Article

Rhythmic-Synchronization-Based Interaction: Effect of Interfering Auditory Stimuli, Age and Gender on Users' Performances

Alessio Bellino (ID)

Facultad de Ingenieria y Ciencias, Escuela de Informática y Telecomunicaciones, Universidad Diego Portales, Santiago 8370109, Chile; alessio.bellino@mail.udp.cl

Abstract: Rhythmic-synchronization-based interaction is an emerging interaction technique where multiple controls with different rhythms are displayed in visual form, and the user can select one of them by matching the corresponding rhythm. These techniques can be used to control smart objects in environments where there may be interfering auditory stimuli that contrast with the visual rhythm (e.g., to control Smart TVs while playing music), and this could compromise users' ability to synchronize. Moreover, these techniques require certain reflex skills to properly synchronize with the displayed rhythm, and these skills may vary depending on the age and gender of the users. To determine the impact of interfering auditory stimuli, age, and gender on users' ability to synchronize, we conducted a user study with 103 participants. Our results show that there are no significant differences between the conditions of interfering and noninterfering auditory stimuli and that synchronization ability decreases with age, with males performing better than females—at least as far as younger users are concerned. As a result, two implications emerge: first, users are capable of focusing only on visual rhythm ignoring the auditory interfering rhythm, so listening to an interfering rhythm should not be a major concern for synchronization; second, as age and gender have an impact, these systems may be designed to allow for customization of rhythm speed so that different users can choose the speed that best suits their reflex skills.

Keywords: interaction techniques; human factors; users' performance; novel interfaces; age and gender effect

1. Introduction

In the UbiComp era, the control of devices is an ongoing challenge. In recent years, many interaction techniques were developed—also for remote control—to meet the different needs of users. In this context, different interaction techniques were studied. Some of these techniques employ rhythmic synchronization [1–4]: they work by displaying multiple animated controls that show different rhythms in visual form, and the user can select one of them by synchronizing with the corresponding rhythm. Controls can be physical (e.g., in Figure 1) or virtual (i.e., shown on a screen, such as those used in this study, see Figure 2). An example of interaction based on rhythmic synchronization is available at the link https://youtu.be/ri7iQ-nJspg, (accessed on 21 January 2022).

Rhythmic-synchronization-based techniques originated somehow from movement-correlation techniques [5–10], where controls show different motion patterns and the user can select one of them by mimicking the corresponding motion.

From the point of view of the input device, rhythmic synchronization techniques can be simpler compared to those based on movement correlation. While motion-correlation techniques require sensors that detect movement (e.g., cameras [6,7] or kinect [9]), rhythmic synchronization techniques require sensors as simple as a button, e.g., [1,4]. As matter of fact, previous studies showed that these techniques can support a wide variety of sensors [1,3]; in particular, any sensor capable of generating a binary input through which

users can perform the required rhythm. As a result, these techniques are quite flexible. In addition, motion correlation techniques are not suitable for continuous control because "they require the user to continuously follow the target for prolonged periods" [7]. On the contrary, it was shown that rhythmic synchronization techniques can be used for continuous control [1,4] without major drawbacks.
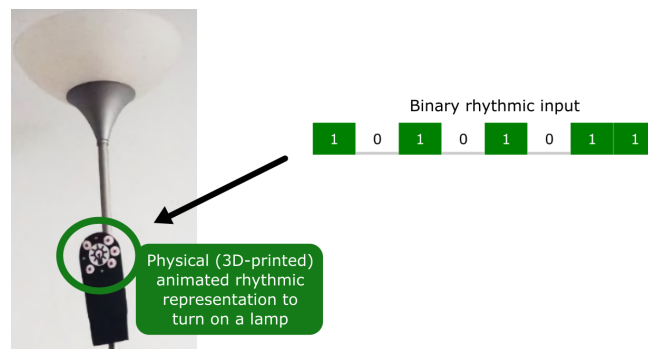

Figure 1. How rhythmic-synchronization-based interaction works: an animated control shows a rhythm, and user can select it by matching corresponding rhythm, e.g., using a binary sensor as a button to mark rhythm. In example, the rhythmic sequence is shown by a physical control using a series of LEDs arranged in a circular way. Once user synchronizes with given rhythmic sequence, lamp turns on.
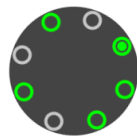

Figure 2. Control visible to user, with which it was required to synchronize.

2. Motivation and Hypotheses

Previous studies focused primarily on the design and user evaluation of rhythmic synchronization techniques (e.g., [1,2,4,11]), while another focused on improving matching by modeling user behavior [3]. However, none of the above studies were conducted evaluating whether any external auditory stimulus rhythmically contrasting with the visual rhythm can somehow alter users' ability to synchronize. In fact, these interaction techniques are often used in scenarios where users may be required to synchronize with visual rhythmic patterns while listening to auditory stimuli (e.g., in a living room, while listening to music through the TV, radio, or media player) whose rhythm interferes with the visual rhythmic pattern. Therefore, we aim to investigate whether the presence of an interfering auditory stimulus may affect the ability to synchronize with the visual representation of a rhythm. The theory of selective attention [12] states that users are able to selectively focus on the stimulus of their interest (e.g., the visual representation of a rhythm) while ignoring interfering ones (e.g., disturbing background music). In addition, visual perception is generally dominant compared to auditory perception [13]. Consequently, the presence of a contrasting auditory stimulus should not significantly affect users' ability to synchronize; however, it is necessary to confirm this hypothesis through an appropriate evaluation with users. Based on the discussion above, the first research question is: to what extent does a contrasting auditory stimulus affect user performance when synchronizing with controls?

In addition, none of the above studies evaluate interaction techniques with a large and varied number of users, and it is therefore unknown to what extent these techniques are suitable for a heterogeneous population in terms of age and gender. For example, three studies [1–3] involved 12 participants each, whose average ages were 27, 23, and 22, and the number of males always outnumbered females (nine males and three females in each study). Another study involved only eight participants [4], six males and two females,

with an average age of 22 years. The last study mentioned above [11], finally, involved 11 participants of unspecified age and gender. Considering then that these studies involved few users, all young, and overwhelmingly male, it seems relevant to us to understand the extent to which interaction techniques based on rhythmic synchronization can be used by a heterogeneous population. In fact, an interaction technique should be usable by users of different ages and genders for it to be truly applicable in the real world. In the case of rhythm-based interaction techniques, however, there are several reasons why users might perform differently depending on age, gender. More specifically, these techniques show a cyclic rhythm pattern, which can have different speeds depending on system settings. Faster rhythms require greater synchronization ability—i.e., faster reflexes—than slower rhythms. Therefore, considering that these techniques require users to have certain reflex abilities to synchronize correctly with the displayed rhythm, we need to understand to what extent these reflex abilities vary depending on age and gender. Previous literature shows that response time to stimuli varies with age and gender [14–16], so one might expect older and female users to have more difficulty to synchronize with visual rhythm patterns than younger and male users. To confirm the above hypothesis, it is necessary to verify the effect of age and gender through a user evaluation. Based on the discussion above, the second research question is: to what extent do age and gender affect user performance when synchronizing with controls?

