BRIGHTNESS IS MORE EFFICIENT THAN DELAY TO INDUCE WEIGHT PERCEPTION IN AUGMENTED REALITY

Louis Lafuma, Guillaume Bouyer, Jean-Yves Didier

Université Paris-Saclay, Univ Evry, IBISC
34 Rue du Pelvoux, 91080
Évry-Courcouronne, France
`firstname.name@ibisc.fr`

Olivier Goguel

Asobo Studio
6 Rue de la Seigliere, 33800
Bordeaux, France
`ogoguel@asobostudio.com`

ABSTRACT

Multimodal feedback is used to convey various and rich information in virtual environments. It can also change users' perceptions of the haptic properties of objects through the pseudo-haptic illusion. While visual feedback has been extensively examined to induce haptic properties like weight in virtual reality, only a handful of studies have explored the use of audio feedback, and even fewer in augmented reality. Our study aims to extend, in an augmented reality context, previous research findings that used sound to enhance the weight perception of virtual objects. Participants were asked to grab two objects that produced sounds with different audio delays and brightness and to determine which of the two was heavier. The results on 38 participants showed that while the delay did not affect weight perception, the brightness did have a significant impact. This is in line with previous studies and could be used to enhance the perception of hand interactions in AR.

1. INTRODUCTION

In virtual and augmented reality, users manipulate most of the time virtual objects. They may want them to replicate the physical behaviors of their real-world counterparts to induce pseudo-natural immersion. The difficulty of emulating those haptic properties can vary but is usually complex and restrictive because it relies on specific devices that are often limited in use cases or require a constrained context.

Pseudo-haptic illusions are a well-studied field that consists in conveying haptic properties using visual or audio feedback instead of haptic devices. Ujitoko et al. [1] recently published a survey of visual-based pseudo-haptic studies. They noticed an increase in interest in the domain during the last two decades, for example with the work of Lecuyer et al. [2] in 2008 on modifying the visual behaviour of a cursor to simulate bumps and holes.

Although auditory pseudo-haptic has received some attention, it has yet to be examined to the same extent as its visual counterpart. Several haptic features have already been simulated with sounds: stiffness [3], temperature [4], and weight [5]. We aim to contribute to this field by focusing on the auditory stimulation of the perception of weight in augmented reality, which could assist the manipulation of virtual objects.

Our goals are as follows:

- Study if we can convey weight with the use of audio in augmented reality
- Determine which auditory feature helps to better perceive weight differences

In order to do this, we have first examined related work on audio feedback and weight perception in virtual or augmented reality to identify audio cues and parameters that could convey the sensation of weight (Section 2). Then we have developed an application based on Microsoft Hololens and specific auditory feedback to sonify the manipulation of virtual objects (Section 2.3). This setup has been evaluated in a user study wherein participants had to choose which of two sonified objects felt heavier when handled (Section 3). We present and analyse our results in Section 4 and discuss them in Section 5 before stating the main conclusions and the possible perspectives of our work.

2. RELATED WORK

2.1. Audio feedback in a virtual environment

The use of audio to convey information, also called auditory display, is a vast field of study that still has much to research. Rocchesso et al. [6] explained how a real object can convey informations about his properties with sound. An in-depth view of the field is presented by Walker et al. [7], which gives four broad categories to describe the function of auditory display, applicable to virtual environments.

The first one is alarms, alerts, and warnings. Alerts are widely used in Virtual Reality and Augmented Reality to inform users that they have interacted with the system. Ariza et al. [8] provide the information that users have touched an object by playing a sound, and those kinds of displays are used to inform users that they have grabbed, or even released an object [9][10][11]. Lugrin et al. [12] give audio feedback when users and their avatar perform an action to see if it can help to give an illusion of virtual ownership. Kang et al. [3] use auditory cues when the user manipulates a cylindric picker to augment the feeling of roughness of the object. Alarms and warnings are types of alerts that are constrained to specific events that carry particular urgency. It is used in addition to visual feedback [4] to warn the users that their hands are burning. Dix et al. [13] modify the alarm signal given by a wire and loop game, where a sound is emitted when the loop touches the wire, by adding a delay on the visual and audio feedback to see if it impacts the user's efficiency.

The second category of auditory display is status, process, and monitoring messages. Status feedback can be used to inform users that they are getting close to an object or even that they are touching it [8], or more globally that they are in the right area and their movement is correct [14]. Haptic properties can be considered as status information, therefore those kinds of auditory displays can be used to convey the stiffness of a cube [3], or the weight, either by adding a sound coming from the cube [15] or by using physiological audio feedback, like heartbeats [5]. Heartbeats as status audio feedback are also used to increase heart rate and anxiety level [16], and other physiological properties can be modified by using different filters on footstep sounds [17]. Status sounds can also be spatialized, giving users information about their surroundings and allowing better collaboration between two remote users [18].

The third category of auditory display is data exploration, which is specific to the transmission of data set with sound.

The last category is art, entertainment, sports, and exercise, as the work from Glushkova et al. [14].

2.2. Modifying weight perception in a virtual environment

Virtual objects, in their simplest form, do not transmit haptic sensations. Real haptic rendering requires a physical device, which can be restrictive or tedious to use, especially in the context of augmented reality, involving mobile or large-scale interactions. Pseudo-haptic feedback is another mean to induce haptic properties using visual or audio feedback [1]. It is less constraining because there is not any physical apparatus, but the effects on perception are less reliable.

