfaces" (p. 7) are necessary. While things like the GUI are important aspects for this, it can be argued that an optimal solution makes use of the full range of human capabilities.

As also mentioned by Mok et al. (2015), "designing robots that can effectively work as a team with human users requires a deeper understanding of the social and non-verbal collaborative cues used during task-based activity" (p. 1). While the study looked specifically at how robots can communicate more effectively using expressive motions, the principles might be transferred or extended to sound. It was found, that while proactive behavior could be helpful, it might also have negative effects on the social perception of the robot. However, these effects could be mitigated, by additionally making the motion expressive, which made the users perceive the robot more on their level rather than as a boss. It could be argued, that adding sounds might make certain movements and actions even more expressive, adding to this effect. Also, it might be a viable alternative, for cases, where altering the motion from the programmed and expected, or optimal path might not be possible, which can often be the case in an industrial setting. This supports the idea of making cobots more expressive and adding features for closer collaboration, using sound as a feedback modality for intuitive interaction.

As sound, in many everyday cases, is the right fit for a feedback and communication modality, in order to achieve a natural and intuitive interaction, it should be investigated also for the industrial robotics context and possibly employed more. Looking at its uses in the context of (industrial) work, firstly, it can be noted that sound naturally is very information rich. The way sound is produced naturally, by for example a mechanism, is influenced by many small factors, like material properties, forces and static properties, environmental conditions, mechanical wear of parts, temperature, humidity, and other external influences. This makes every repetition of it sound slightly different, depending on the starting conditions. These differences can be subtle, and when describing the sounds, no distinction would be given usually, also because these are often hard to articulate. However, to a trained ear these differences can become quite noticeable and informative, and they might also find better ways of characterizing and describing them. While usual listeners might only identify a played note as for example a violin or a piano, a trained musician might hear fine distinctions and auditive qualities differentiating them and might even recognize the specific brand or model. Similarly, a mechanic might listen to the sound of a motor, and make out the type, characteristics, and specific parts of it not running smoothly, just from experience. Alexanderson, P. (2004) more closely investigated such effects in practice, taking process control operators in a chemical factory as a subject. These could make many useful conclusions for their work, simply by listening to the workplace soundscape. Specifically, the author identified, that they "use sound for identifying things and places, notification about status of surrounding artifacts and for maintaining social awareness" (p.281).

While Alexanderson, P. (2004) focuses mainly on the current state and the use of natural sounds, he also starts to discuss how these principles could be employed and extended with intentionally designed sounds, but also notes, that this "is a complex undertaking and the richness in a natural sound environment should not be underestimated" (p.284). His work mostly pertains to the identification of "settings suitable to augment with sound events" (p.284), following a phenomenological approach, mainly focusing on identifying currently used sounds in a given setting. While this is useful to augment existing sounds or to inspire the addition of thematically or conceptually similar feedback, critical new additions might be missed. If no comparable feedback is present at all in the soundscape yet, it might be hard to imagine the use of new signals giving new types of information. This is especially the case, if no clear need for this feedback has been identified and vocalized by users yet, which might only lead to slight variations, and simple notifications for existing messages, rather than new use paradigms being proposed. For this reason, the later chapters will try to also look at the work and information needs of users in detail, to identify additional scenarios for auditive augmentation. It can however be said, that gaining detailed insights into the current soundscape, and its use in practice is a very helpful basis for further developments and might also inspire new ideas, and a focus on the specific context is advised by multiple researchers on this topic (Alexanderson, 2004; Macaulay & Crerar, 1998).

In this regard, Macaulay & Crerar (1998) also stress the importance of properly analyzing the existing and potential work practices, and existing workplace soundscape, to design auditory displays effectively, and not just "driven by technology[…] and […] specific cognitive issues" (p. 9). While they mention that many of the established design methods in fields like HCI still need to be adapted or developed to fit the design with audio instead of GUI's, they propose especially mapping the existing soundscape, and "developing a shared language with which to talk about the auditory aspects of work and the workplace soundscape" (Macaulay and Crerar, 1998, p. 9) as effective strategies.

In this way the identification of useful additions to the auditive design can be guided. Based on the identified scenarios, as also proposed by Alexanderson, P. (2004), "use scenarios and function prototypes will be developed and evaluated" (p.284), which is defined as a goal of this thesis.

Macaulay & Crerar (1998) also makes similar findings of the usefulness of audible information in the soundscape of a busy newsroom, where concept of "outlouds" plays an important role in facilitating collaboration within the team. The relevant quality of these outlouds here is the fact that they can be shouted or sent out without a clear recipient in mind, and then picked up by whoever it is relevant to. This can allow for a more intertwined, while still efficient collaboration.

This points to another advantage of sound, as emphasized by Macaulay & Crerar (1998), lying in its immediacy and ability to convey information without demanding visual attention. This would require recipients to specifically turn their focus towards the information, especially when considering textual information. While some of the periphery is still visible, in order to consciously process the visual information, the eyes need to be focused on the subject, allowing only one such subject at a time. Audio on the other hand, supports peripheral awareness to a much larger extent.

While Macaulay & Crerar (1998) describe the use of spoken outlouds in practice, containing quite complex and specific pieces of information, the analysis of Alexanderson (2004) focuses especially on non-speech audio. Here the effect of "rich and unfiltered character of natural ([3] In this sense, natural refers to the original sound of an artefact, not designed with any specific intention.) sounds" (p.283) becomes much more apparent, and a number of examples from the industrial context are mentioned.

The ability to monitor these sounds in the background, often unconsciously, could add a lot to users' peripheral awareness, without overloading their attention. In a collaborative industrial setting, where visual attention might be occupied with tasks like programming or manufacturing activities, or with monitoring the robot's actions, auditory cues can play an important supporting role. This might not only enhance safety but also ensure smoother collaboration between the human and the robot. However, as the amount of auditive information coming from this largely digital system of a cobot is limited, compared to many, more mechanical machines, these signals might be extended or added in any areas useful to the operators. This opens up a large number of options for what to convey, and how to map sounds appropriately, but will just as well pose the challenge of doing this in an appropriate manner. When executed well, these signals can be learned quickly or understood intuitively, and give users a much better insight into the robot's processes to support smooth collaboration. For safety reasons, this becomes especially important as the user works more closely with the cobot and is within its reach. Additionally, as noted by Zaatari et al. (2019), "As cobot autonomy increases, an operator is more likely to feel unease due to the cobot's decreased predictability" (p. 175). As a satisfying user experience was defined as one of the goals, this effect shall be mitigated as much as possible.

With these kinds of interactions becoming more prevalent in collaborative robotics, the need for appropriate supporting signals also becomes apparent. However, while several elaborate, smart interaction features are presented, there is still a relatively big discrepancy between the academic state and the application in the industry (Zaatari et al.; Michaelis et al., 2020). Practical and intuitive communication methods might be especially necessary to support more complex interactions but can also simplify existing ones. Michaelis et al., 2020 also mentions, as a proposed solution to the discrepancy, to "support template-based and generalized programs that better afford workers, as end-users, to interact with and program the sys- tem for truly flexible and collaborative applications" (p. 10). To be able to make these higher level HRC programs usable and understandable to the common operator without in depth knowledge, supportive signals that are in line with their capabilities are needed. While extensive simplifications of the GUI and programming interaction are happening already, it can be argued that these can only go so far in creating a natural interaction. While trying to implement such features, the challenges for these in industrial environments, as well as under strict safety requirements need to be considered. Zaatari et al. (2019) suggest that "until natural speech and gesture understanding reaches a reliable level for industrial use, it is advisable to stick to a fixed set of verbal or non-verbal commands which are easier to recognize" (p. 175). Here, also auditive signals, as discussed earlier, offer an immediate and unintrusive way of conveying information, and thus supporting understanding and safety.

Furthermore, the authors mention that even when "other technologies are used to create/alter the cobot's program, a UI is still necessary to override any of them since it is the most reliable means of controlling the cobot" (Zaatari et al., 2019, p. 176). While this confirms the focus on visual communication, it also again suggests that the visual attention can be occupied in many situations, even in future versions.

As described in a number of examples, rich and unfiltered sounds can give a lot of information to workers that have grown accustomed to it. Even an example from practice at KUKA is known, where operators over time learned to listen for the click of the motor breaks of the cobot to get an understanding for what it was doing. This clearly hints at the need to understand the cobots internal processes more deeply, for working collaboratively and harmoniously with it, and to understand what to expect. While previous mechanical devices used to be much richer with these kinds of small sounds, a robot arm doesn't naturally give off much noise besides motor sounds and possible pneumatic equipment. Implementing such feedback artificially gives many options, but appropriate situations and information content needs to be chosen for an effective design. Otherwise, the added audio could just become a distraction, rather than improving the experience.

The work environment can already be stressful, with many different influences to be aware of, different machines to monitor, and multitasking being required in some cases. This can be demanding to the attention so additional signals should be added in a sensible way. However, audio can also be part of a remedy to this problem, as it can offer unintrusive monitoring in the background, which doesn't take up as much attention an allows to focus on other things at the same time. This becomes especially apparent when comparing to textual notifications, which are common today for communicating all kinds of information. As

Macaulay and Crerar (1998) also note in this regard “problems of living with the huge amounts of information now available to us have been the subject of much debate recently, with the phrase 'information sickness' entering the vernacular. But rather than thinking of the problem in terms of 'information overload', we can think of the problem in terms of 'word overload'. The problem might not be too much information, but crucially, the way that the information is presented (typically as text).” (p. 8)

Also (Alexanderson, 2004) mentions peripheral awareness to these soundscapes, and mentions examples in which despite the big amount of information present, the relevant information can still be easily filtered out without needing to pay additional attention to it or being distracted. Using these effects could offer an effective solution to such problems.

In addition to simply offering an additional communication channel, sound carries a unique set of properties, which might also be used to enrich HRI. While our day-to-day engagements with technology mainly utilize the visual modality, sound introduces another dimension, with unique advantages, aswell as the possibility to be combined and enrich existing communication, making them more intuitive and immersive. Additionally (Henkelmann, 2007) summarizes the following “advantages over a visual display:

> ***Free movement***: *When we watch data, we always have to keep our eyes on the screen (or another output device). If data is presented to us as sound, we can move freely about.*
>
> ***Background monitoring***: *When some quantity is permanently supervised aurally, we can focus our attention on other tasks, as long as the changes in the data create a monotone feedback. But as soon as the data changes drastically (which should be represented in a drastic change in sound), the sonification automatically comes to our attention.*
>
> ***Temporal resolution***: *The temporal resolution of hearing is about twice as high (20-30 ms) as the temporal resolution of vision (50-60 ms). When we take spatial location into account, human hearing can differentiate time intervals of down to 1 ms![Warren, 1993]*
>
> ***High dynamic ranges***: *We hear over a large range of amplitudes and pitches which allows for a high resolution of presentation.” (p.7)*

Especially the advantages of free movement and background monitoring allow for a good fit to collaborative industrial tasks, where many processes might have to be monitored at the same time, and there might be different workplaces. Also, this might offer a solution to the demanding textual notifications and high amount of required attention, through the background monitoring mentioned before. The other two can become especially relevant when looking at more dynamic uses than simple notification sounds, which will be investigated more deeply in the following chapter.

# 3 Using the Medium of Sound

To make efficient use of Sound in this context, not only the possibilities of manipulating this medium need to be known, but also the specific challenges for it and the given context need to be considered. Further the perception of different kinds of signals to users should be looked at, as this can vary, and clear distinguishable categories should be found for this. In the end, also methods to appropriately map these dimensions to a certain parameter should be given. For this, there are several general methods available, which will be discussed. The concrete uses it can be mapped to in this context can then be examined in the next chapter with these in mind.

While, as pointed out, there are several distinct advantages of sound as a communication medium, there are of course also new challenges that can arise. Interactions with various existing devices, as well as interactions with other actors, were looked at, to identify and classify the most important categories of challenges. These can then be kept in mind for the further design process, and either avoided, or purposefully designed against.

## 3.1 Acoustic Factors

Related to audio, these are of course the first to consider, as they will play a role in how the signals are transmitted and perceived by the listeners. Here we can differentiate between factors related to the environment, often directly related to the sender of the signal, and then there are also effects happening at the receiving end.

Starting with the *sound creation*, several methods are available. The most readily available to us is the one, traditional musicians also employ when playing their instrument. Physical movement excites the active element and subsequently, the air, creating the sound. In this setup, there are many factors influencing the resulting sound. These include the way in which the movement is executed, material properties like flexibility and size, as well as further resonance of the sound. All these factors are learned by musicians over time through hours of practice, until they can manipulate them, without conscious consideration. The same principles would be at play for sounds like the whirring of an electric motor, which will create sound through vibrating parts and friction. An analysis of these sounds in the robotics field is also given by (Tennent et al., 2017), where their effect on the user experience is pointed out. While these consequential sounds usually just result from the product's construction and functional elements, there are still many cases where these parameters are consciously optimized, to for example make the sounds more pleasing or silent. For this, careful manipulation of the construction and materials, and regular checks of the resulting sounds are necessary. An additional option could be to design additional sounds in accordance with the existing consequential sounds, for example by matching the frequencies harmonically or avoiding them.
When we want to create sounds electronically and with higher consistency than acoustically, there are various synthesizers, which can be divided into analog and digital, as well as the option to store the resulting soundwave digitally for later use. Analog synthesizers use circuits to create and modify the vibration electronically and then output it through speakers. Here, dedicated components are necessary for each modification done to the sound, meaning it can usually produce a limited range of sounds. An easier and cheaper solution nowadays, giving access to any imaginable waveform easily is digital synthesis, or the use of sampled waveforms. Here the limiting factor is the sample rate of the file or synthesizer, which determines how many data points per second are available to represent the waveform. Similarly, the bit depth determines the resolution of the amplitude. While generally it can be said that higher rates will give higher quality, more accurate sound there are some considerations to make. Firstly, the lowest sample rate to use to represent a waveform is at least two times higher than the highest frequency to represent. This will result in two data points per wave cycle, allowing you to read out the data accurately as the actual intended wave frequency. This should be considered to give an accurate representation of the signal as desired. In the end, since higher rates do not only increase file sizes and transmission rates, this is a balancing act, and it can be useful to choose these values lower, and design the sounds in accordance, to minimize the effects of quality loss.

#### *Sound Dimensions*

In the following, a number of principles for the modification of audio as also common in the sound design and musical field will be collected. These can become relevant to give a clearer overview, and better understanding of the options in modifying potential sounds, and communicating these appropriately:

- **Pitch**: frequency of the sound wave, perceived highness or lowness of a sound

- **Volume/Loudness**: the perceived intensity of a sound, determined by the amplitude of the sound wave
- **Rhythm**: the pattern of regular or irregular pulses created by the arrangement of sounds and silences in time
- **Timbre**: the quality or character of a sound that distinguishes it from other sounds, which is determined for example by the spectrum of overtones present in the sound
- **Tempo**: the speed at which a piece of music is performed
- **Duration**: the length of time a sound or group of sounds lasts
- **Envelope**: The curve representing the volume profile of a sound over time, including its attack, meaning the speed at which the volume increases, and the release, indicating how long the sound will take to fade out
- **Harmony**: the combination of simultaneously sounded musical notes to produce chords and chord progressions
- **Dynamics**: the degree of loudness or softness in music
- **Spatialization**: the placement of sound in a three-dimensional space, influenced mostly by the balance of the sound between the left and right channel
- **Reverb(eration)**: the persistence of sound in space after the original sound has stopped
- **Echo**: the repetition of a sound due to reflection of the sound waves
- **Flange**: the combination of two identical signals, one of which is delayed by a small and gradually changing amount, creating a swirling effect
- **Phasing**: the shifting of the phase of one sound relative to another, creating a sweeping effect
- **Chorus**: the combination of a sound with one or more delayed, pitch-modulated copies of itself
- **Tremolo**: the rapid repetition of a single note or chord, creating a wavering effect
- **Vibrato**: a rapid, slight variation in pitch, creating a pulsating effect

The next step is <u>*sound transmission*</u>, starting with outputting the sounds into the analog, physical world. At this point, the sound will be affected by the limitations of the loudspeaker, or other chosen way of outputting. There are other methods available, like headphones, piezo elements or even bone conduction technology, which can deliver the vibration directly to your skull, or can even be used on other materials to make them vibrate and thus sound. The focus will however be put on traditional speakers as they are most widely used and would likely also find application in a cobot, especially as its location independent.
For the chosen output method, the various limitations, and effects it carries with it need to be considered, to design around them when necessary.
Firstly, the frequency response varies depending on the components. Some frequencies are boosted, while others will be reduced in volume, distorted, or will not be represented at all. Generally, speakers are rated for a specific frequency range, usually roughly fit to the human hearing spectrum of about 20Hz to 20kHz. However, even within this spectrum, the reproduction isn´t accurately flat. While in musical applications a very accurate representation can be important for the most pleasant listening experience, the requirements in these settings are usually different. Here likely the most important effect will be designing audio specifically for smaller, possibly cheaper speakers. For this again equalization can be useful for minimizing the effects and avoiding unpleasant sounds.

Next, environmental conditions will affect the sound, including acoustic properties of the environment like hall or resonance to certain frequencies. Especially considering the industrial environment hall is favored by the big spaces with large, hard, reflective surfaces, instead of having absorbers that dampen the sound. This can cause increased reverberation as well as echo, which might negatively affect the experience of the sounds. Since changes to this environment are usually not feasible, it might be necessary to adjust the audio accordingly. For this firstly an appropriate sound level needs to be set so it is properly audible to the intended recipient, but not too loud or annoying for them or other people in the area. At the same time, background noise levels need to be considered so signals are not masked. Additionally, taking care in adjusting the speaker direction and position properly can help focus the sound to where it is needed. This can either be done by placing it close to and facing the usual work area, or by actively manipulating these factors based on the recipient's position or maybe the type of message. Some signals might be sent only in the direction of the worker directly interacting with the cobot, while others, like important alarms, can be played loudly, and in all directions to make sure someone will pay attention to it.

An additional option, also to improve clarity over background noises might be an active noise cancellation, which can cancel out some noises, and make the intentional ones clearer. However, the applicability and feasibility in the industrial context is unclear and needs to be further investigated.

At the end of the transmission is the *sound reception*, in this case in the form of the worker hearing a notification signal. At this stage, several personal factors also come into play. Personal conditions include factors like personal hearing range and sensitivity, and any hearing impairments. Generally, the human audible spectrum ranges from about 20Hz to 20kHz, but over time the range decreases, making especially high-pitched sounds less and less audible to older people. Also, considering the industrial context hearing protection might be worn, reducing the recipient's sensitivity to potential signals. When this is known, it might be possible to adjust the audio and loudness accordingly, however with safety-relevant sounds a sufficient volume, as well as supportive signals like LEDs will likely be required.

The threshold of the hearing curve is also not constant over the frequencies, pointing to another effect of human hearing. Volume is perceived differently depending on the pitch of the tone, so if an equal perception is the goal adjustments need to be made accordingly. Since this effect is well recognized, these Loudness adjustments are common. It is to investigate, how hearing protection might influence this further and if this can be considered too.

Further, the human hearing is affected by so called masking effects, which should be considered to ensure audibility and recognizability. Namely two types of masking effects will be relevant as they could suppress the designed signals, as described by Herre & Dick (2019).

