AALBORG UNIVERSITY
STUDENT REPORT

Study Board of Computer Science
Selma Lagerløfsvej 300 DK - 9220 Aalborg East
Phone +45 99 40 99 40
computerscience@sict.aau.dk
www.sict.aau.dk/computer-science

Title: Sound Zone Interaction: A Study of Designing and Evaluating Interaction Techniques
Semester: 10th Semester
Semester theme: Master's thesis.
Project period: 1. February 2019 – 18. June 2019.
ECTS: 30
Supervisor: Jesper Kjeldskov
Project group: hci1016f19

[Alexander Friis]

[Dong Hyun Kim]

[Emil Skovgaard Mikkelsen]

SYNOPSIS:

The concept of having a personal sound zone dates decades back, but has been an active research topic ever since. Many methods and proof of concepts have been proposed, for how to practically create personal sound zones. An unexplored aspect of sound zones is how users interact with these, more specifically how users experience different interaction techniques for sound zone controlling. This paper reports the findings of an experimental study with 60 participants using three interaction techniques; two TUIs and one VUI, in an attempt to understand how users experience these interaction techniques in terms of effectiveness, efficiency, and user responses of the COOL questionnaire and an exit interview. From our findings, we conclude that when designing for sound zone controlling, tangible artefacts should be considered to achieve an efficient and effective interaction, while also bringing out enjoyable and fun aspects. Furthermore, sound zone controlling through voice interaction was the slowest interaction technique, with a relatively high incomplete task rate, and a low COOL rating.

Number printed:	[0] Pieces
Pages:	[11] Pages
Appendix:	[1]
Enclosures:	[0]

By signing this document, each member of the group confirms participation on equal terms in the process of writing the project. Thus, each member of the group is responsible for the all contents in the project.

Preface

This project is conducted by the 10th semester group hci1016f19 studying Interaction Design at Aalborg University 10th semester as their master's thesis in the spring of 2019. The projects research area is sound zone interaction with the focus on designing and evaluating interaction techniques. The project is divided in one main scientific paper and an appendix. The scientific papers format is CHI's scientific standard 2019, and the reference format is indicated by the Harvard method. Citations from interviewees are indicated with quotations and italic, and statistic data are indicated with parentheses. The project group wants to thank Dr. Jesper Kjeldskov and Prof. Dimitrios Raptis for their supervision. We would also like to thank the participants for their participation. The following students have participated in this project:

Alexander Friis Dong Hyun Kim Emil Skovgaard Mikkelsen

Resumé

I dette kandidatspeciale har vi udarbejdet et eksperimentelt studie omhandlende design af håndgribelige og stemmestyret interaktions teknikker til styring af lydzoner. Konceptet af lydzoner dateres årtier tilbage, men har siden været et aktivt forskningsemne. Adskillige metoder og proof of concept af hvordan man praktisk og matematisk skaber en personlige lydzone er blevet udarbejdet, men eksisterer endnu ikke i en fuldstændig løsning. Et uudforsket aspekt af lydzoner er, hvordan brugerne interagerer med disse, mere specifikt hvordan brugerne oplever forskellige interaktionsteknikker til styring af lydzoner. Vi har udarbejdet tre forskellige interaktionstyper ved hjælp af design funnel metoden. I denne proces er adskillige ideer indenfor områder som applikationer, gestures, stemmestyring og håndgribelige artefakter blevet udarbejdet. Ud fra disse, blev tre forskellige interaktionsteknikker udvalgt på baggrund af at være forskellige i forhold til fysiske aspekter som størrelse, form og interaktions metode. Den første interaktion teknik var CubeZone, som var en kubeformet håndgribelig artefakt, som kunne styre en lydzone ved at vende den rundt og placere den på en overflade. Det anden håndgribelige artefakt var CylinderZone, som var en cylinderformet artefakt med fire ringe der kunne roteres for at styre lydzone. I forbindelse med det eksperiment, designede vi en simuleret lydzone i et usability laboratorium.

I denne artikel rapporteres resultaterne af et eksperimentelt studie med 60 deltagere som interagerede med en af tre designet interaktionsteknikker; to tangible user interfaces og en voice user interface i et forsøg på at udforske, hvordan brugerne oplever disse interaktionsteknikker med hensyn til fejlrate, effektivitet og brugerrespons fra COOL-spørgeskemaet og et exit interview. Ud fra vores resultater kunne vi konkludere at når der skal designes til styring af lydzoner, bør håndgribelige artefakter overvejes for at opnå en effektiv og hurtig interaktion, samtidig med at skabe en underholdende interaktion. Vi kunne yderligere konkludere at styring af lydzoner gennem tale interaktion, var den mest langsomme interaktions teknik med en relativ høj fejlrate og lav coolness vurdering.

Sound Zone Interaction: A Study of Designing and Evaluating Interaction Techniques

Alexander Friis
Department of Computer Science
Aalborg University
Denmark
afrii13@student.aau.dk

Dong Hyun Kim
Department of Computer Science
Aalborg University
Denmark
dh13@student.aau.dk

Emil Skovgaard Mikkelsen
Department of Computer Science
Aalborg University
Denmark
esmi13@student.aau.dk

ABSTRACT

The concept of having a personal sound zone dates decades back, but has been an active research topic ever since. Many methods and proof-of-concepts have been proposed, for how to practically create a personal sound zone. An unexplored aspect of sound zones is how users interact with these, more specifically how users experience different interaction techniques for sound zone controlling. This paper reports the findings of an experimental study with 60 participants using three interaction techniques; two TUIs and one VUI, in an attempt to understand how users experience these interaction techniques in terms of effectiveness, efficiency, and user responses of the COOL questionnaire and an exit interview. From our findings, we conclude that when designing for sound zone controlling, tangible artefacts should be considered to achieve an effective and efficient interaction, while also bringing out enjoyable and fun aspects. Furthermore, sound zone controlling through voice interaction was the slowest interaction technique, with a relatively high unsuccessful task rate, and a low COOL rating.

KEYWORDS

Sound zone; VUI; Tangible artefacts; COOL Questionnaire

1 INTRODUCTION

In a household with multiple house members interested in listening to only their own sound source, earphones seem like the obvious choice [7]. When it comes to situations where individuals are interested in watching the same movie with different audio output, and still able to converse, no simple solutions exist.

