Bustoni, Isna Alfi (2025) *Toward manipulating user perceptions of objects by altering interaction sounds.* MPhil(R) thesis.

https://theses.gla.ac.uk/84878/

Toward Manipulating User Perceptions of Objects by Altering Interaction Sounds

Isna Alfi Bustoni

Submitted in fulfilment of the requirements for the
Master of Philosophy

School of Computing Science
College of Science and Engineering
University of Glasgow

January 19, 2025

Abstract

Humans naturally use interaction sounds emitted from physical contact with objects to gather information about their properties, such as size, texture, and quality. Augmenting these interaction sounds may manipulate our perception of the object's properties. Furthermore, such augmentation could be useful in conveying extra information. For instance, augmenting the perceived weight of a cup by manipulating the interaction sound could help users track their daily food and carbohydrate intake to aid them in their diet. While previous work has demonstrated the effect of reverb on room size perception in various contexts, such as VR gaming, few studies have explored the potential of manipulating interaction sound, such as adding reverb or changing the sound, to alter user perception of object properties.

In two experiments, this thesis explored the effects of different auditory feedback (surface materials, musical instruments) and reverb parameters (mixing, depth) on perceived simultaneity, unity, agency, and object properties (object size, material, and room size), along with their recognition. Through the experiments, it was discovered that modifying the sounds emitted from tapping an audible object can affect our perception of its size and the size of the room in which it is located. In regards to the user ability to recognize auditory feedback , it was discovered that auditory feedback did not considerably impact the accuracy of audio recognition, whereas reverb played a more significant role.

This thesis presents a discussion of the findings, along with the limitations of the current study and suggestions for future research. Lastly, it provides guidelines on augmenting interaction sounds based on our experiments and findings.

Contents

List of Tables

List of Figures

Acknowledgements

To my husband, Indra, and my two angels, Freya and Kaiden, for giving me all the love, strength, and support.

To my mom, dad, brothers and other family, who are always there for me.

To my dearest friends, Dilla, Inna, Kiki, Marlita, Nirna, Arina, and many others, who all helped me through it.

To my supervisors, Stephen Brewster and Mark McGill, who gave me the opportunity.

To my colleagues, Shaun, Paddy, Ammar, Melvin, Basma, Nasreen, and others, who offered advice, empathy, and advocacy.

Thank you all for helping me..

Declaration and Contributing Papers

The research presented in this thesis is entirely the author's own work. Experiment 1-2 in Chapter 4 is currently under submission to the ACM journal *Transactions on Computer-Human Interaction* under the title *The Perception of a Tap: Using Auditory Augmented Reality to Alter the Contact Properties of a Physical Object*.

Chapter 1

Introduction

1.1 Motivation

Humans use physical actions like tapping, knocking, stirring, or dragging to interact with nearby objects. We expect to receive auditory feedback from these actions [93], which convey meaningful information about the objects and their surroundings. This feedback can include details about the object's shape and material, density [26], quantity [117] and even the size of the room in which it is located [101]. While tactile feedback is primarily used to determine or identify these object's details, other senses such as visual and auditory are often used simultaneously [57]. The presence of visual cues can helps humans distinguish stiffness among objects [104]. However, we also use our auditory senses when vision cannot provide adequate information, such as for object material and density, given the same appearance (i.e., surface colour, thickness). For instance, a hollow-ringing sound from a watermelon is a way to determine its ripeness [1], which cannot be deduced from its external appearance alone.

The relation between touch and audio feedback has been linked with the neural mechanism; for example, one may feel body sensation when listening to a heavy bass acoustic [57]. Furthermore, the watermelon example demonstrates how auditory feedback emitted from user contact with an object (interaction sound) plays an essential role in providing information and inducing the perception of its properties. Thus, changing this interaction sound in real time may change the user's perception of the object. Additionally, interaction sounds provide valuable insight of our environment, allowing us to gauge a room or an object's size: dropping an object in a spacious room produces a much more noticeable reverberation than in a smaller one. Despite previous research into the effects of reverb on room size perception in various contexts, such as VR gaming, few studies have explored the potential of manipulating interaction sound such as adding reverb or changing the sound, to alter user perception of object properties.

Manipulating user perception of object and room properties could provide many benefits for HCI. In the development of future AR/VR interactions, augmenting user perception of object

---

[1] Brüel & Kjær: https://www.bksv.com/en/knowledge/blog/perspectives/ripe-watermelon

and room properties may potentially enhance passive haptic feedback and support casual interaction with objects. Passive haptics uses the alignment of virtual elements with real physical objects to enhance the realism of interactions, e.g., touching a virtual object rendered in AR placed against a real physical surface [46]. For AR-based passive haptics, altering the acoustic properties of the surface can be used to change the perceived size, volume, stiffness, and more [56]. For casual interactions, augmenting auditory feedback could be used to convey information through everyday objects [77]. For example, the reverberation of a tap on a coffee mug could indicate the remaining free time until the next meeting. A coffee mug is an example of an everyday object that is in widespread use and found in various materials and shapes. Additionally, tapping a mug can be more discreet in certain situations, such as when conversing with others in the office.

There are many techniques to alter user perception of object properties, such as using a physical proxy that employs actuators to progressively alter an internal weight [116], wearable haptic interface [21], kinesthetic haptic devices [53], or using visual effects [99]. Kinesthetic mechanics requires a standalone device, which is less portable [53]. In contrast, visual-based user perception alteration necessitates an image rendering device (e.g. a Head-Mounted Display or HMD), which can have a high computation rendering cost, and is less socially acceptable in a public setting. Another approach remains underexplored: augmenting auditory contact properties by adding virtual audio to real-world sound as an extension or complement [**?**]. Real-time auditory contact property augmentation can be done using currently available devices such as acoustically transparent headsets, for example, bone conduction headphones [2], Apple Airpods [3], or BOSE Frames, which are more socially acceptable than the visual head-mounted displays due to their more discreet form factor [4].

Auditory augmentation can be seamlessly integrated into existing perceptual and cognitive processes as humans naturally possess parallel skills when processing auditory information with other senses (e.g., visual) [73]. The concept of auditory augmentation itself could be seen as a part of Augmented Reality (AR) [10, 58] as some studies refer it as Audio Augmented Reality (AAR) [58, 59, 112]. Thus, to develop a successful auditory augmentation system that can manipulate the user perception of object and room properties, we should consider the aspects of AR itself. AR should have three characteristics as defined by Azuma[5]: (1) combines real and virtual; (2) interactive in real time; (3) registered in 3D. To design auditory augmented interaction, other than intermixing digital audio and real sound, we must maintain the perceived interactivity in real time. Interactivity in augmented reality refers to the ability of users to interact with and manipulate virtual objects that are overlaid in the real-world environment [67] and increase the experience immersion [82]. The system also should be able to sense interaction and give feedback in real-time. Yet, how can we design such real-time AAR for interaction?

---

[2] https://www.techradar.com/news/best-bone-conduction-headphones

[3] https://www.apple.com/uk/airpods/

[4] The Guradian: https://www.highsnobiety.com/p/bosesframeshowtostyle/

When designing interactive experiences, designers must consider the user's perception of unity, agency, and simultaneity. The unity assumption is a condition where a user believes that multisensory cues (e.g., visual, haptic, auditory) *"come from the same object or event"* [20]. Agency refers to *"the feeling of having produced the sound"* [18], including a sense of belonging and control over interaction [12]. Meanwhile, based on [39], perceived simultaneity is *"an illusion of two events appearing to be simultaneous and happening at the same time"*. These three concepts are crucial for creating an immersive and realistic experience [5], a key characteristic of augmented reality systems as defined by Azuma[5].

However, interaction sound augmentation, such as altering audio feedback, modifying acoustic properties, or adding effects (e.g., reverberation) may impact the perceived unity, simultaneity, and agency. For example, when tapping a coffee mug but hearing animal sounds, the interaction may be perceived as not coming from the tap, which could reduce the perceived interaction unity. In addition, when multiple augmentations are used, for example, replacing the audio, manipulating the new audio's frequency, and adding a reverberation effect, is the auditory feedback still identifiable? Retaining identifiability of the auditory feedback is important especially when used to convey information.

To summarise, auditory augmentation of interaction sounds could potentially be used as an effective medium to alter user perception toward object and room properties. This approach can be applied in various scenarios, such as leveraging passive haptic feedback and facilitating casual interaction with everyday objects within an augmented reality system.

This work examines acoustic manipulation of interaction sound emitted from user contact with a common, everyday object – a coffee mug. The choice of a coffee mug is due to its widespread use. To answer the main research questions, this thesis enhanced the audio feedback produced by tapping the mug. A prototype of auditory augmentation was built using a microcontroller and contact microphone. The prototype enabled participants to experience a near real-time audio interaction where they could hear the manipulated interaction sound with 8ms of latency from their finger tapping. The findings revealed that various types and intensities of auditory feedback could be used to augment interaction sound without compromising the perceived agency and unity of the tap. Furthermore, the auditory feedback generated from the augmentation process was easy to identify and could significantly impact the perception of the mug's properties. For example, a metal contact sound could make the mug surface appear harder. To provide additional flexibility, reverb could be incorporated, allowing manipulation with a second independent variable. Two experiments in this study explored various reverb configurations, including echo mixing and room size, to provide multi-level audio feedback. Based on the results, guidelines for design of acoustic properties that can be manipulated to alter user perception or convey information were created, supporting the notion of future usage of auditory augmentations.

1.2 Thesis Statement

This thesis explores the impact of interaction sound augmentations on user perception of object properties and augmentation recognisability. Various auditory feedback is used, including material sound and non-material sound (music), along with various reverb adjustments. This work also explores technical challenges in the current development of such AAR interaction to deliver a prototype to augment interaction sound in real-time whilst retaining interaction unity, agency, simultaneity and recognisability.

1.3 Research Questions

To develop an AAR system, it is crucial that the auditory augmentation must be presented in real-time and interactive, retaining its interaction perception (unity, simultaneity, and agency) along with its recognisability when used to convey information. However, currently, there are no comprehensive guidelines for:

- **RQ1**: How can we augment real-world interaction sound emitted from an everyday object whilst preserving:
    (a) the perception of the interaction sound coming from the same object or event (unity)
    (b) the perception of the tapping and augmented sound happening at the same time (simultaneity)
    (c) the perception of having produced the interaction sound (agency)
- **RQ2**: How does acoustic manipulation (replacing the interaction sounds, adding reverb) impact:
    (a) the perception of object properties such as size and stiffness
    (b) the perception of environmental properties such as room size
    (c) the recognisability of augmented auditory feedback

1.4 Contribution and Impact

This thesis makes contributions in the following areas:

1. Demonstrates that augmenting real-world interaction sounds from tapping on a coffee cup can preserve the feeling of connection (unity, simultaneity, agency) with the interaction;
2. Identifies the impact of acoustics manipulations (replacing the sounds, adding reverb) on object size and stiffness, and environment (room size) properties, along with the augmentation recognisability;

3. Delivering practical design guidelines to be applied in AAR use cases.

1.5 Thesis Outline

This thesis is composed of the following chapters:

- **Chapter 2** - Literature Review: This chapter provides an overview of previous research and the fundamental theory of auditory augmentation. It includes a classification system, definitions of key terms, a review of related research, and a summary of the essential aspects of auditory augmentation for interaction sounds. Additionally, it presents the context of the research questions and the contributions of this thesis.
- **Chapter 3** - Augmenting Interaction Sound in Real-time: This chapter describes the technical aspects of developing real-time auditory augmentation. It presents two approaches for real-time tapping recognition and audio alteration: hand gesture recognition using Hololens 2 AR/VR/MR HMD and touch vibration detection with a contact microphone and microcontroller. A comprehensive discussion of these approaches will also be presented.
- **Chapter 4** - Impact of Auditory Augmentations: This chapter investigates user perception of auditory augmentation. Two experimental studies have been conducted to evaluate the impact of various auditory feedback and reverb adjustments.
- **Chapter 5** - Discussion and Auditory Augmentations Guideline: This chapter provides a discussion and a general overview of the thesis, highlighting the answers to the research questions, and presenting auditory augmentation guidelines based on the findings.
- **Chapter 6** - Conclusion and Future Work: In this chapter, a highlight of this thesis contributions, summarise of the experiments, and recommendations for future work, is presented.

Chapter 2

Literature Review

This chapter reviews existing studies and the fundamental theory of auditory augmentation. The review begins with a taxonomy, followed by definitions of terminology related to auditory augmentation. Moreover, related work around acoustic manipulations used in previous research is presented. Then this chapter summarises the important aspects of auditory augmentation for interaction sounds, motivates the research questions, and provides context for the contributions of this thesis.

2.1 Auditory Augmentation Taxonomies

As a subset of AR, in which the majority of research has focused on visual augmentation, auditory augmentation has been investigated under various terms. Auditory augmentation or Audio Augmented Reality (AAR) is defined as *"a blend of virtual auditory content into the physical world to augment the user's real acoustic environment"* [112]. Due to its broad scope, many scholars worked on aurally augmenting reality even though they did not clearly use the term AAR [58]. For example, research on tangible music that augments users' soundscapes [110] can also be seen as an AAR application. Thus, researchers tried to combine common use cases of AAR and built AAR taxonomies.

Krzyzaniak et al.[58] present a taxonomy of AAR consisting of (1) using digital sound to augment mute physical things; (2) mixing real world acoustic and digital sound; (3) placing sound into a real-world location; (4) making real-world sound richer by conveying additional information; and (5) real-time modification. Furthermore, another terminology for AAR was also proposed by Schraffenberger[85]. Based on [85], AAR can take different forms such as (1) Extended reality, scenarios where the virtual supplements the real environment; (2) Diminished reality, cases where virtual content removes real elements from the real environment; (3) Altered reality: environments where the virtual information changes the apparent qualities of the real world; (4) Hybrid reality: scenarios where the virtual completes a physical environment that would be considered incomplete without the virtual additions; and (5) Extended perception:

cases where unperceivable but real aspects of the real world are translated into virtual information that can be perceived.

2.1.1 Enchanting Mute Physical Object with Auditory Feedback

Whilst the auditory augmentation of physical objects is still rarely studied, some works have shared characteristics, such as interactive surfaces, Interaction Design with Audio (IDwA) [16], and Tangible Auditory Interaction (TAI) [10]. In interactive surfaces, research mainly focuses on the technology that can help achieve touch interaction on various surfaces, from projecting visual cues into a home appliance [100] to augmenting the surface with tactile feedback [6]. These augmenting surface concepts can be viewed as a part of enhancing physical object concepts in augmented reality [58], yet, arguably, it can also be seen as a part of developing Tangible User Interaction (TUI) when the perspective is the control and interaction toward the object. For example, Potts et al.[78] proposed TangibleTouch, a toolkit that enables users to control digital content by touching the surface of the a cube. Ullmer and Ishii[103] also mention how these terms share characteristics. For example, McGee and Cohen[65] developed RASA, a sticky-notes-based interface in military command posts, linked to digital content enhancing current military command in practice: *"..while these efforts can also be viewed as augmented reality systems, their usage of physical objects as computationally mediated artefacts also holds much in common with tangible interface approaches"* [65].

Jacob et al.[48] tried to unify these concepts as a single framework: Reality-based Interaction (RBI). RBI aims to unify emerging human-computer interaction styles such as virtual, mixed and augmented reality, tangible interaction, surface-based interaction, and ubiquitous and pervasive computing while keeping the characteristics of each concept. For example, URP [103], a TUI model, is a part of the naive physics (common sense knowledge on the physical world) group in the RBI framework [48]. We argue that our work, augmenting interaction sound, is also part of the naive physics group in the RBI framework. Thus, we could say that the concept of auditory augmentation of interaction sound is indeed a part of augmented reality whilst also can be seen as TUI.

In regard to interaction with physical objects, the research on TUI also often overlaps with AR research. The use of physical objects for interaction, also known as artefact-based interactions, is also common in mixed reality research due to their manipulability [78]. This phenomenon is unavoidable because augmentation, which is also a part of augmented reality-mixed reality, is one of the TUI characteristics defined by Ishii and Ullmer in their original work [47]. In TUI research, a physical object is used to manipulate or represent digital information [33] such as demonstrated by [43]. The objects used in TUI research varies, ranging from generic 3D geometric shapes such as a sphere [115], cube [78], or cone [52], to Lego shapes [40, 4, 71]. Gohlke et al.[40] used Lego to produce music based on the position, orientation, colour and shape of pieces.

TUI research by Zhiglova[118] used an interactive carpet as a textile interface for children on the autism spectrum. In this interactive system, different inputs (touch, pressure, placing objects onto a matching pattern, stroke) produced different outputs (light, vibration, sound). These TUI works can be seen as having shared characteristics with the concept of augmenting silent physical objects in AAR, where the digital sound is added as the output of the object, disregarding the sound produced by the interaction if available. Our work with interaction sound is trying to augment real-world sound emitted from interaction with physical objects, to alter user perception or to convey digital content. This concept, augmenting interaction sound, also could be seen as "auditory augmentation" defined by Bovermann et al.[10], which is also a paradigm that evolved from TAI. Bovermann et al.[10] defined auditory augmentation of a physical object as *"a paradigm to vary the objects' sonic characteristics such that their original sonic response appears as augmented by an artificial sound that encodes information about external data"*. Thus, it concluded that although there is a shared similarity between AAR and TUI with sound as output (i.e. TAI), AAR with interaction sound, or auditory augmentation of physical objects, highlight the use of real-world sound generated from user interaction.

2.1.2 Altering Interaction Sound Emitted from Physical Object

Augmenting auditory contact properties of physical objects can expand the horizon of interaction design offering a more natural and connected user interaction linked with prior knowledge and expectations. As briefly described in the Introduction, human interact with everyday objects and expect auditory feedback [93] that may reveal object properties such as material, quantity [117], and room size [101]. Another example, as provided by Hermann and Weger[44], is shaking a box to investigate the contents, which is a natural way we try to get information of unseen detail. Weger et al.[106] describe this process (actively expecting auditory feedback after interaction) as active listening. This active/conscious listening to interaction sound, which connects the user's action and their expectation, could also be used as actionable cues and bring up our subconscious reactions or change our behaviour. For example, we adapt our walking style to the sound of the floor, e.g. when the wooden floor is perceived as thin we may walk more slowly. Such auditory feedback becomes an extension of the user's body, connecting the interaction sound with the user's speculation and body response. Creating a natural interaction means designing an interface which able to take full advantage of the user's prior knowledge [107]. Wigdor and Wixon[107] defined, a perceived natural interaction also has to create an experience that could possibly be the extension of the user's body [107]. This is strongly coherent with the idea of augmenting interaction sounds that could be used to enhance prior experience with everyday objects (i.e. tapping), for example, to accentuate information or to alter their perception of object properties. Therefore, this thesis will investigate the idea of augmenting the interaction sounds of everyday objects.

2.2 Non-speech Audio Feedback for Augmenting Interaction Sound

Many scholars have investigated converting digital information with auditory feedback using speech-based feedback [2] and non-speech auditory feedback such as Earcon, Auditory Icons, etc [79, 13]. Speech-based interactions have been widely used for example voice assistants such as Amazon's Alexa or Google Assistant to provide weather or news updates. Still, this type of interaction might not be suitable in some cases. For example, in a public setting where privacy and information discretion are needed [69]. Moreover, speech-based interactions as input/output rely on the user understanding and speaking the language used by the system, which can be a barrier for users who are non-native [111]. In addition, language barriers may yield bigger problems, such as speech recognition errors, when used as input. Speech-based interactions may not always accurately recognize spoken input (which can be triggered by user's accents): thus, it can lead to errors or misunderstandings and user frustration [23].

Furthermore, speech output may have a high cognitive load, as the user may have to wait for the spoken output to finish to get the information [62]. Meanwhile, compared to speech-based auditory feedback, non-speech auditory feedback, such as Earcons or Auditory Icons, offers some benefits. For instance, it is language independent [79], offers speed of access [79], supports multitasking [105], and could be part of product branding [74]. Earcons are a non-speech auditory display that uses sounds, music or other non-speech audio to convey information or instructions to the user [14]. Earcons can be used to convey information regardless of the user's language, as they use symbolic representation rather than human language, making them useful in multilingual or general settings. Considering all the benefits above, this thesis will use non-speech auditory feedback to augment the interaction sounds.

2.3 Augmentation Recognisability

Auditory feedback that has been manipulated should be identifiable by the user. For example, in the passive haptic feedback use case, without clearly identifying what auditory feedback is presented, the user may misjudge the object properties, thus breaking the realism of the interaction. *"Determining materials is more difficult as they produce changes in the acoustic signal that are confounded with accompanying material changes- two bars of different material can produce perfectly identical sounds"* [63]. The misinterpretation of material perception due to similar sounds may break the immersion of virtual content. For example, when the user is expected to feel wood but instead hears a foam sound. This situation may happen when the acoustic manipulation of the wooden sound is inappropriate. In the case of casual interaction, recognisability is a part of usability, which should be retained to create a successful system. For example, to ensure the user can precisely determine the quantity of digital information. However, as we used non-speech feedback whose identifiability may be affected by the masking effect of acoustic augmentation [55], what acoustic manipulations can we use that retain the recognisability of the auditory feedback?

2.4 Changing the Perception of Object Properties by Manipulating Interaction Sound

As humans naturally gain information from contact with physical objects (e.g., tapping a watermelon to know its ripeness, tapping a wine glass to know its quality), augmenting this auditory contact property may influence perception of the object and/or its surroundings, as it may give users cues about the texture, weight, and density of the objects as well as surroundings. For example, presenting a soft, muffled sound, for example, may suggest that the object is composed of a soft substance, such as foam or rubber. In contrast, a sharp, metallic sound may indicate that the object is made of a tougher material, such as steel. In the surrounding context, an echo may indicate a large room. Whilst there are many techniques to alter user perception of object properties, such as using a physical proxy that employs actuators to progressively alter an internal weight [116], wearable haptic interface [21] or using visual effect [99], employing AAR to augment this perception is under-explored.

The ability to augment object properties could be beneficial in many cases, such as passive haptic feedback and casual interaction. In passive haptic feedback, augmenting object property may enhance immersion when using the object. Furthermore, it may also change the perception of the user's own body and possibly impact the user's behaviour, as demonstrated by [97]. For example, changing the object's texture, weight, size and stiffness according to the presented digital object, such as portraying a virtual button on the object's surface, then altering its contact sounds to a metallic sound to make the user tap harder on the surface.

In casual interaction, augmenting object properties could present various information or digital content. Object size could represent the quantity of the information, while stiffness may represent information/content density. For example, tapping the laptop surface and a hollow sound may imply an empty slot in the working schedule to arrange the new meeting. The texture could convey more salient digital content; for example, tapping pillows and sensing a crispy texture instead of a fluffy one may inform the user they forgot to lock the door. Weight also could be possible interesting property. For example, to inform daily food/carbohydrate intake to help users in their diet. The more they consume carbohydrates, the heavier the object.

2.5 Auditory Interaction Perception

Rapid development in audio-AR-enabled personal devices, for example, acoustically transparent pers onal audio devices such as the BOSE Frames, Sony ambie sound earcuff [5], to Hololens which supports built-in spatial sound [64], make the future use of AAR as the interactive audio-based system practical. However, how could we develop a successful AAR device that makes the auditory augmentation seemingly connected to the user's contact with objects?

In AAR with interaction sound, multiple senses are used (e.g. visual, haptic, auditory). To create a successful interaction, retaining the perception of simultaneity [51], unity [20], and agency [18] in multimodal interaction is crucial because it helps to create a more immersive and believable experience for the user. This notion also in line with the idea of an AR system [5] which one of the key aspects is real-time augmentation. In addition, auditory feedback that has been manipulated should be identifiable by the user. For example, in the passive haptic feedback use case, without clearly identifying what auditory feedback is presented, the user may misjudge the object properties, thus breaking the realism of the interaction. *"Determining materials is more difficult as they produce changes in the acoustic signal that are confounded with accompanying changes in material- two bars of different material can produce perfectly identical sounds"* [63]. The misinterpretation of material perception due to unnoticeable differences between sounds may break the immersion of virtual content. For example, when the user expects to feel wood but hears a Styrofoam-like sound. This situation may happen when the acoustic manipulation of the objective is inappropriate. In the case of casual interaction, recognisability is a part of usability, which should be retained to create a successful system. For example, to ensure the user can precisely determine the quantity of digital information. However, as we used non-speech feedback whose identifiability may be affected by acoustic augmentation [55], what acoustic manipulations can we use that retain the recognisability of the auditory feedback?

2.5.1 Real-time perception

Real-time augmentation is important in AAR with interaction sound because it allows the user to have a sense of agency in producing the sound (hearing sounds that are generated in response to their actions or movements). This can enhance the sense of presence and interactivity, making the experience more engaging and believable. Zamborlin[114] demonstrated a real-time modification of audible objects, Mogees. However, this work only alters the sound without conveying additional information. An interesting work by Kari et al.[54] synchronises in-car music with a journey in real-time, altering the acoustic properties of the music and creating an unobtrusive interaction providing situational awareness for the user (i.e. informing tunnel entrance with high pitch). The music played in a car will be mixed with a particular digital sound to convey navigation information, in this case, a shift in the music represents event changes.