To conclude, only by answering the two research questions above we can understand the extent to which similar techniques are usable in contexts where interfering auditory stimuli may be present (first research question) and by a heterogeneous population (second research question). In the following sections, we present the related work, the user evaluation, which involved 103 participants, and the results. Finally, we discuss our results and the implications for the design of rhythmic synchronization techniques.

3. Related Works

In this section, we first present some rhythm-based interaction techniques that were previously investigated by other researchers. Next, we discuss studies on selective attention, a theory that states that users are able to selectively focus the stimulus of their interest (e.g., the visual representation of rhythm) while ignoring interfering ones (e.g., disturbing noise, or background music). Finally, we discuss some studies that discuss how reaction to stimuli varies by age and gender.

3.1. Rhythm-Based Interaction Techniques

Rhythmic patterns are not widely used as an interaction technique, but their potential is promising. Ghomi et al. [17] discovered that rhythm-based interaction is not only easy to understand and easy to use by new users, but rhythm memorization is as simple as keyboard shortcuts memorization.

In recent years, several rhythm-based interaction techniques were studied covering different usage scenarios such as user authentication, device pairing and intelligent object/environment control. Among these techniques it is necessary to distinguish between those that use rhythm in a generic way and those that are based on rhythmic synchronization—where controls show different rhythmic patterns, and users can select one of them by synchronizing with the corresponding rhythm.

Among the techniques that use rhythm in a generic way, we could mention TapSongs [18] and RhythmLink [19]. The former uses rhythm patterns to authenticate users, allowing them to use a single binary sensor to establish self-created rhythm patterns as passwords. Once the patterns are established, the rhythm entered by the user is compared with the previously established pattern, allowing user authentication. The latter allows devices to pair through predefined rhythm patterns that must be known by both devices.

The rhythmic synchronization techniques discussed below, however, are the most related to our study. Tap-to-pair [2] allows users to pair devices by performing a rhythmic coupling associated with a device without the need for hardware or firmware modifications.

Rhythmic Menus [11] allows users to select menu items. These are highlighted cyclically at a given period and users can select one by clicking anywhere on the screen when the target item is highlighted. Synchrowatch [4] allows users to control smartwatches simply by matching rhythm patterns on the screen using a passive magnetic ring as a rhythm detection device. Finally, SEQUENCE [1] uses a new design that represents the rhythmic patterns through eight animated dots arranged in a circular way around the target. This facilitates target selection as the associated rhythm patterns are fully displayed to the user [3].

To conduct our user evaluation, we used the SEQUENCE design to visually represent the rhythmic patterns.

3.2. Studies on Selective Attention and Stimuli Combination

Selective attention is the ability to select a stimulus in the presence of distractors [12]. It allows us to ignore what is irrelevant to the task at hand and focus only on what is relevant. One of the most widely recognized theories is that of perceptual load theory [20–23], which indicates that the success or failure of selective attention depends on the processing load required by the task at hand. It also proposes that stimuli that are relevant to the task in question should be processed first, rather than those that are not.

In regards to the combination of auditory and visual stimuli, where the first one works as a distractor—which is our case, since the user should follow the visual stimuli ignoring the any auditory stimuli—Molloy et al. [24] suggest that observers fail to note auditory stimuli under high visual perceptual conditions (inattentional deafness). A similar conclusion was found by Macdonald and Lavie [25], which demonstrated that people fail to note the presence of acoustic stimuli when they are focused in a high-visual-load condition task. This means that if users are concentrated towards a high-load visual stimulus (e.g., the visual rhythmic representation in our case) they tend to ignore auditory stimuli (e.g., an interfering rhythmic auditory stimulus).

On the whole, the previous literature shows that, presenting contrasting auditory and visual stimuli, the latter tend to override the former as in some audiovisual illusions (e.g., ventriloquist effect [26], McGurk effect [27], Colavita effect [28]). Thus, vision is generally dominant in humans. However, there are particular conditions in which visual perception is altered by auditory stimuli [13,29].

3.3. Effect of Age and Gender on Reaction Time

There are numerous studies that analyze the effect that age has on people's reaction time, which is also used as an indicator of age deterioration. Forzard et al. [14] studied changes in reaction times over an eight-year observation period, finding that, starting at approximately age 20, reaction times increased by about 1ms per year. In addition, both the number of errors and reaction time increased for both sexes, and males showed shorter reaction times than females in any observed age group. Der and Deary [15] observed that reaction times have a curvilinear relationship with age, and grow faster at older ages. On the other hand, females showed slightly higher reaction times. Thompson et al. [16] studied reaction times through the performance of video game players, determining that players' cognitive-motor skills start to decrease at the age of 24, regardless of previous experience in video games. Finally, differences in performance were observed depending on the type of reaction to be performed by the user. In particular, Siegler [30] observed that, as we age, decision times decrease, while movement times increase.

4. User Evaluation Method

4.1. Participants and Apparatus

For this study, we selected participants from all sociocultural backgrounds who had at least minimal familiarity with computers. Participants with physical disabilities that would not allow them to press keys on the keyboard were not considered. Accordingly, we involved 103 participants (51 males and 52 females) aged between 18 and 73 (M: 34, SD:

14). None of them had previous experience with the interaction technique covered in this study. Due to the COVID-19 contingency, we conducted online experiments. Therefore, participants were required to have a personal computer capable of emitting sound, to plug in headphones (if possible), and to conduct the experiment while making sure to avoid distractions. Due to this condition, the operating systems and web browsers of the participants differ in each instance of the experiment. Concerning the operating systems, most of the participants used Windows (67), while the rest used macOS (17) and Linux (5). With regard to web browsers, most of the participants used a Chromium-based browser (76), while the rest used Safari (5) and Firefox (8). We were unable to identify browser/OS information in 14 cases.

We excluded three participants because they claimed, in retrospect, that they did not fully understand the experiment; accordingly, we considered the data of 100 participants.

4.2. Evaluation Application and Collected Data

A web application was developed for the execution of the experiment. At the beginning, the application explained what the participants had to do (see procedure section below). Then, participants were allowed to declare gender and age so that the experiment could start.

The experiment consisted of 60 trials, and for each of them, the application showed a rhythmic control (Figure 2) consisting of green and gray circles and a smaller dot moving within them, along with seven other hidden controls whose rhythms were shifted by one unit of time each. Therefore, there were eight controls active at the same time, seven of which were hidden to the user, while one randomly chosen was visible to the user for each trial (Figure 3).
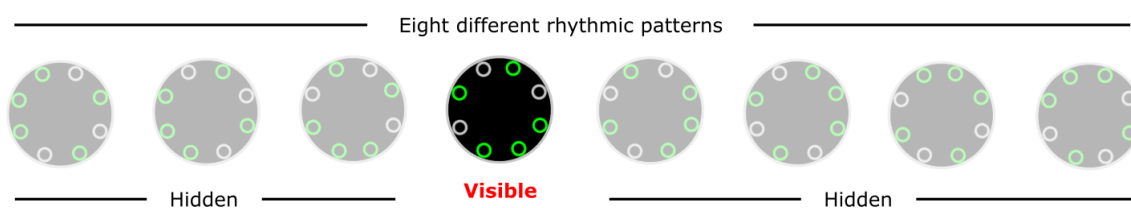

Figure 3. Eight controls with different rhythm patterns, of which one was made visible to user while seven were hidden. Visible control changed randomly during experiment for any trial. If one of hidden control was unintentionally activated, a false positive would be recorded.