The ability to record sound and playback has become increasingly omnipresent in our everyday lives [4]. The technology for spatial sound reproduction has made great progress in reproducing sound fields over large regions of space using loudspeakers aligned in arrays. This has led to the idea of establishing personal sound zones where interface-free audio can be delivered to multiple listeners in the same room, without the need of a headphones, and the result of not feeling physically isolated. Personal sound zones have drawn attention due to the numerous applications, from being able to control sound radiation from individual users to creating sound zones in mixed environments such as cars, shopping centres or open offices. The methods for creating personal sound zones are close to becoming a reality, but how users interact with sound zones is yet to be investigated, as some complications within understanding sound zones arises, such as, how sound zones are created, controlled or modified [1]. Furthermore, how different types of interaction techniques affects users' experience and performance with sound zones remains unexplored.

The interface of technologies is key to how users make use of it, and therefore researchers and designers continuously seek to design interfaces, that are user-minded. In recent years, tangible user interfaces (TUIs) have gotten a lot of attention in both learning and HCI communities, because of the distinguishing characteristics of TUIs [10]. Contrary to TUIs, another growing subject within HCI has shown to be voice user interfaces (VUIs) [21]. The grown interest of using VUIs is grounded in the advances of voice recognition technology, which has led to virtual assistants being integrated into devices such as smartphones and smart speakers. VUIs allow users to verbally interact with a natural spoken language, which can be used for tasks such as playing music, retrieving weather information or playing games [27]. The culmination of sound zones being realised and the current interest within the HCI community to develop new interactions techniques are the driving factors for this study, in order to explore sound zone interaction.

In this paper, we have investigated three interaction techniques designed for sound zone controlling. We conducted an experiment in a usability lab with 60 participants, measuring effectiveness, efficiency, COOL questionnaire and exit interview responses. Initially we provide an overview of related work within the areas of sound zones, tangible artefacts and voice interaction. We then present our three interaction techniques, CubeZone, CylinderZone, and Voice interaction. Lastly, we report and discuss our findings.

2 Related Work

Sound zones enable multiple individuals to enjoy different audio content within the same acoustic environment, without disturbing each other, see Figure 1. This effect can be obtained using headphones, however, they can hinder verbal interaction between individuals and can be inconvenient at times [18].

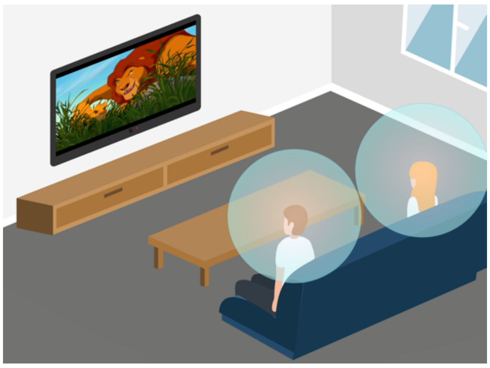

Figure 1: Illustration of two individuals with separate sound zones within the same acoustic environment.

Druyvesteyn and Garas [11] proposed the initial concept of a personal sound zone, that is by reproducing sounds in a desired area of space, while reducing sound levels elsewhere. The underlying concept was demonstrated in 1967 at Illinois Institute of Technology, where an array of loudspeakers was distributed on a surface, enclosing a select region, hereby enabling sound radiation control [4]. Microsoft presented a similar concept called "Personal Audio Space" where an array of 16 drivers was used to enhance sound in one area, while cancelling sound waves in another area. As of 2019 Lee et al. [19] presented a proof-of-concept of optimizing sound zones, shaped according to the human auditory system, bringing the concept of sound zones closer to a reality.

In 2013, Clever and Elliot [9] investigated the effects of a car cabin sized enclosure for different listening zones. With their system, they attempted to produce two independent listening zones; one in front and one in the rear. However, Clever and Elliot concluded that the proposed setup required additional speakers and further optimization.

Simon et al. [12] described an approach to boosting high frequencies for elderly or individuals with hearing loss, when watching television, by placing an array of speakers pointing towards an individual's specific location. Simon et al. were able to generate bright acoustic zones, where the sound was amplified, and dark acoustic zones where individuals with a normal hearing would not be interrupted by the amplification.

2.1 Tangible and Voice Interaction

The HCI community has for a long time been dedicated to better, and in a more naturally way, promote how information is transferred from humans to machines. The most natural form of communication for humans is through speech, which is also one of the most difficult processes to be understood by machines [23]. In HCI, voice inputs are simply spoken commands with no use of graphical user interfaces (GUIs), TUIs, or gestures for that matter. Unlike other interaction forms, VUIs pose a deep-rooted problem, namely they do not afford a given functionality to the user, and the other way around, the user is not able to perceive the capabilities of VUIs [17].

In contrast to hands-free interaction, TUIs are another popular topic within the HCI community [6]. With TUIs, the idea is to afford digital information in physical space and make use of human's capabilities to sense and manipulate objects. Contrary to VUIs, TUIs are not as machine heavy, but more user and task oriented, which opens up for different ways of interacting with medias [29].

Despite the difficulty of machines recognising speech, a previous study by Sciuto et al. [27] showed that participants were still willing to use technology that supported this type of interaction. Their findings even suggested that despite situations where their digital assistant would not understand their commands, the participants remained patient, and these situations did not retain the them from using it. Furthermore, the study showed VUIs are becoming more integrated in the everyday life of the participants, and were used for controlling smart home equipment such as home lighting, music and televisions.

In some of the earliest work of TUI, a numerous of prototypes have been developed within the areas of learning, programming, problem solving and entertainment. One example is the AlgoBlock [28] which is a tangible educational tool to support forming and maintaining a community of learners. Their findings showed that AlgoBlock was very effective in facilitating interaction among learners, due to the use of tangible physical blocks.

Newton-Dunn et al. [24] introduced the concept of Block Jam, inspired by interactive toys and sound devices, which consisted of physical artefacts for manipulating an interactive music system. They demonstrated that their system succeeded as a TUI, elicited positive experiences and evoked collaboration between users.

O'Malley and Fraser [25] asserted that tangible interfaces could be beneficial within the area of learning, and if carefully designed tangible interfaces could simplify problem solving tasks. O'Malley and Fraser pointed out that objects are not necessary easier because it is concrete, but the representational mapping is built by the physical activity.

3 EXPERIMENTAL STUDY

While no research regarding sound zone interaction exists, we have developed three interactions techniques for sound zone controlling and conducted an experimental study, and analysed collected data for further discussion. Through a usability evaluation, we sought to explore users' interaction with different interaction techniques for controlling a sound zone.