As mentioned in the introduction, there are numerous haptic properties that can be simulated. An interesting property to emulate in the context of AR manipulation is the weight. Much research has been conducted using visuals to alter weight perception. A common pseudo-haptic technique is manipulating the C/D ratio, the difference between the user's hand movement and representation in the virtual space. Rietzler et al. [19] and Samad et al. [20] add an offset between the hand displayed in VR and the real one to induce weight to a bowling ball and a small cube. Taima et al. [21] do the same thing, but in video see-through AR, they use the fact that the hand is displayed through a screen to modify its position. Jauregui et al. [22] also use video see-through AR to induce weight by modifying the reflected avatar of the user on the screen while lifting a dumbbell. It is also possible to combine C/D ratio modification to haptic feedback, like in [23], where a physical ball is lifted by the participants while the display of this ball is changed, or in [24] where the C/D ratio modification is deported from the virtual object with a weight to a real-world effector manipulated by the user. C/D ratio modification can also be mixed with other types of feedback like visual vibration, change of color of the hand, and audio feedback [5]. Herbst et al. [15] do not use C/D ratio modifications but mix haptic vibrations, object color, and audio feedback to convey weight.

Another commonly used visual feedback is the modification of the size of the object. Heineken et al. [25] display different-sized cylinders in VR or AR without changing their weight to see if visuals can impact the feeling of weight, and Kim et al. [26] do the same thing but with purely virtual cubes in VR. Omosako et al. [27] use the object's size to change a briefcase's perceived center of gravity in VR with a physical double. Some exotic visuals convey weight, like displaying a moving liquid inside the object [28] or using a separate visual indicator to convey weight information [29].

Audio can be another pseudo-haptic way to convey weight. By modifying the footstep sound of users when they walk, they can feel as if their weight is different [17]. Takashima et al. [30] use sounds with different loudness and pitch when the participant is wielding an object to modify the perceived weight, while Kaneko et al. [31] try to modify the pitch, the delay, and the volume of a sound played when the user clicks on a button to add a sensation of weight to it.

Another way to give weight to a virtual object is to play on the detection of the grab by the user, like changing the force or the finger distance needed to grab the object to give the user the impression that the object is heavier by forcing it to exert more effort [32].

2.3. Auditory parameters for weight simulation

Several parameters or processing options are available for modifying an audio signal. In previous works to induce weight with audio feedback, the parameters used were filters on audio frequency [17], loudness and pitch [30], or the addition of delay [31]. In the results of those studies, it seemed that loudness is not a suitable parameter to induce weight, as it already has some relation to other properties of the object, like its distance to the user. The results also showed no significant link between a change in loudness and a change in weight perception.

After considering the promising results showcased by both [30] and [31] regarding the significant impact on perceived weight, we opted to incorporate the pitch as a variable parameter in our study. Additionally, [31] highlighted the efficacy of utilizing delay to induce weight, leading us to include it as a second parameter. While filters also appeared interesting, we ultimately refrained from using them as an experimental variable to focus on two parameters in the experimental study: pitch and delay.

To choose the values of those parameters, we did a preliminary experiment based on the previous studies. For the pitch, Kaneko et al. [31] use a base sound with a frequency of 300 Hz, a lower pitched frequency of 200 Hz (which is around -33%), and a higher pitch frequency of 400 Hz (which is around +33%). Our base sound being a generated noise with a base spectral centroid of **500 Hz**, it was rather a modification of brightness than a modification of pitch. We tried different variations on the spectral centroid to see which one an average human ear can detect. While a variation as small as +-5% can be detected, we choose **+-15%** to ensure that users can hear the difference while still being small enough not to change the nature of the sound. Thus our low-brightness spectral centroid is **425 Hz** and our high-brightness spectral centroid is **575 Hz**.

For the delay, Kaneko et al. [31] use 5 values: 0, 100, 200, 300, and 500 ms, even if some of those values decrease the sense of agency between the user's action and its effect on the environment. Toida et al. [33] studied 5 different delay intervals: 19 to 253 ms, 119 to 353 ms, 186 to 419 ms, 19 to 119 ms, and 286 to 519 ms. They found out that the average Delay Detection Threshold (DDT) among participants, which is the delay after which 50% of the population think there is a discrepancy between their actions and the sound produced, was respectively for the first 4 intervals 136.3±31.6, 208.9±31.5, 309.1±45.6, 89.1±13.4. They also computed the average Just Noticeable Difference (JND), which is half the difference between the lower (25%) and upper (75%) bounds

Figure 1: A participant uses our experimental setup. The right cube has been lifted and maintained up. It need to be released on the table before doing the same with the left one.

of the threshold. They found values of 18.7±11.8, 28.7±14.4, 30.8±7.8 and 18.2±4.5 for the first 4 intervals, respectively. From those values, we decided to vary the delay by **intervals of 50 ms** to have variations over the higher JND plus the uncertainty, which would be 28.7+14.4=43.1 ms. Starting from no delay, we ended with delays varying from **50 ms to 250 ms**, giving us our 5 different delay values. Our pre-tests showed that when a delay of 250 ms is almost always detected, a delay of 50 ms is very hard to detect. However, we kept those values to see if the delay can influence the weight perception even if the participant does not consciously detect it.

