Comparison of the Memorability of Text and Visual (Background) Sounds During their Fade-In in Short Scenes With Subtitles

[1]Tamin Rahi, [2]Prof. Dr. Karsten Huffstadt, [3]Selma Al

[1] Enrolled student at the University of Applied Sciences Würzburg-Schweinfurt ,[2] Head of the Institute of Design and Information Systems at University of Applied Sciences Würzburg-Schweinfurt, [3] Enrolled student at the Julius-Maximilians-University Würzburg,
[1]tamin-rahi@outlook.de,[2] karsten.huffstadt@fhws.de, [3] selma.al@gmx.net

Article Info
Page Number: 914 – 925
Publication Issue:
Vol. 71 No. 3s (2022)

Article History
Article Received: 22 April 2022
Revised: 10 May 2022
Accepted: 15 June 2022
Publication: 19 July 2022

Abstract

To test memory performance between nonverbal stimuli in the form of moving images (GIFs) and verbal stimuli in the form of written-down words, three hypotheses were investigated. Short scenes from the cult series The Simpsons were enriched with the above stimuli. The first hypothesis was based on the assumption that GIFs would score higher than words on free recall. This could not be proven. The same applies to the hypothesis that GIFs would be remembered better than words during serial recall. The last hypothesis to be investigated also proved not to be significant. This was based on the assumption, that words are inferior to moving images in terms of recognition. It is assumed that the superimposed scenes overwrote the GIFs during the memorization phase, since they were the same stimulus form. This meant that the attention of the subjects was distracted by the scenes. The high playback values of individual stimuli were also striking. It was assumed that these stimuli were associated with emotions and that the subjects therefore focused their attention on them, as a result of which they were remembered more strongly.

Keywords: GIF, Subtitle, The Simpsons, Working Memory

I. INTRODUCTION

Subtitles have become indispensable in the broadcasting of films, series, and real-time programs via television, as well as in the consumption of this content via streaming services and social media platforms such as YouTube. In addition to the dialogues and monologues that take place in specific actions, sounds and noises that take place in the background of the action or during the action are also rendered in the form of words to convert the auditory variant of the stimulus into a visual symbolic form so that this information can be grasped by viewers in visual form when the sound of the medium is muted and the subtitle is activated.

Subtitles are, as already mentioned, common practice in everyday life. Now the question arises, which findings would result, if sounds, tones or noises were not transformed into a visually symbolic form, but simply the source of the causing sound was faded in with the help of moving images at the moment of the occurrence of this stimulus in the scene and at the same time the sound of the entire scene to be captured was muted. Furthermore, the subtitle should be inserted in the form of sentences when monologues or dialogues occur.

II. RELATED WORK

Multi-memory model

In the 1960s, Richard Atkinson and Richard Shiffrin developed the multiple memory model (MSM), which divides memory into different subsystems. The MSM divides memory into sensory memory, short-term memory, and long-term memory [1].

Sensory memory captures information first and stores it for fractions of a second or seconds. Ultra-short-term memory has a short storage period. However, its storage capacity is enormous. Stimuli that have passed through the perceptual and attentional processes are then transferred to short-term memory. The storage capacity of short-term memory is limited. Maintaining the content can take between seconds and minutes [2, p. 4]. Research has shown that when learning sequences of digits up to 7 ± 2 units of information can be stored in memory [3]. The content can then either be deleted or overwritten by new stimuli. Otherwise, this content can be transferred to long-term memory through repetition [2, p. 5]. Long-term memory (LZG) stores information that has been recoded in short-term memory. It includes information received a few minutes ago as well as information gathered over a lifetime. The LGZ has an unlimited storage capacity and a potentially lifetime storage period [4].

Long-term memory

The LZG is divided into a declarative and a non-declarative memory. In declarative memory, the content is consciously transferred. This process can take place, for example, through speech, writing or images. Episodic, semantic and perceptual memory are classified as declarative memory [5, p. 8]. Implicit memory is the part of the LZG in which memory contents are reproduced unconsciously and automatically [2, p. 39]. The system acquires and stores habits, knowledge, learned experiences and reactions. Implicit memory is divided into priming and procedural memory. The individual memory systems work together [5, p. 8].

Working Memory

Working memory is different from short-term memory. The latter is only responsible for maintaining the content in the short term. In contrast, working memory is additionally involved in processing this information. Working memory accesses content in long-term memory [5, p. 6]. According to Baddeley, short-term memory forms one of the components of working memory [6]. In the phonological cycle, the linguistic content is stored and changed. This suggests a match with short-term memory. The visual-spatial notepad performs the same functions, but only in terms of visual and spatial information. Episodic memory combines information from different sources into an episodic form. The central executive controls attention and coordinates content from the phonological loop and the visuospatial notepad [7]. Working memory is characterized by limited storage capacity and short retention time [4, p.

370].

Retrieval methods

When retrieving content from long-term memory, two methods are used. In free-recall playback, the content is played back without cues or retrieval aids. Recognition involves presenting stimuli that are thought to have been perceived before and testing whether they can be recognized again. An example of this is multiple choice tasks [4, p. 379]. Recognition tasks show better recall results due to the representation of cue stimuli than tasks that have to be solved by free reproduction [8].