The experiment was conducted as a between-subject design, with one independent variable, the interaction techniques, namely the CylinderZone, the CubeZone and voice interaction. The four dependent variables were duration of task completion measured in time, the number of unsuccessful tasks, user responses of the COOL questionnaire and an exit interview. To avoid influencing results, two variables were experimentally controlled for each participant; the randomization of tasks and the beginning of time measuring, which will be described later. For the experiment, 60 participants used one of the three interaction techniques, thus 20 participants for each technique.

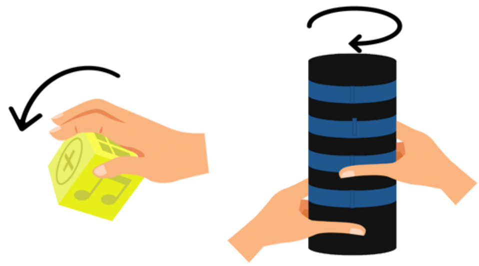

Figure 2: The two tangible interaction techniques: (a) the CubeZone and (b) the CylinderZone.

3.1 Interaction Techniques

The three interaction techniques were designed and inspired by prior studies of tangible artefacts and voice interaction. By following the design funnel process, we created numerous of different interaction techniques within the areas of tangible artefacts and voice interaction. Several different geometrical shapes for the form factor were considered. Three techniques were selected on the criteria of being diverse in attributes such as physical appearance, including size and shape, the interaction technique and realisation of a prototype. This process resulted in the design of two tangible interaction techniques and one voice interaction technique. All three interaction techniques were designed to perform the task shown in Table 2 in section 3.5.

3.1.1 Design of the CubeZone

A number of prior studies have focused on a tangible cube design within different areas, such as remote controls by Block et al. [5] with a playful approach to changing between TV-channels by rotating a cube. Matviienko et al. [22] designed the CubeLendar as an interactive cube which provided multiple types of information, such as weather and events using rotation. Camarata et al. [8] presented the Navigational Blocks for retrieving multiple historical information and the relationship between these by using physical cubes. Roudaut et al. [26] designed the Rubikon, an augmented Rubik's Cube which allowed users to interact with common user interfaces by rotating the sides, for example for navigation or 3D manipulation. Goh et al. [13] designed the i-Cube, a cube-shaped digital manipulative for music composition and users to learn spelling.

We created the CubeZone, see Figure 2a, with inspiration from these studies, as a simple cube shaped interaction technique, with the benefits of being easy to be understood, handled and manipulated by people [3]. The CubeZone has seven functionalities mapped to each side of the cube, which are activated when placing the cube on a surface with the wanted function facing upwards and also when lifted up and down on the same side for unmuting. Each function is illustrated through an individual icon on each of the cube's sides. The prototype was 3D printed in PLA and consisted of a five-sided cube, with a cavity and a cut-out for wiring, and a lid. Inside the cavity, an Arduino and a gyroscope were placed to detect rotation and acceleration of the CubeZone, hereby determining the chosen side, which was then passed on from the Arduino to a computer.

3.1.2 Design of the CylinderZone

Multifunctional cylindrical tangible objects are not prominent within the HCI community. However, a number of rotary tangible knobs have been designed and evaluated for different purposes, such as menu selection, comparing virtual and physical knobs with a focus on performance in speed [30]. Hilliges et al. [15] presented the Photohelix, consisting of a physical knob for browsing, sorting and sharing digital images within an interactive system.

The CylinderZone was designed to be different from the CubeZone in every way possible including its shape, size and interaction technique required to use it, see Figure 2b. The CylinderZone was inspired by the rotational interaction type of the tangible knobs and their shapes. However, to make all the required functions as clear and comprehensible as possible, we decided to expand the one rotatable ring to four rotatable rings placed on each other. Each ring was divided into two functionalities, placed on either side of the ring, which were initiated when turned towards the centreline of the CylinderZone. The CylinderZone was 3D printed in PLA and consisted of a main tube with two ascending rods and a cavity for a circuit at the bottom part, four rotatable rings, three dividing rings, a non-rotatable top ring and a bottom lid. One rotary encoder was locked in place on the centre of each rotatable ring, and the wires were led down to an Arduino placed in the bottom cavity. Through the rotary encoders, changes in rotation of the rings were detectable and passed on from the Arduino to a computer.

3.1.3 Design of the Voice Interaction

Speech input devices have been on the market for a number of years, and quite a few studies have been conducted, comparing other input devices such as keyboard and mouse. Karl et al. [16] used VUI for word processing and showed its improved performance over the use of a mouse. Gordon and Breazeal [14] designed PANDA, a driving assistant for reducing distractions for drivers while also engaging entertainment and education for children in the backseat. Bernheim and Johns [3] build and deployed the Add-in SpeechToast for Outlook to handle notifications, and found that speech input appealed to some participants, while others indicated a conditional willingness to have it. Ashok et al. [2] made a speech-enabled screen reader for web browsing to perform browsing actions. Lopatovska and Williams [20] designed a qualitative study with online diaries, and concluded that the most used aspect of VUI devices was for media purposes.

We chose to create multiple voice commands for each of the specified task. This decision was in response to the common voice assistants' difficulties in understanding users' commands, hereby creating a variety of solutions to solve each task. An emphasis was put on indexicality to ensure the different commands for each functionality were suitable for the specific setup during the experiment. In total, 23 different commands for 7 functionalities were created. These commands were created as conditional statements for a Google Home Mini by using the web-based service IFTTT, which allowed for custom responses to custom commands.

3.2 Designing an Alternative Sound Zone

As the practicality of sound zones is currently limited to specialised setups with low flexibility, we designed an alternative setup and scenario to simulate being in an environment with multiple sound zones and the ability to control different sounds within this.

3.2.1 Design of the Setup

The setup consisted of the two tangible interaction techniques connected to a computer running Arduino, Auto Mouse Clicker and Audacity. Through Audacity five different soundtracks were played through a headset which was to be worn around the participant's neck, simulating being in a sound zone. The five soundtracks were the following: Norwegian soundtrack, English soundtrack, Danish soundtrack, a music album, and a notification sound for creating a sound zone. The playback of the soundtracks was synchronised with another computer, connected to a TV through HDMI, which played the movie The Lion King without sound output.