---

[5] Sony Ambie earcuff: https://www.sony.com/en/SonyInfo/design/stories/ambie/

Another notable example is the auditory contrast enhancement model to help user guess the content inside a closed box [44]. PebbleBox and CrumbleBag demonstrated capturing real-world interaction sounds, manipulating its granular synthesis [29]. Meanwhile, the study by Williamson and Murray-Smith[108] used granular synthesis to convey continuous information. These studies demonstrated how augmenting audio, in general, could be coupled with digital content (e.g., navigation, object internal detail), offering a richer experience with everyday reality. This finding could be used to explore the wider area of augmenting interaction sound with physical objects in real time. For example, to add digital information by augmenting the acoustic properties (frequency, grain) of the contact sound with the object.

A real-time acoustic manipulation should have low-latency audio processing. The study by Kaaresoja[51] stated that <30 ms latency is required for perceived simultaneous interaction. Current AR headphones with hand gesture recognition to detect interaction with objects produce latency around 35 ms [1]. Thus, even though previous studies emphasized how AAR augmenting interaction sound could be established, there is little knowledge on how these augmentation could be benefitted. Audio processing technology such as Bela (https://bela.io/). could be used as it has latency under 10ms. Therefore, can we make a proof-of-concept of AAR interaction sound system with a currently available device?

2.5.2 Perception of Simultaneity

Latency, the time delay between an action and the corresponding feedback, negatively impacts the quality of interaction [51]. The study found that perceived simultaneity, the perception that two events happen at the same time [39], is crucial for interaction quality. Bimodal interactions (tactile and audio) latency must be kept under 100 ms for the feedback to be perceived as simultaneous [51]. For unimodal feedback, the latency should be within 20-70 ms. Another study [96, 94] demonstrated a near-real-time interaction with auditory feedback and an average latency of 10.7 ms, but did not investigate the perceived simultaneity. With advancements in technology, it is becoming possible to create AAR systems with minimum latency, making it achievable to develop AAR interactions that are perceived as real-time and simultaneous. For example, using a contact microphone and embedded computing platforms to manipulate acoustic signals in real-time, such as Bela. However, it is not clear if auditory feedback simultaneity alone is sufficient to prove that the interaction is convincing.

Considering the Hass Effect in Augmenting Interaction Sound

The Haas precedence effect and the echo threshold concept are closely related to the perception of simultaneity in audio. The Haas effect is a psychoacoustic phenomenon discovered by Dr. Helmut Haas in 1949 [37]. Also known as the “Precedence effect,” the law states that when one sound is followed by another with a latency of approximately 40 ms or less (below humans’ echo threshold), the two are perceived as a single sound [37] – even though coming from different direction. The Precedence effect refers to a group of phenomena that are thought to be involved in resolving competition for perception and localization between a direct sound and a reflection . Based on study by Brown et al.[15], the Precedence effect decreased when the delay was more than 10ms, meaning the listener is able to better determine the source or direction of the audio. In our experiment, the user will wear headphones to listen the audio feedback. Thus, it is important to maintain the latency under 10 ms [15], or should no longer than 40m s [37] to ensure the listener perceives the altered audio feedback as coming directly from the audible object's direction instead of from their headphones.

In the Precedence effect, as the delay increases, the direct sound and the reflection become more distinct. This event is often referred as the echo threshold [15]. A percussive sound will have shorter echo threshold compared to a non-percussive musical instrument [11]. In our experiment we will use mug as audible object for interaction. This mug can be considered as a percussive instrument. Thus, the latency must be short. However, if we can modify the audio feedback to have the similar characteristic as non-percussive instruments (has dangling effect), we may be able to change the user perception about the simultaneity of the interaction based on several types of audio feedback. Study by Braasch et al.[11] reveals that the logarithm of the attack time correlate best with the echo threshold. From the same study, duration also has a positive correlation with the echo threshold.

It is worth considering how these audio properties, along with echo threshold, can affect other auditory perceptions such as unity and agency. By examining audio properties like attack and duration, we can gain a detailed understanding of which properties can be manipulated to achieve the desired perception. Thus, in the experiment, this thesis will include the details of the audio properties and evaluate their impact.

2.5.3 Perception of Unity

When designing an interaction, it is important to consider that, even if auditory feedback is presented simultaneously with an action, the user may still perceive them as separate events due to spatial perception. Simultaneity is when two or more events are perceived as happening at the same time, even though they are unrelated. For example, we may hear a car alarm outside the building when we accidentally hit a water bottle on the desk. Even though those two events happened simultaneously, we were able to perceive the alarm sound as an unrelated event to hitting the water bottle. The Unity assumption is when a user believes that multiple sensory cues (e.g. visual, haptic, auditory) are *"..coming from the same object or event"* [20]. This assumption can create a more immersive and realistic experience; however, few studies have clearly compared different types of stimuli to see if they can emphasize the unity assumption [20]. For tapping interaction with auditory cues as feedback, it is not clear if it is possible to create an interaction that is perceived as unified. Additionally, it is not clear if different stimuli can impact the perception of unity.

2.5.4 Perception of Agency

To develop a convincing interaction, the designer also needs to evaluate the perceived agency of the interaction. In the context of augmenting interaction sounds, *“agency is the feeling of having produced the sound”*[18], including a sense of belonging and control over the interaction [12]. It is important to retain the perceived agency because it can enhance the user’s sense of engagement and immersion in the AR experience. When users feel that they have agency, they are more likely to feel invested in the virtual world and to be motivated to explore and interact with it. The perceived agency can also improve the usability of an AR system by making it more natural for the user to interact with. For example, when users can directly manipulate/ access digital content by tapping objects. To be able to retain our sense of agency, the latency should be under 50 ms for discrete interaction (e.g. tapping/ pressing button), and under 150 ms for continuous interaction (e.g. stirring, dragging) [113].

2.6 Acoustic Manipulation

Generating auditory feedback with characteristics that the user identifies well can be done with synthetic sounds using an audio processing program, such as Supercollider [6]. By modifying the acoustic characteristics of sounds through adjustments in pitch or the addition of sound effects, a more dynamic and real-time augmentation of sound can be achieved.

2.6.1 Manipulating Acoustic Properties (Modifying Real Audio without Additional sound)

Duration, amplitude, and frequency content have been widely used to create material properties in acoustic computation [38]. Lederman[60] tested tactile and aural cues for surface roughness (i.e., the texture). Variations in groove width or distance between neighbouring grooves impacted perceived aluminium plate roughness [60]. Their study opened a discussion on how pitch and amplitude may also impact perceived roughness. For example, a surface is perceived as rougher when there is a louder (high amplitude) tapping sound [60]. Furthermore, for general material sound discrimination, frequency is the most influencing acoustic property [63]. This study found that most participants rely on the frequency to discriminate the material composition and determine the object’s perceived hollowness (density) [63].

Figure 2.1 illustrates an audio envelope consisting of attack, transient, onset and decay. Bello et al.[7] describe audio event terms as follows: (1) attack, on the note is the time interval during

---

[6] Supercollider:https://supercollider.github.io/

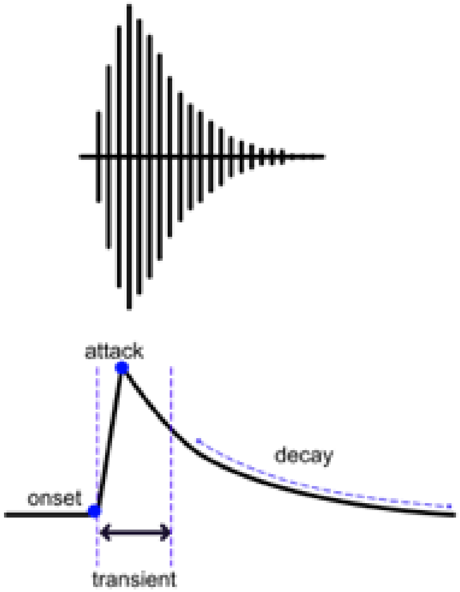


Figure 2.1: Top: waveform. Bottom: audio envelope started with onset, followed by attack, transient, and decay. A note changes event - attack, transient, onset, decay [7]

which the amplitude envelope increases; (2) transients are short intervals during which the signal evolves quickly in some nontrivial or unpredictable way; (3) onset is a single instant chosen to mark the temporally extended transient. In most cases, it will coincide with the start of the transient; (4) decay is a drop or reduction of amplitude over time.

Perceived hardness/stiffness was impacted by attack [34]. In the context of perceived size, Rocchesso et al.[80] concluded that frequency and amplitude impact the perceived size of sounds produced by dropping four balls onto different plates. All these studies were conducted evaluating just the perceived acoustic properties of the sound without directly interacting with it. Little is known about how these acoustic properties impact the perception of object properties in the case of multimodal interaction, i.e., using haptic and auditory senses. Furthermore, can we use these properties to augment the perception of object properties?

2.6.2 Adding Acoustic Effects - Reverberation

Reverberation or reverb is formed by a combination of decay and the collection of echoed acoustic signals [36]. It may reveal an object's weight, texture, size, and stiffness. For example, if a metal object is struck, the reverberation will be bright-ringing indicating that the object has a stiff and dense surface. On the other hand, if a wooden object is struck, the reverberation will be longer and more diffuse, indicating that the object has a rough, porous surface. Additionally, the weight of an object can be inferred from its reverberation, as heavier objects tend to produce longer and more sustained reverberation than lighter ones.

Reverberation (or reverb) is an auditory property that enables the sense of depth and space [90]. This characteristic is often beneficial for Virtual Reality (VR) to compensate for distance compression effects [45]. Reverb can change the perception of sound sources surrounding and potentially offer multi-level auditory feedback [36]. Many studies investigated the use of reverberation in sonification; for example, [102] proposed using reverberation to inform the space available in an office room: small reverb represents a fully staffed room, while large reverb means an emptier space. Reverb also could be tied to the object to complement information conveyed by another auditory parameter, such as the object's underlying sound. For example, a metallic sound may indicate an urgent email message, while a long reverb could reflect the number of such alerts. Frissen et al.[36] found no effect of different audio feedback types on reverb discrimination. However, their study only looked at four variations of audio feedback (speech, singing, drums and piano) categorised into vocal and non-vocal sounds. It did not investigate different material sounds that would be needed for effective auditory interfaces to see how reverb would affect them. In addition, little is known about how reverb may impact the perception of the object properties (i.e. size, stiffness, texture, weight).

In addition, [45] noted that in-congruent reverb degree, for example, high reverb for very close distance, may trigger sensory segregation,*" wherein users are unable to unify visual and auditory sensory stimuli in their perceptions of depth"*. Thus, it is important to find the appropriate amount of reverb to retain the perceived connected experience – perceptual unity.

2.6.3 Replacing Auditory Feedback

The user will expect particular audio feedback while in contact with an object, for example, when tapping the surface of a home appliance, the user expects a metal sound to indicate its quality [3]. Thus, altering these original auditory contact properties by replacing the audio feedback with different properties/characteristics having low/no association with the object may shift the user experience of perceived unity, simultaneity, agency and interpretation of the object.

There is growing evidence that auditory cues can substantially affect the user [95, 97], changing body movement [97] and surrounding perception [89]. Tajadura-Jimenez and Bianchi-Berthouze[94] work demonstrated how manipulating tapping sound strength would alter the user's perception of the object they interacted with. However, this work only looked at one material sound type (wood) and did not explore other potentially useful alternatives, such as fluid, metal, or fabric as output. A study by Tigwell and Crabb[100] explored user experience and preferences when interacting with 15 different surface materials (varying from carpet to tile samples) selected based on their lightness, reflectivity, and texture. This study revealed that, when users interact with different objects, they not only tap on the object but also feel the underlying object texture. Tigwell and Crabb[100] suggesting that providing different material feedback could be promising follow-up research to broaden user experience in interaction with the surface, for example, to express roughness [89].

For casual interactions, it is possible to use a different sound to represent changes in digital content. Such sound should be distinct enough to be noticed. A study by [80] examined how fluid sound, such as filling and emptying a bottle, can help convey continuous information in the background, which supports multitasking. The study found that participants could recognize 91% of emptying sounds and 76% of filling sounds. However, accuracy in determining whether the bottle is empty or half-full was only 40%, mainly due to less friction and interaction sounds. Therefore, users may forget the state of the bottle. To enhance this interaction, one could augment the interaction sound of a shaking bottle. By shaking the bottle, the user could know the state of the bottle and get a cue of the quantity of the presented information or digital content. In terms of varying the auditory feedback, a common type of audio used in the existing studies is musical instrument sound. Musical instrument features (i.e., pitch, rhythm, timbre) have been widely used, from supporting information visualisation by data sonification [83] to creating Earcons (i.e., by augmenting the timbre of the organ or brass sound [13]). The musical instrument sounds also has been used as a non-intrusive notification delivery [50]. However, in contrast to material sounds, using an unrelated sound such as musical instrument sounds to alter the auditory contact properties of an everyday object may feel unnatural, as it could be against the user's prior knowledge. Nevertheless, using musical instrument sounds as audio feedback would allow us to test the extent to which we could stretch the alteration of auditory contact properties whilst retaining the association with the interaction with the object.

2.7 Conclusions and Research Questions

Augmenting the auditory feedback of user interactions with everyday objects has the potential to influence our perception of these objects and their surroundings. Furthermore, this enhancement can be coupled with digital information (e.g., notifications) to expand interaction techniques in the field of human-computer interaction (HCI). Previous studies have showcased various auditory augmentations in different research areas such as TAI, IDWA, or sonification. While techniques like granular synthesis have been explored, other augmentations such as replacing auditory feedback and adding effects like reverb are underexplored. There is still limited understanding of how these augmentations may impact human perception of objects and their environment, as well as their recognizability, to ensure their feasibility in conveying information.

This thesis focuses on the auditory augmentation of real-world sound from interaction with physical objects. The works investigates how auditory augmentation techniques (adding reverberation, replacing auditory feedback) can be applied to the interaction sound from the object, delivering necessary guidelines around acoustic manipulation of interaction sound for designers exploring AAR use cases such as passive haptics and casual interaction. We explore the feasibility of the technology supporting future AAR interaction sounds, creating a real-time AAR interaction proof-of-concept, and measure the perceived interaction (simultaneity, unity, agency) of our proposed concept of auditory augmentation. In addition, we also evaluate its augmentation recognisability to provide designers comprehensive information on using various acoustic manipulations. The key research questions are as follows:

- **Research Question 1 (RQ1)**: *How can we augment real-world interaction sound emitted from an everyday object whilst preserving:*

  (a) *unity*

  (b) *simultaneity*

  (c) *agency*

  Research Question 1 will determine the feasibility of augmenting interaction sound in real-time while retaining the perception of unity, simultaneity, and agency of the interaction. This is a foundational requirement before any subsequent acoustic manipulation is developed. We will develop an AAR prototype to answer this question.

- **Research Question 2(RQ2)**: *How does acoustic manipulation (replacing the interaction sounds, adding reverb) impact:*

  (a) *the perception of object properties such as size and stiffness*

  (b) *the perception of environmental properties such as room size*

  (c) *the recognisability of augmented auditory feedback*

  Following the development of the AAR system, research focuses on exploring acoustic manipulations that potentially alter user perception of object properties and its surroundings, along with these augmentations's recognisability. Two approaches will be taken to answer RQ2: replacing auditory feedback with material and nonmaterial sounds and adding reverb.

Chapter 3

Augmenting Interaction Sound in Real-time

For the development of a successful interactive auditory augmentation system, one important aspect is ensuring that the auditory augmentation can be perceived as real-time as it allows the user to feel a sense of control in producing the sound (agency) as well as their perceived unity and simultaneity toward with their tapping toward object. Hearing sounds that are generated in response to their actions or movements can enhance the sense of presence and interactivity, making the experience more engaging and believable. The focus in this chapter is the technical aspect of real-time auditory augmentation, addressing RQ1.

Previous studies mentioned in the literature review chapter [29, 108, 114, 54] showcased the potential of augmenting audio with digital content such as navigation [54], object internal detail [29], and other everyday reality experiences [114]. This discovery could pave the way for exploring new possibilities in augmenting sound interaction with physical objects in real-time, such as adding digital information by augmenting the acoustic properties (frequency, grain) of the contact sound with the object.

Furthermore, auditory augmentation allows for a more interactive [88, 22] and informative personal audio experience [8, 59, 24], while also maintaining awareness through the use of transparent acoustic headsets [66]. The development of headsets such as BOSE Frames, Sony Ambie sound earcuffs , and the Hololens which supports built-in spatial sound [64], makes the future use of auditory augmentation as an interactive audio-based system possible. This chapter explores real-time auditory alteration using AR/VR/MR headsets and microcontrollers with contact microphones as sensors.

3.1 Developing Real Time Auditory Augmentation using AR Headset

We investigated the technical aspects of AR/VR by exploring the Microsoft Hololens 2, which is a Microsoft Mixed Reality Headset. It has the capability of recognizing hand gestures and providing 3D spatial audio feedback. We created a Hololens 2 application that augments the sound of a knocking a drink can when the user's hand gesture is detected near the can's area. The details of the system are shown in Figure 3.1.

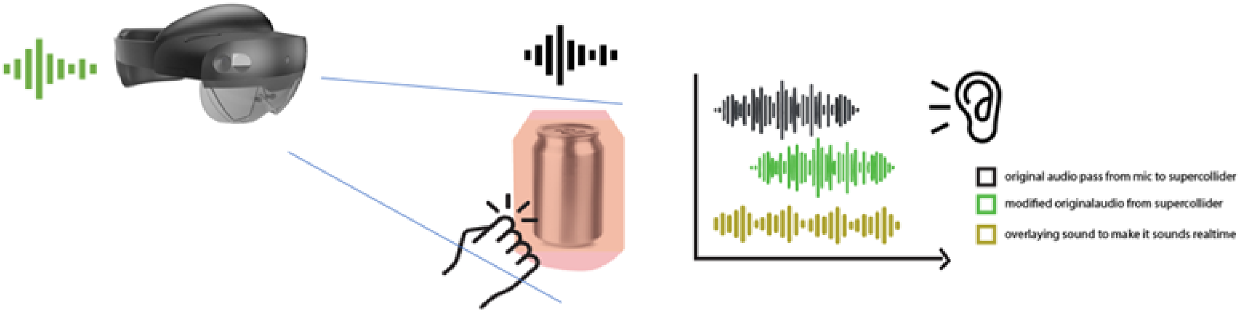


Figure 3.1: AR/VR Based Auditory Augmentation signal processing . Hololens 2 device recognise hand gesture. To reduce the latency, the auditory feedback played immediately after the finger touch the trigger area. Black signal represents original signal emitted from the tapping interaction. Green signal represents digital auditory feedback from hololens. Yellow signal represent mixed signal perceived by user.

In Figure 3.1 the HoloLens 2 camera tracks the user's hand movements, while the built-in microphone captures the sound of knocking (the black signal in Figure 3.1). This method allows for audio feedback to be provided in near real-time, alongside the actual sound produced when knocking the can. We used Supercollider to manipulate, process, and mix the audio from the real-world interaction recorded by the HoloLens microphone with digital sound (the green signal in Figure 3.1). This mixed sound was then played through the HoloLens speakers, shown as a yellow signal in Figure 3.1.

3.1.1 Hand Gesture Interaction

Numerous researchers have investigated the use of hand gestures as an input in interactive systems, from commercial platforms like head mounted displays (e.g., Hololens) to research products such as wristcam [19, 87] and even smart speakers [61]. These studies have demonstrated the potential of hand gestures in advancing the field of augmented and virtual reality by providing intuitive and natural ways for users to interact with virtual environments and manipulate digital content.

Hololens 2 provides five primary hand gestures: touch, hand ray, gaze, air tap, and air tap with hold. Users can interact with displayed holograms by using touch gestures. When the user's hand is detected by the Hololens 2's camera, a floating pointer similar to a mouse pointer appears close to the tip of the user's index finger. This pointer helps in targeting elements, as shown in Figure 3.2. For hand-ray, users should hold their hand with palm facing away to activate hand ray. A laser pointer appears which can be used to target an item. Once an item is targeted, users can perform actions on it. In the case of gaze gesture, gaze is used to make the selection and air tapping is used to confirm it. To perform an air tap gesture, the user needs to hold their hand straight out in front of them with a loose fist. Then, they should point their index finger straight up towards the ceiling, tap it down momentarily, and quickly raise it back up again.

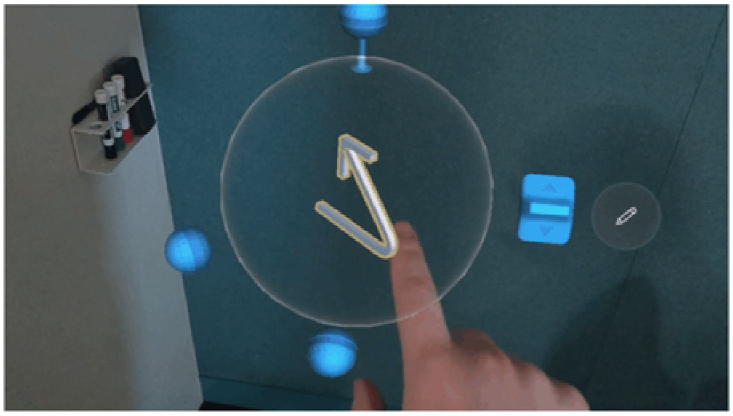
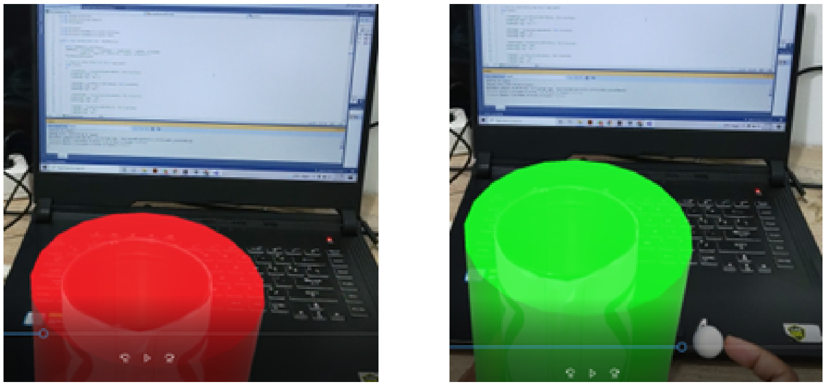


Figure 3.2: Hand gesture in Hololens 2 showing finger pointer (https://learn.microsoft.com/en-us/dynamics365/mixed-reality/guides/authoring-gestures-hl2)



(a) Standby condition

(b) Triggered Condition

Figure 3.3: Figure (a) shows red color indicate trigger area for user to tap. Figure (b) shows a tapping event on the mug with a fingertip. The red area changes color to green.

Figure 3.3 illustrates the functionality of our AAR Hololens app. We used the Mixed Reality Toolkit (MRTK) 2 library to provide augmentation. The user dons a Hololens device and is able to view a cylindrical can. Upon knocking, the Hololens detects whether the user's hand enters the detection region (as depicted in Figure 3.3a). If the hand is within range, the app triggers audio feedback from Supercollider. The audio is then presented in 3D with the speaker from the Hololens 2.

3.2 Audio Processing

Audio processing for this prototype is conducted in Supercollider with input detected by the Hololens' microphone and played on its speakers, with a 44100Hz sample rate. Supercollider code detects audio if it passes the onset threshold and triggers unique audio feedback. The onset function is used to process the audio signal, detecting the beginning of notes/drumbeats/etc. The output of this onset detection is a control-rate trigger signal which is 1 when an onset is detected, and 0 otherwise [92]. The PlayBuf function is then called when the trigger onset is 1 to play the sample sound in memory. Furthermore, a Freeverb function then applied to the buff sound. Freeverb [7] is a function in Supercollider to produce artificial reverb. The algorithm itself is replicated from Faust's [8] reverb library which uses a Schroeder reverberator [25]. The Schroeder reverberator is a widely used reverb effect in the free-software world. It was primarily developed by "Jezar at Dreampoint" and is known for its exceptional tuning. This reverb effect consists of four Schroeder all-pass filters connected in series and eight parallel Schroeder-Moorer filtered-feedback comb filters for each audio channel. The processed audio is then played over the speakers.

```
// setting default server
o = Server.default.options;
o.inDevice_("Windows WASAPI : Jack Mic (Realtek(R) Audio");
o.outDevice_("Windows WASAPI : Speakers (Realtek(R) Audio)");
Server.default.reboot;

//setting buffers for prerecorded auditory feedback
d = Buffer.read(s, "D:../material_fluid.wav");
a = { (PlayBuf.ar(1, d, loop: 0) * 0.1).dup }.play;
a.free;
(
x = {
   // declaring variable in use
   var sig, chain, onsets, pips,out;
   // define input microphone as stereo
  sig=SoundIn.ar(0!2);

   chain = FFT(LocalBuf(512), sig);
   onsets = Onsets.kr(chain, 0.9, \rcomplex);
   // playing auditory feedback
  //playing buffer when triggered
```

---

[7] https://doc.sccode.org/Classes/FreeVerb.html

[8] https://faustlibraries.grame.fr/libs/reverbs/

```
  z=PlayBuf.ar(2, d,trigger:onsets) * 0.1;
  out=FreeVerb.ar(z,mix:0.75,room:0.8);
}.play;
)
s.freeAll;
s.quit;
Server.killAll;
```

The code snippet is being called repeatedly in the Unity application for HoloLens 2. This code will record the input sound from the microphone and mix it with audio in the buffer. The pre-recorded augmented auditory feedback in the buffer is played when the user's finger reaches the detection area.