3. EXPERIMENTAL EVALUATION

3.1. Design

Our experimental study aimed to evaluate the effect of audio brightness and delay on weight perception in a grasping task in augmented reality by asking participants to lift different objects while hearing a sound varying with their movement, as in [30], but with an emphasis on manipulation (Fig. 1). We used a between-subjects design, where participants performed 68 attempts for 6 different delays and 3 different spectral centroid. The experiment was divided into 4 series, going over all combinations in a randomized order while assuring that all combinations will be presented the same way equally.

3.2. Participants

38 people (5 female, 33 male) aged between 22 and 47 participated in the study—33 of them were right-handed, 4 were left-handed, and 1 was ambidextrous. One of our participants did not speak French, but we adapted all our questionnaires to be available in English. All of our participants filled in an informed consent form. Each participation lasted around 30 minutes.

3.3. Setup

The AR headset was a Microsoft HoloLens 2. A computer with an Intel i7-9750H processor and an Nvidia GeForce 2080 graphics card was connected to the headset to be able to reset a task if needed and to provide the questionnaires at the end of the experiment. The AR application was developed using the Unity engine. Hand tracking relied on the headset's sensors, and QR codes ensured all participants had the exact same virtual object placement. The audio feedback was provided with the built-in speakers of the HoloLens 2.

3.4. Stimuli

The sound emitted while moving the object was generated using free and open-source software [34], specifically a pre-configured SFX called SF Wind. It is a white noise with a spectral centroid value of around 500 Hz, modified by a band-pass filter that remains constant throughout the experiment, to make it sound like wind noise. In the base sound, there was also a sinusoidal variation of the volume, as well as a reverb and a distortion effect, that we removed. In our experiment, we vary the volume linearly depending on the object's speed in m/s, reaching its max value at a speed of 1 m/s and 0 when the object is not moving. The participant sets the built-in speaker volume value during the training task to be easy to hear but not too loud. The audio feedback is continuous while the user is moving the object. The sound has been tuned to make the user feel as the weight of the object was generating some opposition from the wind.

The first audio feature that we decided to vary is the brightness. We used 3 different spectral centroid, one being the base spectral centroid of our audio (**sc_base**) and the others being the base spectral centroid minus 15% (**sc_low**) and plus 15% (**sc_high**), as explained in Section 2. The brightness is set at the beginning of each trial, and while the volume will vary throughout the trial, the brightness will not.

The second audio feature that we varied is the delay between the variation of the speed of the object and the variation of the volume of the sound. We used 5 different delays, from 50 ms with a 50 ms step according to section 2, in addition to the baseline without any delay. As for the brightness, we set the delay at the beginning of each trial and do not change it until the beginning of the next.

Combining the 3 different brightness with the 6 different delays leads us to 3 x 6 = 18 different combinations, including the baseline combination with an unchanged brightness and no delay.

3.5. Procedure

3.5.1. Setup

In order to place the holograms in the same place for all participants, the application is automatically calibrated by scanning two QR codes placed on the table using the headset's cameras.

3.5.2. Grasping task

The task consists of grasping an object located approximately 50 cm away on one side of the user, lifting it over an indicator located 35 cm over the table for a duration of 400 ms, of depositing it back on the table, and then doing it again on a second object placed 30 cm to the other side. After that, the user has to say which cube is heavier by pressing a virtual button. Then, two other cubes appear, and the subsequent trial begins. The participant is asked to use the same hand throughout the experiment and to limit to one up-and-down movement to be sure that every participant have the same test conditions. The training trial contains two identical cubes with the baseline combination. It can be done a second time on demand, for example to adjust the speaker volume. After that, the participant is asked to perform the task 68 times, divided into 4 series of 17 trials, each comparing one of the 17 combinations with the baseline one. The order of each trial within each series and the location of the baseline object between left and right are randomized, while still assuring that each case is proposed an equal number of time.

3.5.3. *Questionnaires*

At the end of all the trials, the participants fill in two questionnaires. The experiment questionnaire is intended to: know if the participants noticed the auditory variations between the two objects, detect their ability to hear the difference between three audio recordings with the 3 brightness, and know which variations they preferred between brightness and delay. The demographic questionnaire contains some additional questions about the musical habits of the participants. Finally, we have collected their comments.

3.6. Measures

Several spatial and temporal quantitative data were measured:

- The elapsed time between each step of the task (appearance of the objects, grabbing of each object, reaching the height indicator, putting the object back on the table, and pressing the answer button)
- The objects' positions during the manipulation.

Qualitative data are the users' 68 choices and the responses to the questionnaires.

3.7. Hypotheses

Based on earlier studies, we formed the following hypotheses:

- H1: changing the brightness of the audio should influence the perceived weight of the grabbed object.
- H2: higher brightness of the audio should reduce the perceived weight of the grabbed object.
- H3: lower brightness of the audio should increase the perceived weight of the grabbed object.
- H4: adding a delay to the change in audio volume should influence the perceived weight of the grabbed object.
- H5: higher delay should increase the perceived weight of the grabbed object
- H6: the audio delay should sound as instinctive as the pitch to represent the weight of the grabbed object.

Kaneko et al. [31] showed that the pitch of the sound can impact weight perception. Takashima et al. [30] showed similar results, in particular, an increase in the perception of weight when the pitch decreases. Based on those works, we expect a decrease/increase of the perceived weight with an increase/decrease of the spectral centroid (**H1** and **H2/H3**).