The interaction with each tangible interaction technique was registered through the Arduino which sent simulated keyboard presses for each interaction possibility to the first computer. The keyboard presses were registered through either one of seven Auto Mouse Clicker programs which initiated a mouse click on a specific location on the screen in Audacity. For example, when the cube was placed with the music icon pointing upwards, the Arduino sent the keyboard press "1", which initiated the auto click program 1, which resulted in a mouse click in Audacity on the solo track of a music album.

3.2.2 Design of the Scenario

The scenario for the experiment will be based on Lee et al.'s [19] scenario in which two users want to watch a movie with two different language outputs. In order to increase the scenarios extent, three different scenarios were designed, as shown in Table 1.

Table 1: The three scenarios with description.

	Scenario description
Scenario 1	Two persons watching a movie, person A with Norwegian soundtrack and person B with English and later Danish soundtrack
Scenario 2	Two persons placed on a couch, Person A watching a movie with Norwegian soundtrack, person B listening to music from a speaker only
Scenario 3	Two persons placed on a couch, person A watching a movie with Norwegian soundtrack, person B not wanting to watch or hear sound from the movie

3.3 Participants

In total, 60 individuals took part in the experiment and were recruited through a combination of posters placed at Aalborg University and the online schedule service Doodle. Of the participants 42 were male and 18 were female, between the ages of 19 and 54 years old (M: 24,4). 58 of the recruited participants were students at Aalborg University, and of the other two participants, one of them worked as a consultant and the last participant was a senior citizen.

3.4 System Setup

The experiment was conducted in a usability lab and the setup included a couch placed in front of a table and a TV connected to a computer through HDMI placed in the belonging observation room, see Figure 3. On the table an iPad were placed next to a marked area in which one of the three interaction techniques were placed during the experiment. A camera was placed above the participants with the purpose of recording the interaction with the three interaction techniques. The two tangible interaction techniques were individually wired to a computer placed in the observation room.

Figure 3: The experimental setup.

3.5 Tasks

After the participants had signed the consent form, they were given an oral introduction of the experiment which included the purpose of the study, an explanation of sound zones and the scenario. The participants were explained that they were to solve eight tasks for five rounds using one of the interaction techniques to control different aspects of a sound zone. For each interaction technique, seven different possible types of interaction were implemented, shown in Table 2, leading to one task being repeated twice for the five rounds, which led to a total of 8 tasks x 5 rounds = 40 tasks, as recorded data. The inclusion of a recurring task resulted in an increase in the number of randomised sets of tasks, and in total 30 sets of tasks were randomised between the 60 participants. Each participant underwent a short practice round with the interaction technique at least once to become familiar with it. This practice round included all seven tasks. The participants were also instructed to only interact with the prototype while it being lifted from the marked area and placed within the area afterwards, in order for a consistent time measurement.

Table 2: List of the seven different tasks of interacting with a sound zone.

	Task description
Task 1	Create a sound zone
Task 2	Connect sound zone to the TV
Task 3	Connect sound zone to the speaker
Task 4	Set the language for the movie to Danish
Task 5	Set the language for the movie to English
Task 6	Mute the sound zone
Task 7	Unmute the sound zone

3.6 Data Collection

Throughout the experiment four different types of data were collected to explore differences between the three interactions techniques. The data collection included logging of effectiveness, efficiency, COOL questionnaire and exit interview responses.

3.6.1 Logging of Data

As the two tangible interaction techniques, each were controlled by an Arduino and connected to a computer, data for each participant were collected through the Arduino program. This data included a step by step logging of the participants' interaction with the respective interaction technique to determine if each task was solved correctly. For the voice interaction technique, one observer monitored unsuccessful attempts by hand, which were confirmed by the Google Home rejecting the specific command. In parallel to this logging, the observer measured the participants' time used for each task through a computer program. From these two measures two metrics were logged and used for the analysis, and these were defined as effectiveness (all unsuccessful attempts for each task) and efficiency (the time spent for each task). Furthermore, each of the user's interaction was video recorded and used for backup.

3.6.2 User Feedback

After completing all five rounds all participants answered the COOL questionnaire on the iPad, which consisted of 21 randomly mixed statements to which the participants had to rate on a 7-point Likert scale. Lastly all participants took part in an exit interview consisting of seven questions concerning the participant's experience, most likeable aspects of the interaction technique, surprises that arose, and the perception of being in a sound zone.

4 FINDINGS

The collected data were analysed in regard to usability with emphasis on effectiveness and efficiency. In addition to these, COOL questionnaire ratings and exit interview transcribes were analysed. The 60 participants all used one interaction technique for 5 rounds, each round consisting of 8 tasks leading to 800 time stamps and error loggings for each interaction technique. All collected data were included for the analysis, as no outliers were found in the data set.

4.1 Effectiveness

The effectiveness is defined by the individual participant's number of unsuccessful tasks for each of the five rounds, thus the maximum number of unsuccessful tasks for each participant was eight per round. The average number of unsuccessful tasks per participant for each interaction technique are shown in Table 3. We performed a one-way ANOVA due to the dependent variable being collected in intervals.

Table 3: Means of unsuccessful attempts for each interaction technique

	CylinderZone	CubeZone	Voice
N=8	0.12	0.01	0.19

We identified significant differences among the interaction types (F(2, 237)=5.752, p=0.004). A Tukey Post-Hoc test revealed that all interaction techniques were not significantly different from each other, except *CubeZone* and *Voice interaction* (p=0.003). For *CylinderZone* and *CubeZone* (p=0.082), and for *Voice interaction* and *CylinderZone* (p=0.457). For *Voice interaction*, participants had the highest average of unsuccessful tasks (0.19 out of 8, see Figure 4), while for *CubeZone,* participants had the lowest average of unsuccessful tasks (0.01 out of 8, see Figure 4).

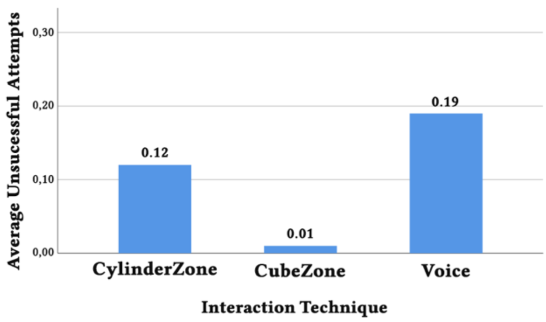

Figure 4: Average unsuccessful tasks for each interaction technique (lower is better).