Firstly, spectral masking, also known as simultaneous masking, describes the suppression of a weaker sound by a louder masking sound in a similar frequency range when played at the same time. This is especially relevant in loud industrial settings and might influence the chosen frequencies for the signals. In the presence of a masker, the curve describing the audibility threshold changes into a masking threshold, characterized by a bell-shaped increase in frequencies close to the masker. Sounds beneath this threshold are masked by the louder signal and become inaudible.

The tonal characteristics of the masking sound influence the strength of the masking effect. Noise-like maskers can mask tonal signals effectively, while tone-like maskers have a weaker masking effect on noise.

Secondly, temporal masking affects the audibility of sounds that are not playing at the same time as the masker. In this way, through a reduced sensitivity of the ear, sounds can be suppressed approximately 100-200ms after the masker ended, which is called post-masking. However surprisingly even a pre-masking effect is recognized, although shorter. A softer sound will be masked by a louder sound playing about 20ms later, likely due to the way they are processed differently in the brain.

The integration of these masking effects into the sound design for cobots is necessary to consider for enhancing the audibility and intelligibility of auditory cues, especially in industrial or otherwise loud environments, but also so signals themselves don't affect each other negatively. Considering frequency-dependent thresholds and adapting sound signals to account for spectral masking effects will optimize human-cobot interaction and communication.

In general, a strategy for designing sounds with this in mind could be to analyze the soundscape frequencies and focus the signals on less busy parts of the spectrum. Considering the abundance of noise in the industrial context, which can also be hard to predict and cover a wide range, the combination of various frequencies in one sound might be useful to ensure audibility, nonetheless. Using equalization, relevant frequencies might also be lowered where possible to not add to the noise even more.

## 3.2 Cognitive Factors

When the sound is successfully transmitted, it still doesn´t mean the message has successfully arrived and been further processed. There are still factors related to the cognitive processing of them at play, influencing for example if the content is understood correctly.

Firstly, the attention of the recipient needs to be considered. While some sounds, especially when they are loud can get users' attention, this might not always be the case if they are focused on another activity, especially when they are already listening to something else.

Especially with busy workplaces it might be a challenge to create a solution that is well recognizable in the busy industrial soundscape. Workers might be too focused on other things and in the busy soundscape sounds can be missed or misheard if they are not standing out enough. On the other hand, adding too many, frequent or loud sounds, might overload their attention, and annoy them, or stop paying attention to them altogether.

In this regard, some personal factors, like the individual perception of importance, urgency, or the attention paid to a given signal come into play. This can change over time as they learn the importance of the signals and the connected messages. However, the aim is of course to make this intuitively understood and noticeable through the sound design. Also affecting this will be the general focus of the recipient. For this, there are several relevant outside factors, like possible distractions or mental preoccupations. The worker could generally not be focused on the task at hand, or parts of the task or environment can take up the attention, like having to watch out for hazards. Further in a workplace also social distractions need to be considered. If it is possible to recognize and take these into account in cobot systems, appropriate moments for the notifications can be selected more effectively. These strategies, finding opportune moments for notification, according to Adamczyk & Bailey (2004) can be a more user-friendly approach. This also includes using the right modality at the right time, which again suggests focusing the use of audio on the most relevant areas.

Mentally there are a few steps necessary after noticing a signal. Starting with recognizing and distinguishing the signal from other signals, which isn't always straightforward. Some signals might be too silent or unclear and might blend in with other sounds making it hard to make them out clearly. Also, if many similar sounds are used, their differences might also be hard to notice. Next, in case it can't be intuitively interpreted, the sound and the according message need to be recalled from memory. Here again, having too many signals can have a negative influence.

The learnability of the signals can also be supported in some ways, making recall easier. Skeuomorphism can be mentioned as an effective strategy for this, to give intuitive understanding and support fast learning for all robotics skill levels. At least in visual and GUI design, this method has been applied broadly, however, the idea of using established recognizable principles for new interactions might be applied to sound design too. In this way, actions can be easily associated with their sound, by taking inspiration from other fields and matching the sound to what users are already familiar with in their interactions. It can make especially complex digital information relatable to all users without industry-specific knowledge, as they are familiar with the parallel or connotation from the physical word and everyday life. In that way, it can be an effective strategy, fit to the requirements of cobot interactions.

Then for certain sounds Temporal resolution or even pitch perception could become important. These can differ among people, so proper limits should be employed to make sure the solution will be accessible to everyone. A general guideline as provided by Barrass (1997) states, that "The fundamental frequency of a warning should be within the range 150-1000 Hz. There should be four or more component harmonics to help avoid maskings. The overall spectral range of warnings should be 500-5000 Hz" (p. 30) For important sounds like warnings, sticking to the middle range around 500-3000Hz is advised, while lower frequencies can be used for signals that must travel further and through obstacles, and still be audible.

Considering the temporal resolution could become especially relevant, when going beyond simple notification sounds. The natural interaction with expert musicians and their instruments described earlier, as also explored by Ghomi (2012), points to the potential of sound design. Especially in combination with other modalities, embodied interaction with the product could be achieved, allowing for closer collaboration. Regarding instrumentality, the author describes making use of the natural relationship between actions and their associated sounds in a consistent way, which is then applied to HCI in general, as well as digital instruments. This relationship is found to be less embodied in usual computer-based music setups or other interfaces. This kind of materiality, in acoustic instruments, is achieved through consistency in temporal, spatial, and causal dimensions, meaning they are matched multimodally, and the sound is coming from the actual logical place where it would be produced. This creates a logical accountability of the product for actions and the accompanying reactions like sound in this case. Some of the strategies proposed by Ghomi (2012) to introduce these principles to digital interactions are the use of metaphors and metonymies. Integrating these can enhance accessibility and engagement, aligning with the principles of tacit knowledge. Especially metonymy, substituting based on material or causal links to the principle itself, instead of only metaphoric relations, empowers expert interactions, and deepens engagement, as it lets users explore the behaviors interactively. Metaphors on the other hand can be especially suited for supporting learnability and recall for new users.

| |
|---|
| **Powerful** - empower everyone |
| **Functional** - accomplish your goals |
| **Delightful** - exiting to use |
| **Supportive** - guidance at any given time |
| **Anticipative** - your companion |

*Table 1: Design Principles Used by KUKA to guide the development of the new Robot Operating System iiQKA (IDZ International Design Center Berlin, 2023)*

Another consideration regarding the recognition and the association of the sounds is brand coherence and fit to the general corporate design and goals. While sound isn't usually the first consideration, a well-designed concept can support the emotional connection. For this, the material already produced by KUKA, especially for advertising purposes can be a great source of inspiration, as it often tries to represent the general goals and principles the company stands for and make them experientable to customers. Also looking at the used design principles can be helpful, in order to align the sound design to the same goals. Matching the company values desired presentation to customers in all aspects can make for a more rounded user experience and coherent company perception. A good example, and also the most well-known one released by KUKA so far, is the promotional video (*The Duel: Timo Boll vs. KUKA Robot*, 2014). While in this type of material the goals are different to the actual requirements during work, and a lot of the tension and excitement creating sounds might need to be removed, the general sound design can still give important, rich insights into the desired perception of the brand and product. Connotations that come to mind for this example are powerful, technological, precise, elegant routine and control, driven, and percussive.

In addition to fitting the sound design to the corporate design, it also might be desired to match it with the specific device it's used on. Similar to how visual appearance influenced the perception of the capabilities of social robots (Dennler et al., 2022), auditive appearance is an important part of the perception too and could have similar effects. Fitting the sounds to the specific cobot embodiment can also make it easier for users to attach the sound to the cobot and recognize it as belonging together. To avoid things like the Uncanny Valley effect (Mori, 2012), or generally false expectations, it should be paid attention, that the sound design represents the actual capabilities of the robot honestly and accurately. The same goes for displaying honesty in the representation of things like the urgency of the used audio messages, to not affect users' perception negatively.

### *Sound Perception Groupings*

While the classification of sounds is of course very subjective, it still helps to identify some themes, that can generally be applied. While the sound dimensions mentioned before, related to acoustic parameters, are useful for varying the sound, especially when mapped to a parameter, the subjective perception of them will vary and can be influenced by the combination of numerous factors. For users to recognize these sounds and to classify them correctly, being able to clearly describe their perception is important. Also, a way to separate and group certain sounds in an understandable way will be necessary, so all the possible directions that the sounds could go in can be described for later use. In the following a selection of these perceptual groupings will be given as opposing poles on a spectrum, giving a good basis for further working with them.

| Mechanical vs Organic | |
|---|---|
| machinery or robotics, e.g.: sound of gears turning, a machine running, a moving robotic arm, short attack, repetitive, consistent | Related to natural phenomenon, or human physiology, e.g.: Heartbeat, water drops, breathing, softer decay and attack, dynamically changing |
| **Low-Energy vs High-Energy** | |
| Low-Energy: data related to low levels of activity or energy, e.g.: gentle breeze, a stream flowing, soft background music | High-Energy: data related to high levels of activity or energy, e.g.: thunderstorm, car racing, rock concert |
| **Abstract vs Concrete** | |
| data that is more conceptual or abstract in nature, e.g.: synthesizer, computer processing, robotic voice | more concrete or tangible, e.g.: hammer striking a nail, saw cutting wood, machine assembling parts |
| **Bright vs Dark** | |
| data that is positive or upbeat in nature, and/or higher pitched, e.g.: bell ringing, bird singing, person laughing | is negative or ominous in nature and/or lower pitched, dull/muffled, e.g.: deep thunder rumble, door creaking open |
| **Textured vs Smooth** | |
| data that is complex or detailed in nature, e.g.: person talking, machine running, orchestra playing | data that is simple or streamlined, e.g.: stream flowing, person humming, wind chime |
| **Natural vs Artificial** | |
| related to the natural or physical world, e.g.: bird singing, water running, person speaking, drum sound, hitting metal | Synthesized artificially, digital, technological, computer processing, robotic voice, beeps, sinewave, |
| **Soft vs Jarring** | |
| subtle or delicate in nature, e.g.: gentle humming, whisper, soft beep | powerful or intense, e.g.: sudden sharp noise, loud unexpected alarm, loud screeching noise |
| **Fast vs Slow** | |
| high speed, rapid movement, e.g.: the sound of a race car, a bullet whizzing by, a fast typing on a keyboard | low speed or slow movement, e.g.: the sound of a snail crawling, a gentle swaying of a pendulum, a slow and steady heartbeat |
| **Melodic vs Percussive** | |
| related to pitch and melody; e.g.: a singing voice, a piano playing a tune, a birds song | related to timing and beat, e.g.: the sound of drums, a ticking clock, metronome |

*Table 2: Comparison of Some Relevant Sound Groupings*

## 3.3 Sonification/Mapping Methods

The literature comes down to two especially applicable approaches to the design of sounds supporting information processing activities. Firstly, a number of general mapping and design approaches are given, taking different perspectives on selecting appropriate sounds. In the following the collected principles will be looked at one by one and applied to the given context.

| | |
|---|---|
| "The **syntactic** approach focuses on the organisation of auditory elements into more complex messages. | Combine different audio elements in a way that resembles a language. Combine messages from different blocks: [point][updated], [program][delete][really?], [program][delete][done!], [program][create][done!], [program][save][done], [program][automatic][run!] |
| The **semantic** method focuses on the metaphorical meaning of the sound. | Relates closely to the skeuomorphic principle described before, for example motor or gear sounds can be used to represent movement of the arm or a crashing sound can be used when an obstacle was hit |
| The **pragmatic** method focuses on the psychoacoustic discrimination of the sounds. | Fitting the sounds to the practical user experience with them to find the optimal fit. It optimizes intensity, rhythm, frequency, etc. Takes into account acoustic and cognitive factors also discussed in previous chapters. It ensures clear, recognizable communication even in the busy industrial context. |
| The **perceptual** method focuses on the significance of the relations between the sounds. | Focusses on how sound perception principles relate to the represented area. Tempos and rhythms can represent the speed of actions like saving. The pitch of a movement sound can be related to the motor speed. In this way it uses principles similar to parameter mapping sonification, adding a component similar to skeuomorphism to create intuitive recognition of the principle. |
| The **task-oriented** method designs the sounds for a particular purpose. | Focusing on the needs and use of the info in context. Considering how they are perceived, e.g. important warnings requiring attention could be played repetitively until attended to, and multimodal combinations used, to ensure recognition of actions like successful or failed saves. |
| The **connotative** method is concerned with the cultural and aesthetic implications of the sound. | Address the users' subconscious reactions to sounds, taking a general sound design perspective. Using a pleasant sound for successful completion for a feeling of accomplishment, while using jarring sounds for failures, leading to negative connotations |
| The **device-oriented** method focuses on the transportability of the sounds between different devices, and the optimisation of the sounds for a specific device. " | Taking into account the possibly differing output devices and how they affect the perceived signals. Considering parameters like volume, equalization, and stereo position, as well as frequency response and ambient noise, especially considering the flexible and changing environment to still create a recognizable signal can be important. These considerations are mainly discussed in the acoustic factors subchapter. |
| (Barrass, 1997, p. 46) | Design Approaches Applied to the Cobot Context |

*Table 3: Application of Auditory Information Design Principles after (Barrass, 1997) to Industrial Cobots' Auditory Interfaces*

In addition, the field of sonification might be applied to represent more elaborate data. It also comes with several recognized approaches, of which a few are especially relevant in this context. Listed by (Henkelmann, 2007) are:

- Audification
- Earcons and Auditory Icons
- Audio Beacons
- Model-based Sonification
- Parameter Mapping Sonification

Earcons, and Auditory Icons representing pieces of information through audible sound, using certain sound patterns or icons to represent specific information or distinct events, or emitting sound signals as pointers, but they are not discussed in detail here due to their limited capacity for immediate, rich feedback in the given context. However, the use of certain notification sounds can sometimes fall under these categories.

In contrast, Audification and Parameter Mapping Sonification are valuable tools to represent more complex and continuous data streams, or even more in-depth, derived information using model-based sonification because these methods provide intuitive, nuanced feedback by mapping data to complex sound models or relevant parameters, enhancing worker understanding and facilitating rapid, accurate responses in human-cobot interactions. In this context and under the given requirements especially the first and last seem to become particularly relevant, for their immediate and rich, but still easily understandable feedback, which can possibly create a much more intuitive interaction and informative communication. The fast subconscious way the information is perceived can let users have fast and accurate reactions for more close interactions, without requiring too much domain-specific knowledge.

Mapping a relevant parameter to a continuous sound can let the worker gain an intuitive tactile understanding of it and the direct experience supports fast, intuitive understanding, as also applied by Stienstra et al. (2011). This can for example be useful for measuring tasks, to combine the accurate sensors and actuators of the cobot with the nuanced understanding and perception of the human worker, letting them work closely and effectively.

Mapping such parameters through an even more complex model to influence certain sound characteristics can especially become relevant when taking into account the already existing digital representation of all the robot's actions in the 3D space. This model can become the basis for further calculations to represent certain aspects through sound. For example, the movement speed of the end effector in space, rather than just the individual motor speeds can influence the pitch of a movement sonification giving more relevant feedback. Similarly for collisions or for deviations from the expected, modeled behavior sound can be played in a reactive way, to give rich feedback about its behavior, rather than simple notifications about events recognized by the system itself.

# 4 Contextual Research

After the theoretical foundation on the possibilities, challenges and practical characteristics of audio has been laid, this chapter shall deepen the insights from a more practical, applied standpoint. Since the real-world application often can differ from the theoretical propositions, which is also evident in the main applications of cobots in the industry thus far (Zaatari et al., 2019; Michaelis et al., 2020), this analysis seeks to help bridge this gap. In this way several factors relevant to the user experience will be investigated in this chapter, to make sure future developments are in line with the actual use patterns and conditions in practice. For this the current state of cobot applications will be evaluated from the company’s perspective, however also considering competitors’ approaches and the overall field. In addition to desk research on the products, and use cases in practice, the experiences of experts within the company were asked and explored throughout the project. This included experts on the engineering side, safety, sales, and robot trainers, as well as experiences gathered through work within the UX design team. In addition to the already mentioned related work, the case studies about applications of various cobot systems by KUKA (KUKA AG, 2023a) showed a good variety of realized applications.

While this user-centric analysis of course also includes a look at the various new potential users themselves, firstly, the conditions affecting their work and learning experience will be analyzed. Starting with the environmental conditions, which were already discussed a bit in the related work section, a comprehensive overview will be given, focusing on concrete effects and influences relevant to the user experience and designing new audio.

Next, even though the goal is to get to universally applicable principles for cobots, the specific characteristics and use patterns of the KUKA system, as it currently is in use, shall be investigated. By examining the specific system interaction design characteristics, functions, and how they are used in practice, as well as its limitations, insights into potential supporting audio cues can be made much more relevant and applicable. While the design of these cues and the underlying principles won't be limited to use on this system, having it as a basis will not only allow for much more focused testing later but also ground the development in concrete practical interactions rather than broad concepts. While no in-depth analysis of other systems specifically will be given, it is important to not be limited by the current company paradigm, but instead keep an open mind for different solutions, be it realized features from competitors or just other proposed novel concepts for cobot interaction. While cobot systems currently for the most part don't differ fundamentally in their interaction, there might still be some aspects of other cobots to consider.

Finally, the new diverse group of cobot users will be investigated, to be able to match sound solutions to their specific needs and consider the specific challenges they can face in their work. While recognizing that there are more stakeholders involved in the cobot ecosystem, this analysis will mainly focus on the operators themselves, as the cobot concept, as envisioned by their manufacturers, also entails a much more self-reliant approach. It aims to empower operators to take on multiple tasks, such as robot programming and safety engineering. In this way, operators could take over all these tasks themselves, given a system that is designed to facilitate their learning and operation of the cobot seamlessly. By understanding the operators' roles, expertise, and preferences, we can design sound cues that align with their workflow and enhance their efficiency.

## 4.1 Environmental Conditions

The environmental conditions under which cobots operate pose several challenges to the perception of auditory cues. These need to be considered for a successful implementation.

### *Environmental Noise*

As uncovered in the literature, the industrial setting comes with several important influences on the use of sound. Environmental conditions, like noise can obviously affect the audibility and recognizability of any new signals, and the added signals themselves might even become a source of further distracting or annoying noise. While they need to stand out in the possibly loud industrial soundscape, they still should not overload the auditory channel and attention. With the risk of missing important signals like alarms, and the necessity for clear, recognizable cues, balancing these aspects is critical for creating a smooth, safe human-cobot interaction. As found in the acoustic factors chapter, strategies like using well-audible frequencies, and frequencies differing from the noise can be effective. Also, Equalization and the use of additional harmonic frequencies can increase the audibility in these conditions. Also, directional audio with a volume adjusted to the needs in context can help in reducing the noise level created by the solution itself. Also, given the often already stressful and attention-demanding environment, with many things to pay attention to, these factors need to be considered before adding any new signals, users need to pay attention to.

*Adapt Sound to the Environment*

Also looking at the possible already existing sounds like safety alarms and other signals from various machines, possible relations, and similarities with the cobot sounds need to be avoided. A strategy might be to limit the sounds to a group of distinct timbres and tonal qualities, that can be recognized as belonging together. However, making them too similar will limit their distinguishability and expressiveness.