Participants were required to select the visible control by pressing the space bar when the dot was inside the green circles, making all green circles disappear in sequence. The green circles were restored to the visible state if the space bar was pressed when the dot was inside a gray circle or if it was not pressed when the dot was inside a green circle. The trial was considered completed when one of the following events occurred: (1) all the green circles disappeared (i.e., control successfully activated); (2) 10 s passed after the start of the trial (timeout interval after which a missed activation was recorded); or (3) an unwanted control was activated (i.e., activation of an incorrect control after which an error was recorded). Upon completion of a trial, a new trial was started. The rhythm speed of the controls was initially set to 2.8 s, as suggested by the previous study on SEQUENCE [1]. However, according to a pretest on 17 students (16M/1F) aged between 21 and 28 years (M: 24), we decided to accelerate the rhythm slightly since the error rate was very low. In fact, eight of them did not make mistakes, while the others made very few. Therefore, to make synchronization more difficult, we set the rhythm to 2.4 s. Note that in the previous study on SEQUENCE, the rhythm speed at 2.4 s caused an error rate of 20.66% [1].

The application collected data on 60 trials split into two conditions consisting of 30 trials each, namely "Sound", in which the application played an auditory rhythm (drum loop) interfering with the visual rhythm, and "Mute", in which the application displayed

the visual rhythm without playing any sound. The condition changed every three trials so that learning, i.e., carryover effect, was distributed equally among the conditions, without one condition being dominant over the other. In addition, the application required the user to complete nine additional trials, whose data were not recorded, at the beginning of the experiment. This allowed participants to familiarize themselves with the experiment before collecting any data. During the 60 trials actually recorded, the application collected data on (1) activation time, (2) false activation (accidental activation of one of the 7 hidden controls), and (3) missed activation (no control activated within 10 s—timeout interval).

At the end of the experiment, the application asked to fill in a questionnaire about the ease of synchronization in both conditions (evaluated on a six-point ordinal scale) and a text field where they could write a comment on the experience. Finally, the collected data were sent to the server and saved as CSV for further analysis.

4.3. Interfering Rhythm Design

To make synchronization as difficult as possible, the auditory rhythm of "Sound" condition was conceived to induce users into error while synchronizing with the visual rhythm. To do so, we designed a drum loop of 10 elements where the accents follow an irregular pattern such as 1,2–1,2,3–1,2–1,2,3. In addition, the auditory rhythm was stretched to match the duration of the visual rhythm to alternate coincidences and contrasts between visual and auditory rhythm as shown by Figure 4. Finally, the video at https://youtu.be/upZIMWzQQLo (accessed on 21 January 2022) shows a comparison between the two different conditions where the contrast between visual and auditory rhythm can be seen.
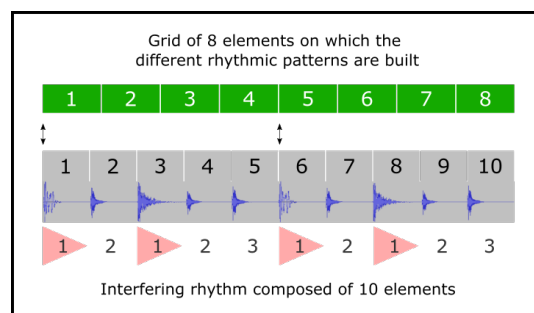

Figure 4. Coincidences and contrasts between visual rhythmic pattern grid (eight elements in green) and auditory rhythm (10 elements in gray). First and fifth beats of visual rhythm correspond to first and sixth beats of auditory rhythm, respectively (see black arrows), while other beats do not correspond. This alternation between coincidences and contrasts, in addition to irregular pattern of accents (red triangles), was conceived to induce users into error when synchronizing with visual rhythm.

4.4. Procedure

We generated several links containing a unique identifier for each participant through which it was possible to take part in the experiment. Each link was sent to the corresponding participant via email, Skype, WhatsApp, or any other platform considered suitable by the same participant. Once the link was opened, an introductory video (https://youtu.be/RLozq5ROPUw (accessed on 21 January 2022)) reminded participants to turn on the computer audio and, when possible, to use headphones. It was also recommended to avoid distractions by informing other people living in the same house to avoid disturbing the participants during the experiment. Whenever possible, participants were advised to set the airplane mode on their smartphone to avoid being interrupted by any calls or notifications. For greater clarity, the same recommendations were also recalled in text form before the experiment directly in the web application described in the previous section. A simple informed consent was then shown to participants. The informed consent stated that, by participating in the experiment, users would agree that the data collected would be used for research purposes and would not be associated with specific individuals.

Then, the participants filled in two fields, i.e., gender and age, and the experiment could start. Once the last trial of the experiment was completed, each participant could fill in a questionnaire on the ease of synchronization for both conditions using a scale from 1 (very difficult) to 6 (very easy) and a complementary text field where they could write any comments on the experiment.

4.5. Data Processing

For each participant and condition ("Sound" and "Mute"), we extracted the number of errors, which is the sum of missed activations (i.e., no activation within 10 s) and unwanted activations (i.e., incorrect activation of a control).

Moreover, we extracted the average activation time for each participant, which could be calculated on successful trials only. Therefore, in case on high number of errors, activation time may be not a reliable indicator. For example, if a participant made 28 errors out of 30, it means that the average activation time could be calculated only on two correct activations, which makes the average unreliable. Consequently, as the number of errors increases, activation times become less reliable. That is why we consider the number of errors as the most reliable indicator for this research.

4.6. Stratification and Data Analysis

To begin with, we point out that age distribution of males and females was quite different, and this could distort the results regarding between-subject comparisons. For example, if we were to find differences between genders, they might be due to the fact that males were on average younger than females, and not because of any real differences between genders. This is a common case of confounding variable: in fact, age behaves as a confounding variable in relation to the comparison between males and females. To control for confounding variables, we use stratification. Broadly speaking, stratification lets researchers control "for confounding by creating two or more categories or subgroups in which the confounding variable either does not vary or does not vary very much" [31].

As stratification strategy, we divided participants in males and females and then in different age ranges. To identify age ranges, we looked for the best compromise to obtain (1) at least 10 participants for each gender, and (2) similar ages between the different genders. As shown in in Table 1, the mean age between males and females is nearly equivalent in any age groups (see mean difference column in Table 1), and this let us to control the effects of confounding variables. In fact, even assuming that age differences are statistically significant, practical importance is not relevant as reaction times increase at a rate of only ≈1 ms for each year of aging [14], and, in our case, the highest average age difference between males and females is only 2.5 years, corresponding to the age group 26–40. In this regard, please note that the concepts of statistical significance and practical importance (as referred in [32]) or substantive significance (as referred in [33]) are quite different.

Table 1. Stratification of participants in age ranges and genders with standard deviation (SD) and absolute difference between means.