Kaneko et al. [31] also showed that adding a delay to the sound produced by the participants clicking on a button can increase the heaviness sensation, so we hypothesize that delay should make the object feel heavier (**H4** and **H5**).

Finally, varying two auditory parameters at the same time allows us to emit the comparative hypotheses that both parameters should feel equally instinctive to the user (**H6**).

Table 1: Mean choice values of the 4 series for each user and combination. A value close to 1 mean that users tend to perceive the modified combination to be heavier than the baseline.

	d1	d2	d3	d4	d5	d6	All delay mean
sc_base	NA	0.52	0.45	0.46	0.5	0.43	0.47
sc_low	0.80	0.80	0.77	0.74	0.74	0.76	0.77
sc_high	0.22	0.22	0.18	0.17	0.18	0.22	0.19
All brightness mean	0.51	0.51	0.46	0.46	0.47	0.47	0.48

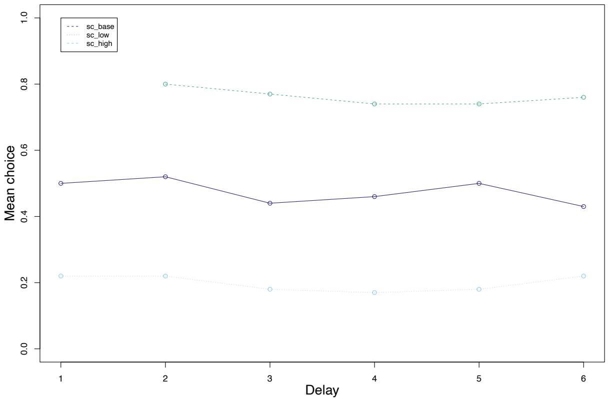

Figure 2: The mean choice of all users for each combination.

4. RESULTS

4.1. Weight perception

To study the weight perception for the different combinations, we assigned a value of 0 if the participant chose the baseline combination and 1 if the modified combination was chosen. We then computed the mean of the 4 series for each user and combination, worth 0, 0.25, 0.5, 0.75, or 1. To visualize a possible link between our two variables, we computed the average for all users for each combination before further analysis (Table 1 and Figure 2).

We proceeded to a two-way repeated measures ANOVA on the data per user to study simultaneously the effect of the delay and the brightness on the user's weight perception. In order to do that, we had a data collection containing the mean between our 4 series for each participant and each combination. We did not compare our baseline (base brightness and no delay) with itself because the participants did not have the option to say that both objects have the same weight. Therefore, we can not compare our combinations with no delay, so we removed all combinations with no delay from our tests to be able to perform our ANOVA (Figure 3).

The ANOVA gave both the interaction effect and the two variables effects. When the brightness has significant effect (F(2,540) = 279.99 $p < 0.0001$, n2g = 0.509), the delay does not (F(4,540) = 1, p = 0.41, n2g = 0.007) and there is no interaction effect (F(8,540) = 0.42, p = 0.91, n2g = 0.006).

On the effect of brightness, we performed a pairwise t-test on our data, and we found significant effects between all different

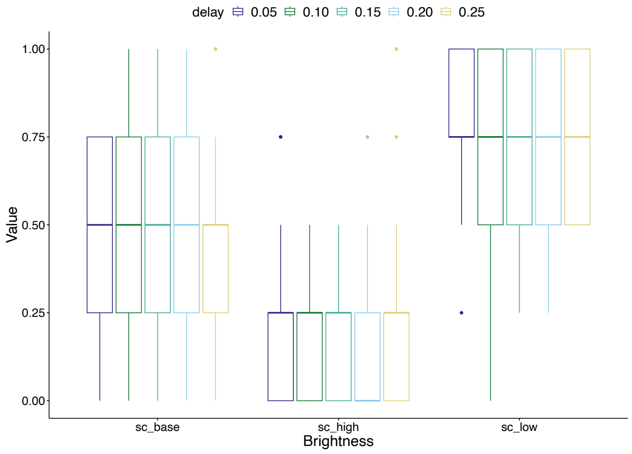

Figure 3: Box plot of the choices for each combination between all users.

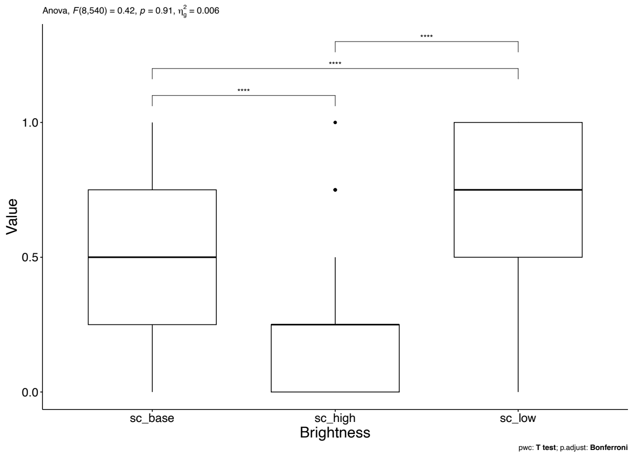

Figure 4: Box plot of the choices for each brightness between all users.