Considering the possibly flexible use cases of cobots, and thus varying conditions like noise level or the types of sound present, adaptions might become necessary. As manual volume adjustments carry a safety risk, they might not be the favored solution, and (Barrass, 1997, p. 30) also mentions avoiding these in a safety-relevant scenario. Therefore, ideally, smart features might be integrated, to recognize relevant factors like user position, soundscape, and use context, to optimize their communication accordingly. For example, the conditions in a cleanroom compared to an industrial manufacturing hall can be quite different, requiring a different volume level, and possibly even modified sounds to keep their recognizability in the context. The diverse and flexible workplace settings under which cobots could operate, in some cases even being mobile, make the design of distinguishable and meaningful sound cues even more important, and they should be easily attributed to the cobot and to the right message. Care should however be taken to not modify too much, losing recognizability.

*Fitting the Sound to the Task*

An important requirement for effective auditory interaction is the alignment of sound cues with the specific cobot actions represented. Auditory cues must logically correspond to actions and provide users with clear and meaningful feedback. In many cases, this can depend on the task and might differ. Ensuring distinctively recognizable sound cues is important for the usability and reliability of auditory communication. In the ideal case, sounds are thus connected to other perceivable actions, like a movement. Ensuring a logical match between sound and cobot action in this way can make it clearer where the sound is coming from, and what it is related to. To be able to navigate the busy environment with different signals from different sources influencing users, the audio should be fitting for and attributed to the cobot easily. As mentioned, sound can be given a more distinct character, rather than simple, interchangeable beeps to differentiate them more easily. Additionally matching other related feedback channels multimodally can make the connection instantly clear, which can be helpful for new as well as experienced users alike.

*Requirements:*

- Audibility and Distinguishability: Crafting sound cues to stand out from ambient noise, and from other machine sounds. Use (a set of) distinct timbres and tonal qualities to prevent confusion, and keep the sound attributed to the cobot. Additionally align sound cues logically with cobot actions and other modalities for clear user understanding
- Seamlessly integrate into the existing auditory environment without causing confusion or overload
- Prevent auditory cues from contributing to distracting or annoying noise by focusing it on relevant areas and limiting volumes according to specific circumstances
- Keep basic audio principles understandable, even when varying them, to make the most important parts and especially safety related messages intuitively understood, as the potential of misunderstanding can carry risks in this context

## 4.2 KUKA Cobot Interaction Design

With the cobot system, industrial robotics is starting to depart from the traditional approach of programming, operation, and setup. While significant progress has been made towards greater user friendliness, the design and interactions with this system are being further improved. The goal is to enable flexible and spontaneous assistance, to be able to fully use the advantages of the human and robotic capabilities where they are needed, in a fluent and natural collaboration. This requires perfecting existing functions with the user in mind and making more complex actions quickly accessible. Simplified programming interfaces and setup routines are used, making it possible for users to operate them with little prior knowledge. Therefore, the focus lies on the new operating system developed for these modern applications, like the use on cobots, called iiQKA (KUKA AG, 2023b).

### Using the smartPAD

KUKA has implemented this system in their new cobots iiwa and iisy, which are primarily operated through the accompanying smartPAD. The focus lies on creating natural interactions and intuitive feedback to ensure a smooth and efficient work process for both new and advanced users. The programming interaction has been redesigned with a visual, block-based interface, allowing users to select commands from a palette of options, and program them using drag-and-drop. Each element can be modified with a parameters window on the right side. In addition to the touch controls, there are a few physical Buttons to access certain functions more quickly. Most notable among these, are the jogging buttons, which allow individual manipulation of axes and coordinates of the end effector, as well as the safety release buttons. These are used to activate a program, either in automatic mode or test mode. While in test mode speeds are limited, and the action will stop once the safety button is released, in automatic mode it will keep running until manually stopped. For both it is necessary to press two buttons, one safety switch, which needs to be kept in a middle position for additional safety, and the start button.


*Figure 2: KUKA SmartPAD Being Used to Program and Adjust a Cobot Application*

### Direct Manipulation

An additional feature becoming essential for the cobot operation is the handguiding mode, which allows users to manipulate the robotic arm directly, by grabbing the Commander and moving it into a desired position. These positions can then be added to the program, using the so-called touch up function. In this way the programmed positions can be related directly to the relevant constraints in physical space, rather than a digital representation, or abstract coordinates. This does a great part in making it more understandable and improving the learning experience for new users but can also make the experience of experts much more efficient and fluent. It allows the users to interact in a much more tangible, visual way, making it much more natural to program the desired path for the robot.

### Audio

While as mentioned some intuitive interaction features are already implemented, for an even more natural experience, another major sense is still open, being the hearing. Currently on the KUKA system no audio is employed by default. The only sound signals that might be used with them in a typical application is a siren connected to the overall production line, but this is not part of the basic product as sold by KUKA Robotics.

While no specifically designed sounds are used yet, the natural sounds this system creates just like any other electromechanical device should not be disregarded. Considering this there are a few sounds noticeable, some of which can be used by experienced operators to make conclusions about the cobot. The loudest sound audible on the system is the motor brake release, which happens when starting the motion, either by activating an automatic program, or when pressing the Commander buttons to start the handguiding mode.

Not directly part of the cobot as sold by KUKA, but still an essential part of the standard system are the grippers, or potentially other tools. These can also have quite noticeable sounds, depending on their type. Especially pneumatic grippers will create relatively loud hissing, but also electric grippers are usually noticeable. These existing noises might also need to be considered, to make sure the sounds fit this and are not masked, or otherwise influenced by them. The final sound noticeable is some silent whirring of the processor and cooling, which however won't be noticeable with additional noises covering it.

As found before, trained operators can make use of a lot of small signals they start to notice and understand through their work experience. There are not many natural sounds emitted by a system like this, compared to for example largely mechanical machines, and thus a lot of the internal workings can't be concluded and understood that deeply. To support this intuitive understanding as it was also described by (Alexanderson, 2004), artificial sounds can be added. For this, useful situations for employing them and useful messages to communicate and support the collaborative work need to be identified. By increasing the recognizability and effectiveness of signals, more information can be conveyed without overwhelming the user.

### *Learnability of the Interface*

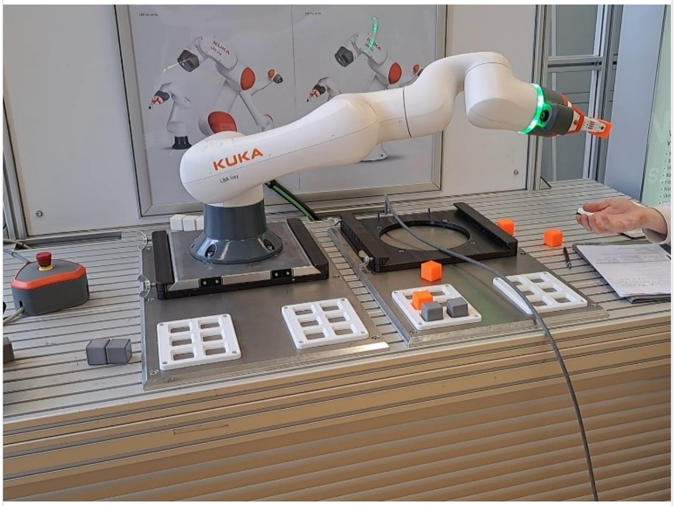


*Figure 3: KUKA LBR iisy setup used for self-test*

As the interaction and programming of cobots is meant to be learnable by everyone within hours, a self-test was used to evaluate this claim, as well as gathering as much first-hand experience about the general experience as possible. In addition, manuals and further material is available, as well as a quick guide series available in video from on YouTube, for users to get started easily. Any open questions could be answered by consulting company experts and analyzing the manuals and further internal or publicly available materials on the development and use of these systems. In this session it was possible to find out all the small obstacles in the learning process and general user interaction. As the cobot system is intended to also be intuitively understood by untrained users, this learning experience is important to consider. Any points that were not initially clear, might open up possibilities for supporting audio signals.
This session also played a substantial part in understanding the current system and its intricacies as described above. In the following the objectives of the session, and topics to be tested are summarized. Further concrete findings about the system interaction, that came up during the learning and use experience are also included:

- Programming Collaborative Tasks:
  - Writing Test Scripts: Creating test scripts to program various actions for the cobot, such as moving the arm to a specific position or gripping and releasing objects. A handoff was also programmed to try some more direct interaction with the cobot, where a cube was dropped in the hand of the user, and is also taken back
  - Utilizing Vision Systems, Incorporating Conditional programming or other smart systems to allow more interesting collaborative scenarios was unfortunately not possible yet on the given system. Using it however could already help inform how an interaction like that might behave in practice.
- Testing and Optimization:
  - Test Mode and Signal Feedback: Enter the test mode to verify possible sound signals and functionalities
  - Error Detection: Identify how the system responds when an action fails due to error messages

    - Safety Zone Verification: Create safety zones and test how the cobot recognizes and responds to them, and how noticeable they are in practice
- User Experience and Efficiency:
    - Cognitive Load Reduction: Explore how sound effects could help reduce cognitive load during programming
    - Workflow Enhancements: impact of individualized sound signals in the programming workflow
    - Improving Efficiency: the effectiveness of the system interface in improving programming speed and accuracy

The main objective however was to identify key moments in which additional communication might have been helpful. These concrete findings from the session, focusing on the interaction with the system for programming and executing collaborative applications, give several additional moments for coordinating between the cobot and worker:

- This action is not available
- Safety release successful
- Movement stopped, ready to continue
- Movement stopped with error, informing about the reason
- Moved back to intended path after deviation, ready to continue
- Movement concluded, target reached

These moments particularly stand out, as they were missed or misread in practice, requiring some additional checks of features like the error messages or other indicators. Checking this caused a delay and additional switching of the context in many of these cases. To keep users focused on the task at hand, audio signals seem useful to support understanding in these situations.

*Requirements:*

1. **Natural Sounds:** consider existing electromechanical sounds of the cobot, and design the sound around them so it gives a pleasant, coordinated impression, and signals are not masked by noises from things like the gripper

2. **Test Mode:** Include additional feedback in the test mode for program verification without influencing the automatic use too much

3. **Error Detection:** Let the user understand the required action quickly by communicating error messages, and info about the error type audibly, without breaking their flow

## 4.3 User

As mentioned, the development of cobots aims to address a much more diverse customer base, adding also Small and Medium-sized Enterprises (SMEs) and, therefore, often customers without dedicated robot programmers. The new configuration of robots focusses more on user friendliness in direct interaction and enabling everyone to program and work with them after minimal training. This also means, that while traditional robot systems required experts from various fields to ensure safety and efficient programming, cobot operators could take over all these roles themselves to some extent. That is why, for the sake of this analysis, the focus will lie on these diverse users themselves, who are working with the cobot on a day-to-day basis. While other stakeholders can have an influence of course, this evaluation focuses on an experiential standpoint, but still it should be ensured to consider all the tasks and requirements coming with this self-reliant approach. While the focus lies on the direct work experience, affecting the direct user and coworkers close by, the other stakeholders like buyers, robot trainers, safety experts, technical engineers, maintenance staff, and regulatory bodies also can play roles in shaping cobot sound design. They shall be kept in mind, and their influences on the interaction design considered, while also considering, that with an improved user-friendly design their roles can get more and more taken over by the operators themselves.

Firstly, the technical proficiency of users will be relevant. Whereas in the past, users had to be extensively trained in the underlying technology and operational guidelines, cobot technology allow for users to have vastly varying levels of technical insight. While some operators may come from industrial or manufacturing backgrounds and often have experience in operating industrial machinery, and some even with robotics, this isn't a prerequisite anymore. These workers will usually have some form of technical training and have gathered experiences throughout their work so many are used to operating more complex systems and machinery, where the user experience is not one of the first design requirements. In fact, in this way a well-designed cobot system might be what gives them an unfamiliar experience at first, as they are used to for example certain safety rules or standard procedures, some of which actively go against the intuitive way. For example, many machines can only be started using both hands for safety reasons, or vehicles going backwards will use a beeping tone as a warning. When some of these safety features and procedures become obsolete, due to the inherent safety of cobots, this might seem like a new situation to these, and steps might be taken to reassure them of the proper working of the cobot. There might be signals that are expected by these users from their previous experiences, and it might be considered to include some of them for the sake of user acceptance and comfort, and to make the behavior more expected. However of course they should still be critically evaluated for their actual use and their advantages and disadvantages. When operators are trained or experiences with traditional industrial robots, rather than collaborative ones, some practices might also need to be unlearned, especially relating to the safety aspect. With it being dangerous to come close to these in many cases, and their employment behind safety fences or light barriers that can stop the process, they might be especially or even overly careful around cobots at first too. Here it might be possible to aid their understanding and feeling of safety in order to facilitate relaxed, focused and efficient work, by making the capabilities and limitations of the cobots clear and understandable.

In general, through their deeper understanding of the technicalities and internal processes of the robot, they might also have a better understanding of and a higher need for more complex and detailed information. Also, this might enable them to filter out the information relevant to them more easily and ignore the rest, compared to novices, which might get confused or insecure when there are many signals, and they can't understand and allocate all of them appropriately.

Less experienced users, especially also considering the high possibility of career changers in these practices, might be influenced by completely different factors. Their expectations for concrete robot behavior might stem mainly from their cinematic portrayal in for example science fiction. While these can serve as a good source of inspiration for actually designing these sounds, as experienced sound designers put in work to realize underlying expectations on their behavior, without being limited by technical factors of their realization, they should carefully be evaluated. The goals in designing these are of course different, focusing more on things like the futuristic feel, emotional value, or engagement. Another more general source of expectations might be the technical devices like smartphones, computers, car infotainment systems, and other appliances from daily life, which have become standard these days. These can be applicable for novel users, but looking at the hardware of newer cobot models, the influence is likely to affect any user. User-centered design practices are also finding their way into industrial applications, making them look and behave more and more like what users are used to from operating a normal tablet. This development is mostly advantageous to the user experience, making it more streamlined and intuitive, as well as taking away the need for some of the training. However, care should be taken to not limit any uses trained operators have developed with these more complex systems in the pursuit of simplification. Related to developing sounds the main influence will be to consider the relevant connotations learned from all these influences and relating to them in an appropriate manner to avoid confusion.

While it can be said that general technological experience is becoming more common, as the generations that have grown up with it are becoming more prevalent in the work environment, there are of course still different levels of experience. The biggest gap, in this regard though, with people who have only been able to get in contact with digital technology later in their life is slowly leaving the workforce year by year as they are retiring. However, it can be said that cobots might also be beneficial for this group in particular, as it can allow them to participate in work life for longer, by relieving them of dangerous, straining, or force intensive tasks. This again underlines the need for intuitively understood interaction and communication, which should appeal to a wide range of users. Similar factors can apply to people with other physical or mental limitations, that might also be enabled by this technology, and generally in this regard making the communication less limited to specific modalities alone can definitely be seen as an improvement and chance for a more universal understanding.

As already mentioned earlier, especially with age limitations in hearing and frequency perception, will become apparent. While this means limiting the frequency, these high frequencies are often unpleasant for many users anyways. Additionally factors like hearing protection must be considered in the industry, to still ensure audibility of signals.

As these technologies are applied worldwide, cultural differences can play a role, for example in musical perception or understanding of certain cultural references. These should be used carefully to create a universally understood solution. Additionally, the use of language is obviously also affected by this problem, suggesting that non-speech audio might be more suitable for many cases.

Finally, the work setting, considering the colleagues present, as well as possible collaborations. For these, the sound notifications need to be considered, as the audio will often also be audible to others. While in some cases this might be favorable, for example by inspiring more close collaboration and help amongst coworkers, if they for example recognize a certain error sound, that they are already familiar with, there are also possible negative influences. For example, a worker might be embarrassed by causing many loud error sounds.

To ensure natural and in-line interactions, several factors must be considered:

1. Modalities: Interaction designs should utilize various modalities that humans naturally use in decision-making processes. Allowing for natural and intuitive interactions promotes trust, cooperation, and effective teamwork between humans and machines.
2. Data Visualization and Sonification: With the abundance of data in the 4th Industrial Revolution, cobots need to process a growing amount of information. Effective output methods, for example through visualization or in this case sonification techniques, complex data, patterns, and trends can be understood more intuitively and directly.

To create a comforting and satisfying experience, attention to small details is crucial:

1. Support flow and remove frustrations: The solution should be designed to support a seamless workflow, minimizing small frustrations and context switches, and ensuring that all elements work together harmoniously.
2. Pleasant sounds: Frequently used sounds should be pleasant and non-straining, while important sounds should be easily recognizable. Matching the audio appropriately with other modalities can also be a way to support this, if done well.

3. Flexibility and adjustment: The system should be adaptable to new user or environmental requirements, allowing for quick setup. It should prioritize supporting user preferences and workflows instead of imposing rigid structures. However, while customization can be beneficial, it must be limited to avoid compromising safety and the recognizability of important signals. Striking the right balance between customization and standardized functionality is essential.

By taking into account the discovered factors relevant about this diverse user group, and especially focusing on the concluded principles considering factors like modalities, data visualization, and a comforting, satisfying experience, the interaction between humans and cobots can be optimized for efficiency, safety, and overall user satisfaction.

#### *Tasks*

In general robots are said to be especially useful for work that is dull, dirty or dangerous, to relieve workers of these tasks. Additionally, their unique set of abilities differs from the human worker, in that it allows for high grades of precision, repeatability, speed, as well as force, depending on the type. Depending on the use case, cobots can also be used for similar tasks as traditional robots (which in practice is often how they are used (Zaatari et al., 2019; Michaelis et al., 2020)), their unique qualities like inherent safety and thus more flexible applicability allows for some new or altered use cases, especially when operating close to workers. In these cases, if the solution works effectively, the capabilities of both the human and the cobot can be used to their full potential. Throughout the research process a list of tasks that are especially suited for collaborative robots was kept and updated to get a good overview of the applications. While in the future new possibilities will most likely open up, this still enables a good overview of the field so far. Some of the common uses include:

- Loading and Unloading (from Machines)
- Pick and Place
- Packaging and Palletizing
- Testing, Measuring, Inspection
- Material Handling
- Performing Overhead Tasks
- Assembly of Automotive Transmissions
- Insertion of Rubber Plugs or Flexible Parts
- Assembly or Gluing Processes
- Sensitive Joining of Gears and other components
- Collaborative Tuning and Adjustment of Products
- Mechanical Machining, Polishing, and Grinding
- Application, Painting, and Gluing
- Welding (Spot and Arc Welding)
- Screwing, Fastening, and Joining Parts Together
- Handling Printed Circuit Boards (PCBs)

#### *User Journey Map*

While the previous list gives a broad overview of the use cases, a more detailed look at the intricacies in such a scenario is also necessary. In order to get a better overview of the interaction with cobots and analyze the potential from a user-centered point of view a type of user journey map was created, adapted from (Sarah Gibbons, 2017). In this type of representation, an interaction and all relevant factors like the user’s actions, and thoughts or info about the product or environment can be mapped out chronologically. It can be used as a point of reference, to get a better understanding of the industrial context, the types of interactions with a cobot, and the actions, reactions, and general experience of the worker using it. It allows for creating a broad overview, considering all phases of the interaction while being able to go into detail on relevant aspects and extend the analysis to find out about how a user might experience a specific situation. In this way the gathered information can be collected and reviewed in a structured way, giving insight into the broad context, as well as detailed insights into the relevant cobot interaction dynamics.

Based on the insights gained up to this point, a scenario was created to represent the typical interactions as well as possible further developments. For this the typical interactions as found in the industry today, for example shown in several case studies with KUKA's customers, as well as the intended further application as envisioned by the company and the academic field were collected. A scenario, representing a variety of the found activities as good as possible, was then created as a working example.

