TILBURG UNIVERSITY

The effect of virtual reality on memory retention using an avatar

Eveline Aurélie Janssen

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
BACHELOR OF SCIENCE IN COGNITIVE SCIENCE AND ARTIFICIAL INTELLIGENCE
AT THE SCHOOL OF HUMANITIES AND DIGITAL SCIENCES OF TILBURG UNIVERSITY

STUDENT NUMBER

2038657

COMMITTEE

Thesis supervisor : Phillip Brown

Second reader : Lisa Lepp

LOCATION

Tilburg University

School of Humanities and Digital Sciences

Department of Cognitive Science & Artificial Intelligence

Tilburg, The Netherlands

DATE

May 21, 2024

WORD COUNT

6376

ACKNOWLEDGMENTS

This research project has been conducted in order to fulfill the requirements of the bachelor's program Cognitive Science and Artificial Intelligence at Tilburg University.

I am truly grateful for my supervisor, Phillip Brown, for his outstanding guidance and support throughout this journey. His optimism and trust enormously helped me with the research process.

I would also like to thank my fellow peer students for their support and motivation. Additionally, I extend my gratitude to my friends and family, who have always been there for me, providing support and believing in me.

The effect of virtual reality on memory retention using an avatar

Eveline Aurélie Janssen

Abstract

With the rapid development of artificial intelligence (AI), technologies such as virtual reality (VR) have been widely used in different fields such as healthcare, education and entertainment. VR does not only offer immersive environments, but also the application of social avatars, which results in a more advanced human-computer interaction. However, there is limited understanding of the effects of an avatar, in comparison to voice assistants, which are more frequently used. These voice assistants apply speech-to-speech communication, commonly featuring a female voice as default. This study examined the impact of VR, involving the presence of an avatar, on memory retention in combination with evaluating the influence of voice gender. A within-subject design is utilized, where participants (n = 49) were instructed to memorize four wordlists containing 30 random English words. Each wordlist was presented under different conditions: either in a virtual environment with an avatar or in a traditional setting without VR and spoken by either a female or male voice. A short delay was followed by the recall task, where participants recalled as many words as possible from the presented wordlist. The result showed no significance difference between the use of VR or not, nor between the male and female voice. Future work is recommended to implement more realistic avatars and to reconsider the duration of the delay in between memorization and recall.

DATA SOURCE, ETHICS, CODE AND TECHNOLOGY STATEMENT

For this study, data is gathered from human participants through the university's participants pool, after approval of the Ethics committee. All the participants were informed about the study, and gave consent, allowing them to withdraw from the experiment at any moment. All the data is anonymous, and the participants were assigned a participant ID. Ownership of the data belongs to the author. Every figure originates from the author, except for Figure 1. This is a screenshot taken from the virtual environment, while using the Meta Quest 3 (Meta, 2024). All code is also developed by the author, using Python 3.10.9 and the following libraries: pandas 1.4.1, NumPy 1.22.2, matplotlib 3.5.1 and SciPy 1.8.0. The code can be found in the corresponding Notebook and on GitHub (https://github.com/aurelieef25/thesis). The tables are generated in Overleaf (Overleaf, 2024). A thesaurus (LLC, 2024) was used as an assistance to vary in word usage. Moreover, the reference tool Mendeley (Elsevier, 2024) was applied.

1. Introduction

On a daily basis, humans constantly perceive a continuous flow of information through multiple senses. Certain matters are more vividly recalled than others. Storing all this information and later retrieving it relates to memory, where the more time passes, the more information will be harder to remember (Zlotnik & Vansintjan, 2019). In 2019, an estimation of 57.4 million individuals globally, were suffering from dementia (Nichols et al., 2022). This number is only likely to increase, whereby a prediction is made saying that in 2050 roughly 152.8 million individuals will have dementia worldwide (Nichols et al., 2022). The implementation of innovative technologies such as VR could be a potential aid in supporting patients who suffer from memory impairment. Mendez et al. (2015) showed some promising results during a pilot study, where patients with behavioural variant frontotemporal dementia were able to interact with an avatar, who interviewed them while using VR. Because of the immersiveness VR has to offer, multiple senses can be stimulated which enhances memory processes such as retention and recall (Amin & Malik, 2013). According to Quak et al. (2015), recalling information happens more easily when exposed to multiple sensory modalities, compared to only one sensory modality. Thompson & Paivio (1994), also observed a higher recall performance when participants were exposed to a picture-sound pair, in comparison to either a single audio or visual stimulus. With the use of VR, it is possible to create a so-called memory palace, where a person is fully immersed in a virtual environment. This resulted in a person being able to recall more information, in comparison to a non-immersive environment (Yang et al., 2021). This is also suggested by Lamers & Lanen (2021), who investigated context-dependent memory effect by conducting three memory related tasks, which resulted in an overall better performance using VR in comparison with no use of VR.

VR is not only applicable in healthcare, but also in various fields such as education, therapy, training and social interactions (Hamad & Jia, 2022). For social interactions, users are able to communicate with each other in virtual environments and have a representation of themselves in a form of an avatar, which resembles a face-to-face interaction (Perry, 2016). Interactions with an avatar could be the future for social interaction (Rogers et al., 2022).

Currently, voice assistants are becoming increasingly popular as a form of human-computer interaction (Terzopoulos & Satratzemi, 2020). Apple's Siri, Amazon's Alexa or Microsoft's Cortana, are voice assistants that are able to process the user's speech as input and provide a spoken answer back to the user as output (Hoy, 2018). Their function is to aid users with their tasks and time management (Myers et al., 2007). One commonality of these tools is that the gender of their voices is female by default (Hwang et al., 2019). According to Nass & Moon (2000), the same social rules are applied by humans on computers, as they would respond to a human. A consequence of these social rules and the default female voice is that harmful stereotypes and biases of women are observed (Sutton, 2020). A gap in literature is observed, given that limited research has been done in varying the gender of voices and the effect of this on the user's memory retention, notably when combined with social avatars in virtual environments.