3.3 Analysis

We experienced a latency of 300ms when using the Hololens 2 approach to recognize user finger gestures and provide auditory feedback. The audio signal emitted from tapping the object was recorded using the Hololens microphone, then processed with Supercollider and played back through the Hololens speaker. The latency measurement involves determining the time duration between the initiation of the real-world tapping sound and the commencement of the output playback audio from the Hololens speaker. AR headphones equipped with camera-based hand gesture recognition, which detect user's hand interactions with objects, have a latency of approximately 35 ms when recognizing the gesture [1]. This latency then increased with audio signal processing latency (e.g., soundcard latency), resulting in around 300ms latency in total.

Although this approach provided flexibility to augment any objects recognized by the Hololens, the latency was too high for real-time acoustic manipulation. Meanwhile, according to a study by [51], a perceived simultaneous interaction can be achieved with less than 20 ms of latency. Despite previous studies emphasizing the establishment of auditory augmentation, our experiment with the Hololens 2 (300ms latency) did not yield desirable latency (under 20ms latency).

Latency has long been a problem for music performers. The issue arises due to the lengthy process required for audio processing. First, sound waves from the musical instrument are converted into electrical impulses, which are then transmitted through a wire to a sound interface. At the interface, the impulses are transformed into a digital signal and stored in the input transport buffer. The audio driver buffer, for example ASIO or CoreAudio driver buffer, receives the signal and waits for the audio interface to utilize it. Once the audio interface is ready, the data are saved in the audio output buffer and sent back to the transport output buffer. Finally, the data are reconverted from digital to electrical signals and transmitted to the speakers.

Reducing the delay that occurs during audio processing can be achieved by decreasing the buffer size. However, another alternative is to consider a hardware-based approach such as using microcontrollers [68, 70, 86, 98]. A microcontroller-based audio processing technology, such as Bela [70], offers audio processing with a latency of under 10ms, making it significantly faster than other methods. As a result, it might be possible to create a proof-of-concept of real-time auditory augmentation of interaction sound with lower latency than using a Hololens device.

3.4 Reducing Latency using Contact Microphone and Microprocessor

In this section, we will explored auditory augmentation using a Bela device with a contact microphone as input, and wired headphones as output. Figure 3.4 presents the hardware used in our setup: (A) Steelseries Siberia 200 headphones, (B) Bela Mini Kit, (C) Asus ROG laptop, (D) Otraki contact microphone, (E) Coffee Mug, and (F) Logitech mouse.

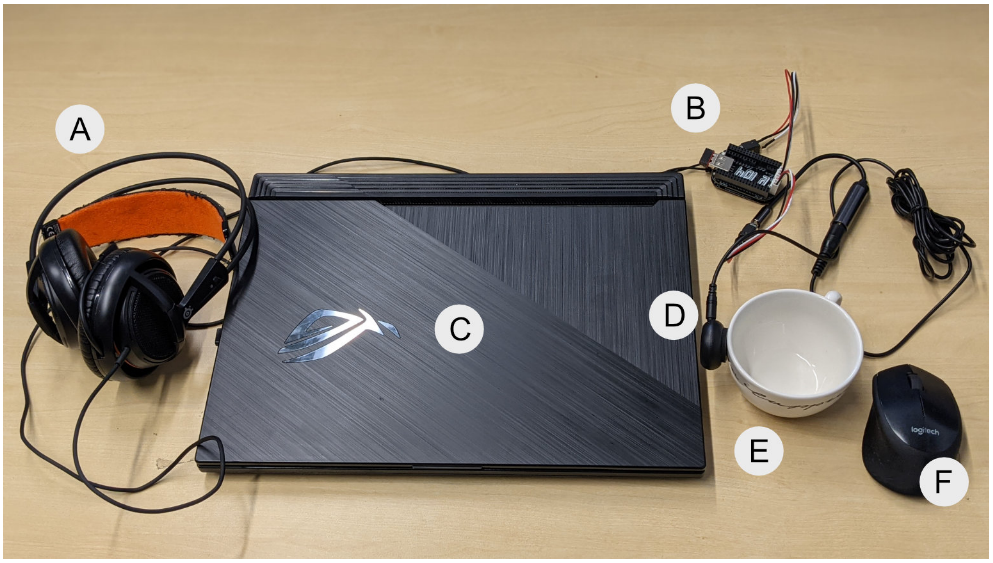


Figure 3.4: Apparatus consisting of the following items: (A) Steelseries Siberia 200 headphones, (B) Bela Mini Kit, (C) Asus ROG laptop, (D) Otraki contact microphone, (E) coffee mug, and (F) Logitech mouse.)

In the prototype phase, we utilized a small white ceramic coffee mug to produce sound. To detect the tapping impact, we affixed an Otraki Adeline AD-35 contact microphone to the mug's interior, a commonly used passive pickup for musical instruments. The microphone was then connected to a Bela Mini. This was linked to a PC laptop operating Supercollider, a platform for real-time audio synthesis and algorithmic composition. Utilizing the Supercollider code, the laptop generated the appropriate audio feedback, which was then played through a pair of Steelseries Siberia 200 headphones. In regards to the audio processing code, we still used the same code as the Hololens 2 setup. Our setup with Bela mini demonstrated an end-to-end latency of just 8 ms with the same measurement technique as Hololens 2. It was well below the suggested 20ms threshold for audio feedback as recommended by Kaaresoja [51].

Despite the low latency produced, our setup with contact microphone and Bela mini, pose challenges in the real-world implementation. For example, to aurally augment multiple objects interaction sound. Our current setup required the contact microphone attached to the object, reducing the flexibility and mobility of the AAR system. Nevertheless, considering a massive development on H-M-D or Frames with spatial audio support and improvement on computer vision method, the future AAR system could be implemented.

Scenarios of multiple interactions or multiple objects could be developed using hand tracking recognition in-built in XR headsets such as the Vive Focus 3 [9], or Meta Quest 3 with 70ms latency [10]. Or using additional accessories such as Ultraleap [11] hand tracking accessory. Moreover, reliably detecting when the finger makes contact with a tracked surface is an area of active research, for example utilizing additional hand/wrist-mounted IMUs to enhance optical tracking [41].

Despite the limitations of our current setup, which still required additional sensor attached to the object, our setup proved that we could build a real-time auditory augmentations for interaction sound with a latency of 8ms. This result also partly answer our RQ1: "How can we augment real-world interaction sound emitted from an everyday object whilst preserving unity, simultaneity and agency?".

3.5 Conclusion and Discussion

3.5.1 Conclusion

The technical aspects of real-time auditory augmentation development are covered in this chapter. Using the Hololens 2 AR/VR/MR HMD for hand gesture recognition and a contact microphone and microcontroller for touch vibration detection, it offers two real-time tapping and audio alteration methods. Our findings show that both approaches successfully alter auditory feedback in real-time, but there is a notable difference in latency. The Hololens 2 HMD has a latency of 300 ms, while our setup with a contact microphone has a latency of 8 ms. Our setup sufficiently addresses the technical requirements of real-time interaction, thus could be used as a test bed for our research aims. We will use this setup to answer RQ2, along with investigating the perceived unity, simultaneity, and agency of our setup to completely answer RQ1.

---

[9] https://www.vive.com/uk/product/vive-focus3/overview/

[10] https://www.roadtovr.com/apple-vision-pro-meta-quest-3-hand-tracking-latency-comparison/

[11] https://www.ultraleap.com/enterprise/pico-neo-3/

Chapter 4

Impact of Auditory Augmentations

In theory, there are limitless opportunities to augment the impact sounds produced by objects. For example, combining various aural feedback mechanisms with digital information generated by user interaction could enhance casual interactions [77]. Other useful potential augmentations may range from altering user perception by manipulating the auditory feedback frequency [95, 96], which could potentially leverage passive haptic feedback, or adding reverberation [36] to create additional parameters for information presentation. However, little is known about how different auditory augmentations may impact the perceived simultaneity, unity, and agency, along with the augmented auditory feedback's recognisability for physical tapping interactions. In addition, we also want to test the possibility of augmenting object and room properties by altering the audio feedback and adding reverb. Thus, in our study, we used the approach with Bela device and contact microphone, as proposed in Chapter 3, creating a near-real-time auditory augmentation of physical object interaction setup with low latency (approximately 8ms, slightly lower thanTajadura-Jimenez and Bianchi-Berthouze's work [94], as a proof-of-concept.

It is crucial to ensure that auditory augmentation is ecologically, and feasibly valid. This step is essential as it acts as a proof-of-concept for auditory augmentation with interaction sound itself. In order to test this concept's feasibility, we examine the interaction's perceived simultaneity, unity, and agency (RQ1).

This chapter comprises of two experiments aiming to answer RQ1, and RQ2. The first experiment demonstrates the feasibility of using auditory augmentation to enhance perception of object and room properties, along with evaluation of the perceived interaction, and its recognisability. The independent variables for this experiment were different material sounds, musical instrument sounds, and various degrees of reverb. The dependent variables were perceived interaction, including simultaneity, agency, and unity, as well as object properties (such as size and stiffness), room size, and recognisability. The second experiment focuses on the impact of reverb mixing on the same dependent variables as Experiment 1.

4.1 Perception of Interaction Simultaneity, Unity, and Agency

To create an interactive experience that is convincing and believable, real-time auditory feedback alone is insufficient. Therefore, we conducted a study to measure the impact of auditory augmentation on interaction sounds using three dependent variables: simultaneity, unity, and agency.

In previous research, the impact of varying auditory feedback has been explored [94, 96, 97]. However, these studies did not take into consideration important factors such as latency [39] and audio properties (such as attack and duration) on the perceived simultaneity. Latency refers to the delay between an action and its corresponding feedback. Audio properties, on the other hand, refer to the physical characteristics of a sound that can influence how it is perceived by the listener.

Despite advancements in technology that have made it possible to develop auditory augmented reality systems with minimal latency, as demonstrated in our study using Bela and a contact microphone, it is still unclear whether we can create an interaction that feels simultaneous to the user. This is a critical factor in determining the quality of the interaction. To accurately assess the quality of the interaction, it is important to measure the perceived simultaneity between the action and the feedback. By doing so, we can gain a better understanding of the factors that influence the user's experience and make informed decisions about how to improve the design of auditory augmented reality systems.

The second dependent variable, unity, is when a user believes that multiple sensory cues are coming from the same object or event. This assumption can create a more immersive and realistic experience. When crafting an interaction, it is important to keep in mind that even if auditory feedback occurs simultaneously with an action, it may still be perceived as separate events due to other factors, such as spatial perception or even user expectations toward the object material: imagine tapping a ceramic mug but hearing a wooden sound. The unity assumption involves the belief that multiple sensory cues, such as visual, haptic, and auditory, are all from the same object or event, creating a more immersive experience. Other than simultaneity perception, measuring the unified perception of the auditory augmentation is important.

Finally, the third crucial factor is agency. Designers must take into consideration the user's perception of agency for a seamless AR interaction. Agency enhances engagement by encouraging users to explore the interaction and making it feel more natural. A previous study by Yang and Yanagisawa indicated that for interaction to be perceived as having agency, the latency should be kept under 50 ms for discrete interactions and under 150 ms for continuous interactions to maintain a sense of agency [112]. The threshold is slightly higher than the threshold for perceived simultaneity (10-40 ms) [51, 37]. In this study, we want to retain both the perceived agency and simultaneity, thus we aimed to get the latency around 10ms

4.2 Object and Room Properties

When interacting with objects, users expect a particular audio feedback. For example, when tapping the surface of a home appliance, a solid firm metallic sound may indicates its thickness and quality. Altering the original auditory contact properties or replacing the audio feedback with different properties (e.g., replace auditory feedback with different material) may change the user experience and interpretation of the object, such as suggested by previous researchers [30, 32, 94, 76, 97, 100] in the literature review chapter 2. Providing different material feedback, emphasized audio characteristic (e.g., psychoacoustics), could broaden the user experience in interaction with the surface, for example, to express roughness.

The audio properties of an object, such as its resonance, can affect how we perceive its temperature [76]. While non-material sounds, such as those produced by musical instruments, have unique characteristics and psychoacoustics (e.g. roughness [31]) that are used in sonification, it is not yet fully understood how these characteristics can affect our perception of an object's properties. It is crucial to consider the effects of non-material sounds, not only to enhance the variability of auditory feedback but also to provide designers with a holistic understanding of auditory feedback characteristics that may specifically influence user perception of object properties, for example, its size.

In terms of room properties, reverb is a property of sound related to the perception of depth and space [36, 90, 45], thus often used to inform amount of information [102]. Considering reverb as a supplementary parameter of other auditory feedback would add a new dimension of information to this approach. For example, a metallic sound could indicate an urgent email message, while a long reverb could represent the number of such alerts. To determine how reverb affects effective auditory interfaces, it is necessary to expand the study to include various materials and sounds. Furthermore, combining reverb with other augmented auditory feedback may result in a cross effect, that warrants further exploration, as it may impact spatial perception and/or object properties such as its material and stiffness.

4.3 Recognisability

Audio processing offers a wide range of possibilities, including changing the auditory feedback type and adding reverb. The combination of audio and reverb has the potential to create multilevel information and alter the user's perception of contact properties. However, it is crucial that these augmentations do not interfere with the recognition of the audio and reverb, and that the sounds remain distinguishable. Despite the promising potential of audio and reverb combinations, little is currently known about how these augmentations can affect the perception of interaction (simultaneity, unity, agency), object properties (size and material), surroundings (room size), and auditory feedback recognition accuracy. Therefore, we conducted an investigation in the context of tapping on an everyday object to better understand these issues.

Our study aimed to explore how the combination of audio and reverb can impact the perception of interaction, object properties, surroundings, and auditory feedback recognition accuracy. By tapping on a daily object, we were able to analyze the effects of these augmentations on the perception of interaction, the size and material of the object, the size of the room, and the accuracy of the auditory feedback recognition.

4.4 Experiment I: Evaluating The Perception of Interaction, Object and Room Properties, and Audio Recognisability

Modifying the contact sound properties of an object or replacing them with a different sound could alter how users perceive their interaction with the object and even influence their behaviour. In Experiment I, we created an interactive auditory augmentation prototype using a coffee mug and assessed how various audio feedback types (material, non- material sounds) and reverb levels affected the interaction perception (simultaneity, unity, agency) in near-real-time addressing RQ1.

There were two independent variables: *Audio feedback type* (10 levels consisting of six surface materials and four musical instruments with audio properties shown in Table 4.1); and *Reverb level* (4 levels consisting of no reverb, low, medium, and high). The study analysed five dependent variables: user perception of the interaction, perception of object properties, surroundings, audio recognition accuracy, and reverb quantity determination. User perception of the interaction was measured by perceived simultaneity [51], unity [20], and agency [12], while perception of object properties was measured by perception of size and material. The perception of the room size was investigated to measure user's perception of the surroundings. The unity assumption was concerned only with the stimuli's perception as coming from the same object, whether from the user tapping on the mug or coming from headphones. To test the study, a within-subjects design was used, where all users tested all combinations of audio feedback types and reverb levels in a 10 x 4 design. The order of conditions was counterbalanced to control for order effects. The experiment was approved by the University ethics committee. We inv estigated the RQ1(a-c)-RQ2(a-c) as mentioned in Chapter 2.

4.4.1 Study Designs

Auditory Feedback

Previous studies [80, 94, 100] did not evaluate augmentation with non material sound. Additionally, considering the Hass effect [37, 11, 81, 42] such as discussed in Chapter 2, it is also important to consider the nature of audio properties such as attack duration. To ensure the diversity of the provided stimuli, we also consider its chroma features variat ion. Ellis[28] defined the chroma feature as *"powerful representation for music audio in which the entire spectrum is projected onto 12 bins representing the 12 distinct semitones (or chroma) of the musical octave"*. A sound with rich pitch variability has a rich colour representation in its chroma. This allows for a simplified representation of the complex pitch structure of a sound. A sound that has rich pitch variability will have a rich colour representation in its chroma, making it easier to visualize and analyse.

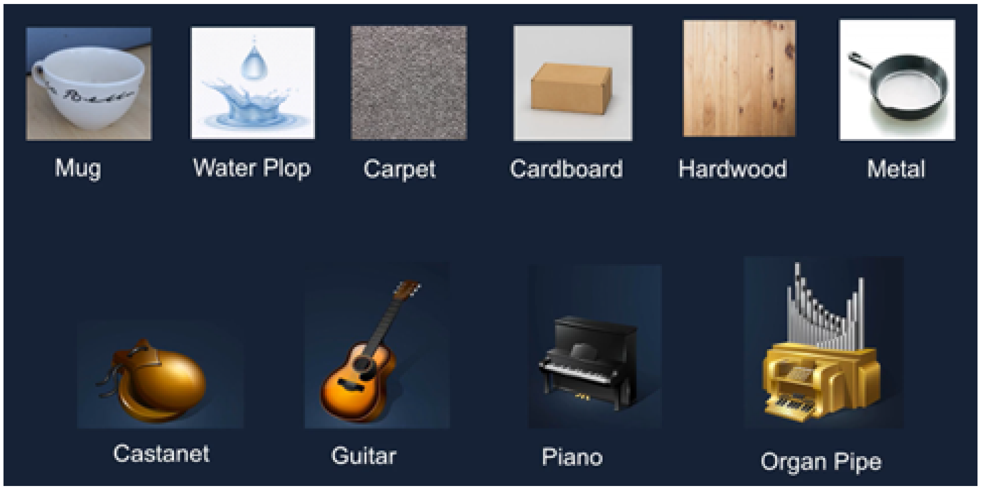


Figure 4.1: Audio Stimuli illustration for Experiment I, consists of: original mug sound, water plop, carpet, cardboard, hardwood, metal, castanet, guitar, piano, and organ pipe.

The experiment uses various types of audio feedback to augment the sound of tapping a mug, from material and non-material sound as shown in Figure 4.1. The following are the types of audio feedback used in the experiment:

1. *Mug*: This is the original sound of the mug and serves as a baseline/control for the experiment.
2. *Water Plop*: This is a fluid sound that is often used for noise masking because it has similar characteristics to brown noise [17]. It is also known to create a pleasant ambiance [49]. The attack duration of this sound is similar to the original Mug sound, but it has a longer duration.
3. *Carpet*: The Carpet sound is similar to white noise and has a uniformly spread power spectrum. It also has rich pitch variability.
4. *Cardboard*: The Cardboard sound has a shorter duration compared to carpet or hardwood and a greater than 10ms attack.
5. *Hardwood*: This sound has a similar wooden characteristic to cardboard but has richer pitch variability and longer duration.

6. *Metal*: The Metal sound has a long duration and less pitch variability than hardwood, carpet, or cardboard.

7. *Castanet*: This is a percussive instrument with a short attack (<1ms) and duration.

8. *Guitar*: This is a representative string instrument with an attack of less than 10ms. Users may have prior knowledge of playing the guitar by pulling the string, but they may also expect a tapping sound instead.

9. *Piano*: This represents both string and percussive instruments with an attack of less than 10ms but a long duration.

10. *Organ Pipes*: This represents woodwind instruments with a greater than 10ms attack.

These audio feedback types cover a wide range of sounds and will help determine the best sounds to augment the tapping of the mug. The attack and duration of the sounds used in this study were calculated using the Essentia Python Library (shown in Table 4.1), which uses the algorithm proposed by Zolzer[119]. Furthermore, in order to visually represent the stimuli used in our study, we have presented a chroma feature diagram in Figure 4.2. Overall, the use of chroma feature diagrams provides a clear and effective way to represent and compare acoustic signals in our study.

Table 4.1: Audio Properties

Category	Audio	Duration (ms)	Attack (ms)
Original (Control)	Mug	40.27	0.43
Material	Water plop	79.09	0.35
	Cardboard	179.91	29.57
	Carpet	388.53	6.6
	Hardwood	474.51	9.56
	Metal	982.04	4.71
Music Instrument	Castanet	37.91	0.25
	Guitar	380.09	5.41
	Piano	722.99	6.76
	Organ pipes	538.14	37.97

*Duration represents the audio duration in a millisecond, and attack represents the duration of the attack in a millisecond, calculated at 70% of the total estimated attack duration, with measurement based on [11].

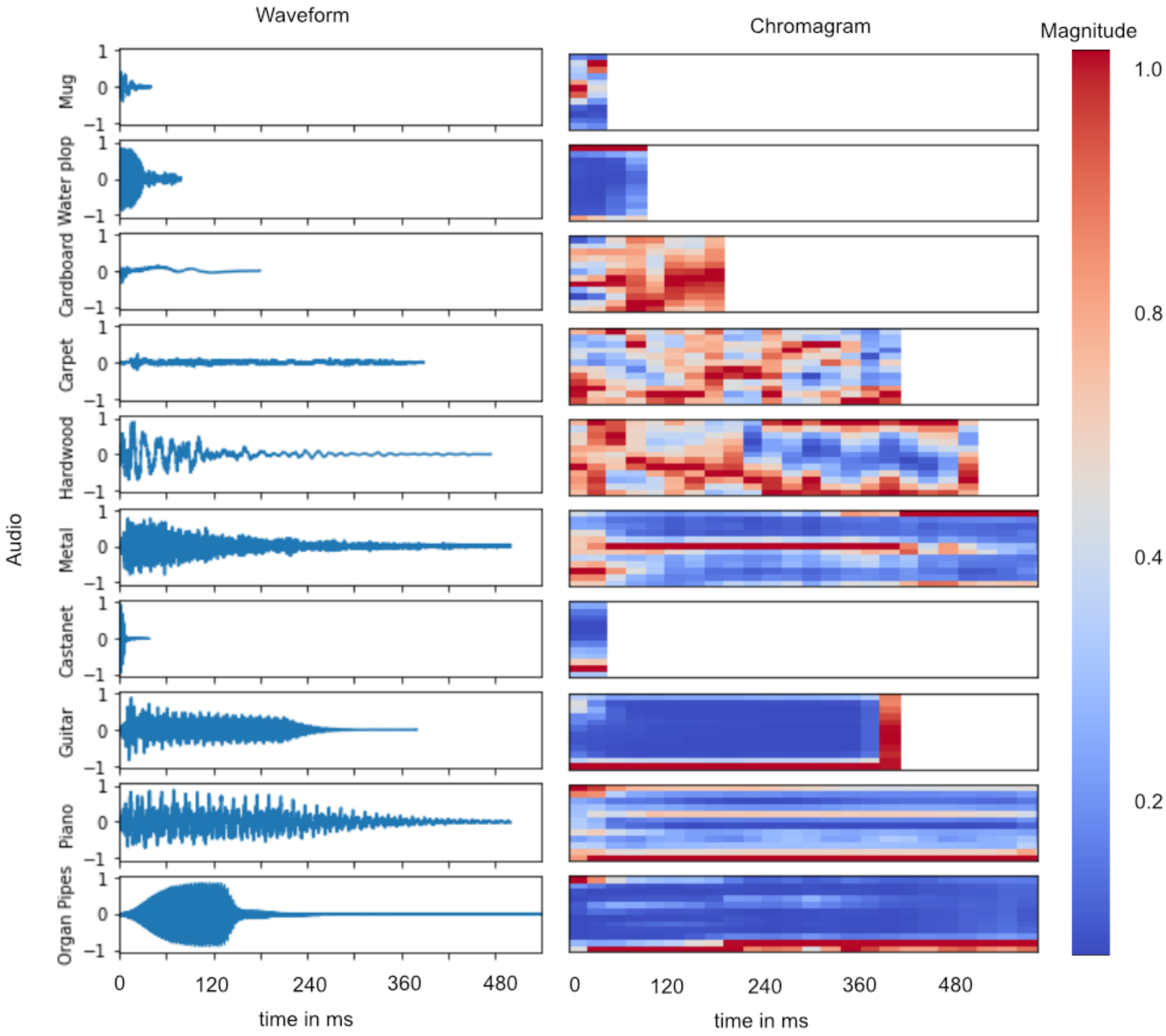


Figure 4.2: Stimuli Chromagram visualizing pitch structure of the stimulus. X axis represent the duration of the audio signal. Color representing the pitch magnitude of the signal.