4.2 Efficiency

The efficiency is defined by the duration of time it takes for the participant to solve one task, thus the measurement starts as the participant lifts or speaks to one of the interaction techniques and stops when the interaction technique is placed on the table again or not spoken to. We performed a one-way ANOVA for each task to examine the efficiency of each interaction technique. Table 4 shows the mean and standard deviation for each completion time per task.

Table 4: Means and (standard deviations) of time spent for each task per interaction technique.

	CylinderZone	CubeZone	Voice
Task 1: Create sound zone	4.035 (4.845)	1.896 (1.038)	3.847 (1.351)
Task 2: Connect to TV	2.811 (1.849)	1.643 (0.804)	4.011 (1.071)
Task 3: Connect to speaker	3.956 (1.914)	1.486 (1.180)	3.810 (0.967)
Task 4: Set Danish	3.377 (1.361)	1.259 (0.555)	3.320 (1.121)
Task 5: Set English	3.575 (2.047)	1.078 (0.593)	3.452 (0.909)
Task 6: Mute	3.730 (2.257)	1.691 (0.961)	3.087 (0.558)
Task 7: Unmute	3.582 (1.366)	0.583 (0.301)	2.429 (0.708)

For task 1, we identified significant differences among the interaction techniques (F(2, 237)=12.767, p<0.001). A Tukey Post-Hoc test revealed that all interaction techniques were significantly different from each other, except *Voice interaction* and *CylinderZone* (p=0.915).

For task 2, we identified significant differences among the interaction techniques (F(2, 237)=64.501, p<0.001). A Tukey Post-Hoc test revealed that all interaction techniques were significantly different from each other (for all cases p<0.001) in relation to efficiency.

For task 3, we identified significant differences among the interaction techniques (F(2, 237)=76.887, p<0.001). A Tukey Post-Hoc test revealed that all interaction techniques were significantly different from each other, except *Voice interaction* and *CylinderZone* (p=0.793).

For task 4, we identified significant differences among the interaction techniques (F(2, 237)=102.203, p<0.001). A Tukey Post-Hoc test revealed that all interaction techniques were significantly different from each other, except *Voice interaction* and *CylinderZ*one (p=0.939).

For task 5, we identified significant differences among the interaction techniques (F(2, 237)=88.513, p<0.001). A Tukey Post-Hoc test revealed that all interaction techniques were significantly different from each other, except *Voice interaction* and *CylinderZone* (p=0.830).

For task 6, we identified significant differences among the interaction techniques (F(2, 237)=41.220, p<0.001). A Tukey Post-Hoc test revealed that all interaction techniques were significantly different from each other (for all cases p<0.001), except *Voice interaction* and *CylinderZone* (p=0.015).

For task 7, we identified significant differences among the interaction techniques (F(2, 237)=223.280, p<0.001). A Tukey Post-Hoc test revealed that all interaction techniques were significantly different from each other (for all cases p<0.001).

In addition to the one-way ANOVA, a mixed repeated measures ANOVA was performed to investigate the combined efficiency of interaction techniques and task completion time with a Greenhouse-Geisser correction, see Figure 5. The Greenhouse-Geisser correction was chosen due to the repeated measures ANOVA assumption of sphericity being violated (p<0.001) and both Epsilon values were below 0.75 (Greenhouse-Geiser=0.475, Huynh-Feldt=0.486). The effect of each task completion time on efficiency was significant (F(2.852, 675.896)=13.542, p<0.001) and so was the effect of the interaction techniques (F(2,237) = 136.171, p<0.001). Their interaction was also significant (F(5.704, 675.896) = 6.284, p<0.001), showing that a combined effect of interaction technique and completion time effectiveness exists.

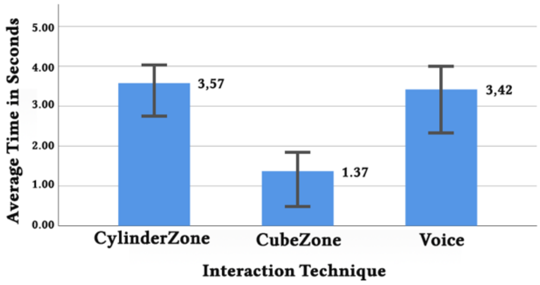

Figure 5: Average time of task completion for each interaction technique (lower is better).

4.3 Coolness

20 participants each evaluated one interaction technique in relation to its individual coolness. This section describes each interaction techniques average scores for the overall coolness, desirability, rebelliousness, usability, classic aesthetics, hedonic quality. Means of the COOL questionnaire responses for each factor are shown in Table 5.

Table 5: Means of COOL questionnaire of each factor for each interaction technique.

	CylinderZone	CubeZone	Voice
Overall coolness	4.55	5.40	4.90
Desirability	4.51	4.98	3.74
Rebelliousness	5.49	5.87	3.83
Usability	6.35	6.60	5.64
Classic	5.08	6.07	5.98
Hedonic	5.29	5.88	4.95

For the overall coolness factor, we have identified a significant difference between the interaction techniques (F(2,179)=5.201, p=0.006). A Tukey Post-Hoc test revealed that *CylinderZone* and *CubeZone* were the only techniques with a significantly difference between each other (p=0.004).

For the desirability factor, we have identified a significant difference between the interaction techniques (F(2,239)=15.398, p<0.001). A Tukey Post-Hoc test revealed that all interaction techniques were significantly different from each other, except *CubeZone* and *CylinderZone* (p=0.102).

For the rebelliousness factor, we have identified a significant difference between the interaction techniques (F(2,239)=58.520, p<0.001). A Tukey Post-Hoc test revealed that all interaction techniques were significantly different from each other, except *CubeZone* and *CylinderZone* (p=0.134).

For the usability factor, we have identified a significant difference between the interaction techniques (F(2,239)=19.391, p<0.001). A Tukey Post-Hoc test revealed that all interaction techniques were significantly different from each other, except *CubeZone* and *CylinderZone* (p=0.266).

For the classic aesthetic factor, we have identified a significant difference between the interaction techniques (F(2,119)=8.735, p<0.001). A Tukey Post-Hoc test revealed that all interaction techniques were significantly different from each other, except *CubeZone* and *Voice* interaction, (p=0.924).

For the hedonic quality factor, we have identified a significant difference between the interaction techniques (F(2,239)=10.683, p<0.001). A Tukey Post-Hoc test revealed that all interaction techniques were significantly different from each other, except *Voice interaction* and *CylinderZone* (p=0.230).