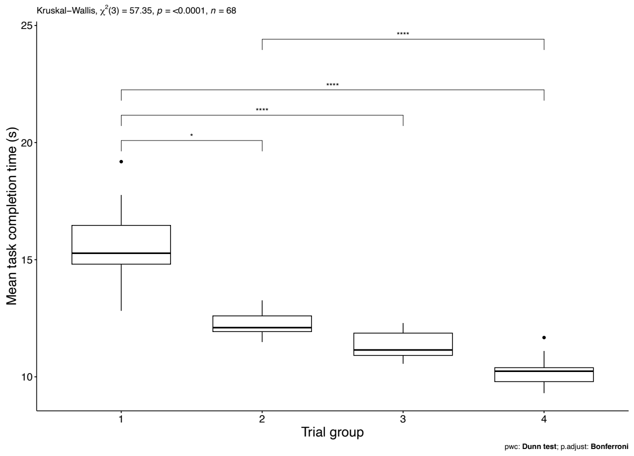

Figure 5: Box plot of the mean task completion time for each trial series.

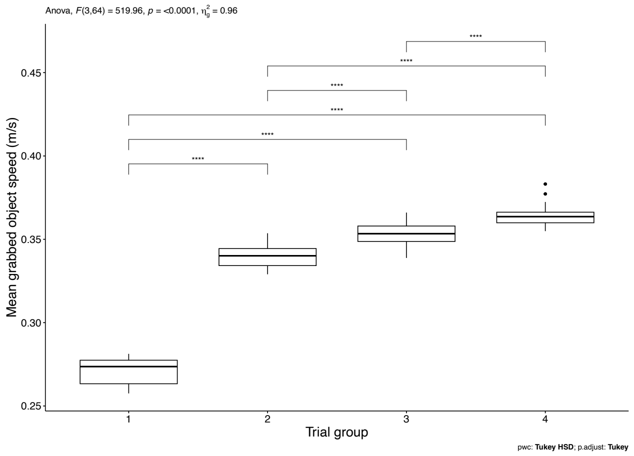

Figure 6: Box plot of the mean manipulation speed for each trial series.

brightness combinations ($p < 0.0001$). We found that the choice rate of the sc_high group was significantly lower than the rate of the other two groups and that the choice rate of the sc_low group was significantly higher than that of the other two groups (Figure 4). In the experiment, it means that participants tend to find the object with the modified combination to be lighter than the baseline more often in all the combinations with the higher brightness, and the opposite with the combinations with the smaller brightness.

As for the delay and the interaction effect, there were no significant results. The results for each different delays combinations were about the same, with no significant differences between the 3 different brightness.

4.2. Temporal data

We used the temporal data to see if there was a training effect during the experiment. We computed the time between the apparition of the objects and the moment when the user leaves the second object on the table. After combining all the data and grouping them by trial series number, we verified the assumptions needed to proceed with an ANOVA to study the effect of the trial number on the experimentation time. There were not any extreme outliers, but the normality assumption was not met. We decided to proceed with a Kruskal-Wallis Test instead, and we observed a large size effect between the first trial, with a mean of 15.6 seconds, and trials 2, 3, and 4, with means of respectively 12.3, 11.4, and 10.2 seconds (Figure 5). We can see a training effect, with the mean time to perform the task decreasing as the tests progress.

4.3. Spatial Data

We used the position data to confirm the training effect observed with the temporal data. We computed the mean speed of the two grabbed objects for each trial, combined all the data, and grouped them by series number. There were no extreme outliers, and the normality and variance homogeneity assumptions were met, so we proceeded with an ANOVA to study the effect of the trial number on the mean speed during manipulation. The results were signif-

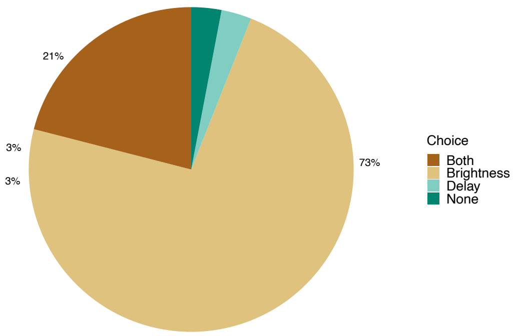

Figure 7: Pie plot of the most significant parameter to differentiate weight according to the participants.

icant (F(3.64) = 309.87, p < 0.0001, n2g = 0.94), and we can see that the mean speed of the objects is increasing as we go along the experiment, with a significant difference between each trial series (Figure 6).

4.4. Questionnaires

When asked to evaluate the ease of weight detection on a scale from 1 to 5, the average score was 2.76, slightly under the expected mean which is 3. We then asked if they detected variations in the sound between the objects, and none of the participants identified the variation of delay. 34 of the 38 participants identified the brightness variation, and the other 4 perceived unexpected variations such as the object's content and the strength of the wind blowing on the object. After being informed of the two parameters that varied during the experiment, participants were asked which one they found the most significant to differentiate between the weight of the two cubes and which one they found the most natural to represent the weight. To the former question, 73% of the participants found that the brightness was more significant than the delay, 21% found that the two were identically significant, 3% that the delay was more significant than the brightness, and 3% that none were significant (Figure 7). To the second question, 61% of the participants found the brightness to be more instinctive than the delay, 26% found that both were equally instinctive, 8% found the delay to be more instinctive than the brightness, and 5% found that none were instinctive to perceive weight. No candidates found the two variations uninstinctive and insignificant in a weight perception task.