Apparatus

We used our setup shown in Section 3.2 Figure 3.4 for the experiment. We developed a PC app using Unity (https://unity.com/) to control and modify the auditory stimuli, provide questionnaires and record user responses. Our main objective was to determine if we could substitute the original impact sound with our modified sounds. For this study, we did not use an AR headset as we needed to keep latency as low as possible. The available AR headsets with hand-tracking were not fast enough, so we devised a custom setup to test our ideas. As AR technology improves over time, our approach can be performed natively in the headset.

Regarding audio feedback, we recorded the original mug sounds using a Samsons Pro condenser microphone. However, the real-time audio quality from the contact microphone was poor, so we decided to use pre-recorded mug soun ds recorded by the researcher doing single

Figure 4.3: Participant tapping coffee mug during the experiment

tip tapping on the mug, triggered by the contact microphone's detection. We sourced the material sounds from Freesounds and the musical instrument audio feedback from Braasch et al.[11] work. To create a reverb effect, we combined echo (dry-wetness balance) and modified the decay properties. We used the Freeverb function in the Faust library in Supercollider to achieve this. In Experiment I, we used the mixing and room size (depth/decay) parameters from the Freeverb function in Supercollider. We tested four different settings: no reverb (mix=0, room=0), low reverb (mix=0.3, room=0.5), medium reverb (mix=0.5, room=1), and high reverb (mix=1, room=1).

Participant Recruitment

We enrolled a total of 20 participants in our study, with 9 identifying as Female, 10 as Male, 1 preferring not to say. The age range of the participants was between 18 to 44 years old, with 20% of individuals being between 18-24 years of age, 55% between 24-35 years of age, and 25% between 35-44 years of age. Among the participants, 85% were students, and 15% were full-time workers. Participants were required to have no hearing impairments and as compensation, were given a £10 Amazon voucher.

Procedure

Prior to taking part in the study, all participants received an information sheet to review and a consent form to sign. Additionally, they were required to complete a demographics questionnaire and undergo training with the audio and reverb. This training involved listening to all of the different audio feedback types and reverb levels that would be used during the study. During the study, each participant was instructed to perform a single "finger tap" interaction with the coffee mug using their finger tip, while wearing headphones to listen to the audio feedback as shown in Figure 4.3. There were a total of 40 different combinations of stimuli, each of which was played three times, resulting in a total of 120 stimuli presented in a randomized order. Participants tapped the mug three times for each stimulus to hear the sound. The details of the experimental stimuli and the trial are as follows:

Number of Audio x Number of Reverb x Number of Trial:

10 audio x 4 reverb x 3 trials = 120 trials

The study lasted approximately 60 minutes, with participants given a break after every 40 stimuli. After each condition, participants answered questions using the same laptop. The average scores for each stimulus were calculated from the scores obtained from the three trials. Table 4.2 presents the specific questions asked during the study.

4.4.2 Result and Analysis

This experiment investigated how the participants perceived various types of audio feedback and different levels of reverb. Results are presented in Table 4.3 and Table 4.4(displaying the raw results), Figure 4.4 and Figure 4.5 (visualizing the audio and reverb effects on dependent variables), and Table 4.6 (presenting the statistical analysis results). To ensure an accurate analysis of the data, we checked the data distribution of all dependent variables using the Shapiro-Wilk test. We found that the data were non-Gaussian. To address this issue, we used the Aligned Rank Transform tool (ARTtool) [109] to transform our data before conducting a two-way repeated measure ANOVA. After the ANOVA, we carried out a Bonferroni *post-hoc* comparison utilizing ARTtool. We also conduct a Pearson's Correlation test to evaluate the correlation of audio properties (attack, duration) with the dependent variables.

RQ1: Interaction Perception

a) **Simultaneity**

Table 4.3 displays that the average simultaneity ratings for different auditory feedback types are mostly around 1, with an overall average of 0.928. This indicates that the participants perceived the auditory feedback from the mug to be synchronous with their tap.

Furthermore, there was a significant negative correlation between attack and simultaneity perception $r(798) = -.088, p = .012$. Meanwhile, no significant correlation was found between duration and simultaneity perception $r(798) = -.002, p = .942$. The carpet sounds produced more varied results, lower than other audio types, with 0.85 as can be seen in the Figure 4.4. However, Table 4.6 indicates that there is no significant difference between audio types in terms of simultaneity $F((9,741) = 2.207, p = 0.020, \eta^2 = 0.03)$, despite.

Table 4.2: Question design in Experiment I and II

Dependent Variables	Questions	Types
The perception of:		
(1) Simultaneity [51]	Was the audio feedback simultaneous with your tap?	Yes/No question
(2) Unity [20]	Were the audio feedback and your tap perceived as a single event?	Yes/No question
(3) Agency ([18]	My tapping produced the audio feedback	five-point Likert scales, recorded as 1-5 : “Strongly Disagree” to “Strongly Agree”
The perception of:		
(1) Mug size	How did the audio feedback affect your perception of mug size?	five-point Likert scales, recorded as 1-5 : “Smaller” to “Bigger”
(2) Room Size	How did the audio feedback impact your perception of room size?	five-point Likert scales, recorded as 1-5 : “Smaller” to “Bigger”
(3) Mug stiffness	How did the audio feedback affect your perception of the surface material (stiffness)?	five-point Likert scales, recorded as 1-5 : “Softer” –“No change”- “Harder”
The accuracy of:		
(1) Recognition	What was this sound?	Multiple choices of audio names
(2) Quantity Determination	How would you classify the audio feedback you heard?	Multiple choices of reverb names

Table 4.3: Experiment I Mean and standard deviation of audio feedback.

Audio	Simultaneity (0-1)		Unity (0-1)		Agency (1-5)		SizeMug (1-5)		SizeRoom (1-5)		MugStiffness (1-5)		AccRecog (0-1)		AccDeter (0-1)	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Mug	0.946	0.145	0.896	0.235	4.483	0.748	3.154	0.797	3.333	0.95	2.983	0.863	0.842	0.26	0.508	0.379
Water Plop	0.958	0.123	0.892	0.204	4.371	0.733	3.275	0.643	3.217	1.043	2.483	0.915	0.979	0.123	0.554	0.356
Carpet	0.85	0.254	0.738	0.366	3.992	1.08	3.092	0.641	3.233	0.919	2.517	0.926	0.675	0.397	0.375	0.353
Cardboard	0.879	0.244	0.846	0.28	4.196	0.924	2.842	0.723	3.258	1.01	2.671	0.821	0.521	0.4	0.454	0.365
Hardwood	0.954	0.138	0.854	0.28	4.225	0.863	3.533	0.816	3.508	0.966	3.554	0.958	0.717	0.319	0.442	0.355
Metal	0.921	0.193	0.892	0.253	4.404	0.818	3.517	0.621	3.312	0.993	3.721	0.848	0.921	0.186	0.392	0.351
Castanet	0.946	0.135	0.921	0.186	4.362	0.767	2.833	0.748	3.204	0.939	2.95	0.952	0.787	0.328	0.571	0.349
Guitar	0.933	0.201	0.904	0.244	4.467	0.711	3.179	0.431	3.062	0.789	2.617	0.875	0.55	0.431	0.421	0.351
Piano	0.963	0.106	0.912	0.223	4.421	0.793	3.55	0.6	3.192	0.917	3.154	0.972	0.696	0.387	0.354	0.341
Organ Pipes	0.933	0.171	0.85	0.3	4.246	0.9	3.154	0.551	3.263	0.923	2.696	0.773	0.896	0.246	0.446	0.352
Average	0.928		0.87		4.317		3.213		3.258		2.934		0.758		0.452	0.928

Simultaneity ranging from (0-1), Unity (0-1), Agency (1-5), SizeMug (1-5), SizeRoom (1-5), MugStiffness (1-5), Accuracy of audio recognition - AccRecog (0-1), Accuracy of reverb determination - AccDeter(0-1).

Table 4.4: Experiment I Mean and standard deviation of reverb score across dependent variables.

Reverb	Simultaneity		Unity		Agency		SizeMug		SizeRoom		MugStiffness		AccRecog		AccDeter	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
No	0.935	0.176	0.888	0.23	4.37	0.834	3.143	0.672	2.828	0.708	3	0.971	0.777	0.347	0.452	0.404
Low	0.945	0.149	0.892	0.231	4.307	0.861	3.157	0.632	3.133	0.858	2.928	0.926	0.802	0.331	0.422	0.333
Medium	0.933	0.177	0.868	0.281	4.318	0.834	3.295	0.687	3.355	0.93	2.885	1.016	0.785	0.331	0.42	0.323
High	0.9	0.211	0.833	0.308	4.272	0.866	3.257	0.82	3.717	1.041	2.925	0.994	0.67	0.385	0.513	0.369
Average	0.928		0.87		4.317		3.213		3.258		2.934		0.758		0.451	

Table 4.5: Attack and Duration correlation table across dependent variables.

Factor	Correlation	Simultaneity	Unity	Agency	SizeMug	SizeRoom	MugStifness	AccRecog	AccDeter
Attack	corr	-0.088	-0.135	-0.114	-0.02	0.004	-0.115	0.002	-0.081
	p-value	0.012	<0.001	0.0012	0.564	0.892	0.001	0.934	0.021
Durration	corr	-0.002	-0.007	0.0005	0.25	0.014	0.253	0.034	-0.153
	p-value	0.942	0.828	0.988	<0.001	0.691	<0.001	0.326	<0.001

Table 4.6: Experiment I ANOVA Analysis result for Audio and Reverb on Audio and Reverb effect.

Factor	Estimate	Simultaneity	Unity	Agency	SizeMug	SizeRoom	MugStiffness	AccRecog	AccDeter
Audio	F value	2.207	4.236	4.116	14.887+++	3.052	21.374+++	30.952+++	3.985
	p-value	0.02	<0.001	<0.001	<0.001	0.001	<0.001	<0.001	<0.001
	η^2	0.03	0.05	0.05	0.15	0.04	0.21	0.27	0.05
Reverb	F value	5.752	2.176	2.163	3.871	79.937+++	0.893	11.411	2.673
	p-value	<0.001	0.089	0.091	0.009	<0.001	0.444	<0.001	0.046
	η^2	0.02	0	0	0.02	0.24	0	0.04	0.01
Audio * Reverb	F value	2.493	2.27	0.938	1.046	1.464	0.573	2.658	2.084
	p-value	<0.001	<0.001	0.556	0.402	0.061	0.961	<0.001	0.001
	η^2	0.08	0.08	0.03	0.04	0.5	0.2	0.09	0.07

The +++ indicates a large effect size measured with η^2. Auditory feedback type has a large effect on SizeMug, Material and Accuracy of Recognition. However, Reverb only has large effect on SizeRoom. There was no interaction between auditory feedback type and reverb for Agency, SizeMug, SizeRoom, material. However, there was an interaction between audio and reverb with a medium effect size on simultaneity, unity, AccRecog (accuracy of audio recognition), and AccDeter (accuracy of reverb determination).

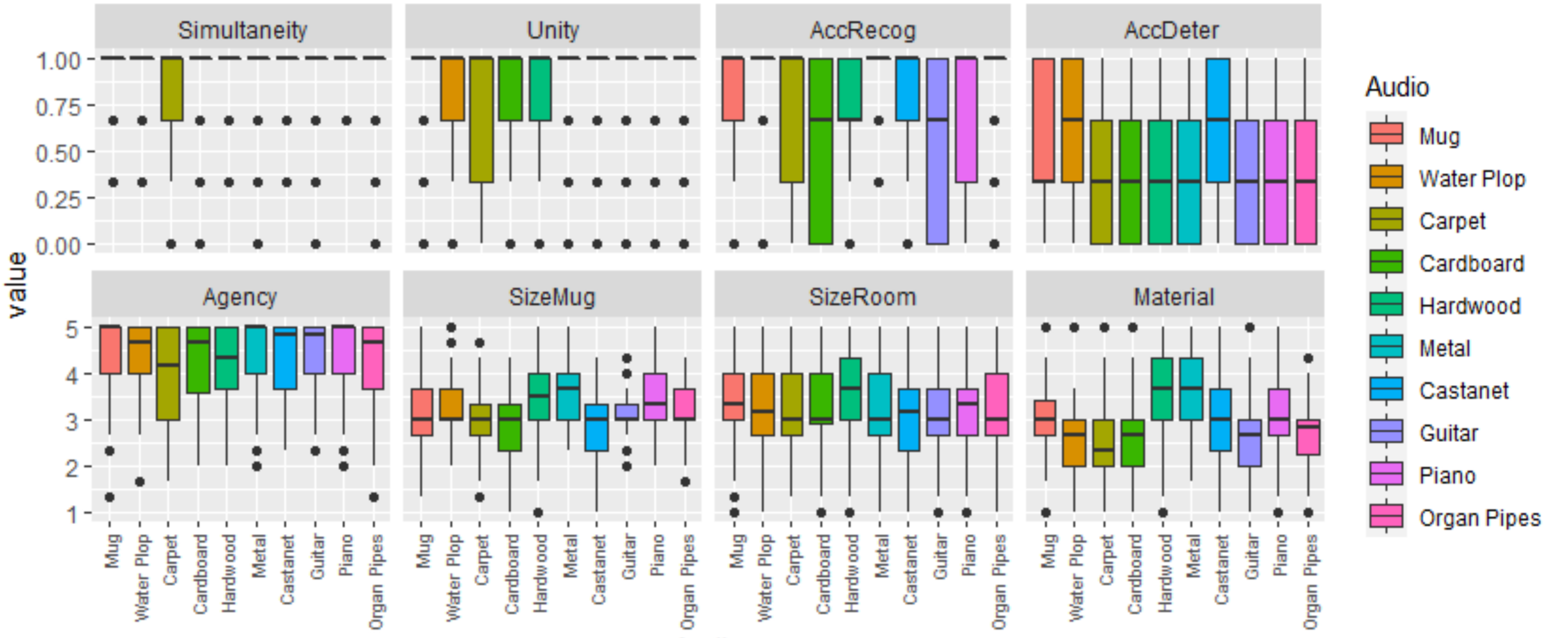


Figure 4.4: Audio Impact on Dependent Variable Experiment I. Y axis represent the score provided by participant. Simultaneity, Unity, scored with binary option (0-1). Agency, SizeMug, SizeRoom, Material scored between (1-5). AccRecog and AccDeter scored 1 if correct, and 0 if incorrect.

In terms of reverb, high reverb had the lowest simultaneity score, and based on the ANOVA result, there was a significant difference among reverb degree but with a small effect size $F((3,741) = 5.752, p < 0.001, \eta^2 = 0.02)$. There is also a medium effect of the interaction between audio and reverb $F((27,741) = 2.493, p < 0.001, \eta^2 = 0.08)$.

Overall, the data suggests that the audio feedback from the mug was perceived to be simultaneous with the tap. The study also showed that there was a significant difference between audio and reverb, but their effect on simultaneity was minimal, with a medium effect of their interaction. The *post-hoc* test revealed that high reverb significantly differed from the low and no reverb.

b) **Unity**

Based on the result presented in Table 4.6, it can be seen that the audio $F((9,741) = 4.236, p < 0.001, \eta^2 = 0.05)$ and audio-reverb interaction $F((27,741) = 2.270, p < 0.001, \eta^2 = 0.08)$ had a significant difference with a noteworthy medium effect size on the perception of unity means. In regards to reverb, no significant difference was found across various reverb settings $F((3,741) = 2.176, p = 0.089, \eta^2 = 0.000)$.

If we look at Table 4.5, a moderate negative correlation between attack and unity is found $r(798) = -.135, p < .001$, the longer the attack duration, the lower the unity perception. In our study, we found that the carpet sound had the lowest unity mean (0.738) compared to other sounds. Table 4.1 showed that the carpet sound has an attack duration of 6.6 ms, which is lower than cardboard and organ pipes (>20 ms). Despite this, the carpet still received the lowest unity score. Still, according to the *post-hoc* test, this result was only significantly lower than Castanet, Organ Pipes, Piano, Water Plop, and Metal, but there was no significant difference in unity for the remaining sounds, including the control condition, mug sound. This suggests that there may be another factor besides attack and duration that influences unity perception.

Furthermore, we discovered an interaction between audio and reverb. Our *post-hoc* test revealed that Cardboard and Carpet with high reverb were perceived to have lower unity than metal with low reverb. Cardboard with high reverb was also significantly lower than mug with no reverb.

c) **Agency**

According to the data in Table 4.6, there was a noticeable impact of audio $F((9,741) = 4.116, p < 0.001, \eta^2 = 0.05)$ on participants' perceived agency, with a moderate effect size. However, no significant differences were observed in regards to reverb $F((3,741) = 2.193, p = 0.091, \eta^2 = 0.000)$ or the interaction between audio and reverb $F((27,741) = 0.938, p = 0.556, \eta^2 = 0.030)$. As for Table 4.3, it indicates that the mug had the highest level of agency among all the objects tested (Mean=4.483), while the carpet scored the lowest (Mean=3.992). Further analysis using the Bonferroni contrast revealed that only

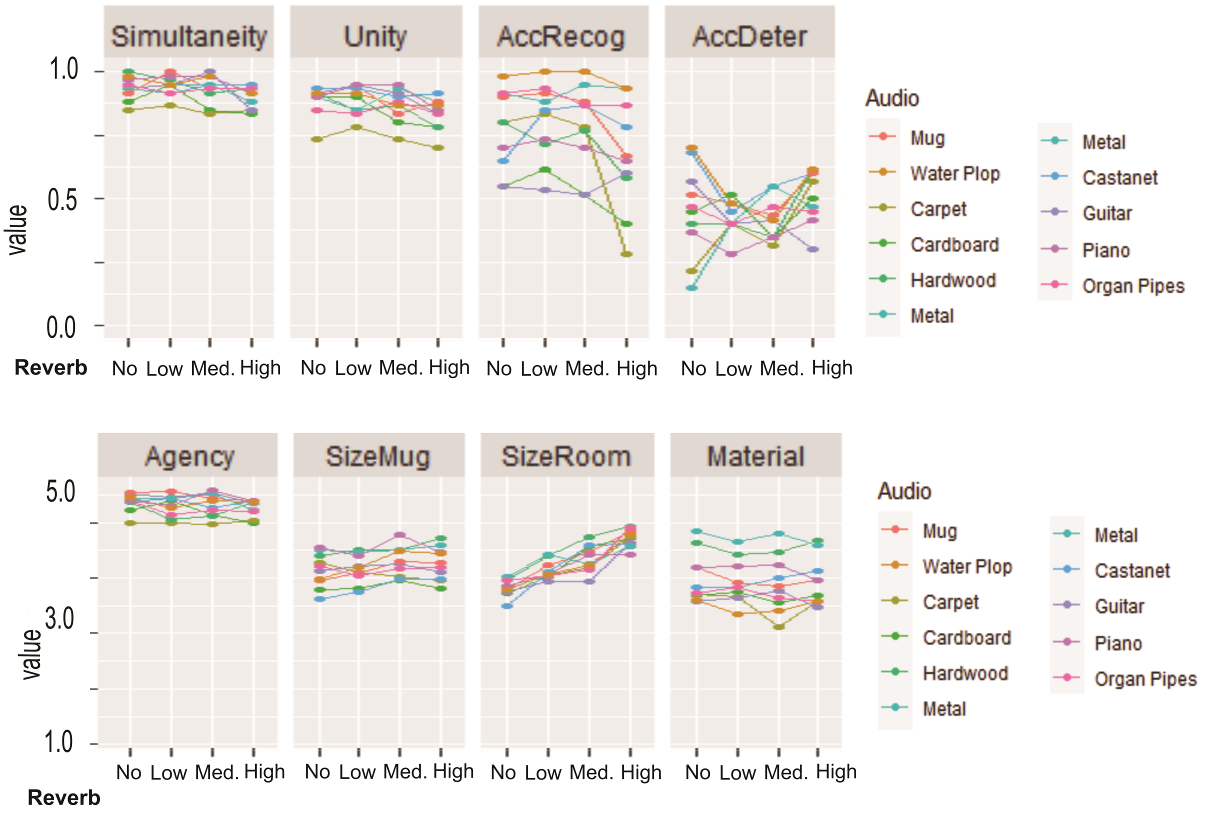
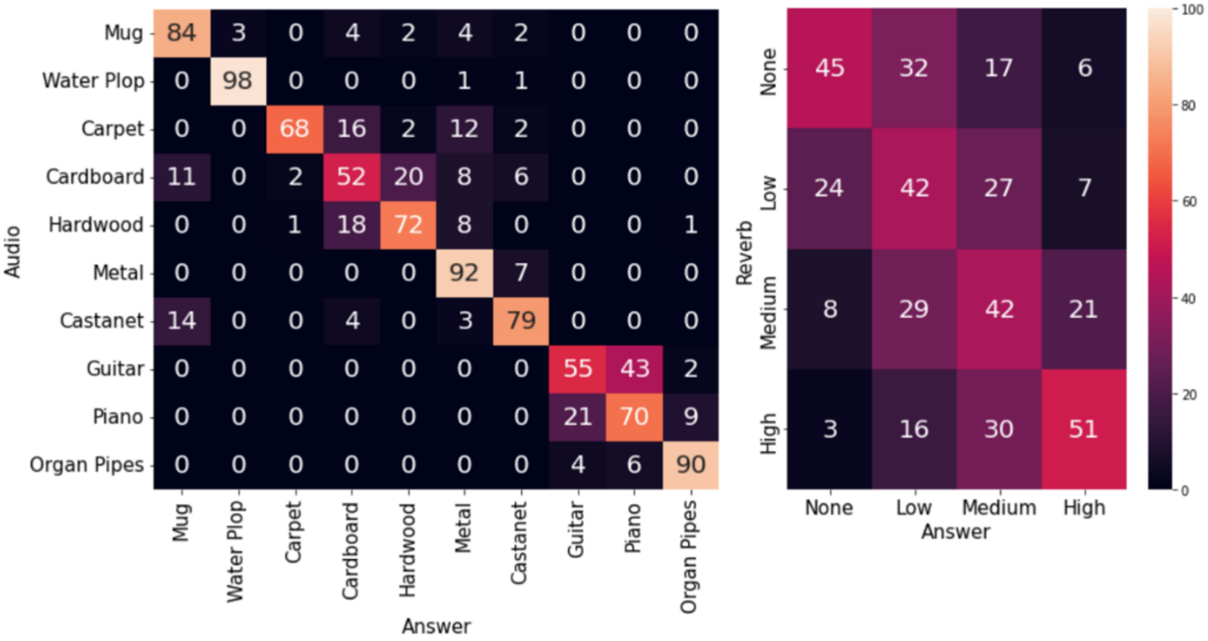


Figure 4.5: Reverb Impact on Dependent Variable Experiment I. X-axis represents the reverb categories. Y axis represents the score for each dependent variable. Simultaneity, Unity, scored with binary option (0-1). Agency, SizeMug, SizeRoom, Material scored between (1-5). AccRecog and AccDeter scored 1 if correct, and 0 if incorrect.



	Mug	Water Plop	Carpet	Cardboard	Hardwood	Metal	Castanet	Guitar	Piano	Organ Pipes
Mug	84	3	0	4	2	4	2	0	0	0
Water Plop	0	98	0	0	0	1	1	0	0	0
Carpet	0	0	68	16	2	12	2	0	0	0
Cardboard	11	0	2	52	20	8	6	0	0	0
Hardwood	0	0	1	18	72	8	0	0	0	1
Metal	0	0	0	0	0	92	7	0	0	0
Castanet	14	0	0	4	0	3	79	0	0	0
Guitar	0	0	0	0	0	0	0	55	43	2
Piano	0	0	0	0	0	0	0	21	70	9
Organ Pipes	0	0	0	0	0	0	0	4	6	90

	None	Low	Medium	High
None	45	32	17	6
Low	24	42	27	7
Medium	8	29	42	21
High	3	16	30	51

Figure 4.6: Confusion Matrix for Recognizability of each audio feedback and reverb. Each audio and reverb, paired with the rest of the stimulus.

the agency of carpet, cardboard, and hardwood were significantly lower than that of the mug (Mean=4.483). Although the carpet sound received the lowest agency mean score compared to the other sounds, it was only significantly lower than the mug, guitar, and piano. Table 4.5 revealed that attack have a significant negative correlation toward the agency $r(798) = -.114, p = .0012)$.

RQ2: Impact of Acoustic Manipulations

a) **Object (size, stiffness) properties**

- **Object Size**:

  The results presented in Table 4.6 demonstrate a significant effect of audio on object size $F((9,741) = 14.887, p < 0.001, \eta^2 = 0.150)$ and a strong effect on the perceived size of the mug, while reverb had no significant effect $F((3,741) = 3.871, p = 0.009, \eta^2 = 0.02)$. No interaction was found between audio and reverb $F((27,741) = 1.046, p = 0.402, \eta^2 = 0.040)$.

  The perceived size of the mug varied from 2.833 to 3.55 refer to Table 4.3, which means slightly smaller to slightly bigger. This indicates that altering audio feedback can significantly influence the user's perception of an object's size. The piano elicited the largest perception of size among all audio feedback types.