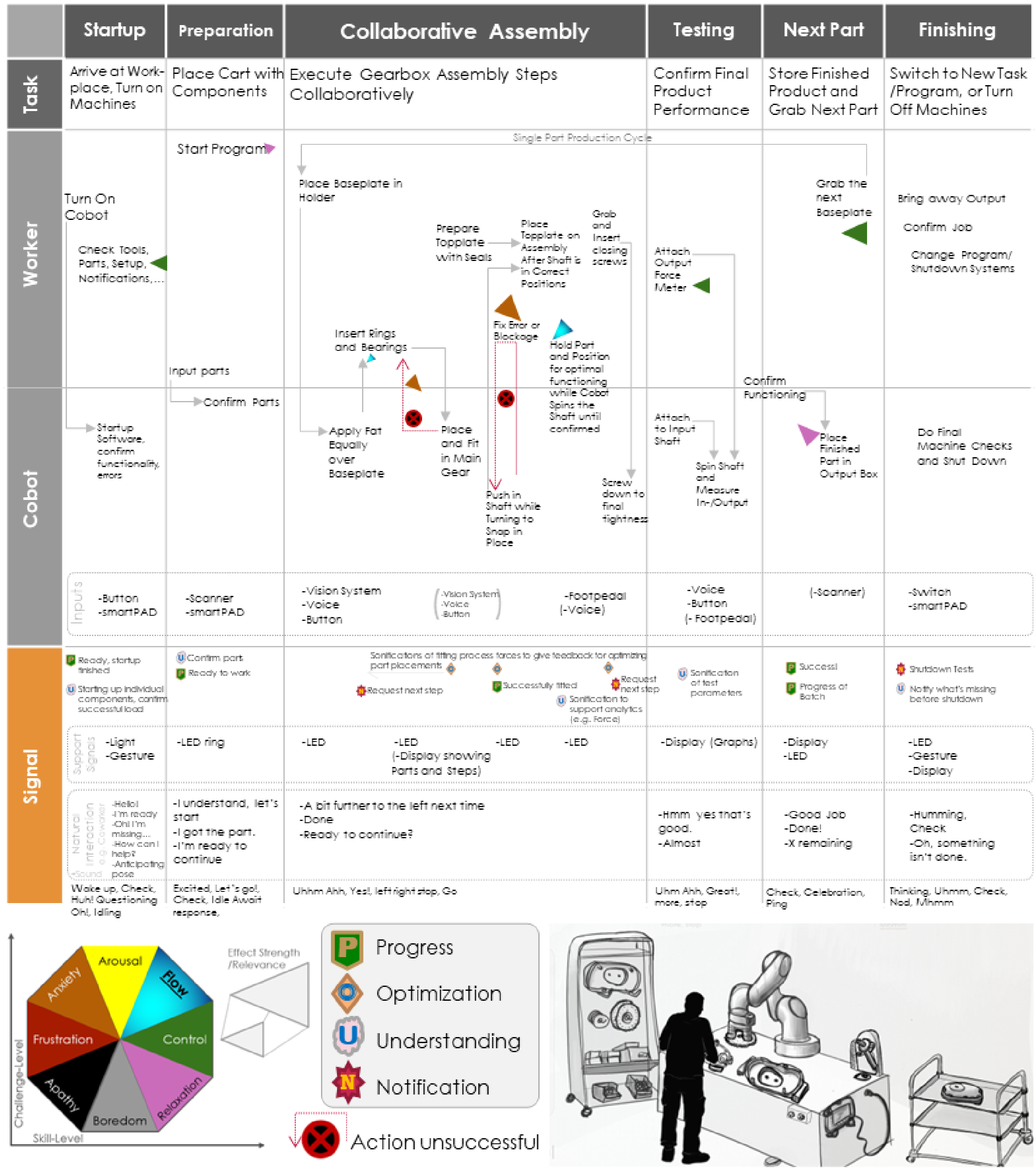


*Figure 4: Customer Journey Map Representing Collaborative Assembly of a Gearbox with a Cobot*

After mapping the general interaction, further associated findings could be added. This includes for example information about the mental state of the user. For this an adaption of the flow model as first proposed by (Cziksentmihalyi, 1990) was used. By connecting the customer journey map with the mental states from the Flow State Model, the link between cognitive and emotional states and the cobot interaction can be investigated. This interplay highlights the significance of user-centered design in optimizing cobot interactions. The integration of the Flow State Model into the map framework aims to aid enhancements in communication and intuitiveness, to craft an overall satisfying user experience.

It begins with the worker arriving at the workplace and getting ready to execute a task collaboratively with the cobot. It then goes through some of the required steps in preparation, before the actual assembly task can start. In this main part a few exemplary assembly steps are executed by the worker, collaborating with the cobot, by splitting the task and giving the workpiece back and forth, as well as collaborating directly by manipulating the part simultaneously. Similarly, the final product is then checked together making use of the cobot's accurate force output and measuring capabilities, and the worker's sensible perception and judgement. The assembly can then be repeated, or the work session at this station can be concluded.

Analysis and interpretation of the results of the laid-out example task could then be added within the map itself and could also be used and added to throughout the project, having it as a context to refer back to for further explorations and insights. Working from the individual situations mapped out, at first the relevant inputs used to communicate with the cobot were added. Here the most suitable modality was selected, fitting with the preferences of the users as found so far, rather than simply going from the current interaction. To get a better feel for the worker's perception and reactions to certain moments in the process, they were marked with the connected mental states from the flow model. It helped make out the relevant moments for communicating more clearly and supported the understanding of how users could react to different situations. For adding new signals taking this into account can be relevant to ensure effective and intuitive communication. For example, in very stressful moments denoted with frustration, anxiety or arousal, the user might not want to be interrupted more or might need additional signals, depending on the cause. In focused situations marked with control or especially flow, the user experience is in an optimal field for effective work, and changes to the signals should only be done cautiously, taking into account the user's points of interest, and what is going right in the situation.

Then as a main result of the analysis, the signals that might be used in each situation to communicate back to the worker were devised. For this, the focus was of course on audio signals, but the supportive signals using other modalities were noted down too to get a more complete picture. An additional step helping with uncovering more of these was to imagine the cobot as a coworker and think of the way they would communicate certain messages, to find the most natural way. Also, as proposed by Bergman et al., 2019, material from fictional works like animations, as well as our behavior and communication with animals can be sources of understandable robot interactions.

The found use cases of sound were then initially divided into several categories. These are signified with icons for Progress, Optimization, Understanding and Notification. These categorizations were then further worked with and modified in the following chapter to find the most valuable use-cases for the given goals.

*Requirements:*

- User Proficiency: Tailor interactions to users' technical backgrounds, enabling both novices and experts to engage with cobots intuitively. Design universally understood sounds using generally known influences to create intuitive understanding
- Safety Communication: Clearly convey cobot capabilities and limitations to users, particularly those experienced with traditional industrial robots.
- Workflow and Pleasant Experience: Create interactions that align with smooth workflows, minimizing disruptions and frustrations.
- Adaptation and Flexibility: Offer customizable interactions to some extent to cover user differences, while ensuring safety and standardized functionality, as well as simple setup.

# 5 Concept Scenarios

## 5.1 Scenarios

Next, the findings made in previous chapters in various areas, like the industrial context, cobot user needs, the KUKA system characteristics, and auditive HMI in general, were used to start defining clear solution concepts, for the found challenges. For this also previously defined goals like fast learning and understanding, safety, or intuitive collaboration came into play again.

The previous analysis of the typical cobot use cases, and the users' role, used a broad view on generally possible tasks, aswell as a closer view on an example case in the user journey map. Combining this with the knowledge about the possibilities and characteristics of using sound as researched before, individual aspects and ideas of the auditive communication could be ideated. By analyzing typical use cases for cobots, and focusing on creating valuable additions to the experience, the sound use scenarios were divided into several categories. For this also the analysis from the Customer Journey Mapping was used, which looked at an example case of assembling a gearbox and the relevant personal and environmental factors. These already pointed to some situations that might require additional communication to be more intuitive. Similarly, the self-test session added some first-hand experiences like the difficulties of understanding some processes or errors.

To identify useful moments for implementation of sound signals, the created insights were then used to refine the cobot use cases again with this focus. Here only one concrete situation of cobot interaction is portrayed, and many variations and different environments or situations are possible, but it still was a good starting point for imagining these more accurately and finding more detailed interactions. Also, the situation was chosen in a way that covers an as wide range of scenarios as possible, while still being logical and consistent and based on common industry practices. In this way an efficient preselection of audio use scenarios could then be made, focusing on key moments that could communicating the cobots internal workings:

- Handing off part
- Monitoring progress
- Identifying errors
- Guiding positioning, orientation, recognition, …
- Quality Analysis Feedback
- Debug & Test Programs
- Optimize Paths etc. (e.g. Power Use)
- Preventative Maintenance
- Axis Mastering

Considering the user needs and the sound interface design principles discovered in literature research, it became apparent that a first audio concept needs to be focused and adding real value towards goals like safety, collaboration, and efficient user experience. While the possible richness of sound shall be made use of as much as possible, a clear focus of the development efforts for this thesis is required, to test these scenarios more thoroughly. For this especially the user centric needs in the context, and the current and envisioned industry relevant cobot applications were considered. The resulting uses could be classified as follows:

1. Alerts
2. Progress
3. Understanding & Optimization
4. Task Guidance
5. Virtual Assistant

The first two categories are already being employed much more broadly than the others. While they should be designed just as carefully as the other categories, they don't offer as much potential for communicating deeper level information, as they are bound to their specific use (i.e., notifying about specific fixed alerts or steps in the process). They will therefore not be the focus of this study, as their implementation is more straightforward. However, of course in the end the whole sound design concept will need to work as one system, as each sound influences the others in terms of for example frequency and distinguishability. Therefore, they will be revisited in the general concept in the end and modified to match the overall system.

On the other hand, the last two are much less implemented and developed yet, and they will be handled in a similar way, as again, they are relatively bound by their use. They might need a separate specific analysis to create a useful and effective system, that will also be more like a separate addon, rather than an enhancement to the current one.

This means, the focus of this study will lie on the *Understanding and Optimization* Category. This offers a lot of options for communicating internal workings of the cobot to the worker and become especially relevant in collaborative applications being developed currently. In this way the efforts can be focused on developing an effective sound design system that caters to the Understanding and Optimization category, which likely will play an even more crucial role in collaborative applications. Here efficient interaction is more crucial, as interactions happen directly and closely to each other and more communication is necessary, compared to a system that is separated from the operator by a safety fence, and only operated from the outside through a GUI. In this way the interactions can become much more natural, and more similar to interactions with a (human) coworker, rather than a machine or computer system, which is why added communication in this category will be most beneficial to start off the development of an audio interaction system.

The ideas coming up in this area to that point could be classified into 3 concept areas, which could offer potential for further exploration:

| C1 - Movement Sonifications | C2 - Data Sonifications | C3 - Notification and Debugging Signals |
|---|---|---|
| adding appropriate sounds to movements of the cobot, and especially to relevant changes and stops in those movements. Communicate the reasons better, or make them more expectable; e.g. safety stop, obstacle was hit, safety area entered, movement started | Focus on the cobots internal sensing and data processing, and making relevant aspects of it audible, to create a better understanding of its workings; e.g. loading different elements, current commands, vision system recognition parameters, presence recognition | simple notification signals informing of system events and states, to get the attention of the user aswell as communicate a given message. Could also be added manually by users inside the program to improve coordination and collaboration in specific use cases |

*Table 4: Concept Scenarios Used for Further Evaluation and Focus of Development and Testing Efforts*

## 5.2 Evaluation

As mentioned, the chosen method of this project includes exploring the potential of sound from two sides, the unique characteristics and possibilities sound specifically provides, as well as the needs of users in this regard. In this way the most favorable, synergetic combination can be found, to start incorporating auditive communication at the points where it is most impactful. After analyzing these points in user and literature research, narrowing down the solution concepts was concretized using an evaluation scheme. This scheme and its main focus areas were created in coordination with the project supervisors, to focus the further, more concrete research on the relevant areas.

Specifically, the areas of effectiveness of the solution to enhance the communication and understanding, feasibility of implementing it, safety aspects, as well as ergonomic and personal factors influencing the user experience were identified. While the related work chapter and research on the usage of audio clearly showed the potential of audio the enhance HRC, the feasibility factor came up through consultation with the project supervisors, and other company experts, as well as the general analysis of the use context. Safety is of course a factor, and especially in the context of cobots which should offer this inherently this factor has been a goal since the start of the project. Additionally, the user research, including the own learning and use experience, as well as colleagues' experiences, made a number of ergonomic and personal factors apparent, which could greatly influence the user experience, and thus in the end also the effectiveness of the whole audio solution. To make the evaluation of these found criteria more nuanced on these areas, sub-questions were developed for each, to fully capture the relevant aspects.

| fulfills this category<br>0 - not at all<br>1 - barely<br>2 - a bit<br>3 - reasonably well<br>4 - well<br>5 - perfectly | **C1 - Movement Sonifications**<br>adding appropriate sounds to movements, and especially relevant movement changes and stops | **C2 - Data Sonifications**<br>make cobots internal sensing and data processing audible; e.g. loading different elements, current commands, vision system recognition parameters, presence recognition,... | **C3 - Notification and Debugging Signals**<br>simple notificarion signals informing of system events and states, or being added manually to improve coordination and collaboration in specific cases |
|---|---|---|---|
| **Effectiveness in enhancing human robot communication** | **4,50** | **3,75** | **4,25** |
| How easy and natural is it for users to understand and use the audio communication system? | 5 | 3 | 5 |
| How effective is the audio communication system in improving understanding of the cobot processes? | 5 | 4 | 4 |
| How satisfied would users be with the function of the audio communication system in the long term? | 5 | 4 | 3 |
| How adaptable is the audio communication system to changing user needs and task requirements? | 3 | 4 | 5 |
| **Feasibility of implementation** | **4,50** | **3,75** | **4,00** |
| How much does the audio communication system improve work efficiency? | 4 | 4 | 4 |
| How compatible is the audio communication system with existing collaborative robot systems, different robot types, machines, and industrial systems? | 5 | 4 | 3 |
| How compliant is the audio communication system with relevant standards and regulations, and does it fit into the oveall safety and communication concept? | 5 | 3 | 4 |
| How feasible is the implementation of the audio communication system technologically? | 4 | 4 | 5 |
| **Safety** | **4,33** | **3,33** | **3,67** |
| How effective is the audio communication system in enhancing situational awareness and safety for both the human and the robot? | 4 | 5 | 4 |
| How reliable is the audio communication system in noisy and dynamic environments? | 4 | 2 | 3 |
| How well does the solution integrate with current relevant standards and regulations, and how well does it avoid confusion with preexisting signals? | 5 | 3 | 4 |
| **Ergonomic / Personal** | **5,00** | **3,25** | **3,25** |
| Is the advantage of audio used in this implementation? Does it use sound in an appropriate way, and is it supplemented in a multimodal way where appropriate? | 5 | 4 | 3 |
| How useful is the audio communication system rather than distracting, annoying, or being ignored when used over longer periods of time? | 5 | 3 | 4 |
| How human-like is the audio output and is it able to convey contents naturally? | 5 | 4 | 3 |
| How well does the audio solution integrate into the work environment overall, while taking into account social considerations, including privacy and collaboration? | 5 | 2 | 3 |
| **Overall Score** | **4,58** | **3,52** | **3,79** |

*Table 5: Concept Scenario Evaluation Results Based on Communication Enhancement, Feasibility, Safety and Ergonomic Factors*

Based on the previous analyses several example scenarios were selected, to achieve a good balance of existing or realistic industry practices and openness for future changes. These will allow considerations about different ways of implementing audio, that are grounded in real practices and well imaginable, while still allowing for new interactions outside of the established paradigm to be considered. For this reason, while a focus area has been selected, the goal is to select scenarios to cover a wide range of applications within that area. In this way, the chosen scenarios open up varied opportunities to explore potential applications of sound. The selection was mainly influenced by the previously executed user research and combined influences from intended cobot applications of the field in a research perspective, as well as practical insights from users, experts, as well as from the self-test session giving a closer relation to the specific KUKA cobot system characteristics. The chosen categories were then judged based on this information, giving them a numerical value from zero to five, and the final evaluation again discussed with the professor.

While C2 and C3 might also be useful, they are separated at this point for two main reasons. Firstly, as mentioned, the system currently doesn't use audio signals by default, so adding many different systems at once without being able to iterate and fit them together might overload users and limit the use and understanding of it in the end.

Secondly, the split is done in a way, that might require a different approach for implementing the other concepts, as the signals represent a whole different type of data. Since the goal of this thesis is to, in the end, go back to the underlying principles, it makes sense to get a deeper insight into the principles related to the one general concept at first. At a later step these could then be evaluated for their universality and fit to the other concepts, and they can be implemented with relation to some already established sounds and principles, allowing for simpler iteration.

The Ideas were then worked out more concretely in the defined area, considering findings like the relevant usage areas, as well as discovered principles on the use of audio, in relation to the movement aspect especially the discussed sonification techniques. Specifically, the following concrete audio concepts could be identified which shall be used for further testing:

- Safety release (successful and unsuccessful actuation of the safety switch to start the program)
- Safety sensor stop (a safety sensor recognizes an obstacle like a worker entering a dangerous area and stops the process)
- Safety Area Speed Limiting(/stop) (cobot enters a defined safety area and slows its movement)
- Handoff (notifying about needing to give or take something from or to the cobot)
- Axis Limit (The cobot approaches its axis limits, indicating the limit of its reach or movement)
- Hand guiding Position Information (Letting you hear digital elements while moving the arm in hand guiding mode)

This selection of concepts is aimed at addressing the found needs, possibilities and techniques optimally, to add value to an efficient, safe and pleasant work experience. These solutions will be developed and prototyped further to enable the testing of the concepts with some users.

# 6 Testing

In this chapter, a mixed testing approach, focusing on open discussion, is used to evaluate several prototype sounds, and test and improve the underlying principles accordingly. The testing strategy aims to measure the initial reactions and interpretations of different participants to these sounds and assess their capacity to transmit the intended messages accurately. The effectiveness of the principles can then be judged on this basis to be able to iterate on them effectively.

For the test, a combination of open interview and blind testing the sound prototypes was executed, and some additional quantitative data was gathered. Through this, ideation and open discussion can effectively be stimulated, while being able to steer and go more into detail where needed. This mixed approach provides a balance between gathering detailed qualitative insights and obtaining quantifiable data for objective comparison.

## 6.1 Methodology

The methodology employed for this testing section was aimed at evaluating the preliminary audio interaction principles discovered so far and gathering more input for refining them according to user feedback. While for this test a number of collaborative scenarios were chosen, the overarching goal was to use the feedback to conclude back to the underlying principles in order to create futureproof principles for enhancing the interaction with cobots.