5. DISCUSSION

Based on the weight perception results, it is clear that altering the brightness of the emitted sound during object movement directly impacts the participant's perceived weight. Lowering the brightness leads to a heavier sensation, with over three-quarters of participants reporting that the object felt heavier than the baseline. Conversely, raising the brightness results in a lighter sensation, with over three-quarters of participants reporting that the object felt lighter than the baseline. Those results **validate H1** and, more specifically, **H2 and H3**, following the results of [30] and [31], who found that a lower (resp. higher) pitch increases (resp. decrease) the perceived heaviness of real objects.

Considering our findings, it would be beneficial to modify the brightness of a sound in order to alter how the weight of the object is perceived. While we used a complex sound and not a sinusoidal pure-tone, and even if the manipulated object do not usually produce sound, we suggest to use sound with different brightness to induce weight on an object.

There was no significant difference between the 6 combinations for the delay, as participants does not tend to answer differently between the different delay combination inside each one of the 3 pitch conditions. Therefore, we **cannot validate H4 and H5**, which does not support the results from [31] showing that the addition of an offset to the sound can increase the sensation of heaviness to a real button.

One possible explanation is that our value intervals were too small to produce an effect. We wanted the delay to be short enough to ensure the users did not feel the sound was dissociated from their actions but long enough to impact them. A longer delay may have produced better results in terms of weight perception without giving a feeling of disconnection. The comments on the experiment show that the delay gave some participants the feeling that the visual representation is moving differently. With better parameters, audio delay could be as effective as the pitch and more natural for some users because it is much more subtle. Another limit of these results could also be the slow attack, which is not present in previous work, making the delay hard to detect.

Concerning the experiment questionnaire, the average score of the ease of detection means that it was hard to differentiate the two objects. This result is likely because the delay was not well identified, so it was challenging to choose when two objects with the same brightness were presented. We can observe that even if no participant noticed that there was sometimes a delay between their movements and the variation of the sound, 9 of them found the delay to be at least as significant as the brightness to compare the weight of the objects, and even more participants (13 of them) found the delay to be at least as instinctive as the brightness to feel the weight of the objects. However, we do not have equity between the brightness and the delay in terms of how instinctive the weight perception is; consequently, it is **impossible to validate H6**.

When we analyse the temporal data, we can see that the more the participants progress in the experiment, the less time it takes them to complete the manipulation of the object, with a particularly significant gap between the first and the second trial. Those results are confirmed with the mean speed data, where the mean speed of the grabbed objects increases as we progress in the experiment, the differences between the 4 series being all significant. Those changes can be due to a training effect, the participants being more accustomed to the task and more efficient. Because weight perception results did not change between the different series, this augmentation in speed did not decrease the user's performance, meaning that the participant can adapt quickly and correctly to the task and the audio feedback.

We did some statistical analysis on other data, like the maximum height reached by the object during each task, and we tried to split our population according to different criteria, like the music-listening habits of the participants. None of those analysis gave significant results.

After analysing the results of the questionnaire in relation to the temporal data, we can observe that even if users found that it was not easy to evaluate the weight of the objects, they tended to be more efficient as the experiment went on. It could mean that as they go along the experiment, they will start to answer more randomly and quickly, but we would not have significant results on the brightness if that were the case. It could also mean that participants tend to acknowledge the fact that a lower brightness means that the object is heavier and will instinctively answer quickly when the brightness is different between the two objects, basing their reflection on audio rather than weight, but even if it is the case, we did not indicate what was expected about the brightness parameter in the experiment, so it means that it is a conclusion that feels instinctive for the participant.

In the free comment section of the experiment questionnaire, some participants mentioned that they felt that the audio was not the only thing to vary between the two objects, with "the impression from time to time that the cubes were not moving in the same way" or a participant being "surprised that only two parameters change," and having "the impression of having more variation between the different cubes," than only the delay and the brightness. One participant said it was pleasing that "you can shake the cube and feel like something is moving inside." A few participants commented that they felt they had improved during the experiment, that their "perception of weight improved as the tests progressed," or that they "felt the differences better in the middle of the experiment." Finally, a participant told us that when the time came to choose between the two objects, "he could not remember the visual difference at all, but he did remember the auditory difference." Even if there was no visual difference in our experiment, the audio modality seemed easier to memorize than the visual one for this participant.

In brief, the two main results of the study are the following :

- modifying the brightness of the sound of an object can influence the perception of its weight
- Users tend to get used to the audio feedback and to be more efficient without performance loss

Our experiment showed that participants' efficiency increased over time as they became more familiar with audio feedback without affecting their performance. This observation suggests that conveying weight through audio feedback feels instinctive and is easily understood by users.

The limitations and possible extensions of the study include modifying the band-pass filter, as in [17], using another basic sound, or modifying the listening environment by using headphones or adding background noise, as suggested in [31], to change the way the user perceives things.

6. CONCLUSION

In this paper, we investigated whether we could induce weight during the manipulation of virtual objects in AR using different sound parameters, according to existing results on the use of sound to provide pseudo-haptic properties in a multimodal virtual environment. Our base sound is a wind sound whose volume varies according to the speed of the object, simulating the fact that the object creates a sound as it moves through the air. The brightness and delay of this sound with the movement of the object are then modified to study whether this can have an impact on its perceived weight.