  The results of the *post-hoc* analysis revealed that the piano sound was perceived as significantly bigger compared to the original Mug, Water Plop, Carpet, Card-

Table 4.7: Experiment I Recognition Accuracy result (audio recognition and reverb quantity determination). The table also shows the preferences rank.

Audio	Accuracy of Audio Recognition(AccRecog)					Accuracy of Quantity Determination(AccDeter)					Preferences Rank
	No reverb	Low reverb	Medium reverb	High reverb	Avg	No reverb	Low reverb	Medium reverb	High reverb	Avg	
Mug	90	91.67	88.33	66.67	84.17	51.67	48.33	43.33	60	50.83	8.32
Water Plop	98.33	100	100	93.33	97.92	70	48.33	41.67	61.67	55.42	7.84
Cardboard	55	61.67	51.67	40	52.09	45	51.67	35	50	45.42	1.58
Carpet	80	83.33	78.33	28.33	67.5	21.67	40	31.67	56.67	37.5	3.63
Hardwood	80	71.67	76.67	58.33	71.67	40	40	35	61.67	44.17	0.95
Metal	91.67	88.33	95	93.33	92.08	15	40	55	46.67	39.17	1
Castanet	65	85	86.67	78.33	78.75	68.33	45	55	60	57.08	3.58
Guitar	55	53.33	51.67	60	55	56.67	40	41.67	30	42.08	4.84
Piano	70	73.33	70	65	69.58	36.67	28.33	35	41.67	35.42	3.95
Organ Pipes	91.67	93.33	86.67	86.67	89.59	46.67	40	46.67	45	44.58	4.42

board, Castanet, Guitar and Organ pipes sounds. As there were significant differences between the piano and mug sounds, we need to examine the audio properties in detail, which can be found in Table 4.1 and Table 4.5. Table 4.5 shows a significant positive correlation between duration and the perceived size of the mug $r(798) = .250, p < .001$, indicating that the longer the duration, the bigger the object appears. Based on Table 4.1, it can be seen that the piano sound has a long duration similar to the metal sound, which is supported by the *post-hoc* tests that revealed no significant difference between piano and metal. On the other hand, Castanet had the lowest size perception, being significantly lower than hardwood, metal, piano and water plop but not significantly different to the mug sound.

- **Object stiffness**:
  Based on the insights outlined in Table 4.3 and Figure 4.4, it is clear that there are discernible differences in the perceived stiffness of various audio cues. The two-way ANOVA analysis in Table 4.6 also shows a significant difference among audio types on the perception of mug stiffness $F((9,741) = 21.374, p < 0.001, \eta^2 = 0.210)$ with a strong effect. The *post-hoc* results suggest that Metal (with a mean of 3.72) is noticeably harder than the other audio cues, except for Hardwood (with a mean of 3.554). Conversely, the other audio cues tend to make the material surface feel smoother or cause no significant change in hardness (with scores of 3 or lower). Additionally, the Water Plop audio cue is perceived as significantly smoother (with a mean of 2.48) when compared to Mug, Castanet, Piano, Hardwood, and Metal. Table 4.5 shows a strong negative correlation between attack and stiffness $r(798) = -.115, p = .001$, and a strong positive correlation between duration and stiffness $r(798) = .253, p < .001$. This suggests that longer duration and shorter attack induce a perception of a stiffer material.

b) **Environtment (room size) properties**
There is a significant difference of room size perception among various audio feedback $F((9,741) = 3.052, p = 0.001, \eta^2 = 0.040)$ and reverb $F((3,741) = 79.937, p < 0.001, \eta^2 = 0.240)$. However, it was found that only the reverb had a substantial effect on the room size perception. Furthermore, a *post-hoc* test showed that participants perceived a significantly larger room size when exposed to the sounds of Hardwood, only in contrast to Castanet, Guitar, and Piano.

The hardwood sound had the highest score, guitar the lowest score for the room size. However, no significant correlation was found between their attack $r(798) = .004, p = .892$ and duration $r(798) = .014, p = .691$ on the room size perception. Figure 4.5 displays data that shows a clear positive slope in "SizeRoom", indicating that the participants perceived a larger room size with the addition of reverb. No reverb has the lowest room size score, whilst high reverb got highest room size perception.

c) **Recognisability**

- **Auditory feedback recognition (AccRecog)**

  The participants demonstrated varying levels of accuracy in recognizing auditory feedback $F((9,741) = 30.952, p < 0.001, \eta^2 = 0.270)$, which had a large effect. Additionally, there was a weak but significant difference in reverb levels $F((3,741) = 11.411, p < 0.001, \eta^2 = 0.040)$. Notably, audio and reverb had an interaction effect of medium intensity $F((27,741) = 2.658, p < 0.001, \eta^2 = 0.090)$, as depicted in Figure 4.6. Most audio feedback experienced a decline in accuracy when the reverb was high (r4). However, some audio cues showed an increase in accuracy as the reverb became stronger, from r1-r3, before experiencing a slight decrease in r4. The guitar, on the other hand, had the opposite effect. Its accuracy declined from reverb r1-r3, but increased when the reverb was r4, as indicated in Table 4.7. Among the cues, Water plop had the highest accuracy score (Mean=0.979), which was significantly different from all other cues based on *post-hoc* test, except for Metal (Mean=0.921) and Organ Pipes (Mean=0.896). Water plop, Metal, and Organ Pipes achieved over 95% accuracy.

  Interestingly, cardboard was only identified accurately as cardboard 52% of the time, while it was identified as hardwood up to 20% and as a mug at 11%. The guitar was identified as a guitar 55% of the time, a piano up to 43%, and as Organ Pipes for 2%. We conduct a more detailed analysis using information transmission analysis, with results shown in the appendix table Table A.1. In this study, the water plop achieved the highest mutual information, 3.142 bits, which indicates that it is one of the most distinctive stimuli. Followed by Metal with 2.937 bits of mutual information. Cardboard has the lowest mutual information (1.355 bits), suggesting it may be confused with other stimuli based on the response pattern.

  Correlation tests revealed no significant correlations for attack $r(798) = .002, p = .934$ or duration $r(798) = .034, p = .326$ with audio recognition. This suggests that another factor, other than acoustic properties (attack, duration), may be affecting audio recognition. For instance, the family of musical instruments (both guitar and piano are stringed instruments) may influence the recognition.

- **Reverb degree recognition (AccDeter)**

  Figure 4.6 shows how accurately users could determine the reverb degree. Overall, reverb determination accuracy was under 50% except for High reverb, which was achieved slightly over it. The type of auditory feedback presented had a statistically significant difference with weak effect $F((9,741) = 3.985, p < 0.001, \eta^2 = 0.050)$ in accuracy of reverb quantity determination, but reverb did not showed significant difference result $F((3,741) = 2.673, p = 0.046, \eta^2 = 0.010)$. However, there is an interaction between Audio and Reverb with medium effect $F((27,741) = 2.084, p = 0.001, \eta^2 = 0.070)$.

Regarding the accuracy of quantity determination, according to the *post-hoc* test, the water plop sound with r1 (no reverb) has the highest accuracy while metal r1 (no reverb) has the lowest score. Table 4.7 showed how different reverb degrees may impact differently across audio feedback types. For example, water plop, and metal showed no significant difference in audio recognition, but in terms of reverb determination (with more reverb added), metal accuracy tends to increase for both its audio recognition as well as reverb determination, while water plop showed the opposite.

Looking closer at the audio properties detail in Table 4.1, metal has the longest total duration with a longer attack than water plop. The Pearson correlation test revealed that both attack $r(798) = -.081, p = .021$ and duration $r(798) = -.153, p < .001$ plays substantial role in the accuracy or reverb determination.

4.4.3 Findings and Discussion

Real-time Auditory Augmentation and its impact on Interaction perception

Our near-real time auditory augmentation prototype with 8ms latency has shown positive results in terms of perceived simultaneity, unity, and agency. According to our analysis, with an average perceived simultaneity of 0.98 out of 1, a unity score of 0.87 out of 1, and an agency score of 4.31 out of 5, our AAR prototype can convincingly augment tap interactions. Additionally, our analysis suggests that audio feedback and reverb, as single factors, do not significantly affect perceived simultaneity, unity, and agency. Therefore, different audio feedback or reverb can be used to convey information without hampering the perceived real-time interaction. These findings also answer RQ1.

According to our findings, we could create a sound that defies expectations without losing the perceived interaction but with exception of some audio feedback types such as the carpet sound. It is important to note that the Carpet sound had significantly lower scores in terms of simultaneity, unity, and agency. Despite sharing similarities in attack and duration with piano or guitar sounds, Carpet sound's perceived unity was lower. This could be attributed to the multi-sensory effect [20], where the eyes see finger tapping, the fingertips feel the object material, and the user expects percussive sounds but hears continuous dragging sounds instead. Additionally, the semantic link between the expected object contact sound and the resultant audio feedback may also influence this perception. For example, a metallic sound, even though it is continuous, could be perceived as more simultaneous than a carpet sound because the user can more easily imagine tapping a metallic mug than tapping a carpet.

This highlights the challenge of using audio feedback that lacks a semantic association with the underlying interaction. However, the carpet sound characteristics (less unified, less simultaneous and lower perceived agency), make it a promising option for representing an abnormality through a discordant but highly noticeable sound. It could be used to alert the user of change on the meeting schedule or provide more noticeable warnings, such as using carpet sounds when tapping a ceramic mug to inform the user that the tea is still too hot.

In terms of the influence of reverb on perceived interaction, research has indicated that solely high reverb (mix=1, room size=1) has a noteworthy impact on perceived simultaneity. Consequently, when developing interactions that incorporate reverb, it is crucial to maintain the perception of simultaneity. This can be accomplished by utilizing a lower value for the mix and room parameter. This findings, contributed to the RQ1(a-c).

Acoustic Manipulation and its impact on object (size, stiffness) and environment properties (room size)

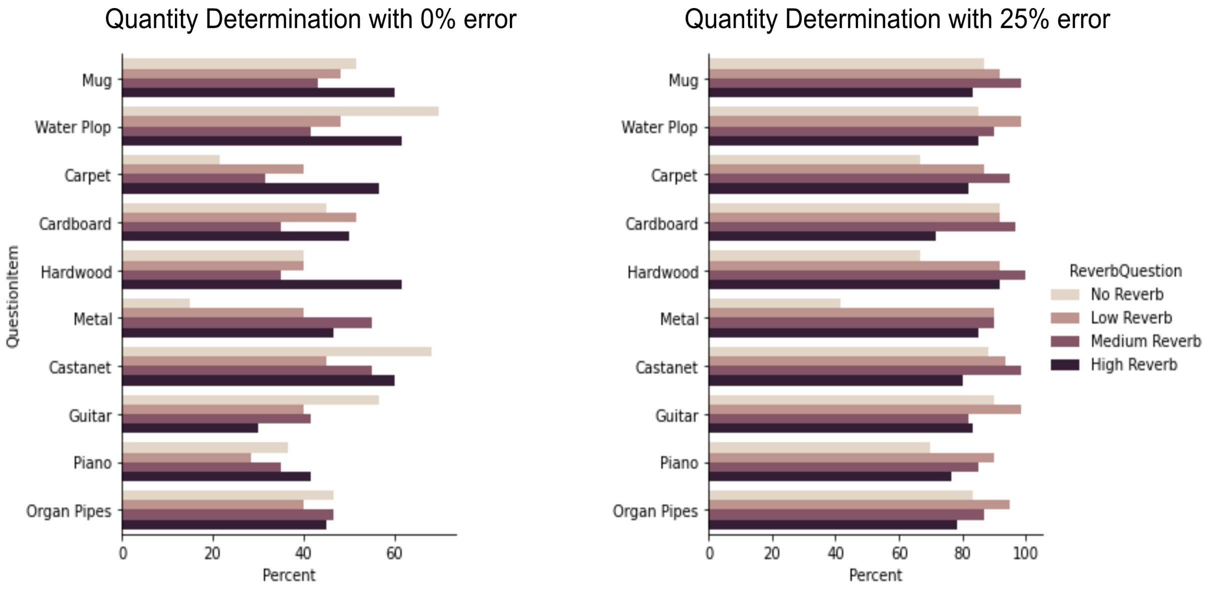
Modifying audio feedback can have an impact on how users perceive the size and surface material of an object and the perceived room size, answering RQ2a-b. Similarly, changing the reverb can affect the perception of the size of the room in which the object is located. However, there is no significant difference or strong effect between different combinations of audio and reverb. As a result, it might be possible to utilize a mix of audio and reverb to modify multiple user perceptions simultaneously. Moreover, there is a strong correlation between attack, duration, object size, and object stiffness. Duration has a positive correlation with object size and stiffness, while attack has a negative correlation with object stiffness.

These findings are significant as they offer new insights into how audio feedback and reverb can impact our perception of room size, which can have important implications on designing auditory interactions, or augmenting object and environment properties for passive haptic feedback. The perceived stiffness of an object can greatly affect the user's experience with a product display, especially in cases where home appliances are being exhibited. People naturally want to test the quality of a product by tapping on it [9], and adjusting the object's stiffness can make the experience more engaging and immersive. Interestingly, by altering the perception of the surface stiffness, we may also be able to change the user's behaviours when interacting with the product.

Recognisability of various acoustic manipulations

As for RQ2c, our study findings reveal that audio feedback causes a significant difference but with a small effect size on reverb recognition, while reverb causes a significant difference with a small effect on audio recognition. Interestingly, our results do not entirely align with Frissen et al.'s study [36], which suggests that audio types do not affect reverb determination. In our experiment, to create reverb, we replicated audio feedback (creating an echo) while adjusting the decay. We argue that reverb may affect audio recognition by diminishing natural audio characteristics, which can make it challenging for users to accurately recognize audio.

The mixing of reverb, known as a mix of wet/dry balance, could potentially diminish the original audio feedback (dry sound), as Sterne[90] points out: *"little reverb added to a dry sound gives it more complexity and depth; a lot of reverb added to a sound washes it out."* This diminished effect of dry-wet balance may have a significant impact on the accuracy of quantity determination and the room size parameter. However, it is still unclear whether mixing and room size (depth) parameters in reverb may impact audio recognition and reverb determination because we did not have a wider variety of reverb mixing compositions in Experiment I.



Quantity Determination with 0% error

Quantity Determination with 25% error

Figure 4.7: Adjusted reverb degree. Left: 0% error rate. Right: 25% error rate

The lack of sufficient samples of the reverb parameter during Experiment 1 may have resulted in users perceiving a similar degree of reverb. We tried to re-evaluated the reverb determination with 25% error tolerance as seen in Figure 4.7. We found that with this adjustment, the user able to recognised the reverb better, with average accuracy 80%. The low accuracy score in the reverb recognition may due to the limitations of human auditory discrimination, which has been extensively researched. Studies have shown that there is a noticeable conflict between human echolocation, which is related to the precedent-effect, and auditory discrimination suppression [72]. This conflict can lead to inaccurate perception of sound and is an important factor to consider in any experiment involving sound perception. Therefore, it is crucial to provide an adequate number of diverse samples to ensure accurate and reliable results.

Moreover, we still have not achieved the control of identifying reverb (where we could enable multiple levels of interaction), thus we have not fully answer RQ2c. Therefore, we need to further explore reverb mixing degrees to enable better auditory contact augmentations. In summary, our study findings provide valuable insights into the effects of audio feedback and reverb on audio recognition and reverb determination, which can help improve auditory contact augmentations in the future.

4.5 Experiment II: Investigating the Effect of Reverb Mixing Parameter on The Perception of Agency, Object and Room Properties, and Recognisability

Experiment I answered some of our RQs. However, some of our findings still not provide a clear and adequate information about acoustic manipulation, especially for the reverb effect. For example, it provides an opposite result to Frissen et al.[36], as based on our result, audio characteristics may influence the reverb determination. Moreover, the determination of reverb was poor in many cases due to the limitations of reverb samples. Additionally, in our previous attempts to use reverb, we have found that it had a negative impact on the recognisability of some types of audio.

To enable designers to use reverb as a property, we need more nuanced ways of controlling it. As we discussed earlier, reverb parameters such as mixing (audio replication or echo) can reduce the accuracy of audio recognition by diminishing the audio characteristics. Therefore, we need to look at the constituent mixing parameters in more detail, taking into account the same measures such as recognisability, agency, and object/room properties. The experiment did not investigate simultaneity and unity as there was no significant impact of either audio feedback or reverb on these measures. However, agency was broadly affected by different types of audio feedback and was therefore investigated. In the Experiment II we investigated RQ1c and RQ2(a-c).

4.5.1 Study Design

Reverb Composition

The experiment involved testing the impact of different levels of reverb (no reverb, low reverb, medium reverb, and high reverb). These levels were created by adjusting the depth (room-size) and mixing the reverb, specifically to answer RQ2c. The details of the parameter composition can be found in the Table 4.8. The experiment measured the impact of audio feedback and reverb level on dependent variables such as agency (RQ1c), and size of mug, size of the room, surface material, accuracy of audio recognition, and reverb determinati on (RQ2a-c).

Auditory Feedback

In this experiment, we used four different audio feedback cues from Table 4.1, namely mug, water plop, metal, and carpet. We reduce the number of auditory feedback types and focused

Table 4.8: Experiment II Reverb Parameter Composition for No Reverb, Low Reverb, Medium Reverb, and High Reverb. Each reverb has 3 different compositions, except for no reverb.

Parameter	No reverb	Low reverb			Medium reverb			High reverb		
		1	2	3	1	2	3	1	2	3
Mixing	0	0.25	0.5	1	0.25	0.5	1	0.25	0.5	1
Depth	0	0.5	0.5	0.5	0.875	0.875	0.875	1	1	1

Mixing and depth parameters varied in continuous scale between 0-1 applied in Supercollider using FreeVerb function.

on the exploring the reverb composition. Similar to Experiment 1, the mug sound was selected as the baseline audio feedback. We then selected other audio feedback based on the findings of Experiment I.

We chose the water plop sound because it displayed the highest accuracy of audio recognition across different reverb levels. This made it the most suitable audio feedback for use in environments with varying levels of reverb. We also selected the metal sound because it displayed an increasing trend in audio recognition as well as reverb determination. This made it suitable for use in environments that require high accuracy in both audio and reverb recognition. Notably, when the reverb is high, the accuracy of recognition tends to drop by over 50%. Therefore, it was important to select audio feedback that could retain high accuracy in such environments. Furthermore, we opted for the carpet sound due to its lower agency when compared to other sounds, as well as its low accuracy in recognizing reverb. Our aim is to investigate whether altering the reverb composition would still result in undesirable outcomes in terms of its agency and recognizability.

In summary, we chose audio feedback ranging from lowest accuracy (Carpet) to highest accuracy (Water Plop), and audio that is capable of maintaining high accuracy for both audio and reverb recognition. Additionally, we selected this audio to represent various chroma features (audio characteristics), which can be seen in the Chromagram at Figure 4.2. This detailed analysis will aid in future experiments that require the use of audio feedback.

Apparatus and Procedure

For this experiment, we followed the same methodology as Experiment I. We used the identical apparatus and procedure to measure the dependent variables, with one exception. Instead of using five-point Likert scales to evaluate object and environment properties, we used a slider with a range of values from 0 to 100 to allow for a continuous range of responses, providing more precise data. By using a continuous range, we can examine details instead of categorization as in Experiment I.

The details of the experimental stimuli and the trial are as follows:

Number of Audio x Number of Reverb x Number of Trial:
4 audio x 9 reverb x 3 trials = 120 trials

Participants

There were 23 new participants, consisting of 12 females and 11 males. As this study uses a within-subjects design, we did not prioritize balancing participant demographics. Additionally, in the previous experiment, we found that participant demographics did not influence the results. The age range of the participants was between 18 and 44 years, with 20% of the individuals being between the ages of 18-24 years old, 55% between 24-35 years old, and 25% between 35-44 years old. Out of the total participants, 65% were students, 8% were homemakers or stay-at-home parents, 20% were full-time workers, and 4% were part-time workers. All the participants were required to have no hearing impairments to ensure that the results obtained from the study were not affected by hearing issues. We also provided a £10 Amazon voucher as compensation for their time and effort in participating in the study.

4.5.2 Result and Analysis

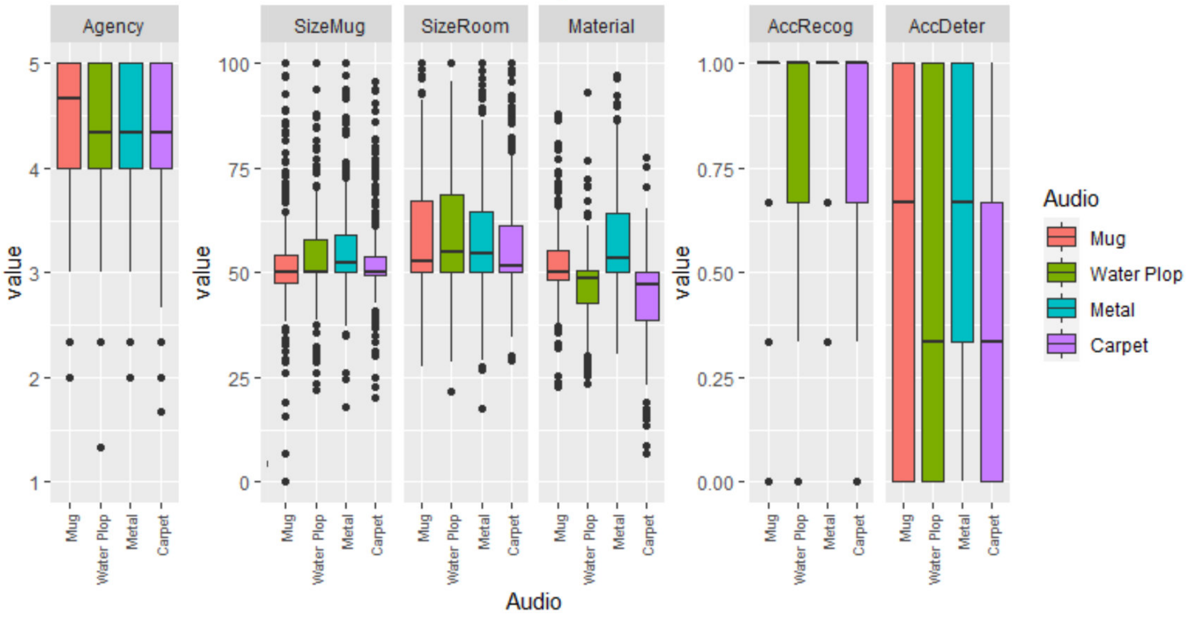


Figure 4.8: Experiment II Audio impact on Agency, SizeMug, Stiffness, SizeRoom, AccRecog and AccDeter

Agency

The analysis of the data collected from the experiment revealed that our prototype was able to retain perceived agency across the reverb conditions, with an average agency score of 4.377 out

Table 4.9: Experiment II Reverb and Audio Effect

Factor	Estimate	Agency	SizeMug	Material	SizeRoom	AccRecog	AccDeter
Audio	F value	9.752	13.475	61.707+++	6.030	0.923	3.659
	Pr(>F)	< 0.001	< 0.001	< 0.001	< 0.001	0.429	0.012
	η^2	0.032	0.044	0.173	0.020	0.003	0.012
Reverb	F value	1.481	42.788+++	1.472	156.408+++	14.854	29.818
	Pr(>F)	0.218	< 0.001	0.221	< 0.001	< 0.001	< 0.001
	η^2	0.005	0.127	0.005	0.347	0.048	0.092
Audio *Reverb	F value	0.875	1.477	0.726	1.684	1.975	3.376
	Pr(>F)	0.547	0.152	0.685	0.089	0.039	< 0.001
	η^2	0.009	0.015	0.007	0.017	0.020	0.033

A repeated measure Two-Way ANOVA with ARTtool was performed to analyse the data. A +++ represent large effect estimated by η^2. There is a significant difference with a major impact on the user's perception of surface material for the auditory cues factor. The medium effect is shown in the size perceptions of mugs between different auditory cues. Reverb has a moderate impact on auditory recognition and reverb quantity estimation but substantially impacts the perception of room size.