Age Group	Gender	Participants	Mean Age	SD	Mean Difference
18–25	M	26	22.8	1.7	1.3
	F	17	21.5	2.2	
26–40	M	12	30.0	4.7	2.5
	F	16	32.5	4.2	
41+	M	11	53.5	9.6	1
	F	18	54.5	9.4	

This study is both (1) within-subject, i.e., each participant has tried both "Sound" and "Mute" conditions to see if interfering auditory stimuli affect user performances

(first research question), and (2) between-subject, i.e., different ages and genders were compared to see if there are differences in terms of performances (second research question). For both within-subject and between-subject comparisons, we used nonparametric tests to detect significant differences since some data were not normally distributed (according to Shapiro–Wilk test). Particularly, when comparing "Sound" and "Mute" conditions, we used the (paired) Wilcoxon Signed-Rank test to detect significant differences in terms of (1) activation times, (2) error rates, and (3) perceived ease of synchronization. Regarding genders, we performed the (unpaired) Wilcoxon Rank Sum test to see any significant differences between males and females in the different age ranges. Moreover, in regards to age, we used the Spearman's rank-order correlation to assess the relation between (1) age and times of activation, and (2) age and error rate for both males and females separately. We also highlight that to identify any differences between different ages, males and females are treated separately to avoid gender acting as a confounding variable. Finally, results are presented using the median (instead of the mean) to avoid skewing possibly caused by non-normality.

5. User Evaluation Results

5.1. Interfering Auditory Stimulus Effect

To investigate whether users are generally able to ignore interfering auditory stimuli when synchronizing with visual rhythmic patterns, we evaluated the differences between "Sound" and "Mute" conditions. In this regard, both for activation time and error rate, the Wilcoxon Signed-Rank test suggests that there is no significant difference ($p > 0.05$) between the distributions, neither by gender nor by age ranges (see Figure 5).
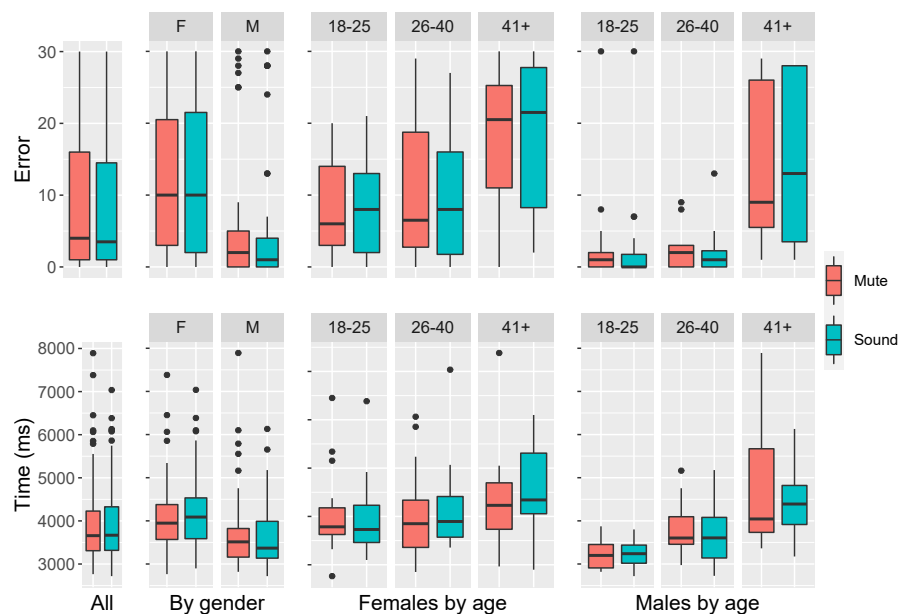

Figure 5. Boxplots showing error rate and time of activation by comparing "Mute" (red) and "Sound" (green) conditions for (1) all participants, (2) by gender, and (3) by gender and age. None of differences are significant.

Regarding ease of synchronization (questionnaire responses) the "Mute" condition is perceived as easier than "Sound" one considering all participants regardless of gender and age. Although the median is five for both conditions, the differences between the distributions is significant ($p = 0.039$). However, when distinguishing between genders, we detect a significant difference within females (Medians: four for "Sound" vs. five for "Mute", $p = 0.038$), but not males ($p > 0.05$). Finally, with regard to age groups, no significant difference was found ($p > 0.05$) (see Figure 6).

Table 2 shows comparisons across all age groups and genders with *p*-values for error rate, time of activation and ease of synchronization.

Table 2. Condition comparison regarding Error, Time, and Easiness of synchronization for all age group. Significant differences are marked with asterisk.

	Group	Sound Median	Mute Median	Sig.
Error	All	3.5	4	0.32
	Males	1	2	0.57
	→18–25	0	1	0.67
	→26–40	1	2	0.49
	→41+	13	9	0.89
	Females	10	10	0.44
	→18–25	8	6	1
	→26–40	8	6.5	0.39
	→41+	21.5	20.5	0.58
Time (ms)	All	3668	3658	0.71
	Males	3368	3513	0.40
	→18–25	3239	3199	0.93
	→26–40	3605	3602.5	0.17
	→41+	4390	4044	0.75
	Females	4088	3947	0.25
	→18–25	3730	3785	0.36
	→26–40	3895.5	3850	0.55
	→41+	4341.5	4229.5	0.14
Easiness	All	5	5	0.03 *
	Males	5	5	0.50
	→18–25	5	5	1
	→26–40	5	5	1
	→41+	4	4.5	0.18
	Females	4	5	0.03 *
	→18–25	4	5	0.70
	→26–40	4	5	0.05
	→41+	4	5	0.20
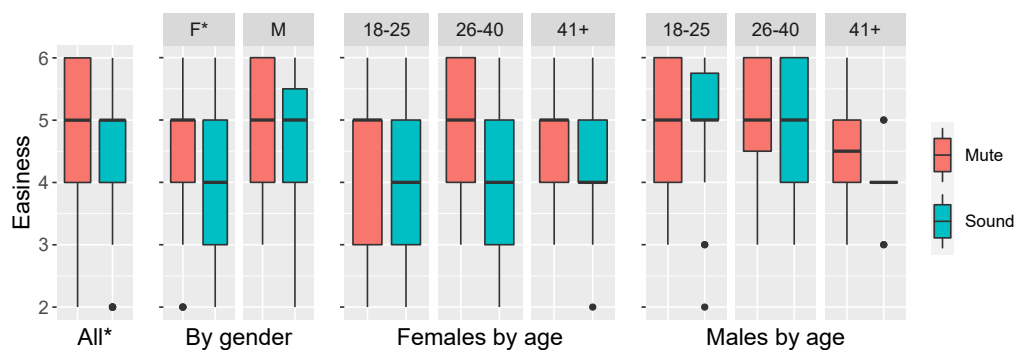
Figure 6. Boxplots showing ease of synchronization by comparing "Mute" and "Sound" conditions for (1) all participants, (2) by gender, and (3) by genders and age. Significant differences are marked with asterisk.

5.2. Age and Gender Effect

To investigate the effect of age and gender on users' performance, we consider the two conditions separately. We begin by presenting data for the "Sound" condition, and next we present data for the "Mute" condition.