Therefore, the main objective of this study is to examine the effect of virtual reality on memory retention, by interacting with an avatar in a VR environment. A comparison will be made with traditional settings, where there is no use of VR. Additionally, by varying the appearance and voice of the avatar, a possible influence of voice gender on memory retention will be explored. Thereupon, a combination of the two previously mentioned questions is investigated, to determine the optimal conditions for memory retention, while focusing on the use of VR and the influence of gender. Lastly, a post-hoc analysis was conducted, to look for a possible relationship between the participants' gender and their performance whilst listening to a particular gender of the voice. Therefore, the main research question is as follows:

1. *What is the effect of having an interaction with an avatar in virtual reality on memory retention?*

The following sub-questions were formulated in order to support the main research question:

1.1 *To what extent does voice gender influence memory retention?*
1.2 *What is the optimal condition for memory retention considering voice gender and use of virtual reality?*

2. Related work

2.1 *Traditional memory*

The process of retaining learned information in long-term memory and being able to retrieve this easily is called memory retention (Bennett & Rebello, 2012). A foundational study from Godden and Baddeley (1975) investigated the effect of context-dependent memory. In this experiment, a group of divers were asked to memorize a list of 36 words presented via an audio tape, either on land or under water. After a delay, they were instructed to recall those words in either the same environment, or in a different environment. They found that participants significantly remembered less words when the two environments were not the same, in comparison to the participants who learned and recalled the words in the same context, and thus performed better. However, other studies have shown that an auditory stimulus is not the most optimal approach for memory retention, but alternatively make use of multisensory modality, where numerous senses are stimulated, which have led to better results regarding memory retention (Ghaemi & Rafi, 2018). Ghaemi & Rafi (2018) observed that when participants were exposed to auditory stress patterns, bold-printed stress patterns and hand gestures simultaneously, they were able to recall more words stress patterns, compared to those exposed solely to auditory stimuli.

2.2 *VR and memory*

Previous studies suggest that studying in a virtual reality environment can result in enhanced memory encoding and retention in contrast to traditional settings (Krokos et al., 2019; Lamers & Lanen, 2021). In an experiment conducted by Krokos et al. (2019), participants were asked to use a head-mounted display (HMD), to let them immerse in a virtual environment. The main point of interest was the difference in performance between a virtual memory palace compared to a non-immersive memory palace, using a desktop. This resulted in participants being able to recall a significantly higher number of faces using the HMD, as opposed to the non-immersive condition, indicating that the virtual environment enhances memory retention.

These results are also seen in a more recent study. Lamers & Lanen (2021) investigated the context dependency effect between a virtual and real environment by conducting three memory tasks. During one task, participants needed to memorize four sentences and recall them, in either the same conditions or two different conditions. The study not only demonstrated the effect of context dependency, but more importantly, the results showed that the process of memorizing and recall in both a virtual environment led to higher performances in contrast to memorizing and recall in a real environment.

Additionally, Yang et al. (2021) suggests that participants are able to recall more information from scholarly articles, while being immersed in a virtual environment. Compared to an image-based and no image at all, participants were asked to retrieve knowledge from the articles, where their recall accuracy score was the highest while using VR.

2.3 *Biases in voice gender*

According to Nass (1997), there is a certain bias among users regarding the gender of the voice used by a machine. During an experiment, the users were instructed to interact with a male- or female-voiced computer, where several stereotypes were tested. One of the outcomes suggested that users reacted adversely, when the female-voice showed dominance and therefore rated it less friendly, in comparison to a dominance tone in a male-voice. Additionally, it appeared that users found the evaluation from the female voice less reliable. These findings illustrate, that prejudices and biases on gender are also present in computers.

Despite the bias and expectations, most voice assistants are characterized with a female voice by default (Hwang et al., 2019). This might be due to the fact that a female voice is considered friendly and comforting (Nass & Moon, 2000). This is also the case, where participants preferred a female agent in an autonomous driving vehicle (Dong et al., 2020). The female gave a warm and trustworthy feeling towards the participants, and found the male agent natural and confident, but more aggressive. In another study, results showed that participants experienced a low male voice as not friendly compared to a low or high female voice (Tolmeijer et al., 2021).

2.4 *Interaction with avatars*

In a pilot study, results have shown the feasibility of VR with patients who suffer from a non-Alzheimer dementia (Mendez et al., 2015). Patients got interviewed by an avatar in a virtual environment using an HMD. In contrast with the real-world interviews, the patients appeared to respond more elaborately towards the virtual avatar. An overall positive experience was reported from the patients using the virtual environment and the interaction with avatar.

Buchta et al. (2022) investigated experiences from users who interacted with a voice assistant in a virtual environment. One condition was an interaction with a voice interface (VUI), and the other with a graphical interface (GUI). Within the VUI it was possible for the assistant to talk and interact with the user. While for the GUI, the interaction was only possible through labels and buttons or a virtual keyboard, meaning there was no actual voice. Overall, the voice assistant left a positive impression on the users in both conditions. The interaction with the VUI was experienced as more natural and realistic compared to the GUI. These characteristics, along with authenticity and fluency, led to an overall preference for interacting with the assistant using the VUI over the GUI.

As mentioned in the above study, the interaction between the user and the avatar is considered positive and preferred if it feels realistic. This is also suggested in a study, where the influence of visual and nonverbal cues on memory retention is examined, by using three types of avatars: a real avatar, a scanned avatar and no avatar (Aseeri & Interrante, 2021). Their research implies that a participants recall performance can be influenced by using a variety of avatar representations. According to the results, a majority favoured the real avatar since it allowed for an effortlessly face-to-face interaction and an engaging involvement. The availability of facial expressions and behavioural cues attributed to this. Additionally, it also appeared that the “no-avatar” condition scored significantly lower on co-presence in comparison to the other two types. Co-presence refers to ‘’the sense of being with another’’ (Biocca et al., 2003). This is related with positive communication outcomes and is also associated with trust, enjoyment and persuasion (Oh et al., 2018).

Based on this theoretical framework, the following hypothesis can be formulated:

1. There is no effect on memory retention between the use of an avatar in VR and traditional settings without VR.

1.1 Voice gender has no influence on recall performance.
1.2 There is no optimal condition for memory retention regarding the use of VR and gender.

3. Methods

3.1 Participants

A G*power analysis was conducted beforehand, to find the minimum required sample size that was determined to be 45 participants with 95% power and an alpha-value of 0.05 (Faul et al., 2009). A total of 49 participants (35 female, 14 male) were recruited for this experiment, with an age ranging from 18 till 32 years old (M = 22.22, SD = 3.00). This study is reviewed and approved by the Research Ethics Committee and Data Management Committee (REDC) at Tilburg School of Humanities and Digital Sciences. In addition to convenience sampling, participants were recruited through the university participant pool.