Table 4.10: Experiment II Correlation between reverb composition and Accuracies

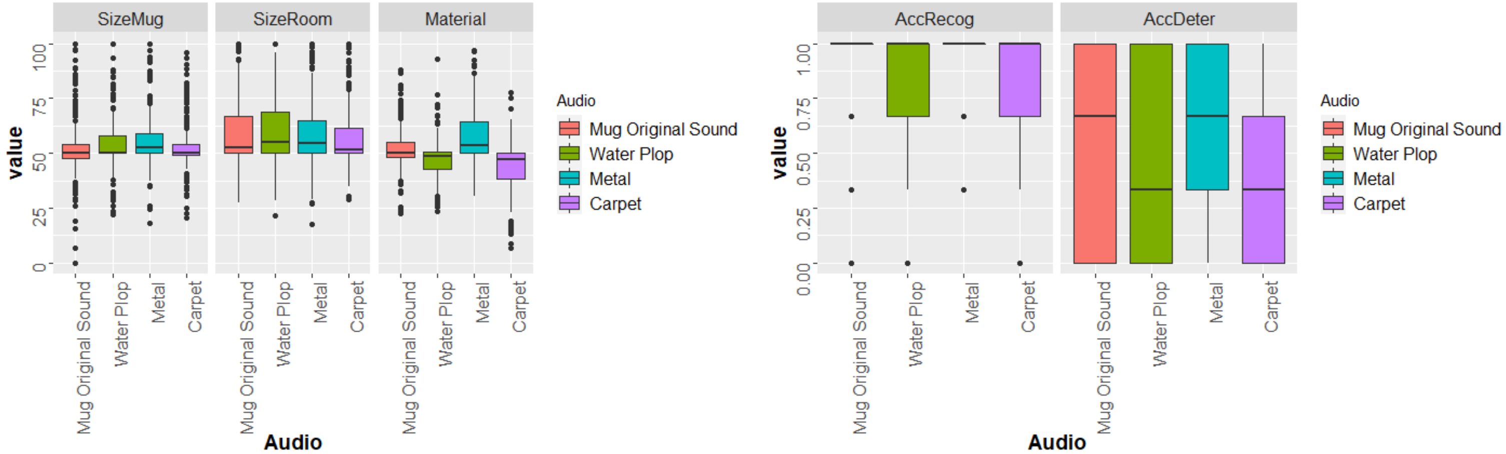
Variable	Factor	AccRecog corr	p_val	AccDeter corr	p_val
Mix	Mix	-0.508	<0.001	0.125	<0.001
Room	Room	-0.149	<0.001	-0.141	<0.001

Table 4.11: Experiment II impact of Mug original sound, metal, carpet, and waterplop sound on Audio and Reverb effect on Agency, SizeMug, Stiffness, SizeRoom, AccRecog and AccDeter

Audio	Agency		SizeMug		SizeRoom		Stiffness		AccRecog		AccDeter	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Carpet	4.216	0.785	53.158	11.786	56.868	13.946	44.012	11.571	0.81	0.334	0.462	0.381
Metal	4.419	0.581	55.53	11.47	58.801	13.355	57.196	13.41	0.983	0.086	0.529	0.368
Mug Original Sound	4.47	0.592	51.926	13.311	58.972	14.737	52.27	10.837	0.89	0.246	0.493	0.392
Water Plop	4.404	0.599	53.572	12.26	59.094	14.683	47.114	9.549	0.767	0.386	0.449	0.394
Average	4.377		53.547		58.434		50.148		0.863		0.483	

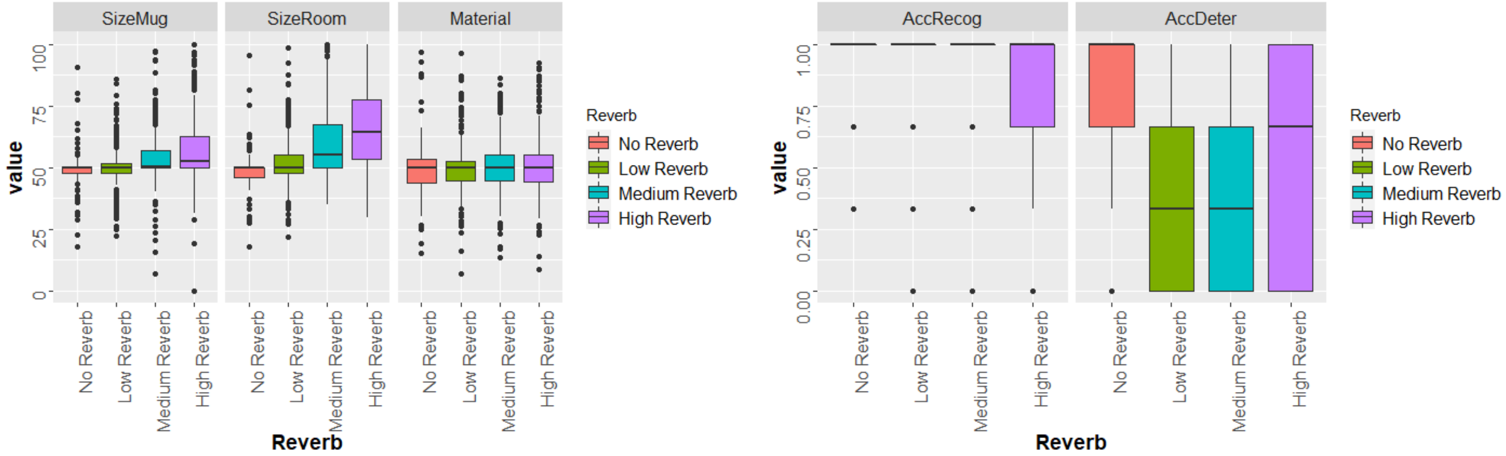
Table 4.12: Experiment II Reverb Impact on Audio and Reverb effect on Agency, SizeMug, Stiffness, SizeRoom, AccRecog and AccDeter

Reverb	Composition		Agency		SizeMug	SizeRoom			Stiffness		AccRecog		AccDeter	
	Room	Mix	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
No Reverb	0	0	4.377	0.722	49.283	9.963	48.732	9.979	50.362	13.873	0.982	0.091	0.754	0.359
Low Reverb	0.5	0.25	4.431	0.634	49.808	8.281	50.786	8.961	50.141	12.04	0.986	0.084	0.344	0.351
Medium Reverb	0.875	0.25	4.413	0.634	53.322	9.306	55.971	9.961	49.851	11.939	0.971	0.117	0.275	0.348
High Reverb	1	0.25	4.38	0.668	54.859	11.247	62.38	12.642	49.678	12.472	0.967	0.111	0.272	0.324
Low Reverb	0.5	0.5	4.457	0.513	50.185	8.98	52.199	9.675	49.645	11.863	0.971	0.107	0.438	0.353
Medium Reverb	0.875	0.5	4.384	0.605	53.71	11.966	59.67	12.196	49.308	12.797	0.96	0.138	0.442	0.357
High Reverb	1	0.5	4.457	0.621	57.989	15.173	66.674	15.002	48.754	13.917	0.949	0.156	0.536	0.363
Low Reverb	0.5	1	4.243	0.706	50.351	9.821	52.801	11.813	48.167	10.281	0.685	0.396	0.533	0.346
Medium Reverb	0.875	1	4.315	0.725	55.67	14.378	64.341	15.435	51.986	11.407	0.583	0.433	0.464	0.376
High Reverb	1	1	4.315	0.653	60.29	16.034	70.786	15.629	53.587	13.328	0.569	0.418	0.775	0.305



(a) Audio Effect on Perception of Object

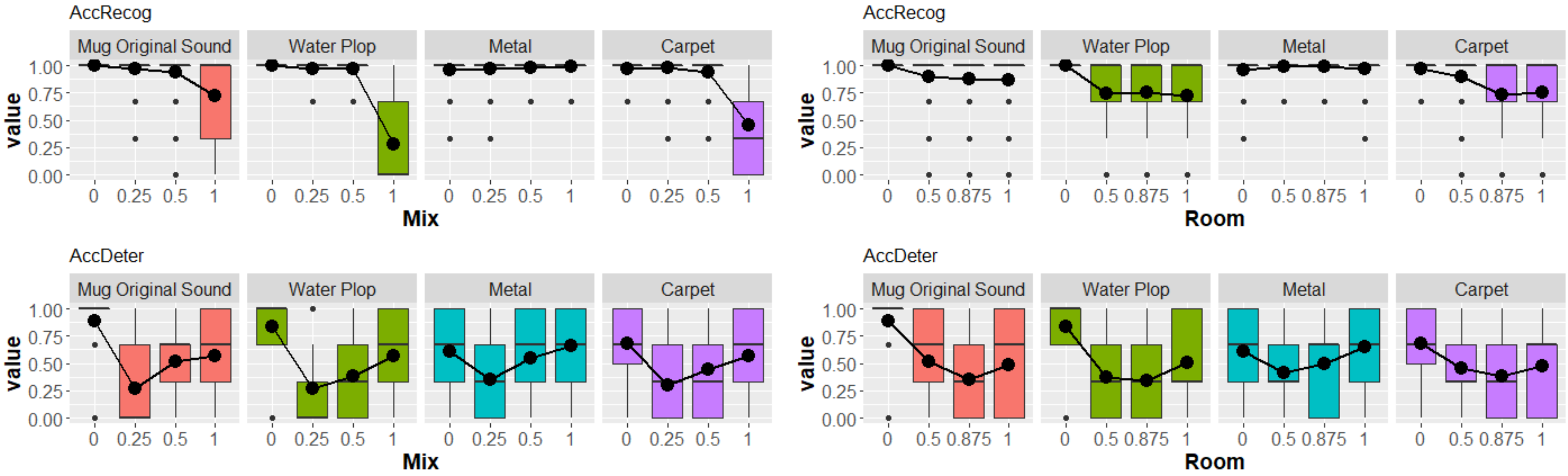
(b) Audio Effect on Accuracy



(c) Reverb Effect on Perception of Object

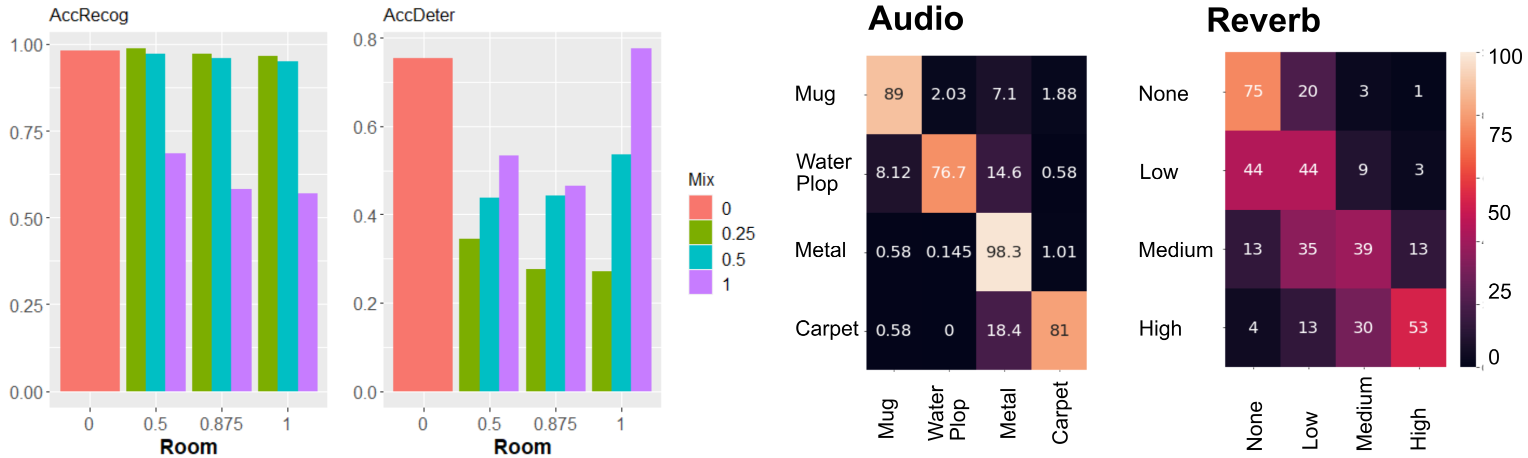
(d) Reverb Effect on Accuracy

Figure 4.9: Exp2: Impact of Audio and Reverb: Audio on Perception **(a)**, Audio on Accuracy **(b)**, Reverb on Perception **(c)**, Reverb on Accuracy **(d)**,



(a) Mixing effect on AccRecog and AccDeter

(b) Room effect on AccRecog and AccDeter



(c) Experiment II Room and Mixing Trend

(d) Experiment II Confusion Matrix

Figure 4.10: Experiment II trends: Mixing effect on Accuracy **(a)**, Reverb effect on Accuracy **(b)**, Mixing and Room Trend on Accuracy **(c)**, Confussion Matrix **(d)**

of 5.0 (Table 4.11). This indicates that the prototype was effective in maintaining a consistent level of agency, regardless of the composition.

The data from Table 4.9 showed that there was a significant difference in the effect of audio on participants' perceived agency, although the effect was small $F((3,882) = 9.752, p < 0.001, \eta^2 = 0.032)$. Upon conducting a *post-hoc* test, we discovered that the carpet had a significantly lower perceived agency than the rest of the audio types. This finding provides more convincing information about the carpet sound as an exceptional sound that affects perceived agency.

Impact of Acoustic Manipulations

Based on Figure 4.8 and Figure 4.9a, it is evident that each auditory feedback, regardless of its reverb composition, has varying effects on the object and room properties. In terms of the accuracy of audio recognition, mug and metal show a more consistent result close to 1 (Table 4.11). Figure 4.8 also displays more divergent results from the participants when determining the reverb degree for various auditory feedback.

1) **Object (size, stiffness) properties**

   As depicted in Figure 4.9a, the audio feedback has diverse impacts on the perception of object and environmental properties. At the same time, the degree of reverb also appears to have an impact on these properties. Figure 4.9b demonstrates an increasing trend in the size of objects and rooms.

   It was found that there is a notable distinction with a minor effect on the perceived size of a mug between audio feedback types $F((3,882) = 13.475, p < 0.001, \eta^2 = 0.044)$, and a significant difference with a strong effect of reverb $F((3,882) = 42.788, p < 0.001, \eta^2 = 0.127)$. This contrasts with Experiment I, in which audio significantly and largely affects the perception of a mug's size, while reverb has a weaker effect. This difference may be due to the sample size of the experiments. Experiment I focused on audio feedback (10 stimuli), while Experiment II focused on reverb (10 stimuli). Among the different audio feedback types, metal resulted in the highest perception of the object's size, and the *post-hoc* result showed that it was significantly larger than metal and water plop. However, metal was not significantly smaller than carpet. With regards to reverb, high reverb caused the largest perception of object size.

   As for stiffness, there is significant difference with a strong effect between audio feedback types types $F((3,882) = 61.707, p < 0.001, \eta^2 = 0.173)$. The *post-hoc* test revealed that metal was perceived as the hardest, while carpet was perceived as having a smoother material compared to the rest of the audio feedback. This result aligns with Experiment I.

2) **Environment (room size) properties**

   The results of Experiment II were similar to those of Experiment I. Both audio feedback

$F((3,882) = 6.030, p < 0.001, \eta^2 = 0.020)$ and reverb $F((3,882) = 156.409, p < 0.001, \eta^2 = 0.347)$ had a significant effect on the perception of the room size. However, only the reverb effect was strong whilst audio have medium effect size. According to the *post-hoc* test, Carpet had the lowest score for size room perception. As expected, high reverb had the greatest score on size room perception, which indicates that an increase in reverb enhances the perception of the surrounding space. These findings strengthen the results from Experiment I.

3) **Recognisability**

a) Audio Recognition

Overall, it can be observed from Figure 4.8c that the materials, mug and metal, exhibit a superior ability to preserve the precision of audio recognition, while water plop and carpet exhibit a slightly lower performance. Figure 4.8d reveals that no reverb, low reverb, and medium reverb have minimal impact on audio recognition, as the score remains close to 1. However, when the reverb is high, there is a notable decline in the audio recognition score.

In contrast to the results of Experiment I, the audio effect did not display any significant variance in audio recognition $F((3,882) = 0.923, p = 0.429, \eta^2 = 0.003)$. Nonetheless, both the reverb $F((3,882) = 14.854, p < 0.001, \eta^2 = 0.048)$ and the interplay between audio and reverb $F((3,882) = 1.975, p = 0.039, \eta^2 = 0.020)$ exhibited substantial discrepancies, which had a small impact on the precision of audio recognition. The consistent and significant differences in audio recognition accuracy, observed due to the interplay between audio and reverb, despite different stimuli in Experiments I and II, prompted further investigation.

In Figure 4.10a, the relationship between audio and mixing is demonstrated in terms of its impact on the accuracy of audio recognition and reverb determination. As shown in the Figure 4.10c, an increase in mixing generally results in a decrease in participants' ability to recognize the audio, with the exception of metal, where the accuracy increased with increased mixing. For all other sounds, the accuracy of audio recognition decreased significantly when the mixing exceeded 0.5. Similarly, Figure 4.10b shows that room composition also has a diminishing effect on the accuracy of audio recognition.

We conducted a two-way ANOVA to examine how mixing and its interaction with audio impact the accuracy of audio recognition. Our analysis revealed a significant difference in accuracy based on the degree of mixing, with a large effect size $F((9,741) = 244.213, p < 0.001, \eta^2 = 0.45)$. Specifically, *post-hoc* tests showed that audio recognition accuracy was significantly lower for mixing (1) compared to the other mixing degrees. Furthermore, we found a significant interaction effect between audio and mixing, with a large effect size on audio recognition $F((9,741) = 77.923, p < 0.001, \eta^2 = 0.45)$. *Post-hoc* tests showed that the carpet with mixing (1) had the lowest audio recognition accuracy compared to other audio-mixing combinations. On the other hand, metal with mixing (1) achieved higher audio recognition accuracy than any other audio.

b) Reverb Recognition/ Determination

Similar to Experiment I, the results of Experiment II indicate that audio $F((3,882) = 3.659, p = 0.012, \eta^2 = 0.012)$, reverb $F((3,882) = 29.818, p < 0.001, \eta^2 = 0.92)$ and their interaction $F((3,882) = 3.376, p < 0.001, \eta^2 = 0.033)$ have a significant impact on reverb determination. However, it was discovered that only the reverb level has a strong influence on the accuracy of reverb determination. The findings suggest that higher reverb levels provide significantly more accurate reverb determination than lower reverb levels.

Figure 4.10a illustrates that the accuracy of reverb determination is inversely proportional to mixing, meaning that the higher the mixing, the more accurate the reverb determination becomes. A similar trend was noticed for the room effect, as demonstrated in Figure 4.10b.

Based on the measurement of reverb accuracy, both mixing and room composition depict a similar trend across various audio feedback, as shown in Figure 4.10a and Figure 4.10b. The results indicate that there is a sudden drop in accuracy from 0 mixing/room towards bigger compositions, but the accuracy gradually improves as the composition size increases.

4.5.3 Findings and Discussion

The perception of Interaction (Agency), Object Properties (size, stiffness), and Room Size

The results of Experiment II suggest that the user's sense of control is not significantly affected by audio and reverb. This means that different types of audio feedback and reverb can be used to convey different types of information without compromising the user's perceived agency. This answered our **RQ1**. These findings reinforce the idea of using audio and reverb as complementary variables to enhance the delivery of information. For example, audio feedback can be utilized to indicate the type of incoming notification, while reverb can be used to convey its quantity or urgency.

Experiment II also found that changes in the audio feedback and reverb can influence the perception of object size. Specifically, using different audio feedback or reverb can make an object appear either bigger or smaller. For example, using metal sounds can create an impression of a larger object, while softer sounds like water plop can make it seem smaller. Furthermore, combining audio and reverb can be an effective way to manipulate the user's perception of object size without having to replace the object's original contact sound.

Furthermore, the study found that audio can also modulate the perception of an object's surface stiffness. For example, using different types of audio feedback can make the user perceive the object as being made of a different material, such as metal or wood. Our findings answered **RQ2a-b.**

Auditory Feedback Recognisability

In our study, we discovered that mixing below 0.5 does not affect the accuracy of audio recognition, while a mixing degree of 1 results in a considerable drop in accuracy. Our research answered RQ2c by showing that audio does not strongly impact the accuracy of audio recognition or reverb determination. However, the accuracy of audio recognition is affected by the degree of reverb mixing, while the accuracy of reverb determination increases linearly with the mixing degree.

However, we did not explore the possibilities of using mixing between 0.5 and 1, so future research could investigate increasing the mixing threshold while retaining the accuracy of audio recognition. Furthermore, we found that the accuracy of reverb determination also increases as the mixing degree increases, and there is no interaction between audio and mixing with regards to reverb determination.

To convey continuous information, such as the amount of liquid inside a mug, high mixing (1) would be fruitful for creating a multi-level reverb. However, the accuracy of reverb determination is still not optimal. Our results in Figure 4.10d showed that the accuracy of no reverb is 75%, while the accuracy of low, medium, and high reverb is only 44%, 39%, and 53%, respectively. Therefore, we conducted further investigations to test whether the reverb itself, regardless of the degree, is identifiable. We grouped the reverb and recalculated the accuracy of reverb determination, which showed that the user can identify the reverb occurrence with an accuracy of 80%. This result implies that it might be more beneficial to convey binary information rather than continuous information to retain its recognisability. Our findings answered **RQ2c.**

Reverb Composition (Mixing and Room) Effect

Reverb plays a crucial role in shaping how objects and rooms are perceived, as well as impacting audio recognizability and reverb level determination. When mixing, it's important to note that a level above 0.5 may decrease accuracy for audio recognition, but increase accuracy for reverb determination. Figure 4.10c illustrates that as the room size increases, accuracy for audio recognition decreases for larger mixes. Conversely, more accurate reverb determination can be achieved with an increase in room size and mixing. However, it is important to keep in mind that increasing the mixing could compromise users' ability to recognize audio feedback.

We suggest that this could be a case where the reverb itself diminishes the original audio properties, such as changing the pitch and shifting the attack). Thus, in order to use reverb efficiently and safely, the designer should considered the mixing and room composition. We recommend using a mixing of 0.5 or under.

Chapter 5

Discussion and Auditory Augmentations Guidelines

In this thesis, we explored the application of auditory augmentation techniques, such as adding reverberation and replacing auditory feedback, to the interaction sound produced by tapping on a physical object. To create these augmentations, prototypes using an AR HMD (Hololens 2) and a custom microcontroller with a contact microphone as input and headphones as output were built. This thesis evaluated the prototypes and found that the microcontroller with contact microphone approach performed better than the HMD device in terms of keeping the latency below 10ms. Using this prototype, two experiments aimed at assessing the efficacy of various auditory augmentation techniques has been conducted.

The first experiment involved replacing auditory feedback with a range of sounds, including material and non-material sounds (such as musical instruments), and incorporating reverbs. This thesis objective was to determine the impact of these techniques on the user's perception of interaction (including simultaneity, unity, and agency), object properties (such as size and material stiffness), environmental properties (such as room size), and recognition of auditory feedback.

This thesis discovered that some of our initial experiment results did not align with existing theories. Specifically, we found that audio feedback impacted the accuracy of reverb determination, which was against Frissen et al.'s findings [36]. To validate the results obtained from the first experiment and to further investigate the impact of reverb, a second experiment was conducted. The main objective in this experiment was to analyze different settings and parameters related to reverb composition, mixing, and room acoustics. The variables measured in the second experiment were identical to those in the first experiment. However, the second experiment did not consider simultaneity and unity , since there was no significant differences when replacing the interaction sound with various audio manipulations.

Based on the two experiments, we have identified several important findings that are discussed in the following section. These findings address the following RQs:

- **RQ1**: How can we augment real-world interaction sound emitted from an everyday object whilst preserving:

  (a) unity
  (b) simultaneity
  (c) agency

- **RQ2**: How does acoustic manipulation (replacing the interaction sounds, adding reverb) impact:

  (a) the perception of object properties such as size and stiffness
  (b) the perception of environmental properties such as room size
  (c) the recognisability of augmented auditory feedback

5.1 RQ1(a-c): The Perception of Interaction

This thesis discovered that, for the most part, the use of audio feedback and reverb did not significantly impact the user's perception of the interaction, answering RQ1. However, there were some exceptions to this finding. It was noted that the combination of audio and reverb could be useful to convey multi-level information without having an undue effect on the user experience. For instance, audio feedback could represent the application, while reverb could represent the quantity of notifications. However, in the experiments, the carpet sound was perceived as less impactful due to its discordant auditory feedback when tapped. This could be because the user expects a percussive or impact sound but hear a more continuous sound. Nonetheless, using carpet as audio feedback could be beneficial in conveying information about abnormalities, such as informing users about allergens in a drink.

Using the latency guidelines proposed by Kaaresoja[51], auditory feedback (materials and musical instruments sound) did not substantially impact perceived agency, simultaneity, and unity. Thus, more material sounds (i.e., rubber, sand) or non-material sounds, such as animal sounds, and environmental sounds (e.g., waves, wind) could be tested in future research. These results provide a fundamental understanding of how much can the underlying audio properties be changed or replaced without losing the illusion that the audio event was connected to the physical touch.

5.2 RQ2(a-b): Manipulating User Perception of Object and Environment Properties

The findings provide strong support for the claims made by [56, 46, 94] that auditory augmentation can significantly alter the user's perception of object properties, such as surface stiffness.

The findings also exposed possibilities to augment other object sizes and their surroundings using various audio feedback. Similarly, reverb can alter the perception of both object and room size. This finding provides valuable insight into how auditory augmentations can impact an object's physical properties and the environment in which it is located, answering RQ2a-b. This revelation also holds great significance for the development of haptic feedback systems and mixed reality experiences, without heavily relying on visual augmentation. For example, it can be utilized to create the illusion of a more spacious room or to alter the perception of real world object properties within a confined area, according to the presented digital object.