Our experiment involved 38 participants, mostly used to virtual and augmented reality technology. The results showed that the brightness impacts the feeling of weight, whereas the delay has no effect. People also tend to be more efficient over time while still being consistent in their weight perception. Given those results, a guideline for AR designers would be to add a neutral sound when manipulating virtual objects and then set various spectral centroid to induce weight differences. Users will likely adapt to the audio feedback even if those objects are not supposed to produce sound.

Even if our results on the impact of delay on the perception of weight were insignificant, greater delays could produce more relevant results. It would also be interesting to study other audio characteristics, such as the type of sound used or the use of filters, and to work on other haptic properties, such as roughness or texture.

7. REFERENCES

[1] Y. Ujitoko and Y. Ban, “Survey of pseudo-haptics: Haptic feedback design and application proposals,” *IEEE Transactions on Haptics*, vol. 14, no. 4, pp. 699–711, 2021.

[2] A. Lécuyer, J.-M. Burkhardt, and C.-H. Tan, “A study of the modification of the speed and size of the cursor for simulating pseudo-haptic bumps and holes,” *ACM Transactions on Applied Perception*, vol. 5, no. 3, pp. 14:1–14:21, Sept. 2008.

[3] N. Kang, Y. J. Sah, and S. Lee, “Effects of visual and auditory cues on haptic illusions for active and passive touches in mixed reality,” *International Journal of Human-Computer Studies*, vol. 150, p. 102613, June 2021.

[4] D. Eckhoff, A. Cassinelli, T. Liu, and C. Sandor, “Psychophysical Effects of Experiencing Burning Hands in Augmented Reality,” in *Virtual Reality and Augmented Reality*, P. Bourdot, V. Interrante, R. Kopper, A.-H. Olivier, H. Saito, and G. Zachmann, Eds. Cham: Springer International Publishing, 2020, vol. 12499, pp. 83–95.

[5] Y. Hirao and T. Kawai, “Augmented Cross-modality: Translating the Physiological Responses, Knowledge and Impression to Audio-visual Information in Virtual Reality,” *Journal of Imaging Science and Technology*, vol. 62, no. 6, Nov. 2018.

[6] D. Rocchesso, R. Bresin, and M. Fernstrom, “Sounding objects,” *IEEE MultiMedia*, vol. 10, no. 2, pp. 42–52, 2003.

[7] B. N. Walker and M. A. Nees, “Theory of sonification,” in *The sonification handbook*. Logos Verlag Berlin, 2011, vol. 1, pp. 9–39.

[8] O. Ariza, G. Bruder, N. Katzakis, and F. Steinicke, “Analysis of Proximity-Based Multimodal Feedback for 3D Selection in Immersive Virtual Environments,” in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, Tuebingen/Reutlingen, Germany, Mar. 2018, pp. 327–334.

[9] R. Canales and S. Jörg, “Performance Is Not Everything: Audio Feedback Preferred Over Visual Feedback for Grasping Task in Virtual Reality,” in *Motion, Interaction and Games*, ser. MIG ’20. New York, NY, USA: Association for Computing Machinery, Oct. 2020, pp. 1–6.

[10] M. A. Zahariev and C. L. MacKenzie, “Auditory, graphical and haptic contact cues for a reach, grasp, and place task in an augmented environment,” in *Proceedings of the 5th international conference on Multimodal interfaces*, Vancouver British Columbia Canada, 2003, pp. 273–276.

[11] M. P. Furmanek, M. Mangalam, K. Lockwood, A. Smith, M. Yarossi, and E. Tunik, “Effects of Sensory Feedback and Collider Size on Reach-to-Grasp Coordination in Haptic-Free Virtual Reality,” *Frontiers in Virtual Reality*, vol. 2, p. 648529, Aug. 2021.

[12] J.-L. Lugrin, D. Obremski, D. Roth, and M. E. Latoschik, “Audio feedback and illusion of virtual body ownership in mixed reality,” in *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, ser. VRST ’16. New York, NY, USA: Association for Computing Machinery, Nov. 2016, pp. 309–310.

[13] A. Dix, C. S. Arlinghaus, A. M. Harkin, and S. Pannasch, “The Role of Audiovisual Feedback Delays and Bimodal Congruency for Visuomotor Performance in Human-Machine Interaction,” in *Proceedings of the 25th International Conference on Multimodal Interaction*, ser. ICMI ’23. New York, NY, USA: Association for Computing Machinery, Oct. 2023, pp. 555–563.

[14] A. Glushkova, D. Makrygiannis, and S. Manitsaris, “Embodied edutainment experience in a museum: Discovering glass-blowing gestures,” in *Companion Publication of the 25th International Conference on Multimodal Interaction*, ser. ICMI ’23 Companion. New York, NY, USA: Association for Computing Machinery, Oct. 2023, pp. 288–291.

[15] I. Herbst and J. Stark, “Comparing force magnitudes by means of vibro-tactile, auditory, and visual feedback,” in *IEEE International Workshop on Haptic Audio Visual Environments and Their Applications*, Oct. 2005, pp. 5 pp.–.

[16] R. Wang, H. Zhang, S. Macdonald, and P. Di Campli San Vito, “Increasing Heart Rate and Anxiety Level with Vibrotactile and Audio Presentation of Fast Heartbeat,” in *Proceedings of the 25th International Conference on Multimodal Interaction*, Paris France, Oct. 2023.