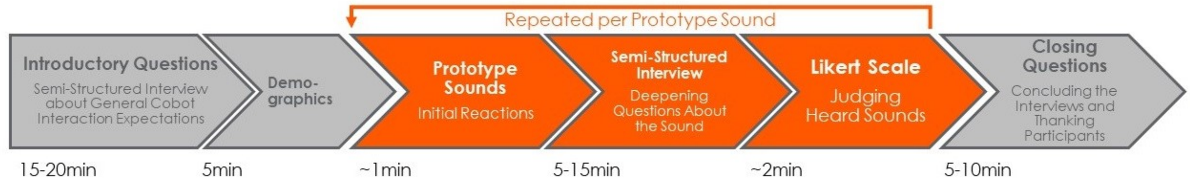



*Figure 5: Testing Procedure Executed With Each Participant*

To initiate the testing, a few semi-structured interview questions were used, to assess the preexisting expectations of participants for how future cobot interaction might look like. During this stage, participants were encouraged to envision future cobot interactions and especially the aspects regarding its communication. The goal was to let them describe their preconceived expectations in a detailed way, imagining their ideal interaction, unhindered by the constraints of current practices or technical limitations. Without telling them details about the project or its goals yet, they could give their ideas in this first part, so they are not influenced by the further parts of the testing.

Next, after gathering some general demographic information for context, relating to their person, work, and technical experience, the main testing part of the interview began.

This main part consisted of blind testing the create prototype sounds, by playing them to participants and letting them express their first reactions. For this, at first, they did not receive any information about the intended meaning or starting situation. This approach aimed to obtain unbiased initial reactions, and their spontaneous interpretation of the sounds, free from any preconceived notions or contextual implications. Additionally, the order of the sounds was altered between interviews, to prevent bias from affecting the results.

After expressing all their first impressions, reaction and interpretation of the message for the heard sound, and answering some deepening or clarifying questions where needed, the intended purpose and context could be revealed to the participant. At this point the open interview could be continued by asking about the perceived match between the sound and message, and suggestions for refining or changing the prototype.

After concluding the questioning for a sound, participants used the Likert scale questionnaire to quantify and note down their reaction and opinion on several aspects, such as recognizability, clarity, effectiveness, and emotional response.

The process didn't take place under laboratory conditions, but rather in a natural environment, similar to, or at their usual workplace. This can have the advantage of making the participants feel familiar and comfortable and could aid them in imagining the described situations more accurately and remembering details. Also, this allowed them to relate to their specific situation and show certain aspects in the real environment, where necessary. This was of course especially the case for the participants who work with a cobot directly at their workplace as this meant they had the option to show certain aspects directly on the actual product. In these environments, listening to the prototype sounds under the influence of the environmental soundscape also made the test closer to the real work situation, possibly influencing their perception regarding for example audibility, and thus resulting in conclusions closer to the actual situation in practice. For the same reason, a small portable speaker was chosen to play the sounds, rather than a bigger HiFi-System or headphones, as this is likely closest to what might be built into future cobots. However, headphones were also brought to allow participants to listen to some sounds again a bit more closely when desired.

To be able to engage meaningfully with participants, and deepen the understanding of their ideas, the process was designed to be relatively open. To be able to engage with participants meaningfully and focus on the conversation, no notes have been taken during the interview. Instead, the talks were recorded, with their permission, to be able to evaluate the information more accurately later on. This process enabled the gathering of more rich, accurate qualitative data, and nuanced insights that may not be captured through structured questionnaires alone. This was especially needed, as the tests were not only intended to verify the concrete sounds but also to make conclusions about the underlying principles.

## 6.2 Hypotheses

Going over to concept testing, this section again recollects the hypotheses focused on the aims of the methodical exploration of audio concepts. The gained understanding of auditory communication principles and their potential impact on collaborative robotics, will again be considered to find the most suitable ways of implementing new signals in the defined area of movement sonification. The hypotheses thus relate to previous research insights and goals, that will guide the and interviews, exploration and testing of the conceptual prototypes:

**H1** Utilizing varying levels of loudness and aggressiveness for sounds can effectively communicate information of varying urgency levels, in different contextual situations, thereby enhancing user awareness without causing discomfort.

**H2** Implementing dynamic sounds guided by sonification principles can enrich the auditory signals with information, making them more comprehensible and prompting appropriate user responses.

**H3** By incorporating movement principles and skeuomorphic sounds corresponding to the communicated message, the sound communication can be made more intelligible, easy to learn and recall for novices, as well as experienced robot operators.

**H4** Users more experienced with cobot systems or specifically the KUKA system might have more specific and more elaborate expectations for the communication.

**H5** Dynamic sounds will be more useful, informative, and well perceived than simple Earcons for some situations.

The second part of the hypotheses includes of course the effective working of the principles that will be used to create the prototype sounds, as well as the usefulness of the signals in the situations chosen, as perceived by the various users. These will be explained in depth in the following chapter. H1 was concluded from the research of acoustic aswell as cognitive factors related to the use of sound, and especially the practical example as described by Alexanderson (2004), which points to the use of such effects also for designed sounds. Similarly, H3 was based on the analyzed cognitive factors of sounds perception, aswell as the transfer of these well known design principles from other areas towards audio. H2 was based on the description and effective application of audio principles as by Barrass (1997) aswell as Stienstra et al. (2011). Through user research and general experiences with users and experts throughout the process H4 was added. The final hypothesis H5 then combines some of the initial observations of problems with audio communication with the researched cognitive and personal factors, as well as successful principles from audio and related areas.

## 6.3 Test Setup

Firstly, in preparation of this test session, prototype sounds needed to be developed. They were created based on the current state of the research executed, and preliminary principles already established during research. These principles did not only pertain to the tonal qualities of future sounds and how they are mapped to the interactions, but also included some ideas on when audio should be employed at all. For this it was also necessary to envision the possibilities for future interaction that are not realized yet, as communication might get more and more relevant, the more intricate and varied the interactions are.

Also, since this thesis focusses more on establishing basic, futureproof principles, rather than specific solutions for the current situation, a wide range of interactions should be evaluated with an open mind. For this reason, the sounds for this test will merely be created as a prototype to further evaluate and refine the interaction principles. These are meant to be formulated in a broad and flexible way, making them widely applicable. Even under possibly unforeseen, changed circumstances, like potential changes or advancements in robot technology and interaction paradigms, their relevance and effectiveness should be maintained. In this way alignment of the sound concept to the interactions with current as well as future collaborative robots can be supported.

The chosen situations were however still somewhat in line with the current interaction paradigm, as it still needs to be understandable and relatable to participants. Future options and enhancements of the functionality were however not excluded, and where possible the open interview methodology was used to stimulate thoughts about new types of interaction functionality. Participants were either interviewed alone, or in pairs, to allow for in-depth discussion of the underlying principles and reactions to the prototypes.

### Audio Prototyping

The sound creation used one main monophonic synthesizer, to create all sounds. This allowed for a variety of sounds through its semi-modular capabilities, while still keeping all sounds somehow related to each other acoustically, crafting a coherent sounds system as intended. It allowed for distinct audio profiles clearly audible in industrial environments, and the exact and intuitive manipulation of a range of audio parameters. A single oscillator generated square or triangle waveforms, which could be manipulated in pitch using a keyboard as well as continuously. An LFO (Low-Frequency Oscillator) introduced pitch and amplitude fluctuations, bringing dynamic variance relating to the timbre or creating a vibrato effect.

Further an effects unit was used, which could be regulated through the mixer: delay added echoes for spatial presence, distortion introduced a rougher timbre, reverb added spaciousness, and modulation effects further varied the tone for texture. Sounds were blended and controlled using a mixer and recorded using and audio input on a computer. Mixer feedback loops were also sometimes intentionally used to create intricate soundscapes with a commonly recognized noise for critical conditions.

Pitch was varied for urgency levels, timbre altered using LFO for organic variation, modulation controlled through effects added complexity, and amplitude adjusted for distinct audibility. This synthesized approach, considering pitch, timbre, modulation, and amplitude, resulted in purposeful sounds for human-cobot interactions. The intuitive control through the various simultaneously available knobs controlling sound parameters allowed for efficient exploration of different possibilities and fast iteration until a suitable sound was found for each scenario. After recording the sounds were normalized and adjusted for a proper volume, as well as removing any unwanted hum frequencies or distortions from recording as far as possible to increase their clarity.

In the following, insights into the concrete development of each sound shall be given, focusing on the used techniques, as well as underlying principles from previous research as applied in practice:

*Safety Release*
(Successful and unsuccessful actuation of the safety switch to start the program or movement)

- Smooth vs textured sound to represent positive and negative action
- Two-tone with upwards movement to support positive message
- First part activated together with the first button and second start sound when the second button is pressed and the program is started
- Short duration and decay to fit with fast and frequent action, and the user is already focused and expecting the signal
- Adding weight to the last sound (notifying about program start) by increasing decay time

*Safety Sensor Stop*
(A safety sensor recognizes an obstacle like a worker entering a dangerous area and stops the process once it gets too close)

- Represent scanning process, higher pitched whirring or hum, fast pulsing or repeating,
- pulsating fitting its frequency, reminding of scanner which could possibly be rotating
- first part dynamically mapped to distance of the workers to the safety zone to give an intuitive relation to their actions
- second part flowing over into more recognizable beeps, similar to a soft alarm, signifying the activation of the safety feature

*Axis Limit*
(The cobot approaches its axis limits, indicating the limit of its reach or movement)

- Slightly negative, deep hum or growl, not safety relevant so more subtle,
- deep and muffled to be not too distracting, but noticeable in direct interaction
- Dynamic mapping to the force/deviation into the limit, like a forcefield, or like straining the motor past its limit. Makes the sound directly relate to the movement done by the user to make them aware of the limit intuitively, through the combination of tactile and auditive signals
- Sound mostly staying low to not bother too much, and rising exponentially in the end when limit is reached, making the limit clearly experienceable

*Safety Area Speed Limiting / Stop*
(The cobot enters a defined safety area and slows its movement)

- Pitch down, aswell as increasing the gaps of the oscillating/repeating sound, like lowering the rotational frequency of the sound source represented, giving a time stretching effect
- Association with deceleration or decreased rotational frequency of the motor
- Recognizable also through the relation to a slowing down vehicle
- Kept short while achieving the slowdown effect, to stay connected to the event of crossing the border
- Rate of slowdown might be mapped to the actual speeds, translated into the pitches. Similarly, a speeding up sound might be created

*Handoff*
(Notifying about needing to give or take something from or to the cobot)

- Intended to be a slight nudge to get attention without being too demanding or annoying
- Shouldn't be too long as users might already know what to do
- Repeated three times was found to be a good balance between making sure its audible and not perceived as a natural sound, and not being too stressful and demanding

- Similar to knocking or tapping on the table

*Hand guiding Position Information*

(Letting you hear digital elements while moving the robot through the area in hand guiding mode)

- No concrete sounds created as this feature in general is more open to interpretation and allows for a number of different options.
- Also, this sound might be the hardest to imagine, or to represent easily, without the actual connected multimodal experience of moving the arm in combination with these sounds.
- Concrete audio prototype for this scenario was left open and used to ask participants for their own ideas and opinions on this.

## Participants

For the test 5 participants from different fields were selected, who are at least somewhat familiar with cobot work, be it through theoretical knowledge or practical experiences. While the participants were all selected from within the company, a variation of different fields and experience levels were chosen to represent the varied potential user group of cobots, that isn’t strictly limited to trained robot programmers anymore. In total 5 Interviews were conducted, with one of the sessions being a group interview with two participants. One of the interviews was conducted in an online meeting, instead of on site, and another one was done in written form instead of face to face, due to availability. The face-to-face interviews were conducted in their work environment, which means there were different robots close by, and the soundscape was like the industrial one, although less loud, as the areas are used for testing and training, rather than full scale manufacturing.

The Participants were all in the age range between 25 and 45 years, which covered the general employee demographic relatively well. Their experiences with robotics ranged from 1-10 years, and their approach to this field is coming from different directions. Two of the participants were experienced with UX design, and each had further insights into specific aspects of the technical system, or cobot vision. Three participants had extensive practical experience in using and setting up various robotic systems, two of which were robotics trainers from the in-house KUKA College. Through this they not only have extensive own interaction experiences, but also can get a good grasp on the assumptions, habits and frequent problems of novice users. The other participant was grounding his experience on deep technical insights gathered through engineering, development, prototyping and set-up work on various robotic systems.

While the small sample size didn’t allow for elaborate quantitative evaluation, it enabled a deeper focus in the open interview part, helping with the understanding of the perception of the underlying sound principles, which is the focus of this research.

## Testing Session

After greeting the participants, giving a very basic introduction about the plan for the session, and starting the audio recording, the interview could begin. As participants shouldn’t be influenced, this explanation only contained basic info about the plan for the session, about the thesis, and the basic topic of cobot interaction. Participants were asked to freely express any ideas and reactions they have, and that anything they think of is a valid response, as this is what the prototype sounds were created for. All participants only received this basic info prior to starting the standardized presentation leading through the interview played to them on a laptop, to ensure consistency in their understanding of the test.

The first, pre-interview part of the interview contained the following questions, intended to get the participants talking, and imagine the future of cobot interaction.

1. What specific sounds do you imagine a future cobot to make when coordinating a collaborative task?
2. Could you describe a situation where info from the cobot (through a sound cue) would have been helpful?
3. In which situations do you think sound can be especially helpful compared to other modalities?
4. What specific sounds and messages would you find comforting, supportive, and encouraging?
5. At which point would sounds become repetitive, annoying or distracting to your work? How would you deal with this, and how would you expect the cobot to avoid this?

This was executed as an open interview part, with deepening questions, checking their underlying beliefs and expectations, and open discussion. Participants were asked to freely express their thoughts, opinions, and expectations related to the sound prototypes and cobot interaction.

In the following Blind audio prototype test, the reactions without prior knowledge of their intended purpose or function were tested. It allowed participants to provide their immediate reactions and attempt to guess the meaning or purpose of each sound. While actually getting the right answer is often hard without context, this wasn't really the goal, as any impression or connotation they could express was helpful as feedback and could still be related to the actual message in some way. Per sound, this interview part took about ten minutes.

The use of the Likert Questionnaire, especially during the interview, while still being able to talk and discuss their thoughts was also sometimes helpful to reconsider their ideas and mention additional thoughts. In the later analysis it also acted as a helpful supplementary tool for some of the more ambiguous replies.

Data was collected by audio recording, to allow for full engagement with the participants, and was later coded in table form to relate and compare users' feedback and look for similarities and differences. This was an important step, as in many cases users answered thematically similarly, however explained in a much different way or in response to another question.

With the last sound concept to present being left open, also stimulated the open discussion again, leading nicely out of the structured testing part into the final discussion of additional ideas.

## 6.4 Results

The testing phase provided valuable insights into the user perception and effectiveness of the designed auditory cues. While only a small number of participants was interviewed, and the quantitative implications thus are limited, the chosen open interview method allowed for in-depth discussion of relevant aspects. In this way a better insight could be gained into the underlying principles and reasons for users' preferences, which is the main focus of this thesis. While many insights, especially about concrete details about each specific sound could already be gained throughout the tests itself, later analysis of the collected feedback of all participants in the style of a thematic analysis revealed additional overall themes and allowed to put their opinions into perspective. In the following the most relevant themes discovered, some of the more disputed opinions, as well as points that found agreement amongst participants are going to be pointed out.

### *Message Frequency*

Firstly, a topic which was already considered before the testing was the balance of message frequency, and some insights could be gained into users' opinion on these. Things that happen too frequently and aren't that useful or annoying are of course not wanted to have a sound. However, the value of small more silent signal noises was also recognized, when there is some value in added understanding. However, these need to be implemented in a sensible way, and users also preferred there to be an option to disable these non-safety-relevant sounds, ate least by an administrator. Also, the thought of using a few basic profiles, which could influence this and similar settings, to quickly set the cobot behavior to an appropriate state for the given environment and task, without needing to go through many individual options. Such a solution is to be favored generally for the use of sound, as sounds are not the main focus of a cobot user, and they shouldn't clutter the system and obstruct users way towards the needed cobot functionalities.

#### *Noticeability and Matching Urgency*

Secondly, a number of sounds were perceived as too unnoticeable for the given environment, by participants. Especially this was the case for more natural sounding signals, which were mentioned to be easily missed or mistaken for a natural, unrelated sound. While in some cases this was intentional, to limit the effect of attention overload for less urgent messages this might still need to be investigated more closely. Specifically, the prototype sounds created for the handoff, which was like a muffled knocking sound, aswell by one participant the safety release sounds were mentioned. While the second sound already had a bit more melodic content than the largely atonal knocking sound, it still used a short decay time making it more likely to miss it. The method of repeating the tone rhythmically three times as done with the knocking sound does increase the recognizability, however depending on the soundscape, context, urgency and distance of the recipient, it might still be too muffled. However, as also recognized by participants, depending on the specific task, the urgency and the frequency of hearing this signal, a less intrusive sound might be preferred. For this reason, especially for this handoff, and maybe similar notification sounds, variations with different urgency might be advisable to balance the amount of attention they demand. For the safety release sound additional tests in context will be necessary to determine the right amount of urgency. As this interaction might happen quite frequently, and it is only a supportive sound to an action on which the user is already focused, over longer use periods, a more subtle sound might be preferred. However, one participant proposed another alteration of this sound that might also improve this, mentioning that the first part of the sound might already have a melodically positive connotation by making it two notes going upwards. As the first part of the sound is played at the first button press, which can either be successful or unsuccessful, it might be good to make this distinction clearer directly. By adding a two-tone for the first part, and keeping the second tone aswell, the overall sound would still sound positive, while the first part standing on its own already points to this meaning, avoiding confusion. As two participants also were unsure about the relation between the positive and negative version of the sound, the negative sound might also be implemented as a two-tone. The opposing consideration for this was, to not make the technical failure be perceived as even more negative by the users, which is why it was kept rather short. This again will need to be iterated more in context, to find a good balance. However especially with the addition of the first two-tone, the differentiability might be reduced again, so keeping it, or possibly adding a second attack to the sound without a gap might be a better solution overall, to balance the mentioned effects. The balancing act for the urgency of a sound thus became apparent again, and in many cases, users were undecided about the matching. The urgency depends on many factors in the specific use case, so it might indeed be required to use different versions of the sounds to represent different levels. As the value of both kinds of signals became apparent in testing, this balance might need further investigation, maybe with codesign sessions in context, before it can be applied concretely. Also, a possibility to vary it per environment or situation can be proposed, to give the users themselves the option to adjust it for their use case, by setting a slider for some sounds to increase their urgency. This might be useful on a global level, as well as making individual instances of sounds more urgent, for example to distinguish between different handoffs.

### *Close Collaboration*

Further, it became apparent, that while using sound still has large importance for situations where the user is further away or otherwise not paying attention to the cobot, the inverse setting also offers many useful situations for sound enhancements. When working closely together with the cobot there are still a number of situations in which communicating through audio is the most natural option, compared to other modalities alone. While some examples like the previously discussed safety release or other errors preventing the users from starting the robot could be considered not paying attention, it was also considered useful as supplemental feedback in some situations. In this regard, especially the possible effect of adding weight to an action that is possible to achieve through sound can be mentioned. While the user is paying attention to the action itself, sometimes the consequences are not immediately visible, and might only be noticed later on. An example it the saving of certain data like coordinates, which can be done especially quickly using the Touch Up function on the KUKA system. It updates a frame/point to the current coordinates of the arm, which is not always wanted as it can cause drastic changes to the program if done at the wrong time. Here an auditive signal, making users aware of the change they just did can make them realize this much more easily, when they get used to the differing sound. While currently a confirmation pop-up is used to make sure it's not activated accidentally, as reported by participants, practice has shown that this will often be accepted instinctively without further thought, out of habit. It might be argued that an additional sound notification will add weight to the action executed, making the user directly aware of the change and thus learning more quickly where mistakes might happen. If done unintentionally it is also much easier for them to correct it directly, rather than only finding out in a test run, and this more immediate feedback could help new users learn the system more quickly, as well as possibly relieving some mental load even for experienced users.