4.4 User Responses

A thematic analysis of the transcribed exit interviews was conducted in order to explore how the participants experienced the three interaction techniques and from this, several interesting aspects emerged. This section will present the participants experience of the interaction technique and their perception of the concept of sound zone.

4.4.1 CylinderZone

Participants responded positively to the interaction of turning the rings on the CylinderZone. 14 participants mentioned that the CylinderZone was simple, and they had a lot of fun interacting with it. In addition, a few participants played with it during the first round of tasks to explore the technique, rather than trying to solve the specific tasks. Participant 9 associated the CylinderZone with a kid's toy which made it fun to interact with, *"It's like those puzzle toy, that makes this very fun to interact with. It's definitely funnier than a single button, because we are not used to turning things, I really love the way these rings can turn and the tactical feel of it"*. In several cases the participants compared the CylinderZone to a remote control, and six participants would prefer the CylinderZone when interacting with a sound zone, *"It [interaction technique] is fine, it's simple, it's easier than a remote control. The most annoying thing on remote is that every button is different and unorganised. The smart thing with this cylinder is that, you can see all the different functionalities on it at all time, there is a clarity in it"* (Participant 12). When asked about the experience of using the CylinderZone to solve the different sound zone tasks, participants felt it was simple and understandable, *"I really love it, because it solves the problem in relation to the desire of multiple language when watching TV, and each person can have one of these in order to have their own sound zone, I think that is pretty cool"* (Participant 8). Participant 7 mentioned that it would be a problem to interact with sound zones using only one remote control, *"I can see the problem when you have multiple sound zones, then it would be difficult with all those buttons on a single remote control"*.

The size of the CylinderZone was a drawback for 18 participants, as they mentioned that the size should be smaller, allowing them to interact with it using only one hand, *"I think a smaller size would be smarter, so that I can control it with one hand, instead of using two hands"* (Participant 16). The participants also had a hard time locating all the functionalities from one angle due to the size and shape of the CylinderZone, making them have to turn it around to locate the specific function, *"it was too big to get a grasp on. And because of its size, I could not see all the elements on each ring"* (Participant 4).

4.4.2 CubeZone

Almost all participants responded positively to the CubeZone in various degrees, and mentioned amongst other topics, that it was easy to use, fun to interact with and easy to remember. For example, participant 37 responded that it was easy to use and learn, and the only thing that would make it potentially difficult to use, was the fact that all functions were not visible at the same time, but something that would be quick to learn. Participant 31 stated similarly, *"It [functions] will probably take time to learn. However, I remembered the position of the functionalities at the fifth round"*. The method of turning the CubeZone to change functionality, compared to an ordinary remote control, was an aspect several participants liked the most, *"What I liked the most about the product was how fast I had the opportunity to switch from one thing to another. You didn't have to hassle with something else [remotes]"* (Participant 30). In addition, participant 22 responded that he also like the feature of having to rotate the CubeZone, and compared to an ordinary remote, the CubeZone was more fun to use, *"I think that rotating is very fine. In comparison to a regular remote, the cube is more fun"*. However, not all of participants responded only positively to the CubeZone, for example, participant 39 did not like the technique in regard to a long-term use, as he commented "*I think it would be cool at first, but it would probably become a little annoying after a few months"*. Participant 29 shared somewhat the same opinion, "*It was fine [the interaction]. Four buttons could have done the same, but it is fine"*.

In regard to the design of the CubeZone five participants had mixed responses to it. Participant 25 mentioned that he did not like the CubeZone because it was clumsy and large in size, and he disliked the chosen colour. In addition, participant 30 mentioned, *"I would prefer a little more colour, so I could better distinguish the individual functions from each other"*. However, participant 21 mentioned that the design of the CubeZone having opposing functions made it easy and simple to use and was an alternative solution to other options he had tried before. Participant 28 mentioned, *"I found that I learned it quite quickly. In the beginning, I turned it around several times, but after a couple of rounds you learn where the symbols are located"*.

4.4.3 Voice Interaction

10 of the 20 participants responded that with voice interaction it was easy to interact with a sound zone, although all 20 participants agreed that it was difficult to remember the exact command for each functionality, *"It (interaction) is very easy, but it's kind of difficult to remember the exact sentences you have to say, otherwise it is pretty easy"* (Participant 50). The participants often felt they wanted to express each command more casually and wanted to shorten the longer commands, *"Some commands should be shortened, especially when connecting to other sources. Like this sentence, when you*

say, 'create sound area here' and afterwards having to say another command 'connect this sound area to TV'. It should ask me which sources I want to connect to right after I have created a sound zone, so then I can just say TV" (Participant 43). A considerable problem was also that the Google Assistant was not able to understand the participants every time, *"It had troubles in understanding a few things and I can become a tiny irritated when it does not understand me"* (Participant 57). Another interesting aspect of voice interaction was how the participants' own voice could be a disturbing element in different situations. Participant 46 felt it confusing to speak to the Google Assistant, while being in a sound zone surrounded by the sound from the movie, *"I feel it way more confusing when it has something to do with sound, that I need to speak louder than the sound played. I don't know whether it is more disturbing for me or it [Google Assistant]"*. Participant 43 felt it being disturbing to everyone else in the context, rather than disturbing for himself, *"If I'm at home, and I'm continuously speaking to it, then it would disturb others around me"*.

4.4.4 The Concept of Sound Zones

All of the 60 participants shared their thoughts on the concept of having a personal sound zone. Despite an overall positive openness of sound zones, five of the participants had difficulties imagining it in reality. Participant 8 mentioned that he liked the idea and that it solved the problem if two individuals wanted to listen to separate sources in the same space, but he was unsure about it being realistic, *"I really like the product. I have doubts about what the product is, how it would work in a real setting"* (Participant 8). Participant 38 stated similar, "*I think it is a little difficult to imagine, because you are yourself, and then to imagine someone else, is listening to something else*".

In regard to the social aspect, two participants mentioned it would be weird being together and listening to different sound sources, *"Maybe a little weird to invite friends over, and you don't want to listen to the movie but rather music"* (Participant 2). In regard to the privacy aspect of a sound zone, four participants could see the benefits of them being able to zone out privately, without having to go to another room. Six participants mentioned it being comforting knowing they would not disturb others around themselves when being in their own sound zone, and that it would open up for a lot of use cases, for example when playing video games, focusing on homework or if their girlfriend talked too much "*It feels reassuring, that you do not disturb other people, when sitting by yourself"* (Participant 21).