3.2 Material

3.2.1 *Wordlists*

Four wordlists were constructed by an online application, that generates random words (Friendly & Dubins, 2019; Paivio et al., 1968). Each list contains 30 English words. Although they are random words, it is possible to adjust the parameters at your preferences. For this study, all variable parameters were set to default, except for the Kucera-Francis word frequency, which was set to a range from 21 to 100. Setting the minimum at 21, based on the median determined by Paivio et al. (1968), reduces the likelihood of participants encountering unfamiliar words, considering that not all participants are native English speakers. All wordlists can be found in Appendix A.

3.2.2 *Voice generator*

The wordlists were pronounced by an artificial male and female voice. These two voices were generated by an AI voice generator (Eleven Labs, 2024). The premade voices “Sarah” and “Josh” by Elevenlabs (2024) were used, where “Sarah” represented the female voice and “Josh” the male voice.

3.2.3 *Technical equipment*

For this experiment, Meta Quest 3 (Meta, 2024) was utilized. From the settings, a default virtual environment was chosen called ‘Paradiso’, which looked like a vacation bungalow, with a sea view (Meta, 2024) Both the female and male avatar were randomly picked from the default Meta avatars, and were for every participant the same, as shown in Figure 1.

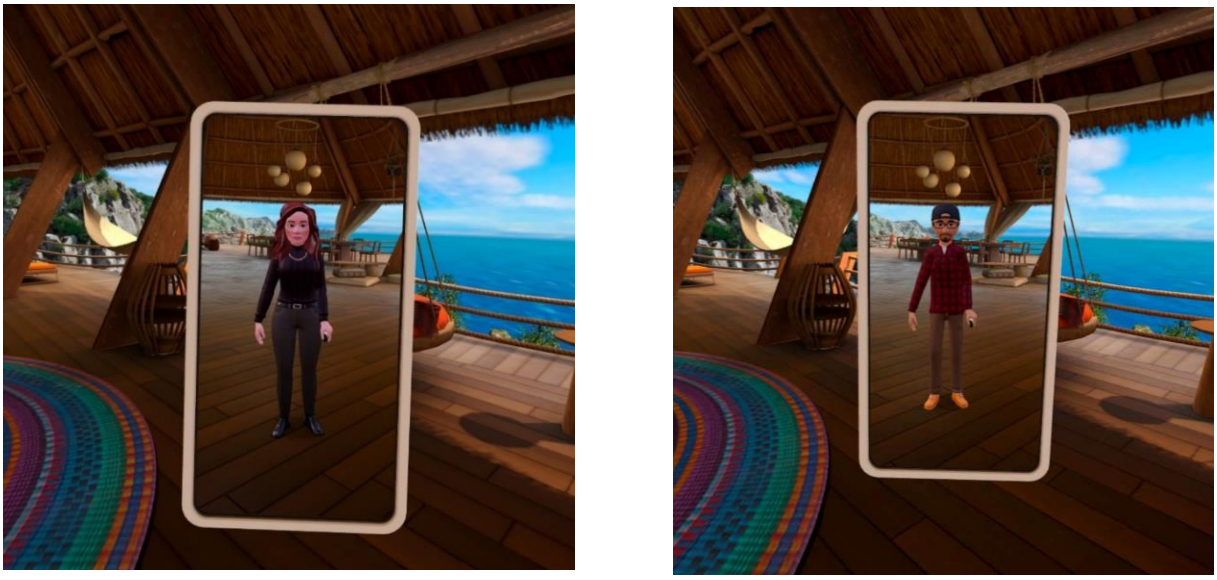

Figure 1: Screenshot taken from the participants' point of view, when exposed to the female avatar (left) and the male avatar (right), in the virtual environment '*Paradiso'* using Meta Quest 3 (Meta, 2024)

3.2.4 *Randomizer math equations*

After each intervention, the participant is asked to solve five arithmetic functions, during the interference phase. These arithmetic functions were the same for every participant and generated by an online application (Hall, 2024), these arithmetic functions were too difficult to solve without external aids, therefore the functions were randomly modified in order to make it possible to solve them mentally.

3.3 Study design

For this experiment, a within-subject design is used. To reduce bias, or other potential factors that could influence the outcomes, the order in which the conditions are exposed to the participants was randomly assigned, by applying the simple random sampling. To minimize cybersickness or other aspects that could affect the performances, a traditional audio condition was followed by the VR condition, and vice versa. This led to eight possible sequences, which are presented in Table 1. Since there is no need for balanced groups, each participant is assigned a sequence by using the *randint* function from the package *random* written in Python 3.10.9. Each wordlist is linked to a condition, meaning wordlist 1 is always executed by the male audio in traditional settings, wordlist 2 by the male-voiced avatar, wordlist 3 by the female audio in traditional settings and wordlist 4 by a female-voiced avatar. Thus, the identicalness of the wordlists is maintained across all participants.

C1: male VR - male audio - female VR - female audio
C2: male VR - female audio - female VR - male audio
C3: female VR - female audio - male VR - male audio
C4: female VR - male audio - male VR - female audio
C5: male audio - female VR - female audio - male VR
C6: male audio - male VR - female audio - female VR
C7: female audio - female VR - male audio - male VR
C8: female audio - male VR - male audio - female VR

.

C1	C2	C2	C3	C4	C5	C6	C7	C8
$n = 8$	$n = 6$	$n = 6$	$n = 6$	$n = 7$	$n = 6$	$n = 7$	$n = 6$	$n = 4$

Table 1: Sample size of participants assigned to a particular condition, which indicates the sequence of the wordlists

3.4 Procedure

Prior to obtaining consent, the participant was given the opportunity to review the study information sheet and ask questions. Afterwards, a short explanation of the whole procedure was given to the participant for clarity. Then the first intervention started, depending on the randomization process of the order of conditions, where the first wordlist was perceived. This was followed by an intermission phase, where there was a break of 10 seconds of silence. Subsequently, the recall phase began, and the participant was asked to recall as many words as possible from the just perceived wordlist in one minute. This was done out loud, so their answers could directly be checked, by ticking the box of the word in the wordlist, if recalled correctly. Afterwards, they needed to rate their performance based on a scale from 1 to 10, where 1 is very poor and 10 indicating a perfect score. Finally, five arithmetic functions were presented on a computer, where the participants needed to fill in the correct answer mentally. This was done with the aim of refreshing the brain, and allowing the participants to think of something else, rather than the words they just perceived. After four repetitions of the tasks, whereby no arithmetic functions were presented after the final wordlist, we conducted an overall survey using a 5-point t scale through Qualtrics. The aim of the survey was to obtain information on the demographics of the participants, their preferences, and how they experienced the use of VR. The participant was required to wear the VR headset during the learning phase, intermission phase, recall phase and while they rated their performance. Furthermore, the participant was only seated while reading the information sheet, solving the arithmetic functions and while filling in the survey.