### *Use in Test Mode*

In general, it was also noticed, that operation in test mode, in many cases, was the more fitting situation for audio feedback, compared to the automatic mode. Here there are much more things to check and keep in mind for users, while in automatic mode things are expected to run smoothly in the background without interruption in many cases. While the study of Tennent et al. looked only on the consequential sounds created by motor noises, the finding might also be extended to additional designed sounds as used in these tests. They found, that "when compared to a silent robot, high-end sound was able to increase perceived competence when the robot interacted with a human, but decrease perceived competence when the robot was not interacting with a human" (Tennent et al., 2017, p. 933). Concluding, supportive signaling might be kept to a minimum in automatic mode and limited to important notifications and warnings requiring the attention of the user. This also fits with the general preference of users to use sound only as much as really needed, which was mentioned by 4 out of 5 participants.

### *User Proposals*

In the open parts of the session, a number of sound use scenarios were also freely proposed by the participants, either in the pre-interview, without knowing about the prototype sounds yet, or after the test as additional ideas. The following findings were made in this regard:

- Error signals should have sounds; different errors should be recognizable through their signal to some extent (e.g. only severity or category to limit amount of sounds)
- Signal the need for interaction / waiting
- Confirmation sound for inputs (especially touchup executed)
- Auralize sensitive joining process (e.g. forces the robot exerts while trying to connect parts)
- Loop count for progress information
- Success sound for a test run without errors
- Gripper sounds
- Other informative warnings or notification sounds too, not just for actual errors

## Comparing Prototype Sound Results

In the main part the prepared sounds were then presented to participants. In the following, the reactions and conclusions for the prototype sounds as tested will be summarized per sound, also considering the validating measures taken from the Likert-questionnaire.

### *Safety Area Slowdown*

While the metaphor representing the slowing down was indeed found to be useful and recognizable, this effect might be extended a bit by making it more pronounced. This could be done by reducing the speed, making the slowdown more noticeable, or by increasing the drop in pitch. Also slowing down might help with the connotation to an animal sound, which was perceived as unpleasant by participants. This was also reflected in the level of comfort reported by participants, which, for this sound, was the lowest score given throughout all sounds and categories. For this sound the largest positive influence on the result was the urgency, being perceived as fitting the situation. For this it should be considered that this sound was preferred to be used only in test mode, as in automatic mode this usually is less important to be informed about. In test mode however a signal like this could support the users' understanding of the digitally set areas in relation to the physical workspace, and the reason for the cobots change in action.

### *Handoff*

With the second highest overall score, and the highest likelihood of use reported by participants, this signal seems to be an important addition to the cobot communication. The use in practice is clear, with some participants even mentioning signals like this in the pre-interview. Especially the trust category scored higher than average, meaning participants would easily recognize the cobots action, message, and their required steps. This is especially pronounced, as a clear action is required by the users in these cases, and also it can be reflected by the current position of the cobot, which could for example hold a finished piece towards the user further supporting the understanding. Ideally, a combination of modalities is used whenever possible. An additional change that might be made is to add variations depending on the importance. These could be increasing over the wait time or be set to a level depending on the task importance, to inform the users about priorities more clearly and

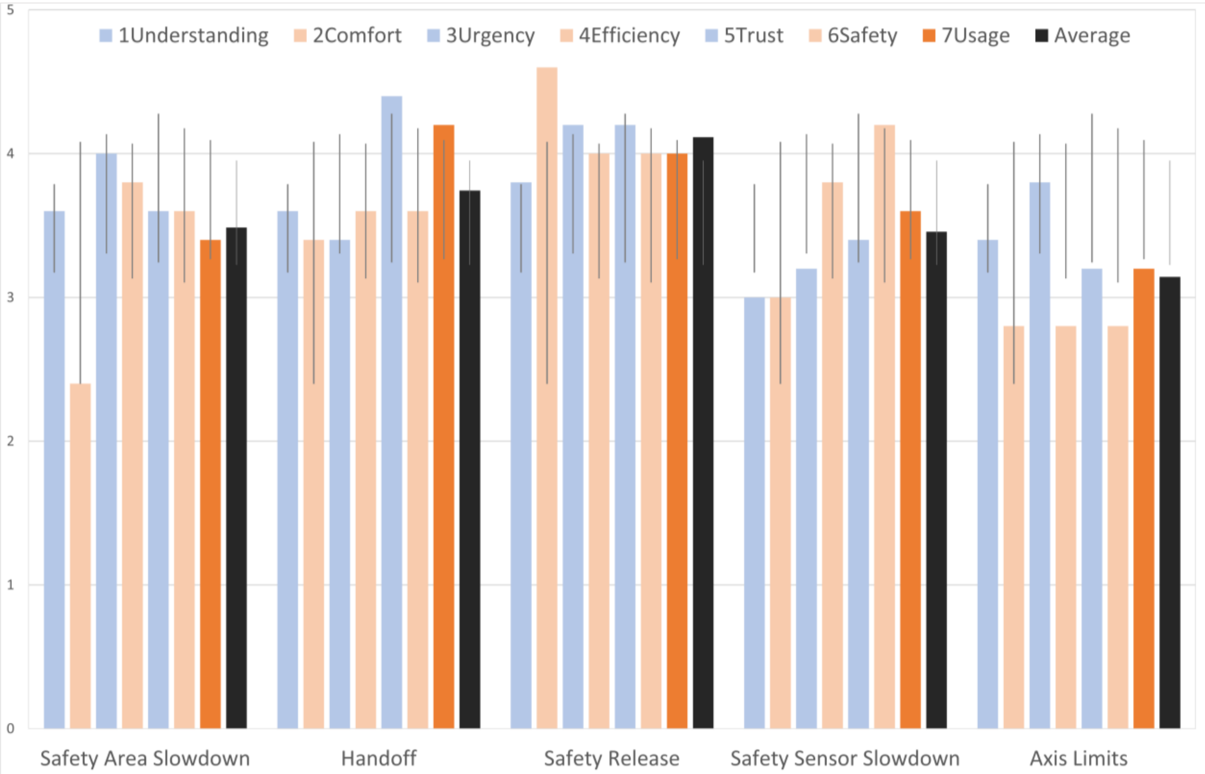



*Figure 6: Bar Chart Representing the Average Results per Interview Question Grouped by Sounds*

to avoid unnecessary disturbances. This could for example be done through volume, adding higher frequencies to the sound, or by increasing the number of knocks. A combination of these elements would allow for an even wider range of importance levels and could ensure that urgent tasks are quickly attended to.

#### *Safety Release*

With the highest value throughout the test given to the comfort category of this sound, and confirming feedback from users, the sound created for this was perceived as the most pleasant of all. With it being the most tonal and clean sound of them all this could point to a general preference, which was also given in the pre-interview by two of the participants. However, it does need to be noted that this is also the most positive sound considering the message content (at least when considering the successful part of it), and the negative sounds will be intentionally designed to have more negative connotations. It might be a consideration to also clean up the other sounds slightly. In general, this sound received positive, and consistent feedback. With understanding being the lowest for this sound, being slightly below average, this could point to the negative effect of its clean sound, making it less characteristic and recognizable. Some alterations like the two-tone might make it more recognizable. Also, through the high frequency of use, especially in combination with the conscious action of pressing the button might balance this out, and make users learn it quickly. Attention does need to be paid here especially when adding more sounds, so the overall sound concept stays consistent and discernible.

#### *Safety Sensor Slowdown*

For this sound especially the safety and efficiency categories have received high ratings. While the responses on the pleasantness of the sound and the fit to its message have been very varied and inconsistent, it can also be noted that one participant specifically mentioned it as the best sounding one. A positive factor discussed with some participants was the dynamic mapping used to relate the sound's aggressiveness and volume through the distance of the user. This might also have been a positive influence on categories like efficiency, while the limitation of not being able to experience this mapping first-hand might have limited the understanding and comfort aspects. For the application of this sound, the usage of a dynamic sound like this is indeed suggested, but further analysis of the pleasantness of the sound in practice might be necessary.

#### *Axis Limits*

For this sound the categories urgency and understanding scored high, while the rest was a bit lower or around the average. The lower sound was perceived as fitting for the limited urgency of the sound, and also the again used dynamic mapping might have helped this aspect. Through this effect the action would also be directly understandable in practice, with a clear, audible reaction to the user action. The limiting factor in its usefulness was that no conclusions can be made about which axis is at its limit, which would also probably be too much to represent through audio alone. For this a combination with for example LED rings would be favorable.

#### *Handguiding Position Info*

While for this concept no prototype sounds were developed, and thus the Likert-Questionnaire wasn't used either some interesting thoughts came from this, some of which have already been mentioned in the user proposals given before. Regarding this sound specifically, especially safety area info would be a useful feature. In this way, it makes the connection between the digital elements and the physical workspace clearer. It lets users experience it first-hand in a tactile way and might even inform them about the different types of areas they go through. Implementing a similar feature for point information was however seen as not feasible, as there is a too large number of them to stay distinguishable usually. Also features supporting the accurate steering of the cobot in Handguiding mode through sound feedback was seen as not necessary, as for accurate movements usually the smartPAD is the quicker and more accurate alternative.

## 6.5 Discussion

In the following, the findings from testing will be evaluated in the context of the defined hypotheses:

**H1**: The utilization of varying levels of loudness and aggressiveness for sounds could indeed be an effective way for conveying information of different urgency levels appropriately in diverse contexts. This approach increased user awareness for important messages, and supporting signals, without causing discomfort through the overload of messages. In practice these levels have to be checked and confirmed again of course, as it largely depends on the number of messages that will be communicated, which can vary per application. The balance between urgency and user comfort emerged as a critical factor, indicating the need for different sound versions to match different levels of urgency.

**H2**: The implementation of dynamic sounds guided by sonification principles significantly was indeed preferred for many cases, for their immediate, informative and understandable feedback. These dynamic sound variations facilitated better comprehensibility and prompted appropriate user responses, highlighting the effectiveness of employing sonification principles in designing robot sounds. These could be employed much more for new use cases coming up, like sensitive joining features or application specific parameters.

**H3**: Incorporating movement principles and skeuomorphic sounds that corresponded to the communicated message enhanced the intelligibility of sound communication. This became especially apparent, when the initial connotations matched the intended message and principle in the blind testing part. In these cases, the intended message could indeed be recognized quite accurately, considering the lack of context. This enhancement made it easier for both novices and experienced robot operators to learn, recall, and interpret the auditory cues, thereby improving the overall user experience.

**H4**: The study recognized that users with greater familiarity with cobot systems, especially the KUKA system, indeed held more specific and elaborate expectations for auditory communication. While the users were all relatively familiar with cobots or at least robotics, their focus areas could be seen to influence their preferences naturally. This finding emphasized the need for tailoring sound design to accommodate the preferences and expectations of experienced users. Contrary to expectation however, the users on the more technical engineering side preferred a more limited sound solution, rather than additional signals in some cases. This could be explained with the fact that for these users, it is comparably easier to find the needed information themselves when they need it, for example by knowing the right menu to check on the smartPAD. As some of the signals might however be relevant to experts aswell as novices, giving options for experts to turn on or off some of these less safety relevant features might be advisable. With a general look at cobot requirements and users, the use of these supportive sounds might lower the barrier of entry for new users.

**H5**: The results indicated that dynamic sounds were generally more useful, informative, and better perceived than simple Earcons, especially in situations where urgency and distinctiveness were crucial. The direct reaction of these sounds to the action makes them well-suited for these scenarios, as it allows users to, in turn, react quickly to them aswell. However, finding the right balance between noticeable signals and not overloading users with excessive auditory cues remained a challenge. Simple Earcons will definitely still have their place for example for simple notifications, however considering these principles in more areas can be advantageous for a smoother interaction.

In conclusion, the tested hypotheses provided valuable guidance for designing effective sound communication in human-robot interaction. The insights gained from testing and user feedback highlighted the significance of balancing urgency, comprehension, and user comfort while tailoring auditory cues to the specific context and user experience. These findings contribute to the further development of auditory communication principles in robotics and underscore the effectiveness of iterative design methods and user-centered approaches, which shall continue to guide these developments.

### Limitations

While, as described, it was useful to have a Likert scale questionnaire for some additional evaluation, it needs to be said that the results are not statistically relevant. They were useful to go through during the interview, as it allowed people to ask additional clarifying questions, and made them consider all points again, leading them to give additional insights in some cases. Also, it helped with the interpretation of some more ambiguous replies afterwards. Further user centered testing or codesign methods with a larger group of participants are advised for the final development of concrete sounds in the future.

Considering the testing of the prepared audio prototypes, some limitation regarding the volume differences needs to be given. It was found that the accurate representation of these levels would require some finetuning and ideally a test in context, to hear the sounds in relation to the actions. While some variation of loudness was given from the sounds themselves and how they are recorded, due to the focus on the actual sound, volume differences could not be represented accurately in this setting. Again, this would need further investigation for deciding final sound designs. However, the executed tests were useful for a validation of the basic principles, which was the focus of this work.

Also, in general, the reactions to sounds could be different when seen in combination with action and in context and might be more understandable through this. Using concrete sounds made the testing much more valuable and the situations more imaginable, than purely interviewing, but of course this is limited by the blind factor. For the given goal of identifying the first unbiased reactions however this method seemed to be fitting. However, the closer to context the better, so future evaluations should include the sound concepts in the full interactions as much as possible.

Finally, revealing the solution after each sound might have influenced their consequent replies. However, the blind testing approach and the randomized order of the sounds were helpful ways to get the users unbiased opinions as far as possible. It does have to be mentioned however that gathering the feedback in person, for the own prototypes can influence the participants to give more favorable feedback. However, measures were taken to limit those effects, for example by asking them explicitly for areas of improvement. The varied feedback received shows that this did seem to work, and they seemed comfortable sharing their opinions also about aspects or sounds they didn't like. Additionally, the comparative measures like the Likert scale results were not taken as absolute measures, but rather just to compare between the different prototype sounds.

# 7 Conclusion

The testing phase of this study aimed to evaluate the effectiveness and user perception of the designed auditory cues for human-robot interaction. A special focus in this test was laid not on the concrete sounds themselves, but rather on the underlying principles, using these more as a prototype and basis for discussion. Through a combination of open interviews, blind audio prototype testing, and Likert-scale questionnaire responses, valuable insights were gained into the participants' preferences, perceptions, and expectations regarding such auditory cues. The results provided further guidance for refining the auditory communication principles, based on a deeper understanding of the possible uses in practice. As discussed in the results section, many of the same sound principles can be employed further in an actual cobot sounds solution, as feedback suggested. While some sounds might still need some audio adjustments for pleasantness and audibility in noisy settings, the general principles applied seemed to be effective. Further, the main findings from Testing were:

- Matching the Urgency: Participants emphasized the importance of balancing message frequency and urgency. While users appreciated informative signals, they also highlighted the need to avoid overwhelming them with excessive auditory cues. Offering variations of certain sounds, with differing levels of urgency was suggested, allowing the more urgent signals to grab attention with distinct sounds, while supportive signals can step in the background a bit.
- Dynamic Sounds and Sonification Principles: The implementation of dynamic sounds guided by sonification principles proved to be effective in enhancing the comprehensibility and user-friendliness of the auditory cues. Dynamic variations in pitch, rhythm, and volume allowed for richer communication, and more direct understanding and recognition of different messages.
- Matching Sound to Message: Participants recognized the use of sound connotations and principles that matched the communicated message, particularly when metaphorical sounds represented specific actions. Skeuomorphic sounds that correlated with the intended action were well-received, as they aligned auditory cues with users' expectations, and were well understandable even without much prior knowledge.

In addition, a number of smaller findings, related to the concrete iteration on the prototype sounds were generated, which could inform the further development of these.

The findings from the testing phase also have implications for enhancing auditory communication in human-robot interaction more broadly. To further refine and apply these findings, the following recommendations are given:

- Contextual Testing and Iterative Design: Further testing should be conducted during actual task execution, to understand how the auditory cues perform in real-world scenarios. Integrating the sounds into the full interaction between users and cobots can provide a more accurate assessment of their effectiveness. In this regard it is also important to stay open for iterations of these sounds based on continuous user feedback, to keep their recognizability and match with actual work practices and needs. To ensure this, methods like codesign sessions are recommended to involve users in the design process.
- Diverse Contexts Come with Different Needs: Provide users with customization options for adjusting sound settings based on their preferences and task requirements. This could include adjusting sound volume, urgency levels, and even enabling/disabling specific auditory cues. However, make sure that this doesn’t overload users with options that might confuse them, so limit the access to experts, or provide simple profiles to group options and allow for efficient adaption to the needs without going through every individual option.
- Integration with Multimodal Feedback: Consider integrating auditory cues with other sensory modalities, such as visual or haptic feedback, to create a more coherent, pleasant and informative interaction experience. Matching the modalities accurately will be important for a smooth controlled experience. This can mean timing lights, existing clicks or button pushes on the cobot, its movements, and any other signals the cobot already uses, to fit with the sound accurately to make it more satisfying, understandable and consistent
- It is still recommended to opt for individual speakers rather than a connected notification system for most cases, as it keeps the directionality of sound so users can spot the source of a notification more easily in a busy workplace. Potentially even multiple speakers might be used on one cobot and the smartPAD, to allow the indication of directions on the scale of the individual workstation.

In conclusion, the testing phase of this study validated most of the hypotheses and provided valuable insights into the design and implementation of auditory cues for human-robot interaction. By addressing the balance between urgency, comprehension, and user comfort, and by incorporating dynamic and contextually relevant sounds, the study contributes to the continued development of efficient communication principles for the robotics context. As technology continues to evolve and human-robot collaboration becomes more prevalent, the lessons learned from this study can guide the design of future auditory interfaces that enhance user experience, safety and efficiency in human-robot interactions, through creating a better understanding between user and cobot.

### *Customizing Signals*

Custom signals to implement by admins etc. give a selection of predetermined sounds to choose from. Should still fit the overall sound design. Maybe reserve a certain sound, which allows programming a variety of melodies, that are still recognizable as a similar signal, but can be fit to different situations. For example, using all of them as a start sound, but being able to distinguish which task came in by the melody. Could be matched with a consistent background sound mixed with the custom melody. While some sounds can be set beforehand as an easily recognizable standard, many sounds in a custom collaborative application might need to be programmed individually. However more individuality in the solution could lead to lowered safety, which is why guidelines need to be set and enforced to keep this within limits. For this reason, it is important to find ways to guide users to employ the same kinds of strategies to achieve an overall useful sound design in the end. Several strategies have already been identified. While of course simply teaching the set of guidelines is an option, it would be advised to ensure this in additional ways. It should be integrated into the programming workflow, for example through context sensitive tooltips, or using a setup wizard, that guides you through the required steps. What's needed to make this feature useful is a simple and fast custom sound feature that lets you select a sound easily from a range of sounds, without going through long lists and trying around. To achieve proper customization easily, especially with the found focus on urgency matching, some sliders or 2D planes can be given to select a sound from to match certain points on the provided dimensions. These could then be limited to certain ranges, allowing users to use only a closed range of sounds for these actions, to keep the fit with the overall sound solution, and not interfere with safety relevant cues.