5 DISCUSSION

Our findings, collected through the experimental study, make preliminary contributions to the research area of interacting with sound zones. We have presented insights in regard to the effectiveness, efficiency, user responses in terms of the COOL questionnaire and an exit interview, of the three interaction techniques.

5.1 Effectiveness

The interaction technique with the lowest average of unsuccessful tasks was the CubeZone, while the technique with the highest average of unsuccessful task were respectively voice interaction and the CylinderZone. We believe that the CubeZone was the most successful for several reasons. Firstly, the geometrical shape of the CubeZone was described by several participants as being simple and easy to remember due to coherent functions placed on the opposing sides of each other. Secondly, the size of the CubeZone was not deemed problematic by the participants in contrast to the problematic visibility of the functions of the CylinderZone, which led to the participants having to awkwardly turn the cylinder to locate each function.

Voice interaction was in many cases problematic for the participants and caused the most unsuccessful attempts, which might be due to several reasons. Firstly, the number of useable voice commands per task exceeded one as for the two tangible interaction techniques, which might have caused confusion for the participants, resulting in them mixing commands and having difficulties in remembering all commands correctly. Furthermore, multiple participants stated that the voice commands felt unnatural to say because of the length and wording off the commands, which might have caused some internal contradictions and led to the higher number of unsuccessful attempts.

5.2 Efficiency

The interaction technique with the fastest completion time was the CubeZone, while respectively voice interaction and the CylinderZone were the slowest. The participants using the CubeZone were significantly faster than the other interaction techniques, and we assume this was due to several reasons, here amongst the simplicity of the CubeZone. The movement of turning the CubeZone's sides was both easy to perform and fast due to the participants being able to use one hand movements. In contrast, the CylinderZone's shape and interaction technique required a combination of two hand movements, for lifting and stabilizing the CylinderZone with one hand, while using the other hand to perform several turning movements of the rings. In addition, the USB cable connected to the CylinderZone caused some problems for the participants in regard to handling and turning of the rings, which may have resulted in a prolonged task completion.

The improvement in time for the voice control was relatively low, which could be due to several reasons. The voice interaction was fairly limited in terms of the Google Home's speech recognition abilities, which were expressed through the numerous of failed recognised commands, leading to the participants having to repeat themselves multiple times. Furthermore, the length of the designed voice commands and the participants ability to speak the commands faster, may have hindered any time improvement during the five rounds. In addition, several participants emphasised expressing the sentence correctly rather than using natural language.

5.3 User Responses

Several of the participants mentioned that they had fun while interacting with both the CylinderZone and the CubeZone. Some participants mentioned that it resembled a toy and further stated that it was fun turning and twisting the tangible techniques. The participants mentioned that the two interaction techniques were enjoyable to use, which may be the reason for the higher score of the tangible interaction techniques, compared to the voice interaction score.

In regard to using the CylinderZone and the CubeZone, some participants found it difficult to initially understand the interaction technique and how to use it, which led to a few participants being slightly frustrated. An explanation for the frustration could be, that not all participants fully understood the purpose of having a personal sound zone, and had their own interpretation of a sound zone, rather than following the description for each task. The frustration could have arisen when they thought they were performing the tasks correctly, but not getting the wanted feedback. In addition, multiple participants experienced that the voice interaction was not able to recognize their commands and explained that it was probably due to their Danish accent. This resulted in many repeated attempts at single commands, which in some cases led to frustrations and the feeling of wasting time. This could be the reason for the lower desirability score of the Voice control compared to other tangible interaction techniques.

The COOL questionnaire responses of the voice interaction were in general lower than the tangible interaction techniques. However, the classic aesthetic aspect was rated in line with the CubeZone, which were the highest rated in terms of classic aesthetic. An explanation for this could be that the tangible interaction techniques were still at a prototype stage, and were not as polished as the Google Home Mini used for the experiment, therefore rated more aesthetically pleasing.

The CubeZone had the highest rated scores for all six coolness factors. An explanation for this could be the simplicity of the CubeZone's design in terms of shape and interaction method, which made it easy for the participants to use and remember the different functionalities.

6 CONCLUSION

We have presented a study of designing interaction techniques for sound zone control. We have compared the use of three different interaction techniques (CubeZone, CylinderZone and voice interaction) for seven different sound zone tasks, to explore their efficiency, effectiveness and user responses in terms of the COOL questionnaire and an exit interview.

Our findings show that the participants performed best using the CubeZone, thus the CubeZone was the most efficient technique in terms of task completion time. In addition, the CubeZone was the most effective interaction technique in terms of the number of successful attempts for each task performed. Lastly the participants had the most positive responses towards the CubeZone in relation to the COOL questionnaire and during the exit interviews.

From our findings, we conclude that when designing for sound zone controlling, tangible artefacts should be considered to achieve an efficient and effective interaction, while also bringing out enjoyable and fun aspects to users. In contrast, voice control performed worst in terms of both effectiveness and user response, while also being the second slowest interaction technique.

7 FUTURE WORK

Future research should expand the scope of interaction techniques and include for example mobile applications or gestures in order to explore users' experience and performance with radically different techniques. In relation to our positive findings of the CubeZone, the area of tangible interaction techniques should be investigated more in-depth and include a large variety of different shapes and interaction methods. Future research should also include additional sound zone functionalities, such as adjusting the sound zone sizes, moving sound zones, or merging sound zones. Lastly, it might be interesting to investigate users' experience with multi-modal interaction techniques.

ACKNOWLEDGMENTS

We would like to thank Dr. Jesper Kjeldskov and Prof. Dimitrios Raptis, for providing the supervision for our project. We would also like to thank the anonymous participants for their participation.

REFERENCES

[1] Aarhus University 2019. Controlling Sound Zones – with Perceptually Optimized Multichannel Signal Processing. Retrieved June 17, 2019 from https://eng.au.dk/forskning/forskningsprojekter/electrical-and-computer-engineering-research-projects/controlling-sound-zones-with-perceptually-optimized-multichannel-signal-processing/

[2] Vikas Ashok. 2014. Capti-Speak: A Speech-Enabled Accessible Web Interface. ASSETS '14. http://doi.org/10.1145/2661334.2661416.

[3] A. J. Bernheim Brush and Paul Johns. 2012. SpeechToast: augmenting notifications with speech input focus. AVI '12. https://doi.org/10.1145/2254556.2254607.