Furthermore, reverb can be used independently or in conjunction with auditory feedback. This can be particularly useful in reducing the computational cost of presenting various material sounds for passive haptic feedback. By adjusting the reverb composition (mixing, room) directly to impact sounds, object properties can be altered without having to simulate different materials. For example, to make the object perceived smoother, more mixing could be applied.

Considering the characteristics of reverb, in which can alter the room size perception, this thesis recommend using reverb as additional information parameter. For example, whilst the waterplop sound could indicate notification from social media app, high reverb can indicate a large number of unread messages during casual interactions by increasing the perception of room size. This can be beneficial in providing users with a quick and easy way to gauge the number of unread messages they have.

5.3 RQ2c: Recognisability of Acoustic Manipulation

The accuracy of audio recognition was not strongly affected by audio feedback, despite the significant difference between audio feedback, answering our RQ2c. In Experiment I, a minimum accuracy of 52% (Cardboard) and a maximum accuracy of 98% (Water plop) was achieved. Cardboard sound had the least accuracy due to its similar properties to hardwood, such as both being percussive sounds and providing a wooden effect. In fact, the results in Figure 4.6 showed that 20% of Cardboard is perceived as Hardwood. The sharing similarity notion is also strengthened by the fact that 43% of Guitar is perceived as Piano, as both of them also share similarities in acoustic properties, such as long decay after short impact.

Experiment II achieved a minimum accuracy of 76% (Water plop) and a maximum accuracy of 98% (Metal) of audio recognition (see Figure 4.10d). Interestingly, some audio feedback is perceived as Metal in both Experiment I and Experiment II. For example, 12% of Carpet was perceived as Metal. In reverb, individual may hear some "metallic" effect as a result of hearing a few decaying tones that correspond to the peaks of the pitch [35]. As shown in Figure 4.2, the Carpet sounds have a richer peak of pitch (showing more amounts of brighter color) compared to the water plop and mug. This thesis argued that many pitches would resonate due to the echo mixing and create a metallic effect. Modulation on the pitch may reduce this metallic effect

[84]. However, rather than reducing this metallic effect, some benefits could be gained from this effect, such as altering many audio types to have metallic sound thus perceived as bigger in size and harder in material perception - all without literally replacing the sounds with a metal sound.

Regarding reverb determination, both Experiment I and Experiment II have not yet achieved a satisfactory result (accuracy under 75% across reverb levels). A further investigation (see discussion in Experiment II) revealed that participants could tell if there is a reverb; thus, reverb can be useful for information clustering such as proposed by Frissen et al.[36]. Furthermore, to enable multilevel reverb, a high mixing could be applied. This thesis found that high mixing (1) could increase the accuracy of reverb determination. Yet, this may result in less accurate audio recognition due to changes in acoustic properties such as pitch and decay. Nevertheless, by using a mixing of 0.5 or under, the accuracy of audio recognition could be retained, supporting information in addressing RQ2c. In the future, more reverb mixing degrees could be explored to increase this threshold.

This outcome helps us comprehend the level of distinctiveness of auditory enhancements and their ability to maintain precise auditory feedback recognition. It also investigates the possibility of manipulating additional auditory parameters, such as reverb, to broaden the potential feedback range that a designer can utilize.

5.4 Proposed Auditory Augmentation Guidelines

Based on our findings, this thesis present sets of guidelines. The first set of guidelines provides recommendations on how to augment the interaction sound in general. The second set of guidelines offers specific parameter suggestions for further auditory manipulation.

5.4.1 Augmentation of Contact Audio using Reverb

a) Rich pitch variability sounds, like carpet sounds, can hinder precise audio identification.

b) When different levels of feedback are needed on the contact event, metallic sounds work well because they maintain accuracy and reverb recognition.

c) Sound effects that have metallic characteristics can also alter how an object appears to the user, making it appear larger or harder on the surface. This demonstrates that manipulating reverb properties affects user perception more deeply than merely producing a distinctly different sound. Sometimes it enhances the sense of the physical characteristics of the object, which experts could also modify.

d) Regardless of the type of audio, reverb can be used in tandem with it to change how big an object appears to the user (because there is no interaction between them).

5.4.2 Parameter Recommendations

a) When combining various auditory feedback with different levels of reverb, the reverb mixing composition should be set at 0.5 or lower for a single interaction. Our findings indicate that maintaining the reverb composition at 0.5 or lower, regardless of the auditory feedback used, retain the accuracy of audio recognition.

b) Alternatively, to maintain the precision of audio recognition while being used in conjunction with audio feedback, reverb can be employed to transmit binary data instead of continuous data. For instance, a user might shake a bottle to check for social media notifications. Hearing a chirping bird sound with reverb could indicate a new Twitter message, while a chirping bird without reverb could indicate no new messages.

c) When using a single auditory feedback with multiple reverb compositions, it is recommended to adjust the mixing parameter closer to 1 in order to effectively discern and differentiate between various levels of reverb. This action can significantly enhance the precision of reverb detection. For example, if you tap a coffee mug and hear a water sound with long reverb, it may indicate rainy weather with a long duration.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

Overall, the thesis provides insights into how interaction sound emitted from tapping a physical object can be augmented by replacing the sound with different types of auditory feedback, and/or adding reverb. This thesis also demonstrated how audio and reverb can be used to modulate user perception of object size, surface material, and room size, contributing prototype demonstration and knowledge as follow.

This thesis presents an auditory augmentation prototype with 8ms latency, demonstrating the feasibility of augmenting real-world interaction sounds from tapping on a coffee cup can preserve the feeling unity, simultaneity, agency with the interaction. Our research found that audio feedback and reverb do not have a significant impact on users' perceptions of their interaction with a system, except when using carpet as audio feedback. Carpet produces discordant auditory feedback, which can negatively affect users' perception of simultaneity, unity, and agency. Furthermore, auditory feedback can change how a user perceives object properties (size, stiffness), while reverb can alter the perception of object's size and room size. Moreover, auditory feedback did not strongly impact the accuracy of audio recognition, but reverb did.

This thesis also outlines a set of guidelines that draw upon the findings of Experiments I and II. These guidelines aim to enable multi-level auditory feedback while ensuring accuracy in audio recognition and reverb determination, as well as achieving a perception of interaction that is characterized by simultaneity, agency, and unity.

6.2 Limitations and Future Study

6.2.1 Limitations

In our experiment, we used non-noise-cancellation headphones to study the perception of simultaneity, unity, and agency. However, we acknowledge that future studies could benefit from incorporating more advanced equipment, such as acoustically transparent headphones or AR devices. These devices allow for the mixing of real and virtual audio, which may affect our findings and the perception of the resulting intermixed audio. For instance, AR devices could provide a more immersive and realistic audio experience that better simulates real-life interactions. Additionally, 3D audio spatialization could enhance the illusion of sound originating from a given contact event, improving the overall accuracy of our findings. As such, we suggest that future studies incorporate these factors as we move closer to realizing the full potential of augmenting contact properties using AR headsets in real-time. Our setup with non-acoustically transparent headphones could demonstrate its interactivity perception (simultaneously, unity, agency). A future study might use acoustically transparent headphones or AR devices. Such devices would mix real and virtual audio, which could affect our findings and the perception of the resultant intermixed audio. In addition, 3D audio spatialization would likewise enhance the illusion of sound originating from a given contact event, thus, future studies could also incorporate this factor as we move closer to realizing the potential of augmenting contact properties using AR headsets in real-time.

6.2.2 Beyond the Coffee Mug – Augmenting Other Objects

In this thesis, a coffee cup was used as a sample object, but our experiments findings suggest that the user's prior knowledge of the object's surface or material does not significantly affect their perception of interaction. This means that our proposed augmentation could be extended to other objects as well. With the ongoing development of sensors, microcontroller, and AR HMD devices, we can even incorporate sound augmentation in the interaction with various objects. Different objects can represent different task groups. For instance, a pen could provide information about work-related app notifications, while a coffee mug could represent personal goals such as sugar or caffeine intake.

6.2.3 Supporting Casual Interaction by Using Multiple Objects

Casual interaction with everyday objects can be a useful alternative to access digital information without solely relying on a mobile phone. By embedding auditory augmentation into the objects around us, we can access information without excessive phone usage. The excessive use of mobile phones has been linked to addiction [75] and reduced productivity [27]. This condition is often triggered by uncontrollable notifications that lead to mind wandering and procrastination [91]. Therefore, using objects around us to access information can reduce distractions and provide more controlled access to notifications.

6.2.4 Supporting Passive Haptic Feedback by Augmenting Various Object's Properties

Various techniques can be used to manipulate how users perceive the properties of an object. These include employing physical proxies that use actuators to change an object's internal weight [116], using wearable haptic interfaces [21] and kinaesthetic haptic devices [53], and relying on visual effects [99]. Although visual augmentation has been extensively researched by augmenting interaction sounds, the use of auditory augmentation to impart the sense of haptic such as temperature, weight, or texture has received less attention. Therefore, future studies could focus on exploring these areas to broaden the horizon of using auditory augmentation to alter user perception of object properties such as object weights, or texture roughness. Kaneko et al.[53] investigated that delay (latency) can influence the perceived weight of the object. Still, our study has yet to explore different latency settings that may impact users' perceived interaction or object weight.

6.2.5 Exploring More Audio Manipulation

In this thesis, we have only investigated two types of audio feedback - material and musical instrument sound. However, future studies should explore a wider range of auditory feedback, such as animal sounds, and add more effects such as flangers. It is also important to adjust the acoustic parameters to convey different information. For example, granular synthesis can be used on sonification to convey information about time-varying probabilistic information. However, the use of other acoustic parameters such as echo and delay to augment audio in auditory AR interaction has not been widely investigated. For instance, manipulating the delay on the audio feedback to represent the duration left. Finally, the reverb parameter itself could be explored further. For example, creating a custom reverb mixing based on given auditory feedback rather than using a single parameter for all types of audio could retain the accuracy of audio recognition for specific cases. Further investigation of AAR parameters is crucial to provide designers and practitioners with a toolbox regarding how we augment our perception of interactions with everyday objects.

Appendix A

Study Materials

A.1 Information Sheet

INFORMATION SHEET

You are being invited to take part in an experiment which will take approximately 50 minutes to complete.

Your participation is voluntary, and you are free to withdraw at any time, without giving any reason, and you are free to omit answering any particular question, without providing a reason. Your participation will help us in conducting our research and is greatly appreciated.

Before you decide whether to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to decide whether or not you wish to take part. If there is anything that is unclear, or you would like more information, or you have any questions, please feel free to raise these concerns with the researcher present, or any member of the study team.

Purpose of experiment

This experiment will be looking for the preferable audio feedback while tapping a mug. The audio feedback will be used to convey information like notification. This experiment will also investigate how the user's perception of simultaneity of the sounds and the tapping, realism, and agency of the audio AR interaction will vary based on different sounds.

Do I have to take part?

Your participation is voluntary, and you are free to withdraw at any time, without giving any reason, and you are free to omit answering any question, without providing a reason.

What is the compensation for taking part in this study?

You will receive a £10 Amazon gift voucher once the experiment is completed. If you are a student, no course credits will be awarded for completing this experiment.

What will I be asked to do?

You will be asked to do a finger tapping interaction to a coffee mug and listen to the audio feedback from the headphones. You will also be asked to fill the questionnaire after every stimulus.

Will my taking part in this study be kept confidential?

Yes, all data collected from you will be treated confidentially, will be seen in its raw form only by the experimenters, and if published will not be identifiable as coming from you.

What will happen to the results of the research study?

The results of the study may appear in research publications. The results may also be presented at scientific meetings or in talks at academic institutions. Results will always be presented in a confidential format where anonymity is preserved.

Contact for further information or queries?

You may ask more questions about the study at any time - before, during and after - the study. You can contact the researchers: i.bustoni.1@research.gla.ac.uk, or stephen.brewster@glasgow.ac.uk or mark.mcgill@glasgow.ac.uk

A.2 Consent Form

University of Glasgow | School of Computing Science
GIST Glasgow Interactive Systems Group

CONSENT FORM

In this experiment, you will be asked to reflect on how the latency on audio feedback affects the user perception on auditory altered reality interaction and how is your attitude toward auditory altered reality interaction. You will be asked to do some interactions with an audible object and reflect your opinion about it.

The experiment should take approximately 90 minutes to complete but there is no set time limit, so take as much time as you need. After each condition, you will be asked to answer a few short questions and at the end of the experiment, the researcher will conduct a short interview to capture any additional thoughts you might have on your experience.

You could also take rests after each condition, and you can pause the experiment at any time. You can withdraw from the study at any time without penalty. You will receive a £10 Amazon gift voucher once the experiment is completed. If you are a student, no course credits will be awarded for completing this experiment.

Lastly, you must meet the following requirements:

- You are at least 18 years old
- You have no hearing disorder and you do not wear any hearing aids
- You agree for the audio to be recorded during the interview part of the study
- You confirm that you have read and understood the participant information for the above study and have had the opportunity to ask questions.
- You understand that your participation is voluntary and that you are free to withdraw at any time without giving any reason and am free to omit answering any particular question, without providing a reason.
- You understand that all data collected from you will be treated confidentially, will be seen in its raw form only by the experimenters, and if published, will not be identifiable if coming from you.

By signing this form, you have read the conditions stated above and agree to take part in the experiment.

FULL NAME: ________________________________

SIGNATURE: ________________________________

DATE: ________________________________

EMAIL: ________________________________

DEMOGRAPHICS	
What is your gender?	
☐ Female	☐ Non-Binary
☐ Male	☐ Prefer not to say
How old are you?	
☐ 18-25	☐ 26-35
☐ 36-50	☐ >50
What is your occupation?	

Project Supervisor: Stephen Brewster (stephen.brewster@glasgow.ac.uk)
Researcher: Isna Alfi Bustoni(i.bustoni.1@research.gla.ac.uk)

Table A.1: Audio Stimulus Information Transmission Experiment I

Stimulus	Conditional Entropy	Mutual Information
Mug	0.948	2.355
Waterplop	0.161	3.142
Carpet	1.394	1.909
Cardboard	1.948	1.355
Hardwood	1.211	2.092
Metal	0.366	2.937
Castanet	1.003	2.3
Guitar	1.111	2.192
Piano	1.146	2.157
Organ Pipe	0.566	2.737

Bibliography

[1] Diar Abdlkarim and Peter Holland. A Methodological Framework to Assess the Accuracy of Virtual Reality Hand-Tracking Systems: A case study with the Oculus Quest 2. doi: 10.1101/2022.02.18.481001. URL `https://diarkarim.com/`.

[2] Ali Abdolrahmani, Ravi Kuber, and Stacy M. Branham. "Siri Talks at You": An Empirical Investigation of Voice-Activated Personal Assistant (VAPA) Usage by Individuals Who Are Blind. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 249–258, Galway Ireland, October 2018. ACM. ISBN 978-1-4503-5650-3. doi: 10.1145/3234695.3236344.

[3] M. Ercan Altinsoy. Knocking Sound as Quality Sign for Household Appliances and the Evaluation of the Audio-Haptic Interaction. In *Proceedings of the 7th International Conference on Haptic and Audio Interaction Design*, HAID'12, pages 121–130, Berlin, Heidelberg, 2012. Springer-Verlag. ISBN 978-3-642-32795-7. doi: 10.1007/978-3-642-32796-4_13. URL `https://doi.org/10.1007/978-3-642-32796-4_13`. event-place: Lund, Sweden.

[4] Jatin Arora, Varnit Jain, Aryan Saini, Shwetank Shrey, Nirmita Mehra, and Aman Parnami. VirtualBricks: Exploring a scalable, modular toolkit for enabling physical manipulation in VR. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery, May 2019. ISBN 978-1-4503-5970-2. doi: 10.1145/3290605.3300286.

[5] Ronald T Azuma. A survey of augmented reality. Technical report, 1997. URL `http://www.cs.unc.edu/~azumaW`. Publication Title: Presence: Teleoperators and Virtual Environments Volume: 6.

[6] Olivier Bau and Ivan Poupyrev. REVEL: Tactile feedback technology for augmented reality. *ACM Transactions on Graphics*, 31(4), July 2012. ISSN 07300301. doi: 10.1145/2185520.2185585.

[7] Juan Pablo Bello, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark B. Sandler. A tutorial on onset detection in music signals. *IEEE Transactions on*

Speech and Audio Processing, 13(5):1035–1046, September 2005. ISSN 10636676. doi: 10.1109/TSA.2005.851998.

[8] Costas Boletsis and Dimitra Chasanidou. Audio augmented reality in public transport for exploring tourist sites. *ACM International Conference Proceeding Series*, pages 721–725, 2018. doi: 10.1145/3240167.3240243. ISBN: 9781450364379.

[9] Leonardo Bonanni, Chia Hsun Lee, and Ted Selker. Cooking with the elements: Intuitive immersive interfaces for augmented reality environments. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 3585 LNCS, pages 1022–1025, 2005. ISBN 3-540-28943-7. doi: 10.1007/11555261_95. ISSN: 03029743.

[10] Till Bovermann, René Tünnermann, and Thomas Hermann. Auditory augmentation. *International Journal of Ambient Computing and Intelligence*, 2(2):27–41, April 2010. ISSN 19416237. doi: 10.4018/jaci.2010040102.

[11] Jonas Braasch, Hari V Savitala, and Jens Blauert. MODELING THE PRECEDENCE EFFECT FOR PERCUSSIVE SOUNDS WITH DIFFERENT ATTACK TRANSIENTS. In *19th INTERNATIONAL CONGRESS ON ACOUSTICS*, 2007.

[12] Niclas Braun, Stefan Debener, Nadine Spychala, Edith Bongartz, Peter Sörös, Helge H.O. Müller, and Alexandra Philipsen. The senses of agency and ownership: A review. *Frontiers in Psychology*, 9(APR), April 2018. ISSN 16641078. doi: 10.3389/fpsyg.2018.00535. Publisher: Frontiers Media S.A.

[13] Stephen A Brewster, Peter C Wright, and Alistair DN Edwards. Experimentally derived guidelines for the creation of earcons. In *Adjunct proceedings of HCI*, volume 95, pages 155–159, 1995.

[14] Stephen Anthony Brewster. *Providing a Structured Method for Integrating Non\-Speech Audio into Human-Computer Interfaces*. PhD Thesis, University of York, 1994. Issue: August.

[15] Andrew D. Brown, Heath G. Jones, Alan Kan, Tanvi Thakkar, G. Christopher Stecker, Matthew J. Goupell, and Ruth Y. Litovsky. Evidence for a neural source of the precedence effect in sound localization. *Journal of Neurophysiology*, 114(5):2991–3001, September 2015. ISSN 15221598. doi: 10.1152/jn.00243.2015. Publisher: American Physiological Society.

[16] N Bryan-Kinns. Interaction Design with Audio: Speculating on Sound in Future Design Education. In *The 4th Central China International Design Science Seminar*, 2017. URL `https://www.researchgate.net/publication/323258898`.

[17] Jun Cai, Jiahang Liu, Nishuai Yu, and Binyang Liu. Effect of water sound masking on perception of the industrial noise. *Applied Acoustics*, 150:307–312, 2019. ISSN 1872910X. doi: 10.1016/j.apacoust.2019.02.025. URL `https://doi.org/10.1016/j.apacoust.2019.02.025`. Publisher: Elsevier Ltd.

[18] Ryan Canales and Sophie Jörg. Performance Is Not Everything: Audio Feedback Preferred over Visual Feedback for Grasping Task in Virtual Reality. In *Proceedings - MIG 2020: 13th ACM SIGGRAPH Conference on Motion, Interaction, and Games*. Association for Computing Machinery, Inc, October 2020. ISBN 978-1-4503-8171-0. doi: 10.1145/3424636.3426897.

[19] Feiyu Chen, Honghao Lv, Zhibo Pang, Junhui Zhang, Yonghong Hou, Ying Gu, Huayong Yang, and Geng Yang. WristCam: A Wearable Sensor for Hand Trajectory Gesture Recognition and Intelligent Human–Robot Interaction. *IEEE Sensors Journal*, 19 (19):8441–8451, October 2019. ISSN 1530-437X, 1558-1748, 2379-9153. doi: 10.1109/JSEN.2018.2877978. URL `https://ieeexplore.ieee.org/document/8509628/`.

[20] Yi Chuan Chen and Charles Spence. Assessing the role of the 'unity assumption' on multisensory integration: A review. *Frontiers in Psychology*, 8(MAR), March 2017. ISSN 16641078. doi: 10.3389/fpsyg.2017.00445. Publisher: Frontiers Research Foundation.

[21] Inrak Choi, Heather Culbertson, Mark R. Miller, Alex Olwal, and Sean Follmer. Grabity: A Wearable Haptic Interface for Simulating Weight and Grasping in Virtual Reality. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, pages 119–130, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 978-1-4503-4981-9. doi: 10.1145/3126594.3126599. URL `https://doi.org/10.1145/3126594.3126599`. event-place: Québec City, QC, Canada.

[22] Laurence Cliffe, James Mansell, Joanne Cormac, Chris Greenhalgh, and Adrian Hazzard. The audible artefact: Promoting cultural exploration and engagement with audio augmented reality. In *ACM International Conference Proceeding Series*, pages 176–182, 2019. ISBN 978-1-4503-7297-8. doi: 10.1145/3356590.3356617.

[23] Benjamin R. Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. "What Can I Help You with?"\: Infrequent Users\' Experiences of Intelligent Personal Assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '17, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 978-1-4503-5075-4. doi: 10.1145/3098279.3098539. URL `https://doi.org/10.1145/3098279.3098539`. event-place: Vienna, Austria.

[24] Stuart Cunningham, Harrison Ridley, Jonathan Weinel, and Richard Picking. Audio emotion recognition using machine learning to support sound design. *ACM International Conference Proceeding Series*, pages 116–123, 2019. doi: 10.1145/3356590.3356609. ISBN: 9781450372978.

[25] Jon Dattorro. Effects Design. *Journal Audio Engineering Society*, 45(9), 1997.

[26] B. Diezma-Iglesias, M. Ruiz-Altisent, and P. Barreiro. Detection of Internal Quality in Seedless Watermelon by Acoustic Impulse Response. *Biosystems Engineering*, 88(2): 221–230, 2004. ISSN 1537-5110. doi: https://doi.org/10.1016/j.biosystemseng.2004.03.007. URL `https://www.sciencedirect.com/science/article/pii/S1537511004000467`.

[27] Éilish Duke and Christian Montag. Smartphone addiction, daily interruptions and self-reported productivity. *Addictive Behaviors Reports*, 6:90–95, December 2017. ISSN 23528532. doi: 10.1016/j.abrep.2017.07.002. Publisher: Elsevier Ltd.

[28] Dan Ellis. Chroma Feature Analysis and Synthesis, 2007. URL `https://www.ee.columbia.edu/~dpwe/resources/matlab/chroma-ansyn/`.

[29] Georg Essl and Sile O'modhrain. PebbleBox and CrumbleBag: Tactile Interfaces for Granular Synthesis. PebbleBox and CrumbleBag: Tactile Interfaces for Granular Synthesis. Technical report, 2004. URL `https://www.researchgate.net/publication/221164809`.

[30] Hugo Fastl. Psychoacoustics and Sound Quality. Technical report, 2005.

[31] Jamie Ferguson and Stephen A. Brewster. Evaluation of psychoacoustic sound parameters for sonification. In *ICMI 2017 - Proceedings of the 19th ACM International Conference on Multimodal Interaction*, volume 2017-January, pages 120–127. Association for Computing Machinery, November 2017. ISBN 978-1-4503-5543-8. doi: 10.1145/3136755.3136783.

[32] Sam Ferguson. SONIFYING EVERY DAY: ACTIVATING EVERYDAY INTERACTIONS FOR AMBIENT SONIFICATION SYSTEMS. In *International Conference on Auditory Display*, 2013.

[33] Kenneth P. Fishkin. A taxonomy for and analysis of tangible interfaces. *Personal and Ubiquitous Computing*, 8(5):347–358, 2004. ISSN 16174909. doi: 10.1007/s00779-004-0297-4. Publisher: Springer-Verlag London Ltd.

[34] Daniel J. Freed. Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *The Journal of the Acoustical Society of America*, 87(1):311–322,

1990. doi: 10.1121/1.399298. URL `https://doi.org/10.1121/1.399298`. _eprint: https://doi.org/10.1121/1.399298.

[35] Jasmin Frenette. *REDUCING ARTIFICIAL REVERBERATION ALGORITHM REQUIREMENTS USING TIME-VARIANT FEEDBACK DELAY NETWORKS*. PhD thesis, University of Miami, Florida, 2000.

[36] Ilja Frissen, Brian F.G. Katz, and Catherine Guastavino. Effect of sound source stimuli on the perception of reverberation in large volumes. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 5954 LNCS, pages 358–376, 2010. ISBN 3-642-12438-0. doi: 10.1007/978-3-642-12439-6_18. ISSN: 03029743.