In Figure 2, the procedure is presented, and a short description of each intervention is described below. Note that this is according to the first sequence of conditions as shown in Table 1.

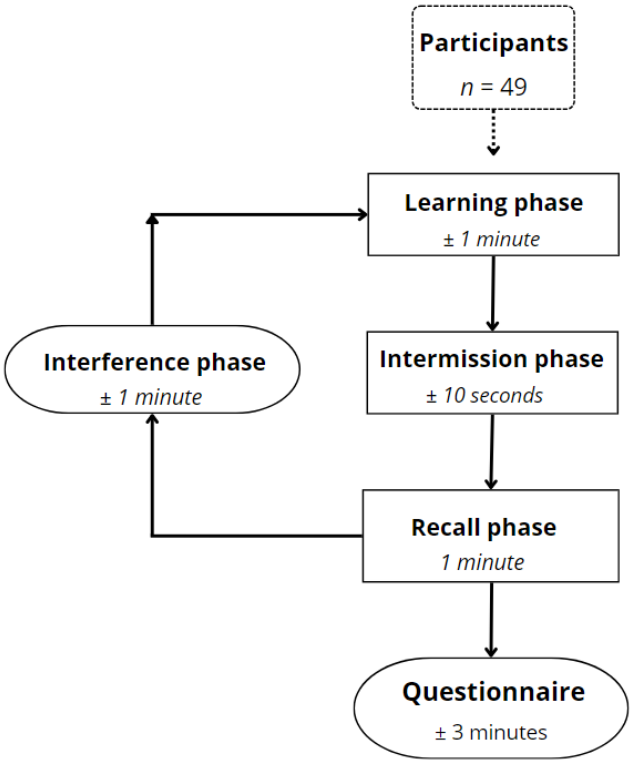

Figure 2: Workflow of the study design.

Intervention 1

While being immersed in a virtual environment, the participant was instructed to remember as many words as possible that were spoken by a male voice through an audio recording. Simultaneously, a male avatar was presented in the virtual environment. After the learning phase, the participant could rest for 10 seconds during the intermission phase. This was followed by the recall phase, where the participant had to recall as many words as possible towards the avatar in one minute. After recalling the words, they had to give themselves a grade for their performance. Finally, the VR headset was removed, and they were assigned to start solving the five math problems on the computer.

Intervention 2

In the learning phase, the participant memorized 30 words via an audio fragment spoken by a male voice. Then, a short break followed during the intermission phase. Thereafter, the participant needed to recall as many as words possible during the recall phase. While saying these words out loud, their answers were checked, by ticking the boxes of the words in the wordlist. Afterwards, they were asked to rate their performance. Finally, they solved the next five arithmetic functions.

Intervention 3

During the learning phase, the VR headset was used again, and the participant was exposed to a female avatar, while simultaneously listening to 30 new words spoken by a female voice via an audio tape. The intermission phase followed, after the end of the wordlist. During the recall phase, the participant said the recalled words out loud towards the avatar. The answers were compared to the words in the wordlist. The rating of their performance was asked, and the VR headset was removed. Thereupon, they solved the final five arithmetic problems.

Intervention 4

During the learning phase, the participant memorized 30 new words by listening to a female voice, while looking at the wall. Silence of 10 seconds followed in the intermission phase. In the recall phase, the participant needed to recall as many as words as possible and did that out loud, so their answers could be immediately examined. On a scale from 1 to 10, they gave themselves a grade based on their performance. Finally, they filled in the online survey on the computer.

4. Results

The data was generated through Qualtrics (Qualtrics, 2024), and then analysed using Python 3.10.9.

4.1 *Main research question*

Firstly, the main research question was evaluated where the effect of virtual reality on memory retention is examined, by comparing the participants' performance on the VR condition versus the non-VR condition. Prior to running a parametric test, the normality and homogeneity in variances were checked. Firstly, a Shapiro-Wilk test was conducted to see whether the data was normally distributed. This resulted in the VR-data sample being normally distributed ($W = 0.9760$, $p = 0.0697$). However, the audio data was not normally distributed ($W = 0.9362$, $p = 0.0001$). Subsequently, the homogeneity of variance was checked, by conducting a Levene's test, which resulted in an equality in variances between the two groups ($p = 0.4883$). Since the normality assumption was violated for the audio sample, it is not possible to conduct a parametric test. Therefore, the non-parametric Mann-Whitney U test was conducted, as an alternative for the Welch t-test, which resulted in a non-significant difference between the number of correct recalled words, when using VR or not using VR ($U = 4719.0$, $p = 0.8345$).

4.2 *First sub-research question*

To answer the first sub-research question, “to what extent does voice gender influence memory retention?”, the number of correct recalled words was compared when listening to a male voice in contrast to a female voice. Before running a parametric test, the assumptions were checked again. Both samples were checked on the normality assumption by using Shapiro-Wilk test. The female voiced sample is not normally distributed ($W = 0.9320$, $p = 7.6396e-05$), whereas the male voiced sample is normally distributed ($W = 0.9825$, $p = 0.2193$). The assumption for homogeneity of variance is not violated between the two groups ($p = 0.3317$). As a consequence of the not normally distributed data of the female voiced sample, the non-parametric Mann-Whitney U test was conducted. This resulted in a non-significant difference between the performance of the participants when listening to a male voice or a female voice ($U = 4845.5$, $p = 0.9133$).

4.3 *Second sub-research question*

In order to look for the optimal condition, where the participants were able to recall the greatest number of words, the performances on each wordlist were evaluated separately. Before running the parametric tests, the normality and homogeneity assumptions were checked. Only the female audio sample appeared to be not normally distributed ($W = 0.8910$, $p = 0.0003$). Whereas the female VR ($W = 0.9573$, $p = 0.0736$)), male audio ($W = 0.9719$, $p = 0.2880$)) and male VR ($W = 0.9801$, $p = 0.5709$) samples were all normally distributed. Additionally, there was a homogeneity in variances found among the four groups ($p = 0.5498$). Therefore, the non-parametric Kruskal-Wallis test was performed as an alternative for ANOVA, since not all data is normally distributed. This test showed that there was no significant difference found in the number of correct recalled words between the four groups ($H(3) = 0.6909$, $p = 0.8753$).