[4] Terence Betlehem, Wen Zhang, Mark A. Poletti and Thushara Abhayapala. 2015. Personal Sound Zones: Delivering interface-free audio to multiple listeners. IEEE '15. http://doi.org/10.1109/MSP.2014.2360707.

[5] Florian Block, Albrecht Schmidt, Nicolas Villar and Hans W. Gellersen. 2004. Toward a Playful User Interface for Home Entertainment Systems. EUSAI '04. http://doi.org/10.1007/978-3-540-30473-9_20.

[6] Mark Brown, Winyu Chinthammit and Paddy Nixon. 2014. A Comparison of User Preference for Tangible Objects vs Touch Buttons with a Map-based Tabletop Application. OzCHI '14. https://doi.org/10.1145/2686612.2686645.

[7] Michael Bull. 2006. Investigating the Culture of Mobile Listening: From Walkman to iPod, Consuming Music Together: Social and Collaborative Aspects of Music Consumption Technologies. 131-149. http://doi.org/10.1007/1-4020-4097-0_7.

[8] Ken Camarata, Ellen Yi-Luen Do, Brian R. Johnson and Mark D Gross. 2002. Navigational Blocks Navigating Information Space with Tangible Media. IUI '02. https://doi.org/10.1145/502716.502725.

[9] Jordan Cheer and Stephen J. Elliott. 2013. Design and implementation of a personal audio system in a car cabin. Acoustical Society of America '13. http://doi.org/10.1121/1.4805230.

[10] Sébastian Cuendet, Engin Bumbacher and Pierre Dillenbourg. 2012. Tangible vs. Virtual Representations: When Tangible Benefit the Training of Spatial Skills. NordiCHI '12. https://doi.org/10.1145/2399016.2399032.

[11] W. F. Druyvesteyn and Ronald M. Aarts. 1994. Personal Sound. Acoustical Society of America '94. https://doi.org/10.1121/1.410932.

[12] Marcos Felipe Simón Gálvez, Jordan Cheer and Stephen John Elliott. 2012. Loudspeaker arrays for family TV. ICSV '12. 1-8.

[13] Wooi Boon Goh, L.L. Chamara Kasun, Fitriani, Jacquelyn Tan and Wei Shou. 2012. The i-Cube: Design Considerations for Block-based Digital Manipulatives and Their Applications. CHI '12. https://doi.org/10.1145/2317956.2318016.

[14] Michael Gordon and Cynthia Breazeal. 2015. Designing a virtual assistant for in-car child entertainment. IDC '15. https://doi.org/10.1145/2771839.2771916.

[15] Otmar Hilliges, Dominikus Baur and Andreas Butz. 2007. Photohelix: Browsing, Sorting, and Sharing Digital Photo Collections. TABLETOP '07. http://doi.org/10.1109/TABLETOP.2007.14.

[16] Lewis R. Karl, Michael Pettey and Ben Schneiderman.1992. Speech versus mouse commands for word processing: an empirical evaluation. International Journal of Man-Machine Studies '93. http://doi.org/10.1006/imms.1993.1078.

[17] Lars Bo Larsen, Kasper Løvborg Jensen, Søren Larsen and Morten H. Rasmussen. 2007. Affordance in Mobile Speech-Based User Interaction. MobileCHI '07. https://doi.org/10.1145/1377999.1378020.

[18] Taewoong Lee, Jesper Kjær Nielsen, Jesper Rindom Jensen and Mads Græsbøll Christensen. 2018. A Unified Approach to Generating Sound Zones Using Variable Span Linear Filters. IEEE '18. https:// doi.org/10.1109/ICASSP.2018.8462477.

[19] Taewoong Lee, Jesper Kjær Nielsen and Mads Græsbøll Christensen. 2019. Towards perceptually optimized sound zones: A proof-of-concept study. IEEE '19. http://doi.org/10.1109/ICASSP.2019.8682902.

[20] Irene Lopatovska and Harriet Williams. 2018. Personification of the Amazon Alexa: BFF or a Mindless Companion. CHIIR '18. https://doi.org/10.1145/3176349.3176868.

[21] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really bad PA": The Gulf between User Expectation and Experience of Conversational Agents. CHI '16. https://doi.org/10.1145/2858036.2858288.

[22] Andrii Matviienko, Christoph Ressel, Sebastian Horwege, Susanne Boll and Lennart. 2016. CubeLendar: Design of a Tangible Interactive Event Awareness Cube. CHI EA '16. https://doi.org/10.1145/2851581.2892278.

[23] Cosmin Munteanu and Gerald Penn. 2015. Speech-based interaction: Myths, Challenges, and Opportunities. CHI EA '15. https://doi.org/10.1145/2702613.2706679.

[24] Henry Newton-Dunn, Hiroaki Nakano and James Gibson. 2003. Block Jam: A Tangible Interface for Interactive Music. NIME '13. http://doi.org/10.1076/jnmr.32.4.383.18852.

[25] Claire O'Malley and Danae Stanton Fraser. 2004. Literature Review in Learning with Tangibles Technologies. NESTA Futurelab Report 12. 4-12.

[26] Anne Roudaut, Diego Martinez, Amir Chohan, Vlad-Stefan Otrocol, Rupert Cobbe-Warburton, Max Steele, Ioana-Madalina Patrichi. 2014. Rubikon: A Highly Reconfigurable Device for Advanced Interaction. CHI '14. https://doi.org/10.1145/2559206.2581275.

[27] Alex Sciuto, Arnita Saini, Jodi Forlizzi and Jason I. Hong. 2018. "Hey Alexa, What's up?": Studies of In-Home Conversational Agent Usage. DIS '18. https://doi.org/10.1145/3196709.3196772.

[28] Hideyuki Suziki and Hiroshi Kato. 1995. Interaction-Level Support for Collaborative Learning: AlgoBlock - An Open Programming Langauge. CSCL '95. https://doi.org/10.3115/222020.222828.

[29] Cristina Sylla, Pedro Branco, Clara Coutinho and Eduardo Coquet. 2010. TUIs vs. GUIs: comparing the learning potential with preschoolers. Personal Ubiquitous Computing '12. https://doi.org/10.1007/s00779-011-0407-z.

[30] Simon Voelker, Kjell Ivar Øvergård, Chat Wacharamanotham and Jan Borchers. 2015. Knobology Revisited: A Comparison of User Performance between Tangible and Virtual Rotary Knobs. ITS '15. https://doi.org/10.1145/2817721.2817725.