5.2.1. Age and Gender Effect in "Sound" Condition

In regards to "Sound" condition, we computed Spearman's rank-order correlation to investigate the relation between age and errors for both males and females separately. Correlations indicates a moderate, positive, significant correlation between errors and age in both males ($R = 0.49$, $p < 0.001$) and females ($R = 0.50$, $p < 0.001$). Accordingly, we can state that as age increases, the number of errors increases for both males and females (Figure 7).
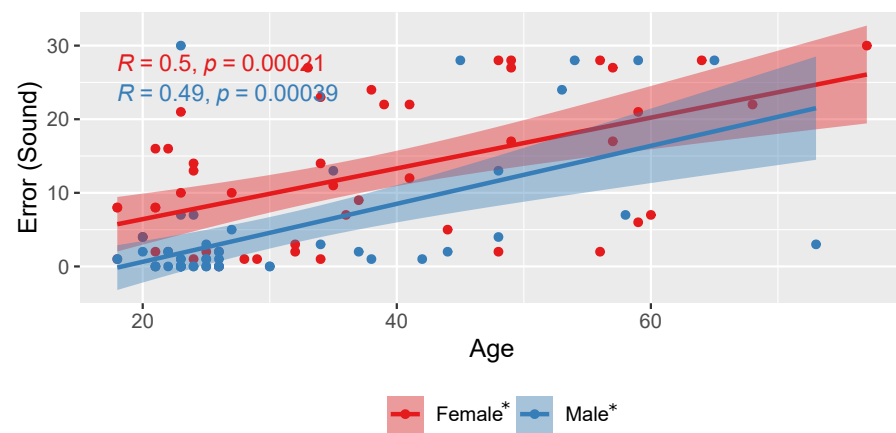
Figure 7. Error versus Age in Sound condition: correlation for males and females displayed using a scatter plot with regression lines and 95% confidence interval. Significant correlations are marked with asterisk.

Focusing on gender differences, Figure 7 suggest that confidence intervals do not overlap in the ≈22–40 age interval. Therefore, we can state there is a significant difference at a confidence level of 95% between males and females in terms of errors in that age interval.

We further analyzed gender differences by comparing the number of errors between males and females in the different age groups (Figure 8). To begin with, in 18–25 age group, males make fewer mistakes than females (Medians: 0 vs. 8), and the difference is significant ($p < 0.001$). Moreover, in 26–40 age group, males also make fewer error than females (Medians: 1 vs. 8), and the difference is significant ($p = 0.012$) In the 41+ age group, no significant differences were found.

Continuing our study of the "Sound" condition, we now focus on activation times. To do so, we computed Spearman's rank-order correlation to investigate the relation between age and times of activation for both males and females separately. Correlations indicate a moderate, positive, significant correlation between times of activation and age in males ($R = 0.59$, $p < 0.001$). Moreover, the Spearman's rank-order indicate a weak, positive, significant correlation between times of activation and age in females ($R = 0.35$, $p < 0.014$). Accordingly, we can state that as age increases, the times of activation increases for both males and females (Figure 9). Focusing gender differences, moreover, Figure 9 also suggest that confidence intervals do not overlap in the 18–35 age rages. Therefore, we can state that there is a significant difference at a confidence level of 95% between males and females in terms of activation time at that age range.
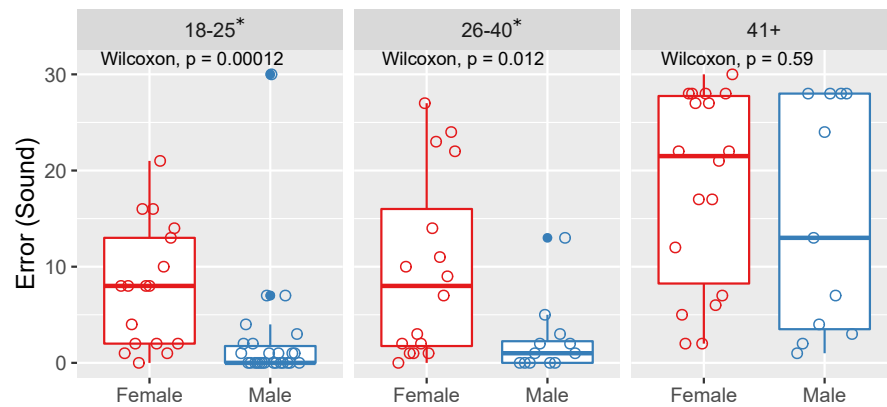
Figure 8. Boxplots comparing age groups of males and females regarding errors in "Sound" condition. Significant differences are marked with asterisk.
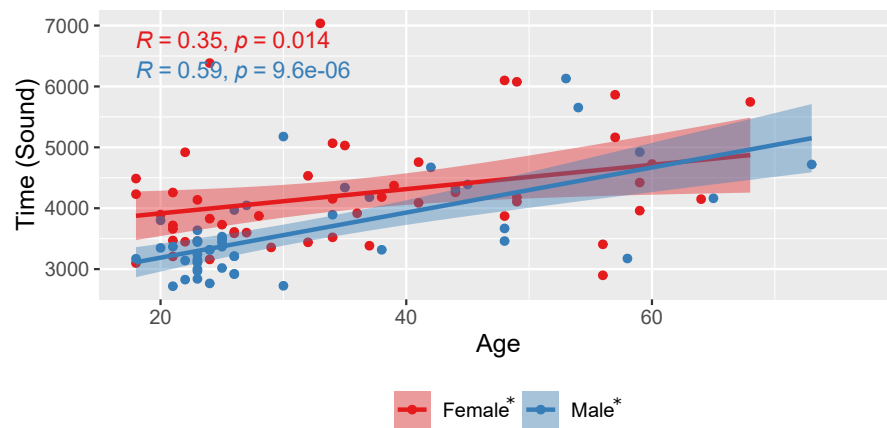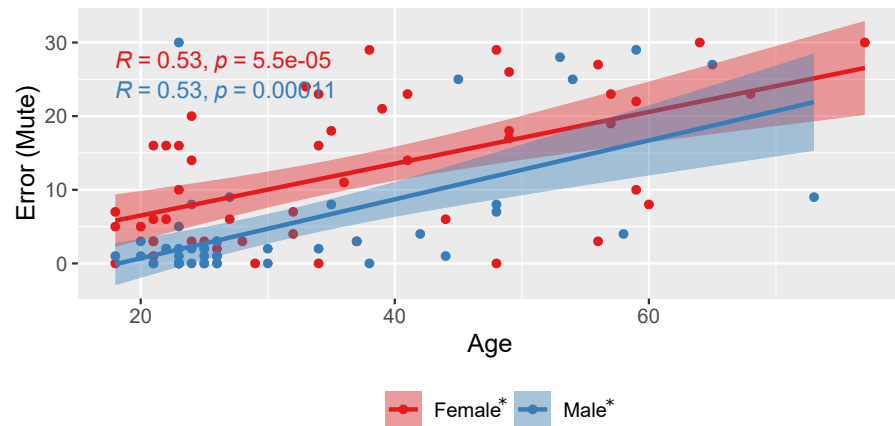

Figure 9. Times of activation versus Age in Sound condition: correlation for males and females displayed using a scatter plot with regression lines and 95% confidence interval. Significant correlations are marked with asterisk.