4.4 *Post hoc analysis*

Finally, a post hoc comparison is performed to observe a possible relationship between the gender of the participants and the voice gender. This is evaluated by looking if the male participants were able to recall more words when listening to a particular gender, and vice versa for the female participants. The sample female and male participants on a female voice were both not normally distributed (male: ($W = 0.9165$, $p = 0.0020$) & female: ($W = 0.9294$, $p = 0.0007$). Whereas the samples on the male voice for both genders were normally distributed (male: ($W = 0.9734$, $p = 0.3276$) & female: ($W = 0.9720$, $p = 0.1180$). A Levene’s test indicated a homogeneity in variances between the four groups ($p = 0.8376$). There is also an equality in variances between the male and female participants, perceiving a male voice ($p = 0.1008$). Since the assumptions were not violated for that comparison, an independent t-test was performed and showed a non-significant difference in the mean of correct recalled words between male and female participants listening to a male voice ($t(3) = -0.0188$, $p = 0.9851$). Because the assumptions are not met for the comparison between both gendered participants listening to a female voice, the alternative Mann-Whitney U test was conducted, that showed no significant difference in performance between male and female participants under these conditions ($U = 1783.5$, $p = 0.7115$). At the end, a Kruskal Wallis test was performed as an alternative for ANOVA, between all four variables. The result showed there was no significant difference in performance between the voice gender and the gender of the participant ($H(3) = 0.1432$, $p = 0.9862$).

5. Discussion

5.1 *Main research question*

The main objective was to investigate whether an interaction with an avatar in VR affects memory retention. The difference in the mean of the number of correct recalled words between VR and non-VR is very minimal, where participants scored slightly better when exposed solely to an audio stimulus (*M* = 8.1837, *SD* = 3.1941), in comparison to an exposure with an avatar in the virtual environment (*M* = 8.0612, *SD* = 3.0942). Furthermore, no significant difference was found between the two variables and thus we cannot conclude whether there is an effect on memory retention using VR. This is in contrast with earlier findings, where a significant effect was found between the HMD and the desktop condition, meaning that the number of recalled faces and names was higher using the HMD and thus performed better (Krokos et al., 2019).

Figure 3 illustrates the distribution of the participants' recall performance in traditional settings compared to being immersed in a virtual environment. The data reveals a widespread distribution, with no significant outcome. Notably, there are a couple of outliers, that suggests a participant recalled more words in traditional settings, while a few that performed very well while using VR. Moreover, the regression line appears to be more linear than expected, indicating that participants generally performed consistently across both conditions. Thus, no large difference in number of correct recalled words is observed between the two conditions.

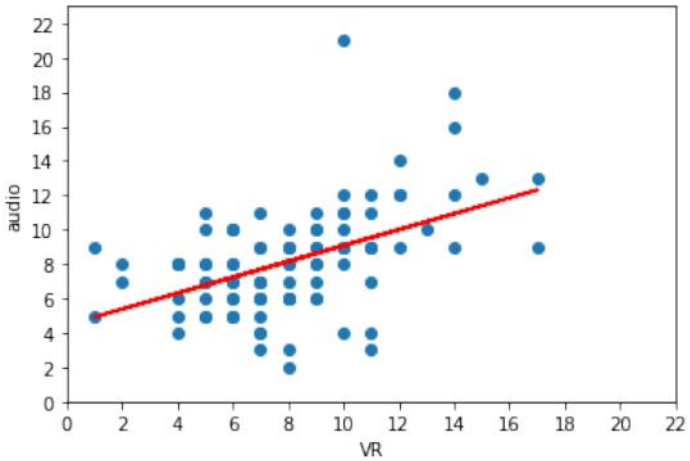

Figure 3: Scatterplot of the participants recall performance while using VR versus traditional settings.

According to our survey, 51% of the participants agreed with the statement that when exposed to solely an audio stimulus they remember better. Whereas only 35% of the participants agreed to the statement to believe that they remember better when exposed to an avatar in VR. The VR environment and the avatar itself could be a possible reason, causing the overall mean of recalled words to be lower than expected, as some participants indicated after the experiment, that the avatar and some moving objects in the environment were distracting and therefore interrupted their focus. As well as the level of realism of the avatar could be reconsidered. Despite the fact that 61% of the participants found the interaction with the avatar realistic, they were not able to recall more words in comparison to the non-VR exposure. This is not in line with earlier findings, that suggested that individuals were able to recall more letter combinations while being immersed in a virtual environment with a full body representation, in comparison to not being in a virtual environment (Pan & Steed, 2019). The lack of realistic traits could be due to the fact that the virtual avatar was a reflection of the participant. The participant stood in front of a mirror, and when they moved, the avatar would move along. Therefore,

the avatar was not able to move independently, nor was there a synchronization of its mouth movements with the pronounced words. These are all factors, which makes the interaction with the avatar less realistic. The level of realism is an important aspect of the interaction, since a more realistic avatar, leads to a higher level of co-presence (Oh et al., 2018). Therefore, future work is recommended in using a more realistic representation.

An alternative of our avatar could be the use of MetaHumans powered by Unreal Engine 5 (Epic Games, 2024). This is a realistic digital human character, initially used for computer games, but can also be implemented in other applications. An attempt to generate such avatar failed and is beyond the scope of this study.

5.2 *First sub-research question*

The first sub-research question is: "to what extent does voice gender influence the performance of recalling?". There is only a marginal difference between the mean of recalled words, where the participants recalled slightly better while listening to a female voice (M =8.2857*,* SD = 3.3858) in comparison to a male voice (M = 7.9592*,* SD = 2.8751). In addition, no significant difference in performance has been found between perceiving a male or female voice. However, this is not in line with Samoza et al. (2015), who found a significant effect of the teacher's voice types on student's performances. They saw that students were able to recall more words while perceiving a high-pitched voice versus a lower pitched voice.