[37] Mark B. Gardner. Historical Background of the Haas and/or Precedence Effect. *The Journal of the Acoustical Society of America*, 43(6):1243–1248, June 1968. ISSN 0001-4966. doi: 10.1121/1.1910974. Publisher: Acoustical Society of America (ASA).

[38] Ed Ginzel and Ben Turnbull. Determining Approximate Acoustic Properties of Materials. *e-Journal of Nondestructive Testing*, 21(12), 2016. URL `http://www.ondacorp.com/tecref_acoustictable.shtml`.

[39] Randy Goebel, Wolfgang Wahlster, and Joerg Siekmann. Multidisciplinary Aspect of Time and Time Perception. Technical report, Greece, 2010.

[40] Kristian Gohlke, Michael Hlatky, and Bram de Jong. Physical Construction Toys for Rapid Sketching of Tangible User Interfaces. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '15, pages 643–648, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 978-1-4503-3305-4. doi: 10.1145/2677199.2687900. URL `https://doi.org/10.1145/2677199.2687900`. event-place: Stanford, California, USA.

[41] Yizheng Gu, Chun Yu, Zhipeng Li, Weiqi Li, Shuchang Xu, Xiaoying Wei, and Yuanchun Shi. Accurate and Low-Latency Sensing of Touch Contact on Any Surface with Finger-Worn IMU Sensor. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, pages 1059–1070, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 978-1-4503-6816-2. doi: 10.1145/3332165.3347947. URL `https://doi.org/10.1145/3332165.3347947`. event-place: New Orleans, LA, USA.

[42] Gabriel Haas, Evgeny Stemasov, and Enrico Rukzio. Can't you hear me? Investigating personal soundscape curation. *ACM International Conference Proceeding Series*, pages 59–69, 2018. doi: 10.1145/3282894.3282897. ISBN: 9781450365949.

[43] Florian Heller. Muffidgets: Detecting and Identifying Edible Pastry Tangibles on Capacitive Touchscreens. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 978-1-4503-8213-7. doi: 10.1145/3430524.3442449. URL `https://doi.org/10.1145/3430524.3442449`. event-place: Salzburg, Austria.

[44] Thomas Hermann and Marian Weger. DATA-DRIVEN AUDITORY CONTRAST ENHANCEMENT FOR EVERYDAY SOUNDS AND SONIFICATIONS Thomas Hermann Marian Weger Ambient Intelligence Group Institute for Electronic Music and Acoustics (IEM) University of Music and Performing Arts. *Th 25th International Conference on Auditory Display*, (June):23–27, 2019. doi: 10.4119/unibi/2935744.2.

[45] Long Huang and Chen Wang. Notification privacy protection via unobtrusive gripping hand verification using media sounds. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, pages 491–504, New York, NY, USA, October 2021. ACM. ISBN 978-1-4503-8342-4. doi: 10.1145/3447993.3483277. URL `https://dl.acm.org/doi/10.1145/3447993.3483277`.

[46] Brent Edward Insko. Passive Haptics Significantly Enhances Virtual Environments. Technical report, 2001.

[47] Hiroshi Ishii and Brygg Ullmer. Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, pages 234–241, New York, NY, USA, 1997. Association for Computing Machinery. ISBN 0-89791-802-9. doi: 10.1145/258549.258715. URL `https://doi.org/10.1145/258549.258715`. event-place: Atlanta, Georgia, USA.

[48] Robert J.K. Jacob, Audrey Girouard, Leanne M. Hirshfield, Michael S. Horn, Orit Shaer, Erin Treacy Solovey, and Jamie Zigelbaum. Reality-Based Interaction: A Framework for Post-WIMP Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 201–210, New York, NY, USA, 2008. Association for Computing Machinery. ISBN 978-1-60558-011-1. doi: 10.1145/1357054.1357089. URL `https://doi.org/10.1145/1357054.1357089`. event-place: Florence, Italy.

[49] Jin Yong Jeon, Pyoung Jik Lee, Jin You, and Jian Kang. Perceptual assessment of quality of urban soundscapes with combined noise sources and water sounds. *The Journal of the Acoustical Society of America*, 127(3):1357–1366, 2010. ISSN 0001-4966. doi: 10.1121/1.3298437.

[50] Ralf Jung. AMBIENCE FOR AUDITORY DISPLAYS: EMBEDDED MUSICAL INSTRUMENTS AS PERIPHERAL AUDIO CUES. In *The 14th International Conference on Auditory Display*, Paris, 2008.

[51] Topi Johannes Kaaresoja. *Latency Guidelines for Touchscreen Virtual Button Feedback*. PhD thesis, 2015.

[52] Mitali Kamat, Alvaro Uribe Quevedo, and Peter Coppin. Tangible Construction Kit for Blind and Partially Sighted Drawers. pages 1–6. Association for Computing Machinery (ACM), February 2022. doi: 10.1145/3490149.3505580.

[53] Seitaro Kaneko, Takumi Yokosaka, Hiroyuki Kajimoto, and Takahiro Kawabe. A Pseudo-Haptic Method Using Auditory Feedback: The Role of Delay, Frequency, and Loudness of Auditory Feedback in Response to a User's Button Click in Causing a Sensation of Heaviness. *IEEE Access*, 10:50008–50022, 2022. ISSN 21693536. doi: 10.1109/ACCESS.2022.3172324. Publisher: Institute of Electrical and Electronics Engineers Inc.

[54] Mohamed Kari, Tobias Grosse-puppendahl, Alexander Jagaciak, David Bethge, Porsche Ag, Reinhard Schütte, and Christian Holz. *SoundsRide : Affordance-Synchronized Music Mixing for In-Car Audio Augmented Reality*, volume 1. Association for Computing Machinery, 2021. Publication Title: The 34th Annual ACM Symposium on User Interface Software and Technology (UIST '21), October 10â•ﬁ14, 2021, Virtual Event, USA Issue: 1.

[55] Gerald Kidd, Christine R. Mason, and Tanya L. Arbogast. Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns. *The Journal of the Acoustical Society of America*, 111(3):1367–1376, 2002. doi: 10.1121/1.1448342. URL `https://doi.org/10.1121/1.1448342`. _eprint: https://doi.org/10.1121/1.1448342.

[56] Roberta L. Klatzky, Dinesh K. Pai, and Eric P. Krotkov. Perception of material from contact sounds. *Presence: Teleoperators and Virtual Environments*, 9(4):399–410, 2000. ISSN 10547460. doi: 10.1162/105474600566907. Publisher: MIT Press Journals.

[57] Lua Koenig and Tony Ro. Sound frequency predicts the bodily location of auditory-induced tactile sensations in synesthetic and ordinary perception. *bioRxiv*, 2022. doi: 10.1101/2022.06.06.495023. URL `https://www.biorxiv.org/content/early/2022/06/07/2022.06.06.495023`.

[58] Michael Krzyzaniak, David Frohlich, and Philip J.B. Jackson. Six Types of Audio That DEFY Reality! A Taxonomy of Audio Augmented Reality with Examples. In *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound*, AM'19,

pages 160–167, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 978-1-4503-7297-8. doi: 10.1145/3356590.3356615. URL `https://doi.org/10.1145/3356590.3356615`. event-place: Nottingham, United Kingdom.

[59] Mark Lawton, Stuart Cunningham, and Ian Convery. Nature Soundscapes: An Audio Augmented Reality Experience. In *Proceedings of the 15th International Audio Mostly Conference*, AM '20, pages 85–92, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 978-1-4503-7563-4. doi: 10.1145/3411109.3411142. URL `https://doi.org/10.1145/3411109.3411142`. event-place: Graz, Austria.

[60] Susan J. Lederman. Auditory Texture Perception. *Perception*, 8(1):93–103, 1979. doi: 10.1068/p080093. URL `https://doi.org/10.1068/p080093`. _eprint: https://doi.org/10.1068/p080093.

[61] Dong Li, Jialin Liu, Sunghoon Ivan Lee, and Jie Xiong. Room-Scale Hand Gesture Recognition Using Smart Speakers. In *Proceedings of the Twentieth ACM Conference on Embedded Networked Sensor Systems*, pages 462–475, Boston Massachusetts, November 2022. ACM. ISBN 978-1-4503-9886-2. doi: 10.1145/3560905.3568528.

[62] Mengqi Liao and S. Shyam Sundar. Sound of silence: Does Muting Notifications Reduce Phone Use? *Computers in Human Behavior*, 134, September 2022. ISSN 07475632. doi: 10.1016/j.chb.2022.107338. Publisher: Elsevier Ltd.

[63] Robert A. Lutfi and Eunmi L. Oh. Auditory discrimination of material changes in a struck-clamped bar. *The Journal of the Acoustical Society of America*, 102(6):3647–3656, December 1997. ISSN 0001-4966. doi: 10.1121/1.420151. Publisher: Acoustical Society of America (ASA).

[64] Charles Patrick Martin, Zeruo Liu, Yichen Wang, Wennan He, and Henry Gardner. Sonic Sculpture: Activating Engagement with Head-Mounted Augmented Reality. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, December 2020. URL `http://arxiv.org/abs/2012.02311`. arXiv: 2012.02311.

[65] David R. McGee and Philip R. Cohen. Creating Tangible Interfaces by Augmenting Physical Objects with Multimodal Language. In *Proceedings of the 6th International Conference on Intelligent User Interfaces*, IUI '01, pages 113–119, New York, NY, USA, 2001. Association for Computing Machinery. ISBN 1-58113-325-1. doi: 10.1145/359784.360305. URL `https://doi.org/10.1145/359784.360305`. event-place: Santa Fe, New Mexico, USA.

[66] Mark McGill, Stephen Brewster, David McGookin, and Graham Wilson. Acoustic Transparency and the Changing Soundscape of Auditory Mixed Reality. *Conference on Hu-*

man Factors in Computing Systems - Proceedings, 2020. doi: 10.1145/3313831.3376702. ISBN: 9781450367080.

[67] Graeme McLean and Alan Wilson. Shopping in the digital world: Examining customer engagement through augmented reality mobile applications. *Computers in Human Behavior*, 101:210–224, December 2019. ISSN 07475632. doi: 10.1016/j.chb.2019.07.002. Publisher: Elsevier Ltd.

[68] Andrew Mcpherson, Robert Jack, and Giulio Moro. Action-Sound Latency: Are Our Tools Fast Enough? July 2016.

[69] Aarthi Easwara Moorthy and Kim\-Phuong L. Vu. Privacy Concerns for Use of Voice Activated Personal Assistant in the Public Space. *International Journal of Human-Computer Interaction*, 31(4):307–335, 2015. doi: 10.1080/10447318.2014.986642. URL `https://doi.org/10.1080/10447318.2014.986642`. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10447318.2014.986642.

[70] Giulio Moro, Astrid Bin, Robert H Jack, Christian Heinrichs, and Andrew P McPherson. Making High-Performance Embedded Instruments with Bela and Pure Data. June 2016.

[71] Thomas Muender, Anke V. Reinschluessel, Sean Drewes, Dirk Wenig, Tanja Döring, and Rainer Malaka. Does it feel real? Using tangibles with different fidelities to build and explore scenes in virtual reality. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery, May 2019. ISBN 978-1-4503-5970-2. doi: 10.1145/3290605.3300903.

[72] Mats E. Nilsson, Carlos Tirado, and Malina Szychowska. Psychoacoustic evidence for stronger discrimination suppression of spatial information conveyed by lag-click interaural time than interaural level differences. *The Journal of the Acoustical Society of America*, 145(1):512–524, January 2019. ISSN 0001-4966. doi: 10.1121/1.5087707. Publisher: Acoustical Society of America (ASA).

[73] Collins Opoku-Baah, Adriana M. Schoenhaut, Sarah G. Vassall, David A. Tovar, Ramnarayan Ramachandran, and Mark T. Wallace. Visual Influences on Auditory Behavioral, Neural, and Perceptual Processes: A Review. *Journal of the Association for Research in Otolaryngology*, 22(4):365–386, August 2021. ISSN 1525-3961, 1438-7573. doi: 10.1007/s10162-021-00789-0. URL `https://link.springer.com/10.1007/s10162-021-00789-0`.

[74] Elif Oźcan, René van Egmond, Alexandre Gentner, and Carole Favart. Incorporating Brand Identity in the Design of Auditory Displays: The Case of Toyota Motor Europe. In *Foundations in Sound Design for Embedded Media*, pages 155–193. Routledge, 2019.

[75] Subramani Parasuraman, AaseerThamby Sam, StephanieWong Kah Yee, BobbyLau Chik Chuon, and LeeYu Ren. Smartphone usage and increased risk of mobile phone addiction: A concurrent study. *International Journal of Pharmaceutical Investigation*, 7(3):125, 2017. ISSN 2230-973X. doi: 10.4103/jphi.jphi_56_17. Publisher: EManuscript Services.

[76] he peng and joshua d. reiss. why can you hear a difference between pouring hot and cold water? an investigation of temperature dependence in psychoacoustics. *journal of the audio engineering society*, October 2018.

[77] Henning Pohl and Michael Rohs. Around-Device Devices: My Coffee Mug is a Volume Dial. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices; Services*, MobileHCI '14, pages 81–90, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 978-1-4503-3004-6. doi: 10.1145/2628363.2628401. URL `https://doi.org/10.1145/2628363.2628401`. event-place: Toronto, ON, Canada.

[78] Dominic Potts, Martynas Dabravalskis, and Steven Houben. TangibleTouch: A Toolkit for Designing Surface-based Gestures for Tangible Interfaces. pages 1–14. Association for Computing Machinery (ACM), February 2022. ISBN 978-1-4503-9147-4. doi: 10.1145/3490149.3502263.

[79] Don Ritter. USING AUDITORY TONES IN A HUMAN INTERFACE: Earcons. 1986.

[80] Davide Rocchesso, Federico Fontana, Edizioni Di, and Mondo Estremo. *The Sounding Object*. 2003.

[81] Enrico Rukzio, Florian Michahelles, Kaisa Väänänen, Mark Billinghurst, DOCOMO Euro Labs, Net mobile AG, SIGCHI (Group : U.S.), Association for Computing Machinery, and ACM Digital Library. *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia : 2012, Ulm, Germany : MUM 2012*. ISBN 978-1-4503-1815-0.

[82] Marie-Laure Ryan. Immersion vs. interactivity: Virtual reality and literary theory. *SubStance*, 28(2):110–137, 1999. ISSN 00492426, 15272095. URL `http://www.jstor.org/stable/3685793`.

[83] Niklas Rönnberg. Musical elements in sonification support visual perception. In *ECCE 2019 - Proceedings of the 31st European Conference on Cognitive Ergonomics: "Design for Cognition"*, pages 114–117. Association for Computing Machinery, Inc, September 2019. ISBN 978-1-4503-7166-7. doi: 10.1145/3335082.3335097.

[84] Urban Schlemmer. Reverb Design. 2011.

[85] Hanna Kathrin Schraffenberger. *Arguably Augmented Reality Relationships Between the Virtual and the Real*. PhD thesis, 2018.

[86] Benjamin G. Schultz and Floris T. Van Vugt. Tap Arduino: An Arduino microcontroller for low-latency auditory feedback in sensorimotor synchronization experiments. *Behavior Research Methods*, 48(4):1591–1607, December 2016. ISSN 1554-3528. doi: 10.3758/s13428-015-0671-3. URL `http://link.springer.com/10.3758/s13428-015-0671-3`.

[87] Nabeel Siddiqui and Rosa H. M. Chan. Multimodal hand gesture recognition using single IMU and acoustic measurements at wrist. *PLOS ONE*, 15(1):e0227039, January 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0227039. URL `https://dx.plos.org/10.1371/journal.pone.0227039`.

[88] Marjan Sikora, Mladen Russo, Jurica Derek, and Ante Jurčević. Soundscape of an archaeological site recreated with audio augmented reality. *ACM Transactions on Multimedia Computing, Communications and Applications*, 14(3), 2018. ISSN 15516865. doi: 10.1145/3230652.

[89] Tasha R. Stanton and Charles Spence. The Influence of Auditory Cues on Bodily and Movement Perception. *Frontiers in Psychology*, 10, January 2020. ISSN 16641078. doi: 10.3389/fpsyg.2019.03001. Publisher: Frontiers Media S.A.

[90] Jonathan Sterne. Space within Space: Artificial Reverb and the Detachable Echo. Technical report, 2015. Publication Title: Grey Room Volume: 60.

[91] Cary Stothart, Ainsley Mitchum, and Courtney Yehnert. The attentional cost of receiving a cell phone notification. *Journal of Experimental Psychology: Human Perception and Performance*, 41(4):893–897, August 2015. ISSN 19391277. doi: 10.1037/xhp0000100. Publisher: American Psychological Association Inc.

[92] Dan Stowell and Mark Plumbley. Adaptive whitening for improved real-time audio onset detection. August 2007.

[93] Patrick Susini, Nicolas Misdariis, Guillaume Lemaitre, and Olivier Houix. Naturalness influences the perceived usability and pleasantness of an interface's sonic feedback. *Journal on Multimodal User Interfaces*, 5(3-4):175–186, May 2012. ISSN 17837677. doi: 10.1007/s12193-011-0086-0.

[94] Ana Tajadura-Jimenez and Nadia Bianchi-Berthouze. Sonification of Surface Tapping Changes. pages 48–57, 2015.

[95] Ana Tajadura Jimenez, Manos Tsakiris, Torsten Marquardt, and Nadia Bianchi-Berthouze. Action sounds update the mental representation of arm dimension: Contributions of kinaesthesia and agency. *Frontiers in Psychology*, 6, May 2015. ISSN 16641078. doi: 10.3389/fpsyg.2015.00689. Publisher: Frontiers Research Foundation.

[96] Ana Tajadura-Jiménez, Maria Basia, Ophelia Deroy, Merle Fairhurst, Nicolai Marquardt, and Nadia Bianchi-Berthouze. As Light as Your Footsteps: Altering Walking Sounds to Change Perceived Body Weight, Emotional State and Gait. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2943–2952, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 978-1-4503-3145-6. doi: 10.1145/2702123.2702374. URL `https://doi.org/10.1145/2702123.2702374`. event-place: Seoul, Republic of Korea.

[97] Ana Tajadura-Jiménez, Joseph Newbold, Linge Zhang, Patricia Rick, and Nadia Bianchi-Berthouze. As Light as You Aspire to Be: Changing Body Perception with Sound to Support Physical Activity. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–14, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 978-1-4503-5970-2. doi: 10.1145/3290605.3300888. URL `https://doi.org/10.1145/3290605.3300888`. event-place: Glasgow, Scotland Uk.

[98] Israt Tasnim, Priscella Asman, Chandra Prakash Swamy, Sudhakar Tummala, Sujit Prabhu, and Nuri Firat Ince. Microcontroller-Based Low Latency Audio System to Study Cortical Auditory Evoked Potentials: Applications with Intraoperative Language Mapping. In *2023 11th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 01–04, Baltimore, MD, USA, April 2023. IEEE. ISBN 978-1-66546-292-1. doi: 10.1109/NER52421.2023.10123904. URL `https://ieeexplore.ieee.org/document/10123904/`.

[99] Bruce H. Thomas and Paul Calder. Applying Cartoon Animation Techniques to Graphical User Interfaces. *ACM Trans. Comput.-Hum. Interact.*, 8(3):198–222, September 2001. ISSN 1073-0516. doi: 10.1145/502907.502909. URL `https://doi.org/10.1145/502907.502909`. Place: New York, NY, USA Publisher: Association for Computing Machinery.

[100] Garreth W. Tigwell and Michael Crabb. Household Surface Interactions: Understanding User Input Preferences and Perceived Home Experiences. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery, April 2020. ISBN 978-1-4503-6708-0. doi: 10.1145/3313831.3376856.

[101] Asia Turato. *Emotions in everyday sounds. An experimental study on the influence of*

colour in emotional knocking sounds. PhD Thesis, Dipartimento di Studi Linguistici e Letterari - DISLL, 2021.

[102] Rene Tünnermann, Jan Hammerschmidt, and Thomas Hermann. BLENDED SONIFICATION-SONIFICATION FOR CASUAL INFORMATION INTERACTION. In *International Conference on Auditory Display*, Poland, 2013. URL `http://www.cinismoilustrado.com/2012/07/mirar-el-celular.html`.

[103] Brygg Ullmer and Hiroshi Ishii. Human-Computer Interaction in the New Millenium. Technical report, 2001.

[104] Myla van Wegen, Just L. Herder, Rolf Adelsberger, Manuela Pastore-Wapp, Erwin E. H. van Wegen, Stephan Bohlhalter, Tobias Nef, Paul Krack, and Tim Vanbellingen. An overview of wearable haptic technologies and their performance in virtual object exploration. *Sensors*, 23(3), 2023. ISSN 1424-8220. doi: 10.3390/s23031563. URL `https://www.mdpi.com/1424-8220/23/3/1563`.

[105] Yolanda Vazquez-Alvarez and Stephen Brewster. Designing Spatial Audio Interfaces to Support Multiple Audio Streams. pages 253–256. Nolo, 2010. ISBN 1-4133-0843-0.

[106] Marian Weger, Thomas Hermann, and Robert Höldrich. AltAR/Table: A Platform for Plausible Auditory Augmentation. In *Proceedings of the 27th International Conference on Auditory Display (ICAD 2022)*, pages 16–24, icad.org, June 2022. International Community for Auditory Display. ISBN 0-9670904-8-2. doi: 10.21785/icad2022.005. URL `http://hdl.handle.net/1853/67378`.

[107] Daniel Wigdor and Dennis Wixon. *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1st edition, 2011. ISBN 0-12-382231-9.

[108] John Williamson and Roderick Murray-Smith. Granular Synthesis for Display of Time-Varying Probability Densities. In *International Workshop on Interactive Sonification*, Bielefeld, 2004.

[109] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 143–146, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1978963. URL `https://doi.org/10.1145/1978942.1978963`. event-place: Vancouver, BC, Canada.

[110] Michael Wolf. *A Tangible Interface that Enables Children to Record, Modify and Arrange Sound Samples in a Playful Way*. PhD thesis, Goldsmith, University of London, 2002. Issue: Juni.

[111] Yunhan Wu, Daniel Rough, Anna Bleakley, Justin Edwards, Orla Cooney, Philip R. Doyle, Leigh Clark, and Benjamin R. Cowan. See What I'm Saying? Comparing Intelligent Personal Assistant Use for Native and Non-Native Language Speakers. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '20, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 978-1-4503-7516-0. doi: 10.1145/3379503.3403563. URL `https://doi.org/10.1145/3379503.3403563`. event-place: Oldenburg, Germany.

[112] Papers J Yang, A Barde, M Billinghurst, Jing Yang, Amit Barde, and Mark Billinghurst. Audio Augmented Reality: A Systematic Review of Technologies, Applications, and Future Research Directions. *J. Audio Eng. Soc*, 70(10):788–809, 2022. doi: 10.17743/jaes.2022.0048. URL `https://doi.org/10.17743/jaes.2022.0048`.

[113] Qiuyu Yang and Hideyoshi Yanagisawa. Effects of Space Discrepancy and Latency on the Sense of Agency with Discrete and Continuous Operations. *International Journal of Affective Engineering*, 21(1):13–22, 2022. doi: 10.5057/ijae.ijae-d-21-00002. Publisher: Japan Society of Kansei Engineering.

[114] Bruno Zamborlin. *Studies on customisation-driven digital music instruments*. PhD thesis, 2014. Issue: October.

[115] Md Zarif Kaisar, Hameem Ahsan, Md Mushfiqus Saleheen, Md Amzad Hossain Jacky, Khandaker Tabin Hasan, Fahad Ahmed, and Md Zahiduddin Ahmed. Ball game controller: A tangible user interface. In *ACM International Conference Proceeding Series*. Association for Computing Machinery, January 2020. ISBN 978-1-4503-7778-2. doi: 10.1145/3377049.3377101.

[116] André Zenner and Antonio Krüger. Shifty: A Weight-Shifting Dynamic Passive Haptic Proxy to Enhance Object Perception in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*, 23(4):1285–1294, 2017. doi: 10.1109/TVCG.2017.2656978.

[117] Zhoutong Zhang, Qiujia Li, Zhengjia Huang, Jiajun Wu, Joshua B. Tenenbaum, and William T. Freeman. Shape and Material from Sound. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pages 1278–1288, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 978-1-5108-6096-4. event-place: Long Beach, California, USA.

[118] Yulia Zhiglova. The interactive carpet-Smart textile interface for children on autism spectrum disorder. In *TEI 2018 - Proceedings of the 12th International Conference on Tangible, Embedded, and Embodied Interaction*, volume 2018-January, pages 712–714. Association for Computing Machinery, Inc, March 2018. ISBN 978-1-4503-5568-1. doi: 10.1145/3173225.3173341.

[119] Udo Zolzer. 7. In *Digital Audio Signal Processing*. John Wiley, Chichester, 1st ed edition, 1997.