We further analyzed gender differences by comparing the activation times between males and females in the different age groups (Figure 10). To begin with, in 18–25 age group, males are faster than females (Medians: 3368 vs. 3730), and the difference is significant ($p < 0.001$). Finally, in the 26–40 and 41+ age groups, no significant differences were found.
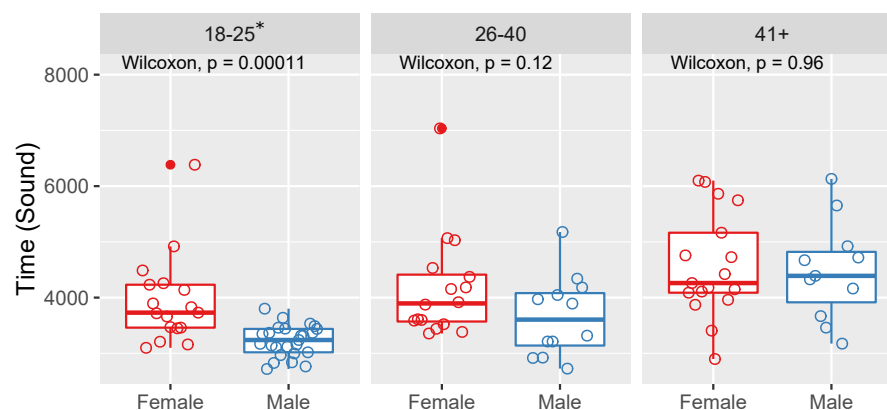
Figure 10. Boxplots comparing age groups of males and females regarding times of activation in "sound" condition. Significant differences are marked with asterisk.

5.2.2. Age and Gender Effect in "Mute" Condition

In regards to "Mute" condition, we computed Spearman's rank-order correlation to investigate the relation between age and errors for both males and females separately. Correlations indicates a moderate, positive, significant correlation between errors and age in both males and females ($R = 0.53$, $p < 0.001$ each). Accordingly, we can state that as age increases, the number of errors increases for both males and females (Figure 11).

Figure 11. Error versus Age in Sound condition: correlation for males and females displayed using a scatter plot with regression lines and 95% confidence interval. Significant correlations are marked with asterisk.

Focusing on gender differences, moreover, Figure 11 suggest that confidence intervals do not overlap in the ≈22–40 age interval. Therefore, we can state there is a significant difference at a confidence level of 95% between males and females in terms of errors in that age interval.

We further analyzed gender differences by comparing the number of errors between males and females in the different age groups (Figure 12). To begin with, in 18–25 age group, males make fewer mistakes than females (Medians: 1 vs. 6), and the difference is significant ($p < 0.001$). Moreover, in 26–40 age group, males also make fewer error than females (Medians: 2 vs. 6.5), and the difference is significant ($p = 0.022$). In the 41+ age group, no significant differences were found.
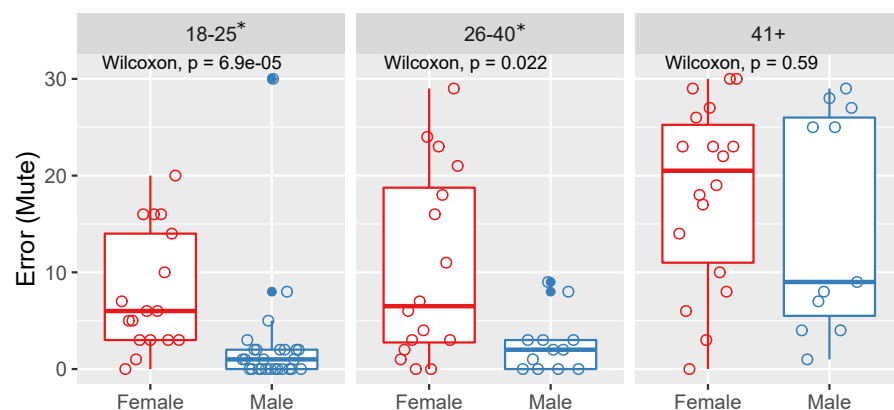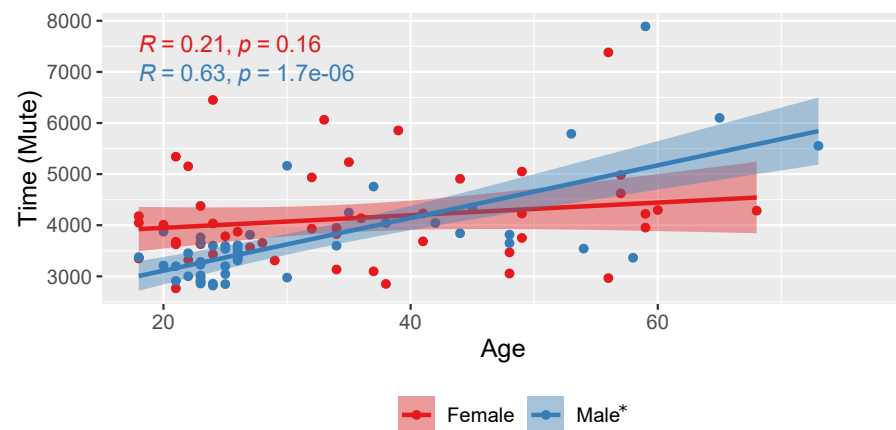

Figure 12. Boxplots comparing age groups of males and females regarding errors in "Mute" condition. Significant differences are marked with asterisk.

Continuing our study of the "Mute" condition, we now focus on activation times. To do so, we computed Spearman's rank-order correlation to investigate the relation

between age and times of activation for both males and females separately. Correlations indicate a moderate, positive, significant correlation between times of activation and age in males ($R = 0.63$, $p < 0.001$). Moreover, the Spearman's rank-order indicate a weak, positive, but *not significant* correlation between times of activation and age in females ($R = 0.21$, $p = 0.16$). Accordingly, we can state that as age increases, the times of activation increase for males. Females have actually a similar tendency, but the correlation is not significant (Figure 13).

Focusing on gender differences, Figure 13 also suggests that confidence intervals do not overlap in the 18–28 age range. Therefore, we can state that there is a significant difference at a confidence level of 95% between males and females in terms of activation time in that age range.

Figure 13. Times of activation versus Age in Mute condition: correlation for males and females displayed using a scatter plot with regression lines and 95% confidence interval. Significant correlations are marked with asterisk.