There was also no result in preference for a particular gender amongst the participants according to our survey. 45% of the participants indicated to have enjoyed listening to a female voice, and 43% enjoyed listening to a male voice. The majority was neutral about having a preference for a particular voice of gender. These findings are in contrast to other studies, that did find a difference between preference for gendered voices. Dong et al. (2020) saw that a majority preferred a non-robotic female voice as a voice agent in an autonomous vehicle. Both the non-robotic and robotic female voices, were considered as warm, likeable and scored low on discomfort, in contrast to the non-robotic and robotic male voices. Lastly, our survey results showed that 43% of the participants were not even aware of the distinction between male and female voices. The voice gender was purposefully not announced prior to the experiment, to examine the awareness of the gendered voice. This lack of awareness could be influencing the outcomes of their preferences. In addition to this unawareness, the implementation of a genderless voice could be considered. Danielescu et al. (2023) evaluated such genderless voice, and found it was favoured by individuals from the non-binary and transgender community. However, it was not the preferred choice for the male and female participants. Nevertheless, according to our survey, a majority of 49% was aware of the distinction between the genders and 8% remained neutral about the matter.

5.3 *Second sub-research question*

Secondly, a combination of the two previously mentioned research question was examined, to look for an optimal condition regarding gender and the use of VR. In Table 2, the mean of recalled words with the corresponding standard deviation is presented. The results demonstrate that participants recalled the most correct words while listening to a female voice in traditional settings (M = 8.5714, SD = 3.5000). The outcome of the non-parametric test Kruskal-Wallis showed that there is no significant difference between the four conditions: female avatar, female audio, male avatar and male audio (H = 0.6909, p = 0.8753). Based on previous studies, this is unexpected, since Lamers & Lanen (2021) found a significant result, where participants recalled more sentences in a virtual environment compared to the real environment. A possible factor that could have influenced the results on the use of VR, is that 35% of the participants indicated that they have never used VR before. Their attention may have been affected by the novelty and excitement. On the other hand, a greater difference between the female and male speaker was expected, since previous studies suggested that when listening to a female voice, individuals tend to significantly forget less information, compared to a

male speaker (Yang et al., 2013). Moreover, 96% of the participants were non-native English speakers. This could have influenced their performance in recalling the words, since they may have encountered unfamiliar words. This is also seen in a study from Kisser et al. (2012), where native English speakers performed better than non-native English speakers, on language related tasks.

Male audio		Male VR		Female audio		Female VR	
Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
7.7959	2.6059	8.1224	3.1400	**8.5714**	3.5000	8.0	3.2787

Table 2: Mean values of the correct recalled words per condition, with the corresponding standard deviations.

5.4 Post hoc analysis

Considering all non-significant outcomes, a post-hoc analysis was conducted to look for a possible relation between the gender of the participants and their performance on the memory task. Both male and female participants recalled the most words while listening to a female voice (male: $M = 8.1071$, $SD = 3.2471$, female: $M = 8.3571$, $SD = 3.4601$). Table 3 shows that both groups of participants scored the worst while being exposed to the condition male audio (male: $M = 7.7143$, $SD = 3.3611$ female: $M = 7.8286$, $SD = 2.2943$) in comparison to all the other mean performances.

Gender	Male audio		Male VR		Female audio		Female VR		Male voice		Female voice	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Male	7.7143	3.3611	**8.3571**	3.2488	**8.3571**	3.4106	7.8571	3.1831	8.0357	3.2601	8.1071	3.2471
Female	7.8286	2.2943	8.0286	3.1388	**8.6571**	3.5805	8.0571	3.3602	7.9286	2.7310	8.3571	3.4601

Tabel 3: For each condition, the mean value of recalled words is presented for the male and female participants, with their corresponding standard deviations. Additionally, the mean performance in both male and female voice is presented, along with the standard deviations.

In Figure 4, the distribution of the male and female participants is plotted, where the female voice has a few outliers, which causes an increment in the overall mean of correct recalled words. Apart from this, the distribution with respect of the female voice stimuli is more concentrated in comparison to the male voice distribution. For the male voice it is more widely distributed, which could result in an overall better performance in recalling. According to the survey, 43% of the male participants preferred listening to a male voice whilst, the preference for a female voice was a bit lower (36%). However, the majority was neutral about their preference regarding the gender. In contrast, the female participants enjoyed the female voice with 49% in comparison to the male voice (43%). Again, the majority of the female participants were neutral. A previous study advocates a relation between the gender of the participant and the gendered voice (Weririch & Simpson, 2018). However, these findings did not contribute to any helpful insights, therefore further investigation on the effect of gendered voices on memory retention is required.

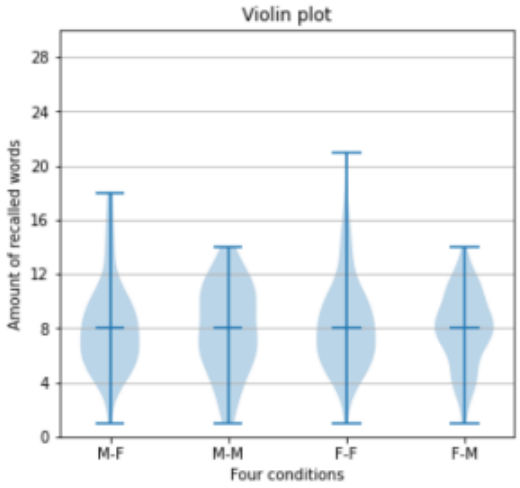

Figure 4: This plot depicts the distribution of male participants listening to a female voice (M-F), male participants listening to a male voice (M-M), and females listening to a female voice (F-F) and lastly the females listening to a male voice (F-M). The minimum, maximum and mean values are depicted as well.

5.5 Limitations

Due time limitations, it was not feasible to use a high-fidelity avatar. This resulted in the use of a less advanced avatar, which could not speak the words itself. Additionally, the delay during the intermission phase was relatively short, and therefore the ability to retrieve the information is less well measured in comparison to a longer delay. Despite the G* power analysis, the sample size may be still too small to observe a significant result and therefore a larger sample size is desired. Finally, an order effect could have occurred, which influences the performance of the participants.

5.6 Future work

Some general implementations are recommended to improve the framework of this study. A further examination of the duration of the intermission phase is recommended. Other studies suggest an intermission phase of 15 to 20 minutes (Koens et al., 2003) or even 24 hours before the recall phase (Godden & Baddely, 1975; Lamers & Lanen, 2021). Furthermore, the number of words can be reconsidered since previous studies have used wordlists that vary from 20 to 35 words (Koens et al., 2003; Hainš et al., 2020). Another reconsideration could be the speed and volume of the words that were pronounced by the artificial voices. A final suggestion is the implementation of a more realistic avatar, that could for example move and talk independently.