Also, the idea to make sounds more related to the individual cobot might be extended and tested more. Slight audio variations like filters, pitch changes or some other effects applied sparingly, could make the cobots more distinct. This could be done to suit the specific model of cobot better, or even be set at random or during setup per individual cobot, making each have a slightly different audio profile. While sounds should keep their distinct characteristics, to stay understandable, these slight audio variations might allow users to distinguish a bit more easily where a notification came from when there are multiple options.

## Matching the Other Concept Categories

Since the usage of sound in this context and product is still a new topic with many decisions necessary, implementing the full sound concept at once is not feasible. To be able to test a focus was set, however this doesn´t mean the rest shouldn't be included in a final solution. Through the decision-making process it was found that the tested solution would likely be most effective at adding to the communication and delivering value. Therefore, it is recommended to start implementing at this point, and add the further, less crucial parts over time, to be able to craft a fitting overall system. This is important as adding more and more new sounds firstly can get hard to memorize, reducing their effect, but secondly even hinders the recognition of the already existing sounds, as they can be confused. For this reason a minimal audio solution should be implemented first, which already allows gathering feedback of the use in practice, and makes it easier to see which parts will actually be used as intended, and where additions and modifications will be necessary.

### *Alerts*

These are of course still important for things like emergency signals. Implementation is rather straightforward, using distinctive signals for critical situations that demand immediate user attention. However, while these signals would be designed to be highly recognizable and convey a sense of urgency, these should not overload the users especially with a look on the high number of error messages that can appear from one action. Here a smart system combining notifications might be necessary to bundle these and stay informative. Some indication about the error type can be given, but not too many categories should be created to stay recognizable. A good measure for separating them is the relevant action required from the user in response is, its urgency, or if possibly an expert is needed. In this way users can quickly identify the nature of the problem based on the auditory feedback, allowing for efficient troubleshooting. The varying levels of severity represented in the sound could also help users accurately assess the urgency of a situation and create a more pleasant experience.

### *Progress*

As also indicated by some users, it might be helpful to add auditory cues updating them on the progress of ongoing tasks. These cues can be implemented to indicate different stages of completion and keep users informed about the cobot's status. One mentioned idea was the use of the repeated loops of a program, which could give a small positive notification after each finished part. Additionally, these might be integrated at certain steps that already require interaction, to get the users attention again in a positive manner. Again they should be implemented in a sensible way to prevent fatigue, which could be achieved with varying levels depending on the importance.  Thinking back to the sonification principles, these might also be varied slightly depending on how well a task was executed, which would connect the progress information back to the understanding and optimization category. In the customer journey example, this could for example be applied at the step where the user needs to insert some rings accurately in the product, after which the next part is fit on top by the cobot. Here the ring placement might influence how much the sensitive joining feature needs to work to get it to fit. Presenting the progress sound in a slightly more positive variation in very successful cases can help the worker optimize the placement and get pleasant feedback. Combining functions in this way can be another way to increase variety and information content, and reduce the amount of signals at the same time.

### *Task Guidance*

While the research and partially also the testing showed that voice might not be fit for the industrial context that much yet, this category might be mentioned as one of the best fits for this. Voice cues could guide users through specific tasks, assembly steps or the solving of errors. These cues can be used to direct users on how to interact with the cobot in a step-by-step manner, or even work conversationally. However, these solutions will require a large amount of additional research. More simply it might be possible to integrate some instructional non-speech sounds. Here especially the skeuomorphic principle, or possibly other mapping strategies mentioned before becoming important again, to create an intuitive understanding without the use of words. These instructions might still provide additional guidance and clarity to users as they perform more complex tasks. With enough repetitions it might still be a helpful nudge for the users to remember the steps and combining this with for example interaction notifications can limit the number of additional sounds.

### *Virtual Assistant*

As mentioned before, starting to implement voice-based interaction opens the door to a whole new research field. However, under the right conditions it can be said that employing features like this might create an even more intuitive natural interaction. At the start some simple voice commands might be added, but in future versions the full integration of voice-based virtual assistants that can communicate using natural language might be researched.

These assistants might provide information, answer questions, alter settings without searching for them and assist in task programming, and flexible applications.

*Conclusion*

While the initial focus of the sound concept was limited to the specific aspects of movement sonification for the prototyping and testing phase, there were a number of further additions coming up throughout the research and interviews, which could enhance the overall interaction experience. Incorporating emergency signals, progress updates, task guidance sounds, and a virtual assistant can expand the functionality and usability of the auditory communication system. However as mentioned the signals all need to be tested against each other before implementing too many signals, limiting their recognizability. Implementing these additions incrementally, focusing on the found principles, will allow for the gradual iteration and improvement of the auditory interface, as well as the general communication of the cobot including all modalities.

## 7.1 Interaction Principles

In the following, the used interaction principles, which emerged through research, prototyping, and testing will be abstracted and collected again. They are intended to serve as a basis for further, more applied research and development of more concrete sounds.

### > Musical Principles

While not all sounds must be melodic or even tonal, musical principles are an important aspect to consider. Musical factors, like harmony and rhythm are important for the emotional responses of users, and thus can be used for creating the intended experience. They can influence the sound perception very generally, by for example creating positive or negative connotations, which are relatively consistent over a range of users. Namely, for example harmony is positively connotated, and the principle of speed can be related to urgency and excitement. These correlations are well studied for their musical application, and also intuitively understood in many cases through exposure to them, so they can often be transferred to sound signal design too. Taking principles, as also outlined by Barrass (1997) as a basis can lay an important basis in understanding how certain sounds will be perceived by the general population and can guide the design of them effectively.

### > Materiality of Sound

The principle of materiality of sound suggests that the physical properties of sound, such as its texture, timbre, and spatiality, should be considered in designing sound for human-robot interaction. In cobot use cases, sound can be employed to enhance communication and interaction in natural ways, by following sounds' logical characteristics known to us from natural interactions, like with instruments. Metaphorical sounds can create a memorable and easily learnable association between the sounds and the actions of the robot, while metonymic sounds can represent materiality and direct reactivity in interactions, making sounds directly related to actions and movements. By using both metaphorical and metonymic sounds, designers can improve the accessibility, learnability, materiality, and expressivity of interactive systems. The sounds used by a cobot should be harmonious and coherent with its physical movements and interactions. For instance, it could use sounds that relate to the speed and change of its movements, enhancing the sense of coordination and connection between its actions and auditory cues. This alignment can be an important factor in keeping the intuitive understanding for the meaning of sounds and causal relationships to their messages, which is a natural property, as seen in Alexanderson (2004), but needs to be paid attention to when artificially creating these sounds.

### > Directionality

Since we are used to recognizing where sound comes from, this basic principle should be used as an advantage, rather than causing additional confusion. This means, wherever feasible, to put the sound source close to the point it wants to inform about. Especially in a bigger system, this might mean adding a separate speaker to each robot, instead of collecting all their signals at a central hub and notifying from there. When working in different places in the production this can help workers quickly identify the source of the sounds and aid them in their understanding. Also, on a smaller scale, this might mean adding the speaker inside the Commander, at the end of the arm instead of or in addition to on the SmartPad. A sound might for example notify about getting close to an obstacle, and as the instinct is to turn towards the sound if that is the first thing that's noticed, it makes this interaction faster if the sound source is as close to the location in question as possible. This can aid in increasing awareness about its position, and thus safety, and make tangible interactions even more natural. Getting such feedback from a more distant location, for example at the base of the robot can create a mismatch, distract from the situation at hand, or even disconnect the sound from the relevant action completely. Also, this ties back in with the materiality principle, when considering the spatial relationship of a sound to its source.

### > Skeuomorphism

Connotations and using known principles to make sounds recognizable and easily learnable are central to the principle of skeuomorphism. By incorporating familiar sounds that users can relate to, such as the sounds of a motor or a camera, or, as mentioned by a user the familiar audio interaction with a parking distance sensor and its characteristic beeps, designers can make use of users' existing mental models to create intuitively understood auditory cues. This principle allows for a bridge between the digital and physical worlds and abstracting some of the known principles in sound design proved to be effective in making the sounds understandable to new users.

### > Cobot Identity

As companies develop unique brand identities through visuals and audio, cobots should seamlessly integrate into the aspired perception. In this way, the perception and personality of the cobot can better align with the overall brand image.

Additionally tailoring the sound design to the specific cobot embodiment it is used on can be a good way of relating the sound to it more closely. Related to the specific model's form and function, or even for every separate cobot used in a space, they could each have their own distinctive auditory identity. Here it is important to note that only slight variations in pitch or timbre should be used, to keep the sounds recognizable. However, these slight variations could, over time help establish a strong and memorable identity, and help especially in more busy environments, with multiple cobots. A certain starting chime or effect on the already used sounds can be used to differentiate between different robots, or robot types, similar to the studied effects how different robots were perceived in Dennler et al. (2022).

In this light also the use of speech needs to be evaluated closely, as the use will drastically change the personality and appearance of the cobot, and it might not represent it in the wanted way. Also adding voice opens up a very large range of possible personalities within different voices, communication styles and intonations, as humans are very adapted to recognizing small differences in this.

### > Honest Representation

The sounds produced by a cobot should give an honest and accurate reflection of its capabilities, states and actions. Here the message urgency is the prime example, as urgent messages need to be recognizable as such even when other messages are heard aswell. While many devices like smartphones play a loud notification sound, no matter the content, this often doesn't accurately represent the actual urgency, leaving users to check this every time. In order to allow a fluent, pleasant interaction, not interrupting the workflow, these should be sent out more carefully, and should allow the user to make conclusions and decisions based on the understanding of the sound alone. If a cobot is conveying an urgent alarm that needs attention, its sounds should genuinely communicate urgency, while other messages that are for example simply informative shouldn't distract from the current task, and should stay in the background, also represented in factors like their volume. This principle encourages transparency and reduces the chances of miscommunication between the cobot and its human counterparts and will serve for a more efficient and pleasant experience. Similarly, its capabilities need to be represented honestly, for example regarding force limits. Here also the use of voice can be mentioned again, also considering the uncanny valley effect. Trying to get closer to human behavior in this way by simply adding voice features is thus not recommended when the underlying linguistic capabilities don't allow an actual conversational interaction, as this could evoke a false sense of its capabilities and might limit the usefulness, as also mentioned by Villani et al. (2018).

### > Natural Interaction

With the goal of making the interactions with a cobot learnable as quickly as possible, it proved useful for the sound development to adapt to natural communication principles that are already familiar. This way focusses on the user experience rather than technical limitations and aims to tailor the communication to human capabilities and the natural ways of interacting with other actors. In this context especially coworkers can be mentioned as examples, but also non-human actors, like pets and work animals, and our interactions with them can be taken as a reference. Cobot interactions can be modelled after these examples, using for example similar rhythms, pitches or conversational structures. Mok et al. (2015) already applied similar principles to the movements of robots, and applying similar expressivity to the audio signals, making them more in line with the natural interactions we're used to could take these interactions further towards being fully intuitive. For example successful actions could be connected to a sound resembling a happy or content hum, while abrupt, loud and shrill sounds can be used to warn of danger, just like shouting to a coworker would.

Mimicking the flow and rhythm of a natural conversation can also make the interactions more pleasant and comfortable in general. Considering for example the back and forth of speaking between two parties, which when applied to cobot sounds makes the user receive a message in moments where he is actually receptive for them.

When these sources of inspiration are exhausted also inspirations from cinematic and animated representations of robots and other actors can be used, to still give a familiar experience, however the more natural and interactive situations are likely more intuitive.

### > Satisfying Experience

To create an overall satisfying user experience, factors like multimodal matching were found to be important. In these, visual, auditory, and tactile cues complement each other seamlessly, which also makes the relation clearer, aiding the understanding and learning of these signals. Additionally, the overall communication should be considered as a whole, to tune it to be relatively positive. While of course notifications are used in most cases to signify negative or unexpected events like errors, as that is when they are especially needed, some methods can still dampen the effects on the user's perception. Positive interactions could also be accompanied by pleasant sounds and perhaps gentle vibrations or lights, reinforcing the user's perception of success and progress. Similarly, when errors occur, the severity of negative experiences can be softened by employing empathetic and helpful sounds, and not making them more dramatic than they need to be. This especially needs to be considered with the possibility of many error messages being caused at the same time, which might make a sound combining these necessary, to not overload the user with auditory cues. Feedback matching the interpretation and state of the user and harmoniously matched over multiple modalities can serve for a coherent and pleasant experience. Striving for a holistic and emotionally pleasant work experience could affect users' overall satisfaction with the cobot and its performance.

# 8 Discussion

Principles were designed with a focus on futureproofing, by applying first principle thinking, instead of merely designing solutions for concrete problematic situations common in the current system. This approach helped clarifying the essential underlying factors at play, resulting in a set of principles that are not simply adaptations of, and additions to existing interaction paradigms. Instead, they aimed to present a general solution aimed at the core of the communication challenges and characteristics in practice, also considering the possibilities for future interactions. By trying to envision and consider the possible new configurations and interaction scenarios as well as emerging technologies and industry trends and designing the testing and principles in an open way, the general applicability should be kept for future iterations. This is necessary, as future communication patterns might become more dynamic, diverse, and elaborate, with features like voice control and artificial intelligence features on the horizon. However, the complex and interconnected developments in the field are often not straightforward to predict and might result in new scenarios that weren't considered in the design.
Formulated in a broadly applicable way, these principles are designed to simplify adaptation to these new scenarios as well, offering long-term value as a robust foundation for evaluating alternatives and informing the decision-making processes.

While it focused on the application of principles found in research and practice, adopting an iterative, user-centric design process helped to start off the optimization of the user experience. Also, the design process incorporated regular feedback loops, with the sound principles and use cases being iterated to fit new findings and focus areas, allowing the principles to evolve and adapt over time.

The focus was not just on adding audio to the messages the system currently delivers using the other modalities, but also on the holistic experience of all potential users, taking into consideration their diverse needs, abilities, or environmental conditions, to craft a well-rounded user experience in practice. This study focused on sound specifically, partially for the reason, that other topics and modalities were already covered more extensively in the literature, as well as internally at the company. For the final implementation and development of concrete sounds, a holistic analysis of the user experience will still be necessary, to integrate all the signals into a coherent concept. A general limitation of the testing done in this thesis is the limited number of test participants. While this allowed for more depth, it limited the validity of the results. The resulting principles of this study should still be applied in a more full-scale study with increased implementation in the cobot workflow, and a bigger group of participants. For this, it is recommended to rely on user-centered development methods, like codesigning sessions, to iteratively find fitting solutions. As this medium proved to be difficult to describe and modify effectively with other people, a tool for creating specific sounds together with users is proposed to be used in such sessions. Similar to a synthesizer, but with parameters that are already limited to certain areas, to fit the defined sound concept, and with additional shortcuts that simplify and speed up the use by unfamiliar users, it could allow the collaborative creation and iteration of sounds, to fit the expectations of users more concretely.

Summarizing the contributions of the work in this thesis, the structured flow from first principles through evaluation, back towards the final principles is essential. Starting with a holistic literature review of the field, viewing it not only from the industrial standpoint but also for audio generally, as well as other related influences, has guided the development of preliminary design principles. This has laid a solid foundation for the following testing of audio concepts and prototypes based on those principles, which made them more practical and applied, and allowed to get feedback from real users as well. Verifying and refining the principles in this way has lead to the main contribution of this thesis, being the interaction principles, which may be used as design guidelines for future development and research.

# 9 References


Adamczyk, P. D., & Bailey, B. P. (2004). If not now, when?: The effects of interruption at different moments within task execution. *Proceedings of the 2004 Conference on Human Factors in Computing Systems - CHI '04*, 271–278. https://doi.org/10.1145/985692.985727

Alexanderson, P. (2004). *Peripheral awareness and smooth notification: The use of natural sounds in process control work*. 281–284. https://doi.org/10.1145/1028014.1028057

Barrass, S. (1997). *Auditory Information Design*. https://doi.org/10.25911/5D7A2ACD2B71D

Bergman, M., de Joode, E., de Geus, M., & Sturm, J. (2019). *Human-cobot Teams: Exploring Design Principles and Behaviour Models to Facilitate the Understanding of Non-verbal Communication from Cobots:* 191–198. https://doi.org/10.5220/0008363201910198

Cziksentmihalyi, M. (1990). *Flow – The Psychology of optimal experience*.

Dennler, N., Ruan, C., Hadiwijoyo, J., Chen, B., Nikolaidis, S., & Mataric, M. (2022). *Using Design Metaphors to Understand User Expectations of Socially Interactive Robot Embodiments* (arXiv:2201.10671). arXiv. http://arxiv.org/abs/2201.10671

Emilien Ghomi. (2012). *Designing expressive interaction techniques for novices inspired by expert activities: The case of musical practice*.

Herre, J., & Dick, S. (2019). Psychoacoustic Models for Perceptual Audio Coding—A Tutorial Review. *Applied Sciences*, 9(14), 2854. https://doi.org/10.3390/app9142854

Henkelmann, C. (2007). *Improving the Aesthetic Quality of Realtime Motion Data Sonification*. https://cg.cs.uni-bonn.de/publication/henkelmann-2007-improving

IDZ International Design Center Berlin. (2023). *IiQKA Design System KUKA AG*. https://ux-design-awards.com/winners/iiqka-design-system#:~:text=The%20KUKA%20iiQKA%20Design%20System,while%20empathizing%20the%20KUKA%20brand.

KUKA AG. (2023a). Case Studies: Automation solutions from KUKA. *Kuka.Com*. https://www.kuka.com/en-de/industries/solutions-database?&filters=7CDBD7898401457CB38964B0637641BD|A5BFFA434F1E45688FD9010800F255AD

KUKA AG. (2023b). iiQKA.OS – The powerful and intuitive new operating system from KUKA. *Kuka.Com*. https://www.kuka.com/en-de/products/robot-systems/software/system-software/iiqka_os

Macaulay, C., & Crerar, A. (1998, November 1). *'Observing' the Workplace Soundscape: Ethnography and Auditory Interface Design*. International Conference on Auditory Display '98. https://doi.org/10.14236/ewic/AD1998.20

Michaelis, J. E., Siebert-Evenstone, A., Shaffer, D. W., & Mutlu, B. (2020). *Collaborative or Simply Uncaged? Understanding Human-Cobot Interactions in Automation*. 1–12. https://doi.org/10.1145/3313831.3376547

Mok, B., Stephen Yang, Sirkin, D., & Ju, W. (2015). A place for every tool and every tool in its place: Performing collaborative tasks with interactive robotic drawers. *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 700–706. https://doi.org/10.1109/roman.2015.7333680

Mori, M. (2012). The Uncanny Valley. *IEEE Robotics & Automation Magazine*, 19, 98-100.