We further analyzed gender differences by comparing the activation times between males and females in the different age groups (Figure 14). To begin with, in 18–25 age group, males are faster than females (Medians: 3199 vs. 3785), and the difference is significant ($p < 0.001$). Finally, no significant differences were found in the 26–40 and 41+ age groups.
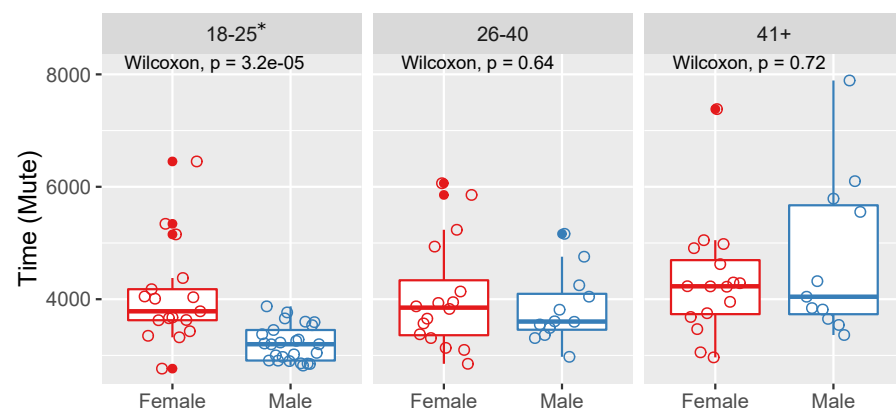

Figure 14. Boxplots comparing age groups of males and females regarding errors in "Mute" condition. Significant differences are marked with asterisk.

5.2.3. Participants' Comments

Forty-four participants wrote their comments on the experiment. We now report the main themes that emerged from the comments.

Twelve participants stated that the sound of the drum loop (Sound condition) did not affect their ability to synchronize with the rhythm, while six participants stated otherwise. Curiously, one participant stated that the drum loop facilitated synchronization. Two users stated that it was easy to synchronize with the visual rhythm because it was always the same-in fact, if we consider the rhythm as a loop, the rhythm pattern is the same for any control, but is shifted by one unit of time to distinguish the different controls [1]. Two other participants stated that their performance could have improved if the image representing the visual rhythm had been larger.

6. Discussion and Implications

RQ1: To what extent does a contrasting auditory stimulus affect user performance when synchronizing with controls?

From a psychological point of view, selective attention works quite well, since users are generally able to concentrate on visual rhythm, ignoring interfering auditory stimuli. In fact, no performances differences were found between Sound and Mute conditions for any gender and age range (see Figure 5), so we can state that the presence of interfering stimuli during interaction with the system should not affect users' performance. However, the results of the questionnaire showed that—at least female—participants thought it was easier to synchronize in the Mute condition (see Figure 6). This, on second thought, would seem reasonable considering that the auditory stimulus, even if it does not really affect users' performance, is in any case a disturbance.

As an implication, it could be said that rhythmic interaction techniques can be used without major concern for background music (for example, while watching television or listening to music) or noise in the environment. In fact, even if a possible auditory stimulus could be considered a disturbance, the effects on the synchronization capacity—at least with regard to rhythmic interference—are practically nonexistent.

With regard to the participants' comments, 12 of them stated that their performance was not affected by the drum loop, and considering that their activation times and error rates were indeed very similar, we can state that their comments reflect the reality of the data. Concerning the other six participants who stated that the drum loop negatively affected their performance, we can state that this was only a subjective feeling, considering that their comment is not reflected in the data. The averages of these participants indicate that there was a negligible difference of 200 ms in the activation time between the conditions in favor of the Mute condition. In terms of errors, they were even less in the Sound condition, and this further confirms that it was a subjective feeling that did not reflect the reality of the data.

RQ2: To what extent do age and gender affect user performance when synchronizing with controls?

Regarding age and genders, our results are in line with previous works, (e.g., [14,15]), which demonstrated that reflex skills vary considerably depending on the age and gender of the users. In fact, we can generally state that activation time and error rate increase with age, and females perform worse than males—at least as far as younger users are concerned.

Particularly, we have enough data to state that males perform better than females in terms of error rate, at least in the 18–40 age range, considering both Sound and Mute conditions. Finally, in terms of activation time, we have enough data to state that males performs better than females, at least in the 18–25 age range.

Accordingly, age and gender differences clearly have implications for the design of rhythmic interaction techniques. Generally, previous studies showed that a slower rhythm speed facilitate synchronization, i.e., users make less errors. Previous studies (e.g., [1–3]) focused on finding the best compromise between lower activation times and error rate, showing that errors decrease at lower speed of rhythm. Therefore, we may assume that slower rhythms speeds can be employed to facilitate the synchronization of less capable users. We then suggest that rhythmic interaction techniques include a brief configuration phase in which the ideal rhythmic speed for the end user is determined in a guided manner—

for example, using a wizard. Broadly speaking, age and gender does not really matter if we look only at the individual skills of the users. What really matters is that users feel comfortable with the given rhythm speed, considering that the optimal speed may vary considerably from individual to individual.

7. Limitations

We identified four limitations of this study. The first concerns the generalizability of the results to other similar interaction techniques; the second concerns the fact that there may be particular auditory stimuli that could worsen synchronization capabilities; the third consists of the use of a single rhythm speed (2.4 s), so our results are not properly generalizable to other rhythm speeds; and the fourth relates to the conduct of online experiments, which does not ensure full control over experimental conditions.

1. Our study may have implications for the design of any interaction technique that requires rhythmic synchronization. However, our study has used the SEQUENCE design paradigm [1] and therefore, in strict terms, its validity could be limited if extended to other similar paradigms (e.g., [2,4,11]).

2. Although vision is generally dominant in humans [13], there could be special conditions, which if intentionally provoked, can cause visual perception to be altered by auditory stimuli. In fact, there may be auditory stimuli that makes focusing on the visual rhythm harder or even subverts that rhythm, as it happens in some auditory–visual illusions [29].

3. Although the results might be generalizable to different rhythm speeds, strictly speaking, the results presented in this study apply only to the rhythm speed corresponding to 2.4 s. Further studies would be needed to investigate whether the results of this study remain valid at other speeds.

4. Broadly speaking, conducting online experiments does not give the experimenter full control over experimental conditions. Therefore, there could were some differences in conditions between participants. Some of the factors we can mention are the following: the quality of the headphones—or loudspeakers—was not equal among the participants, and thus the overall sound perception changes. Although rhythmic perception (which is the critical one for our study) is virtually independent of the type and quality of the headphones or loudspeakers used, there are, however, open headphones, which let through most of the external noise, and closed-back headphones, which limit external noise; thus, each participant could have different perceptions of external noise depending on the quality and type of headphones used. Finally, environmental conditions such as brightness or room acoustics (which may be influential when loudspeakers are used) were not equal among the participants who, therefore, had a different overall experience during the experiment. Statistically speaking, these differences possibly added some noise to the data. However, it is safe to assume that such randomly occurring noise is uniformly distributed across all groups analyzed in this study (male, females, young and old participants) and thus, ultimately, does not practically affect the results.

8. Conclusions

Rhythmic interaction techniques are promising as they require minimal user movement and can be used in multiple contexts (physical object control, intelligent environment, remote control, device pairing, etc.) without major concern for noise or music in the environment. In fact, users' performances are generally not affected by interfering auditory stimuli. However, determining the ideal rhythm speed may be helpful to facilitate user synchronization. Users showed different synchronization abilities depending on their reflex skills, which in turn varied by age and gender. This implies that before using this type of interaction, it might be useful for users to use a setup wizard consisting of some trials through which the system can determine the most appropriate rhythmic speed for them.