6. Conclusion

The goal of this research was to explore the effect of virtual reality on memory retention, by interacting with an avatar. Additionally, we analysed the impact of voice gender on an individual's memory retention by varying the voice gender of the stimuli. No significant difference has been found between using VR and no use of VR. Furthermore, a non-significant result was observed, regarding the extent of voice gender influencing memory retention. We also cannot conclude that there is a significant difference between the four conditions, varying in the use of VR and gender. Therefore, the previous postulated hypothesis, a) "There is no effect on memory retention between the use of an avatar in VR and traditional settings without VR, b) "Voice gender has no influence on recall performance", and c) "There is no optimal condition for memory retention regarding the use of VR and gender, cannot be either rejected or accepted due the lack of evidence. However, it seems that the participants recalled the most correct words while being exposed to the non-VR female condition. Because of various limitations, future work is required to find a significant outcome for this problem statement.

References

Amin, H., & Malik, A. S. (2013). Human memory retention and recall processes. A review of EEG and fMRI studies. *Neurosciences (Riyadh, Saudi Arabia)*, *18*(4), 330–344.

Aseeri, S., & Interrante, V. (2021). The Influence of Avatar Representation on Interpersonal Communication in Virtual Social Environments. *IEEE Transactions on Visualization and Computer Graphics*. https://doi.org/10.1109/TVCG.2021.3067783

Bennett, A. G., & Rebello, N. S. (2012). Retention and learning. In *Springer eBooks* (pp. 2856–2859). https://doi.org/10.1007/978-1-4419-1428-6_664

Biocca, F., Harms, C., & Burgoon, J. K. (2003). Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence*, *12*(5), 456–480. https://doi.org/10.1162/105474603322761270

Buchta, K., Wojcik, P., Nakonieczny, K., Janicka, J., Galuszka, D., Sterna, R., & Igras-Cybulska, M. (2022). Modeling and optimizing the voice assistant behavior in Virtual Reality. *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. https://doi.org/10.1109/ismar-adjunct57072.2022.00086

Dong, J., Lawson, E., Olsen, J., & Jeon, M. (2020). Female Voice Agents in Fully Autonomous Vehicles Are Not Only More Likeable and Comfortable, But Also More Competent. *Https://Doi.Org/10.1177/1071181320641248*, *64*(1), 1033–1037. https://doi.org/10.1177/1071181320641248

ElevenLabs. (2024). *Text to Speech & AI Voice Generator*. ElevenLabs. https://elevenlabs.io/

Elsevier (2024). *Mendeley – Reference Management Software.* Mendeley. https://www.mendeley.com/search/

Epic Games. (2024). *MetaHuman.* Unreal Engine. https://www.unrealengine.com/en-US/metahuman

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160

Friendly, M. & Dubins, M. (2019), Paivio et al. Word List Generator, Online application, http://euclid.psych.yorku.ca/shiny/Paivio

Ghaemi, F., & Rafi, F. (2018). The Impact of Visual Aids on the Retention of English Word Stress Patterns. *International Journal of Applied Linguistics and English Literature*, *7*(2), 225–231. https://doi.org/10.7575/AIAC.IJALEL.V.7N.2P.225

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, *66*(3), 325–331. https://doi.org/10.1111/j.2044-8295.1975.tb01468.x

Hains, V. V., Kucar, M., & Kovacic, R. (2020). Student social media usage and its relation to free-recall memory tasks. *2020 43rd International Convention on Information, Communication and Electronic Technology, MIPRO 2020 - Proceedings*, 731–736. https://doi.org/10.23919/MIPRO48935.2020.9245167

Hall, J. (2024). *Mathsbot Question Generator*. https://mathsbot.com/questionGenerator

Hamad, A., & Jia, B. (2022). How Virtual Reality Technology Has Changed Our Lives: An Overview of the Current and Potential Applications and Limitations. *International Journal of Environmental Research and Public Health 2022, Vol. 19, Page 11278*, *19*(18), 11278. https://doi.org/10.3390/IJERPH191811278

Hoy, M. B. (2018). Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. *Medical Reference Services Quarterly*, *37*(1), 81–88. https://doi.org/10.1080/02763869.2018.1404391

Hwang, G., Oh, C. Y., Lee, J., & Lee, J. (2019). It sounds like a woman: Exploring gender stereotypes in South Korean voice assistants. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290607.3312915

Kisser, J. E., Wendell, C. R., Spencer, R. J., & Waldstein, S. R. (2012). Neuropsychological Performance of Native versus Non-native English Speakers. *Archives of Clinical Neuropsychology*, *27*(7), 749–755. https://doi.org/10.1093/arclin/acs082

Koens, F., Cate, O. T. T., & Custers, E. J. (2003). Context-dependent memory in a meaningful environment for medical education: in the classroom and at the bedside. *Advances in Health Sciences Education*, *8*(2), 155–165. https://doi.org/10.1023/a:1024993619713

Kourtesis, P., Collina, S., Doumas, L. A. A., & MacPherson, S. E. (2019). Validation of the Virtual Reality Neuroscience Questionnaire: Maximum Duration of Immersive Virtual Reality Sessions Without the Presence of Pertinent Adverse Symptomatology. *Frontiers in Human Neuroscience*, *13*. https://doi.org/10.3389/fnhum.2019.00417

Krokos, E., Plaisant, C., & Varshney, A. (2019). Virtual memory palaces: immersion aids recall. *Virtual Reality*, *23*(1), 1–15. https://doi.org/10.1007/S10055-018-0346-3

Lamers, M. H., & Lanen, M. (2021). Changing between virtual reality and real-world adversely affects memory recall accuracy. *Frontiers in Virtual Reality*, *2*, 602087. https://doi.org/10.3389/FRVIR.2021.602087

Mendez, M. F., Joshi, A., & Jimenez, E. (2015). Virtual reality for the assessment of frontotemporal dementia, a feasibility study. *Disability and Rehabilitation: Assistive Technology*, *10*(2), 160–164. https://doi.org/10.3109/17483107.2014.889230

Meta. (2024). *Meta Quest 3 home. Environment Paradiso.* https://www.meta.com/nl/en/quest/quest-3/

Myers, K. L., Berry, P., Blythe, J., Conley, K., Gervasio, M. T., McGuinness, D. L., Morley, D. N., Pfeffer, A., Pollack, M. E., & Tambe, M. (2007). An Intelligent Personal Assistant for Task and Time Management. *AI Magazine*, *28*(2), 47–61. https://doi.org/10.1609/aimag.v28i2.2039