Sarah Gibbons. (2017, November 5). UX Mapping Methods Compared: A Cheat Sheet. *Nielsen Norman Group*. https://www.nngroup.com/articles/ux-mapping-cheat-sheet/

Stienstra, J., Overbeeke, K., & Wensveen, S. (2011). Embodying complexity through movement sonification: Case study on empowering the speed-skater. *CHItaly*, 39–44. https://doi.org/10.1145/2037296.2037310

Tennent, H., Moore, D., Jung, M., & Ju, W. (2017). Good vibrations: How consequential sounds affect perception of robotic arms. *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 928–935. https://doi.org/10.1109/ROMAN.2017.8172414

*The Duel: Timo Boll vs. KUKA Robot*. (2014, March 11). https://www.youtube.com/watch?v=tIIJME8-au8&t=2s

Villani, V., Pini, F., Leali, F., & Secchi, C. (2018). Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications. *Mechatronics*, *55*, 248–266. https://doi.org/10.1016/j.mechatronics.2018.02.009

Zaatari, S. E., Marei, M., Li, W., & Usman, Z. (2019). Cobot programming for collaborative industrial tasks: An overview. *Robotics and Autonomous Systems*, *116*, 162–180. https://doi.org/10.1016/j.robot.2019.03.003

# 10 Appendix

## Digital Supplementary Materials

Further several materials are available digitally:

- Slides from the Prototype testing session, explaining each of the sounds and the accompanying scenario
- Separate audio files of the used prototype sounds
- Some further audio files from the prototyping session which haven't been used in testing
- A high-resolution version of the Customer Journey Map

## Testing Results

In the following the comments from the testing will be given. They were combined already to give a better overview and keep anonymity. The left column gives cases where the results were inconclusive, as opinions differed between participants. The middle column combines the results that were seen as agreed upon, with all or most of the participants mentioning it. When there were only singular responses, which were however not actively disputed by other participants, just not mentioned, they were added to the right column.

| **Disputed** | **Agreed** | **not disagreed** |
|---|---|---|
| | | |
| Pre-Interview | | |
| | abschaltkupplung sound recognized by experts-> crash | |
| | handguiding brake release click sound is recognized->ready to move | |
| voice or just simple beeps | | |
| | error signals should have sound | |
| | | handguiding safety area info |
| | different error signals should be recognizable through their signal to some extent | |
| | | progress or task completion could be auralized |
| | should maybe fit specific cobot too, and be recognizable that its from that one | |

| | | |
|---|---|---|
| | multimodal matching | |
| | signal need for interaction, waiting | |
| | other warnings or notifications as sound too, not only errors | |
| | confirmation signal for inputs | |
| | | feedforward, notify whats possible. For example noise to warn about coming closer to the operator which might be unsafe, or safety countdown before starting |
| | adding weight to actions that have importance but might be missed or activated accidentaly, like touch up, starting the program in automatic mode, or changing the tool or base and coordinates | |
| standards or modifiable | | |
| | dont let everyone change sounds freely, and limit the extent to what can be changed to keep consistency. | |
| | | sound for mobile or large or far away systems where overview is hard |
| | | sound for autonomously acting and hard to predict robots |
| | use sound only where really necessary, not constantly, or in standby/waiting times | |
| | | no unpleasantly high pitched beeps |
| | match importance to loudness, agressiveness and fullness of sounds | |
| Safety Area Slowdown | | |
| | unpleasant | |
| | relatively well understood connotations, going in the right direction | |

| | | |
|---|---|---|
| | will be ignored in automatic, useful for handguiding or programming | |
| | slowdown metaphor fitting, memoarable | |
| Handoff | | |
| | request to do something | |
| | | something switched |
| | | bluetooth speaker ready to receive |
| | | FaceID unsuccessful |
| | | headshake, slight negative, action unsuccessful |
| | : Knockingknocking door, tapping fingers on table | |
| | | : "are you still there?" |
| urgency not decided, more towards fitting, and depends on the task | | |
| | mostly perceived as pleasnant, but maybe make a bit slower, less stressfull | |
| | | make a bit less negative sounding (headshake connotation) |
| maybe annoying in the long term | | |
| | sound very natural and muffled, make more audible for the environment | |
| | | could use speech instead |
| | variations depending on urgency | |
| Safety Release | | |

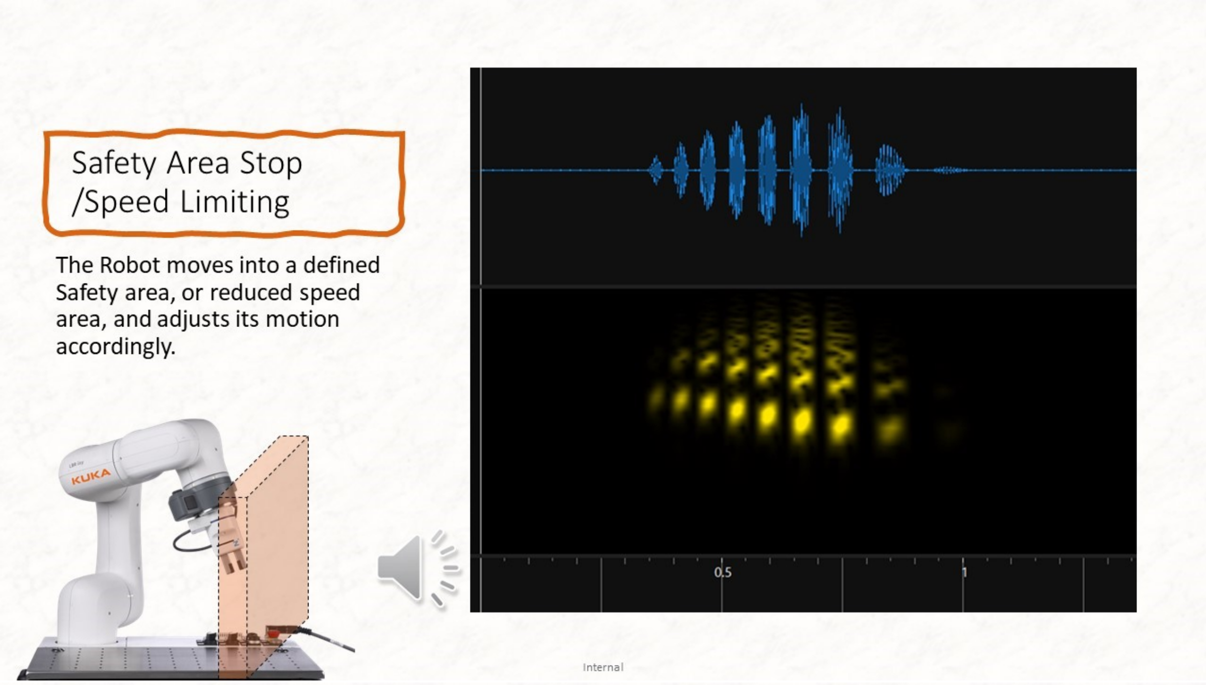
| | | |
|---|---|---|
| | positive negative recognized | |
| | | positive sound: activate toggle switch |
| | | : teaching, saving, accepting |
| | could be more positive | |
| | negative sound recognized, but not as something too dramatic, which fits | |
| | | |
| | | too dull and muffled for work environment |
| | | 3 tone, to make also the first part alone already have a positive sounding 2 tone |
| | | nice playful sound, fitting and not too technical |
| | | negative might be made more expressive |
| | make them more fitting to each other, use double sound for negative too | |
| | | |
| | feedback useful | |
| Safety Sensor Slowdown | | |
| | startup | |
| | beaming, spaceship, healing in a game | |
| | | : interference frequency |
| | | : Data transmission, receival, sending |

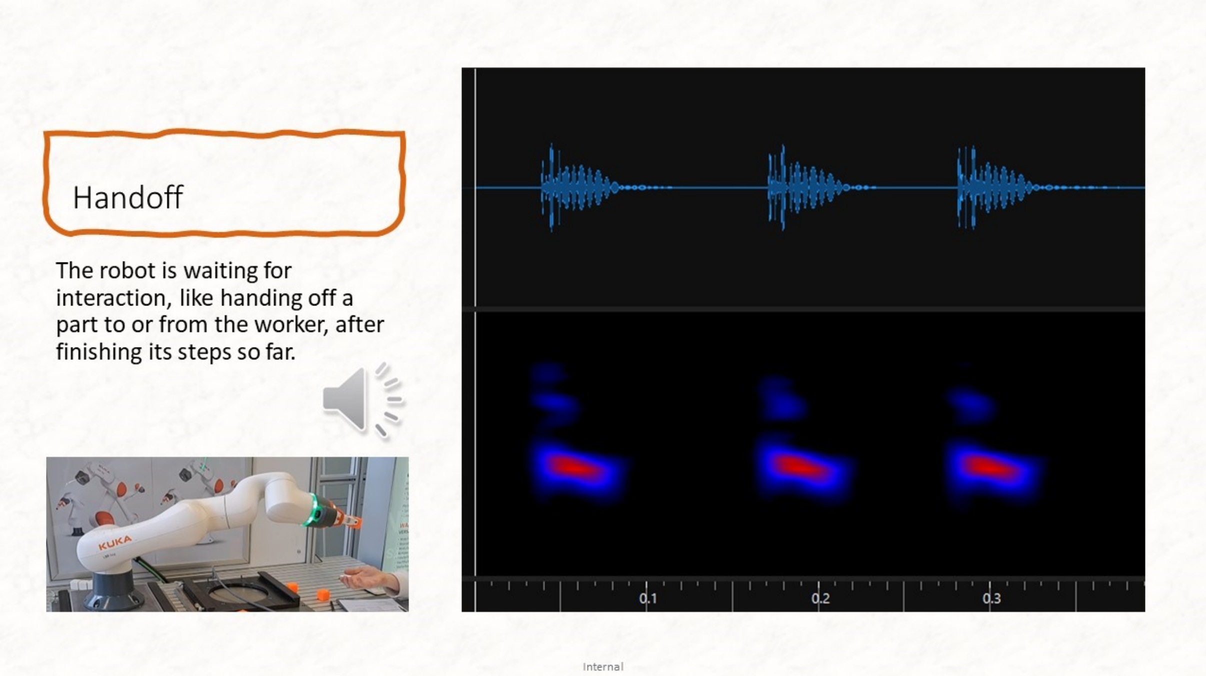
| | | |
|---|---|---|
| | | :  calibration, mastering |
| | | :  searching surroundings |
| very varied response on how well it sounds | | |
| very varied response, some liked first part, some liked second part | | |
| | parking beeps are fitting (similar to second part) | |
| | | maybe make less shrill to make pleasant |
| | could be more urgent, but depends on context | |
| | | |
| | | :  dynamic mapping useful, would be used to play around and experience the area first hand |
| Axis Limit | | |
| | dark:  endboss in a game, negativenegative (not critical as its continuous)dark/somber/grim | |
| | | |
| | | straining growl |
| | | |
| | | startup |
| | getting closer to something | |
| | slightly unpleasant and scary sound, fitting the negative connotation | |
| contested if its fitting | | |

| | | |
|---|---|---|
| short one for hitting the limit only or fitting, not sure | | |
| | additional info might be helpful but might be too much for sound alone | |
| getting close to limit not as important, but liked by 2 participants | | |
| Handguiding Position Info | | |
| | safety area info would be a useful feature, for different types of areas | |
| | | steering through sound not necessary as smartpad is accurate |
| | point info would get too busy | |
| Additional Ideas | | |
| | sensitive joining dynamic sound | |
| | | progress, loop count sound |
| | | handguiding immediate supportive unlock feedback sound |
| | | startup successful or not feedback sound |
| | | feedback sounds for using the Commander as there is no GUI feedback |
| | | : successful test run without errors |
| | | : gripper sound |
| | | : touchup executed |

## Prototype Sounds

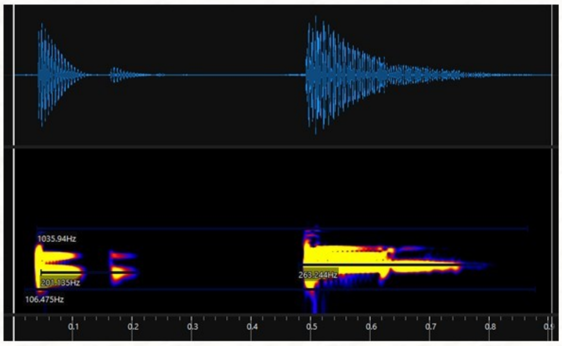
These are the sounds and slides as used in the testing session. For a playable version, supplementary slides are available.

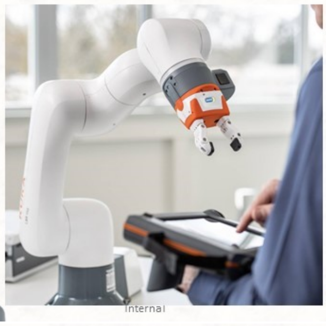


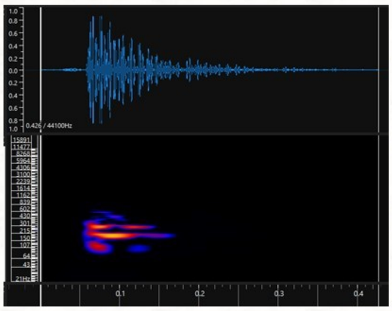



## Safety released /unsuccessful

The operator presses the safety release as well as the start button to run the program. When pressing the safety release, it will indicate if it was successful or if you should release and try again.



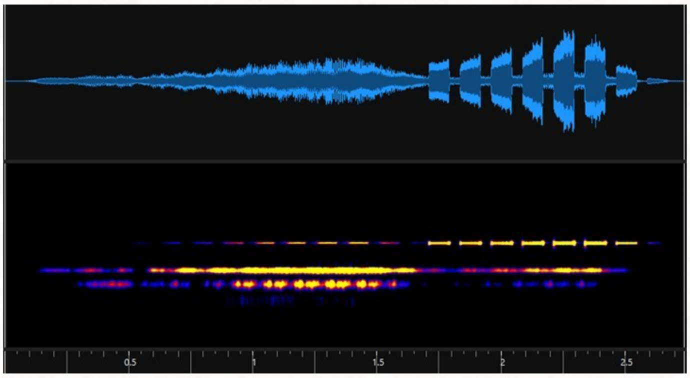




Internal

## Safety Sensor Stop

The Safety sensors, like lidars, vision systems, or a predictive safety system, detect an obstacle, especially a human, entering the path of the robot or a defined safety area around that, and movement is adjusted.



Internal


## Frequency masking

"An important aspect for perceptual audio coding is the consideration of *masking effects* in the human auditory system, i.e., the effect that louder sounds ("masker") tend to suppress the perception of other, weaker sounds ("probe") in the masker's spectral or temporal vicinity. A common example for masking is how the sound of a bird (probe) can be drowned by the sound of a car passing by (masker) in human auditory perception. One category of such masking effects is known as "spectral masking effects" or "simultaneous masking" [25] (see **Figure 3**), for which the following aspects were observed:

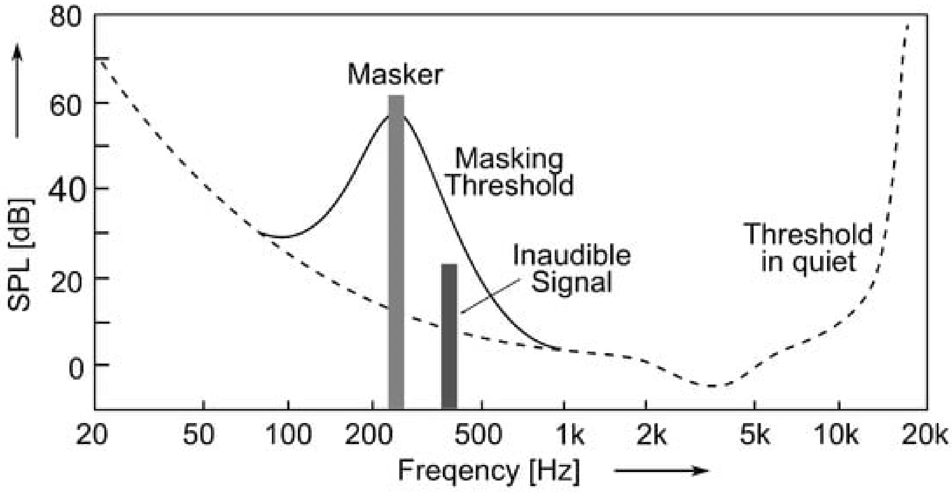



**Figure 3.** Illustration of spectral masking effects. Dashed line represents threshold of hearing in quiet, solid line illustrates the masking threshold due to the presence of a masker signal (e.g., narrow-band noise), due to which weaker signals at neighboring frequencies become inaudible.

- A frequency dependent *threshold of hearing in quiet* describes the minimum sound pressure level (SPL) of a sound to be perceivable in isolation and under extremely quite conditions.
- In the presence of a masker, the threshold in quiet curve changes into a masking threshold, which shows a bell-shaped increase in frequencies in the vicinity of the masker, depending on its frequency, level, and signal type. Any sound beneath this threshold is masked by the louder signal, and thus inaudible for the average listener. In perceptual audio coding, the coding error (i.e., the introduced quantization noise) corresponds to the probe signal in this experimental scenario.
- Masking effects are strongest for signals that are within the critical bandwidth of the masker. Within the critical bandwidth, the masking threshold remains constant. Furthermore, the masking effects spread to frequencies beyond the critical bandwidth (so-called *inter-band masking*). The upper slope of the masking threshold depends on multiple factors, such as absolute frequency and sound pressure level of the masker, whereas the lower slope hardly shows a level dependency.
- Depending on the type of masker, i.e., tone or (narrow-band) noise, the strength of the masking effect varies. While noise-like maskers can mask tone-like signals very well (up to a masker-to-probe level ratio of about 6 dB), tone-like maskers can mask noise only to a much weaker extent [**29**] (about 20 dB).

The second category of masking effects can be described as "temporal masking effects" and describe masking behavior when the masker and probe signals are not present at the same point in time (see **Figure 4**). For "post-masking", quiet sounds that occur after a loud stimulus are masked due to the reduced sensitivity of the ear for approximately 100–200 ms. Additionally, there is also "pre-masking" in a short time window of approximately 20 ms before the masker, where the perception of soft (probe) sounds is masked by subsequent louder (masker) signals. This seemingly non-causal behavior is assumed to be caused by the fact that softer sounds have a longer build-up time for cognitive processing in the brain than louder signals.

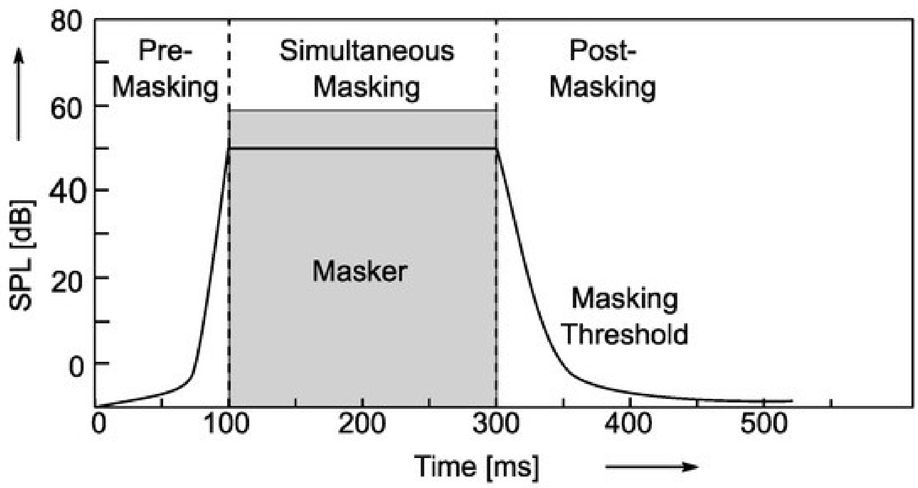



**Figure 4.** Illustration of temporal masking effects.

These effects regarding masking of energy describe the core behavior of the HAS that has to be modeled for perceptual audio coding."

(Herre & Dick, 2019, p. 4-5)