References

1. Bellino, A. SEQUENCE: A remote control technique to select objects by matching their rhythm. *Pers. Ubiquitous Comput.* **2018**, *22*, 751–770. [CrossRef]
2. Zhang, T.; Yi, X.; Wang, R.; Wang, Y.; Yu, C.; Lu, Y.; Shi, Y. Tap-to-Pair: Associating Wireless Devices with Synchronous Tapping. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2018**, *2*, 1–21. [CrossRef]
3. Zhang, T.; Yi, X.; Wang, R.; Gao, J.; Wang, Y.; Yu, C.; Li, S.; Shi, Y. Facilitating Temporal Synchronous Target Selection through User Behavior Modeling. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2019**, *3*, 1–24. [CrossRef]
4. Reyes, G.; Wu, J.; Juneja, N.; Goldshtein, M.; Edwards, W.K.; Abowd, G.D.; Starner, T. SynchroWatch: One-handed synchronous smartwatch gestures using correlation and magnetic sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2018**, *1*, 1–26. [CrossRef]
5. Velloso, E.; Carter, M.; Newn, J.; Esteves, A.; Clarke, C.; Gellersen, H. Motion correlation: Selecting objects by matching their movement. *ACM Trans. Comput.-Hum. Interact. (TOCHI)* **2017**, *24*, 1–35. [CrossRef]
6. Clarke, C.; Bellino, A.; Esteves, A.; Velloso, E.; Gellersen, H. TraceMatch: A computer vision technique for user input by tracing of animated controls. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany, 12–16 September 2016; pp. 298–303.
7. Clarke, C.; Bellino, A.; Esteves, A.; Gellersen, H. Remote control by body movement in synchrony with orbiting widgets: An evaluation of tracematch. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2017**, *1*, 1–22. [CrossRef]
8. Esteves, A.; Velloso, E.; Bulling, A.; Gellersen, H. Orbits: Gaze interaction for smart watches using smooth pursuit eye movements. In Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, Charlotte, NC, USA, 11–15 November 2015; pp. 457–466.
9. Carter, M.; Velloso, E.; Downs, J.; Sellen, A.; O'Hara, K.; Vetere, F. Pathsync: Multi-user gestural interaction with touchless rhythmic path mimicry. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 3415–3427.
10. Vidal, M.; Bulling, A.; Gellersen, H. Pursuits: Spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Zurich, Switzerland, 8–12 September 2013; pp. 439–448.
11. Maury, S.; Athénes, S.; Chatty, S. Rhythmic menus: Toward interaction based on rhythm. In Proceedings of the CHI'99 Extended Abstracts on Human Factors in Computing Systems, Pittsburgh, PA, USA, 15–20 May 1999; pp. 254–255.

12. Murphy, G.; Groeger, J.A.; Greene, C.M. Twenty years of load theory—Where are we now, and where should we go next? *Psychon. Bull. Rev.* **2016**, *23*, 1316–1340. [CrossRef]

13. Hirst, R.J.; McGovern, D.P.; Setti, A.; Shams, L.; Newell, F.N. What you see is what you hear: Twenty years of research using the Sound-Induced Flash Illusion. *Neurosci. Biobehav. Rev.* **2020**, *118*, 759–774.. [CrossRef]

14. Fozard, J.L.; Vercruyssen, M.; Reynolds, S.L.; Hancock, P.; Quilter, R.E. Age differences and changes in reaction time: The Baltimore Longitudinal Study of Aging. *J. Gerontol.* **1994**, *49*, P179–P189. [CrossRef]

15. Der, G.; Deary, I.J. Age and sex differences in reaction time in adulthood: Results from the United Kingdom Health and Lifestyle Survey. *Psychol. Aging* **2006**, *21*, 62. [CrossRef]

16. Thompson, J.J.; Blair, M.R.; Henrey, A.J. Over the hill at 24: Persistent age-related cognitive-motor decline in reaction times in an ecologically valid video game task begins in early adulthood. *PLoS ONE* **2014**, *9*, e94215. [CrossRef]

17. Ghomi, E.; Faure, G.; Huot, S.; Chapuis, O.; Beaudouin-Lafon, M. Using rhythmic patterns as an input method. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, TX, USA, 5–10 May 2012; pp. 1253–1262.

18. Wobbrock, J.O. Tapsongs: Tapping rhythm-based passwords on a single binary sensor. In Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology, Victoria, BC, Canada, 4–7 October 2009; pp. 93–96.

19. Lin, F.X.; Ashbrook, D.; White, S. RhythmLink: Securely pairing I/O-constrained devices by tapping. In Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, Santa Barbara, CA, USA, 16–19 October 2011; pp. 263–272.

20. Lavie, N.; Tsal, Y. Perceptual load as a major determinant of the locus of selection in visual attention. *Percept. Psychophys.* **1994**, *56*, 183–197. [CrossRef] [PubMed]

21. Lavie, N. Perceptual load as a necessary condition for selective attention. *J. Exp. Psychol. Hum. Percept. Perform.* **1995**, *21*, 451. [CrossRef] [PubMed]

22. Lavie, N. Attention, distraction, and cognitive control under load. *Curr. Dir. Psychol. Sci.* **2010**, *19*, 143–148. [CrossRef]

23. Lavie, N. Distracted and confused?: Selective attention under load. *Trends Cogn. Sci.* **2005**, *9*, 75–82. [CrossRef] [PubMed]

24. Molloy, K.; Griffiths, T.D.; Chait, M.; Lavie, N. Inattentional deafness: Visual load leads to time-specific suppression of auditory evoked responses. *J. Neurosci.* **2015**, *35*, 16046–16054. [CrossRef] [PubMed]

25. Macdonald, J.S.; Lavie, N. Visual perceptual load induces inattentional deafness. *Atten. Percept. Psychophys.* **2011**, *73*, 1780–1789. [CrossRef]

26. Howard, I.P.; Templeton, W.B. *Human Spatial Orientation;* Wiley: London, UK, 1966; p. 360.

27. McGurk, H.; MacDonald, J. Hearing lips and seeing voices. *Nature* **1976**, *264*, 746–748. [CrossRef]

28. Colavita, F.B. Human sensory dominance. *Percept. Psychophys.* **1974**, *16*, 409–412. [CrossRef]

29. Shams, L.; Kamitani, Y.; Shimojo, S. What you see is what you hear. *Nature* **2000**, *408*, 788. [CrossRef]

30. Siegler, I.C. Mental Performance in the Young-Old versus the Old-Old Ilene C. Siegler. *Norm. Aging III Rep. Duke Longitud. Stud.* **1985**, *1975–1984*, 232–237.

31. Tripepi, G.; Jager, K.J.; Dekker, F.W.; Zoccali, C. Stratification for confounding–part 1: The Mantel-Haenszel formula. *Nephron Clin. Pract.* **2010**, *116*, c317–c321. [CrossRef] [PubMed]

32. Gelman, A.; Stern, H. The difference between "significant" and "not significant" is not itself statistically significant. *Am. Stat.* **2006**, *60*, 328–331. [CrossRef]

33. Sullivan, G.M.; Feinn, R. Using effect size—Or why the P value is not enough. *J. Grad. Med. Educ.* **2012**, *4*, 279–282. [CrossRef] [PubMed]