Nass, C., Moon, Y., & Green, N. (1997). Are Machines Gender Neutral? Gender-Stereotypic Responses to Computers With Voices. *Journal of Applied Social Psychology*, *27*(10), 864–876. https://doi.org/10.1111/J.1559-18

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nichols, E., Steinmetz, J. D., Vollset, S. E., Fukutaki, K., Chalek, J., Abd-Allah, F., Abdoli, A., Abualhasan, A., Abu-Gharbieh, E., Akram, T. T., al Hamad, H., Alahdab, F., Alanezi, F. M., Alipour, V., Almustanyir, S., Amu, H., Ansari, I., Arabloo, J., Ashraf, T., … Vos, T. (2022). Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an

analysis for the Global Burden of Disease Study 2019. *The Lancet Public Health*, *7*(2), e105–e125. https://doi.org/10.1016/S2468-2667(21)00249-8

Oh, C. S., Bailenson, J. N., & Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Frontiers Robotics AI*, *5*(OCT), 409295. https://doi.org/10.3389/FROBT.2018.00114/BIBTEX

Overleaf. (2024). *Overleaf, Online LaTeX Editor*. https://www.overleaf.com

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, *76*(1, Pt. 2), 1-25.

Pan, Y., & Steed, A. (2019). Avatar type affects performance of cognitive tasks in virtual reality. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST*. https://doi.org/10.1145/3359996.3364270

Perry, T. S. (2016). Virtual reality goes social. *IEEE Spectrum*, *53*(1), 56–57. https://doi.org/10.1109/MSPEC.2016.7367470

Quak, M., London, R. E., & Talsma, D. (2015). A multisensory perspective of working memory. *Frontiers in Human Neuroscience*, *9*(APR), 1–11. https://doi.org/10.3389/FNHUM.2015.00197/BIBTEX

Qualtrics. (2024). *Qualtrics Survey Software*. https://www.qualtrics.com/

Rogers, S. L., Broadbent, R., Brown, J., Fraser, A., & Speelman, C. P. (2022). Realistic Motion Avatars are the Future for Social Interaction in Virtual Reality. *Frontiers in Virtual Reality*, *2*, 750729. https://doi.org/10.3389/FRVIR.2021.750729/BIBTEX

Samoza, P. R., Sugay, J. F., Arellano, E., & Custodio, B. (2015). An evaluation of the effect of various voice qualities on memory retention. *Procedia Manufacturing*, *3*, 1503–1510. https://doi.org/10.1016/j.promfg.2015.07.399

Sutton, S. J. (2020). Gender Ambiguous, not Genderless: Designing Gender in Voice User Interfaces (VUIs) with Sensitivity. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3405755.3406123

Terzopoulos, G., & Satratzemi, M. (2020). Voice Assistants and Smart Speakers in Everyday Life and in Education. *Informatics in Education*, *19*(3), 473–490. https://doi.org/10.15388/INFEDU.2020.21

Thompson, V. A., & Paivio, A. (1994). Memory for pictures and sounds: Independence of auditory an visual codes. *Canadian Journal of Experimental Psychology*, *48*(3), 380–398. https://doi.org/10.1037/1196-1961.48.3.380

Tolmeijer, S., Zierau, N., Janson, A., Wahdatehagh, J. S., Leimeister, J. M. M., & Bernstein, A. (2021). Female by Default? - Exploring the Efect of Voice Assistant Gender and Pitch on Trait and Trust Atribution. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3411763.3451623

Yang, F., Qian, J., Novotny, J., Badre, D., Jackson, C. D., & Laidlaw, D. H. (2021). A Virtual Reality Memory Palace Variant Aids Knowledge Retrieval from Scholarly Articles. *IEEE Transactions on Visualization and Computer Graphics*, *27*(12), 4359–4373. https://doi.org/10.1109/TVCG.2020.3009003

Yang, H., Yang, S., & Park, G. (2013). Her voice lingers on and her memory is strategic: Effects of gender on Directed Forgetting. *PloS One*, *8*(5), e64030. https://doi.org/10.1371/journal.pone.0064030

Zlotnik, G., & Vansintjan, A. (2019). Memory: An Extended Definition. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/FPSYG.2019.02523

Appendix A

Four wordlists that are generated by the online application (Friendly & Dubins, 2019; Paivio et al., 1968).

Wordlist 1	*Wordlist 2*	*Wordlist 3*	*Wordlist 4*
Body	Economy	Cabin	Intimate
Duty	Bos	Institute	Vest
Cell	Explanation	Judge	Belief
Brain	Gif	Effort	Idea
Procession	Excuse	Pries	Arrow
Sovereign	Tribute	Sickness	Nai
Camp	Pride	Child	Hint
Green	Mother	King	Officer
Bloom	Dress	Enterprise	Frog
Slave	Costume	Infant	Golf
Truck	Recognition	Love	Plant
Quantity	Research	Flood	Skin
Background	Fork	Circle	Ceremony
Odour	Sea	Doctor	Joy
Temple	Hope	Honour	Marriage
Oats	Impulse	Anxiety	Horse
Exhaust	Air	Automobile	Tendency
Army	Street	Flag	Steamer
Railroad	Speaker	Paper	Bird
Chanc	Length	Library	Ability
Christmas	Elephant	Occasion	Method
Seat	Mood	Garden	Coast
Law	Moment	Interest	Building
Virtue	Sunset	Machine	Season
Style	Atmosphere	Warmth	Picture
Expression	Board	String	Hall
Form	Gratitude	Distinction	Palace
Bowl	Shore	Confidence	Church
Spirit	Door	Discovery	Product
Candidate	Happiness	Weapon	Origin

Appendix B

Post questionnaire based on a 5-point Likert scale, except for the demographic questions (1,2 and 3)

The last two questions (12,13) are from the Virtual Reality Neuroscience Questionnaire (Kourtesis et al., 2019).

1. What is your gender :
2. What is your age:
3. Are you a native English speaker?
4. I liked listening to the audio tape.
5. I liked having an interaction with the avatar.
6. I enjoyed listening to a female voice.
7. I enjoyed listening to a male voice.
8. I was aware of the distinction between the male and female voice.
9. I believe I remember better when exposed solely to audio stimuli.
10. I believe I remember better when exposed to an avatar.
11. I have used virtual reality before.
12. The interaction with the avatar felt realistic.
13. I felt immersed in the virtual environment (it felt like I was not in the room anymore).