[17] E. Sikström, A. de Götzen, and S. Serafin, “Self-characterstics and sound in immersive virtual reality — Estimating avatar weight from footstep sounds,” in *2015 IEEE Virtual Reality (VR)*, Arles, France, Mar. 2015, pp. 283–284.

[18] J. Yang, P. Sasikumar, H. Bai, A. Barde, G. Sörös, and M. Billinghurst, “The effects of spatial auditory and visual cues on mixed reality remote collaboration,” *Journal on Multimodal User Interfaces*, vol. 14, no. 4, pp. 337–352, Dec. 2020.

[19] M. Rietzler, F. Geiselhart, J. Gugenheimer, and E. Rukzio, “Breaking the Tracking: Enabling Weight Perception using Perceivable Tracking Offsets,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Montreal QC Canada: ACM, Apr. 2018, pp. 1–12.

[20] M. Samad, E. Gatti, A. Hermes, H. Benko, and C. Parise, “Pseudo-Haptic Weight: Changing the Perceived Weight of Virtual Objects By Manipulating Control-Display Ratio,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow Scotland Uk: ACM, May 2019, pp. 1–13.

[21] Y. Taima, Y. Ban, T. Narumi, T. Tanikawa, and M. Hirose, “Controlling fatigue while lifting objects using Pseudo-haptics in a mixed reality space,” in *2014 IEEE Haptics Symposium (HAPTICS)*, Houston, TX, USA, Feb. 2014, pp. 175–180.

[22] D. A. Gomez Jauregui, F. Argelaguet, A.-H. Olivier, M. Marchal, F. Multon, and A. Lecuyer, “Toward "Pseudo-Haptic Avatars": Modifying the Visual Animation of Self-Avatar Can Simulate the Perception of Weight Lifting,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 4, pp. 654–661, Apr. 2014.

[23] L. Dominjon, A. Lecuyer, J. Burkhardt, P. Richard, and S. Richir, “Influence of control/display ratio on the perception of mass of manipulated objects in virtual environments,” in *IEEE Proceedings. VR 2005. Virtual Reality, 2005.* Bonn, Germany: IEEE, Mar. 2005, pp. 19–318.

[24] P. Issartel, F. Guéniat, S. Coquillart, and M. Ammi, “Perceiving Mass in Mixed Reality through Pseudo-Haptic Rendering of Newton’s Third Law,” *2015 IEEE Virtual Reality (VR)*, pp. 41–46, Mar. 2015.

[25] E. Heineken and F. P. Schulte, “Seeing size and feeling weight: The size-weight illusion in natural and virtual reality,” *Human factors*, vol. 49, no. 1, pp. 136–144, 2007.

[26] J. Kim and J. Lee, “The Effect of the Virtual Object Size on Weight Perception Augmented with Pseudo-Haptic Feedback,” in *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, Lisbon, Portugal, Mar. 2021, pp. 575–576.

[27] H. Omosako, A. Kimura, F. Shibata, and H. Tamura, “Shape-COG Illusion: Psychophysical influence on center-of-gravity perception by mixed-reality visual stimulation,” in *2012 IEEE Virtual Reality (VR)*. Costa Mesa, CA, USA: IEEE, Mar. 2012, pp. 65–66.

[28] Y. Kataoka, S. Hashiguchi, F. Shibata, and A. Kimura, “Rv dynamics illusion: psychophysical phenomenon caused by the difference between dynamics of real object and virtual object,” in *Proceedings of the 25th International Conference on Artificial Reality and Telexistence and 20th Eurographics Symposium on Virtual Environments*, Kyoto Japan, Oct. 2015, pp. 133–140.

[29] J. Lee, J.-I. Kim, and H. Kim, “Force Arrow 2: A Novel Pseudo-Haptic Interface for Weight Perception in Lifting Virtual Objects,” in *2019 IEEE International Conference on Big Data and Smart Computing (BigComp)*, Kyoto, Japan, Feb. 2019, pp. 1–8.

[30] M. Takashima, “Perceived Weight Is Affected by Auditory Pitch Not Loudness,” *Perception*, vol. 47, no. 12, pp. 1196–1199, Dec. 2018.

[31] S. Kaneko, T. Yokosaka, H. Kajimoto, and T. Kawabe, “A Pseudo-Haptic Method Using Auditory Feedback: The Role of Delay, Frequency, and Loudness of Auditory Feedback in Response to a User’s Button Click in Causing a Sensation of Heaviness,” *IEEE Access*, vol. 10, pp. 50 008–50 022, 2022.

[32] J. Hummel, J. Dodiya, R. Wolff, A. Gerndt, and T. Kuhlen, “An evaluation of two simple methods for representing heaviness in immersive virtual environments,” in *2013 IEEE Symposium on 3D User Interfaces (3DUI)*, Orlando, FL, USA, Mar. 2013, pp. 87–94.

[33] K. Toida, K. Ueno, and S. Shimada, “Recalibration of subjective simultaneity between self-generated movement and delayed auditory feedback,” *NeuroReport*, vol. 25, no. 5, pp. 284–288, Mar. 2014.

[34] “Helm - Free Synth by Matt Tytel,” https://tytel.org/helm/.