Position effects

Memory experiments have determined that items at the beginning and end of a list are better remembered than those in the middle of the list. It is therefore assumed that the memorized elements of the beginning of the list belong to the memory content of the LZG and those of the end of the list are found in the KZG [9]. The process by which the information at the beginning is more likely to be remembered is called the primacy effect, and the inputs from the list that are better remembered at the end are called the recency effect [2, p. 20].

Subtitle

The use of subtitles is very diverse. Subtitles can be, among other things, a support for learning foreign languages or basically for practicing general learning material [10]-[11] -[12]. Fang et al. compared bilingual subtitles with conventional monolingual subtitles and found that subjects performed significantly better than with monolingual subtitles [13]. Subtitles are much more important for the hearing impaired or deaf. Subtitles enable the aforementioned groups of people to perceive and interpret orally presented content with the help of written words in order to follow the visual content, e.g. in movies or series. In addition, there are already prototypes or approaches that support hearing-impaired people in their daily lives by transmitting verbal content to them in transcribed form in real time [14]-[15]. The study by Teófilo et al. presents an approach for subtitling theater performances in real time. Here, augmented reality and automatic speech recognition systems are used to realize the prototypes [16].

As diverse as the applications are, so are their realizations.

For example, Kushalnagar and Kesavan make suggestions for formatting static subtitles to increase the usability of this type of subtitle in different versions of hardware products [17]. Gorman, Crabb, and Armstrong take the approach of adaptive captioning, in which they allow viewers to personalize the functionality of the subtitles. For example, it should be possible to highlight certain speakers in the subtitling in color or to distinguish the individual speakers by means of color marking [18]. It was mentioned earlier that AR systems and automatic speech recognition systems were combined to develop prototypes to enable simultaneous captioning. Another aspect that will be considered is the search for user-friendly approaches in the field of 360° movies. Static and dynamic approaches are compared here: In the dynamic variant, the texts are placed near the speaker's source, while in the static variant the subtitles are simply inserted in front of the field of view [19]-[20].

GIF

Animated Graphics Interchange Format (GIF) is a digital file format that is ubiquitous in digital communication. GIFs are versatile and, combined with their endless repetitions, can convey multiple layers of meaning in a single image [21].

Memory performance between pictures and words

In one study, Paivio et al. examined images and sounds of objects and their visual and auditory names in free and serial recall. They found that nonverbal items were better remembered in free recall and verbal items were superior in serial recall. In visual modality, pictures were superior to words in both free and serial recall. However, there was no significant difference between sounds and words in free recall [22]. Paivo et al. have shown in a previous study that pictures are significantly better remembered in free recall than the verbal designation of objects [23].

Shepard and Roger determined the recognition scores of words, pictures, and sentences and came to a conclusion in their experiments indicating that the recognition rate is higher for pictures compared to words and that phrases rank last in relation to the other two modalities [24]. A different study compared students' free recall and recognition performance between words, black-and-white pictures, and colored pictures of objects. Here it could be seen that the performance of the subjects increased with increasing age. In addition, a significant difference was found between the color images, the black and white images, and the words among the adults. The performance for color images was higher than for black and white images and words [25]. The study by Snodgras et al. also showed that results were significantly better with pictures than with words alone. The experiments included a comparison between pictures and words and a group in which both stimuli were presented simultaneously. The group in which both stimuli were presented simultaneously did not perform better than the group that saw only the pictures [26].

III. THE AIM AND THE FORMULATION OF THE HYPOTHESES

In Sect. II, different memory models and two types of retrieval methods were presented. In addition, some research content of subtitles was addressed. In both static and dynamic subtitling, the transcribed content was mostly the verbal content of movies, series, real-time monologues or dialogues.

he aim of the present work was to look at sounds, tones and ubiquitously describable noises that occur in movies, series and also in everyday life. It is important to note that the sounds were transmitted to the viewer nonverbally, but not acoustically. This means that the subjects' memory performance was studied when they were presented moving images in the form of GIFs, which were the source of sounds, compared to their verbal variants in the form of written down words or short sentences. To put it compactly: The aim of this work was to find out how the insertion of visualized versus textual (background) sounds in short scenes with subtitles affected viewers' memorability.

For this purpose, the following hypotheses were derived from the Sect. II and examined in this study:

H1: Subjects' recall scores are higher for GIFs during free recall than for written-down words.
H2: Subjects' recall scores are higher for GIFs during serial recall than for written-down words.
H3: Subjects' recall scores are lower for recognition of written-down words than for GIFs.

IV. APPROACH

Experiments were conducted to test the hypotheses. From the aim and the hypotheses to be determined, it was clear that a comparison between two groups should take place. In order to compare the memory performance of the two groups, a separate questionnaire was designed for each group. The detailed description of the questionnaire can be seen in Sect. V. To test memory performance, subjects were asked to memorize the content. These contents were descriptions of sounds, tones and noises. A total of 30 descriptions were selected.

Firstly, a list of sounds and tones was created by listening to them on the site of Salamisound.de and geraeuschesammler.de. Subsequently, the names of the Sounds were sometimes expanded or supplemented. Afterwards, Gify.com was searched for gifs that best represented the source of the sound. As a test, a person was asked to describe these gifs in order to determine whether the selected sound was recognizable. Subsequently, it was decided to choose The Simpsons as a reference for the causative noise sources. Scenes from the series were selected for each of the 30 sounds. The selected scenes were cut to a length of about 12s and numbered consecutively. In total, an approximately 7-minute video was created for each group, including sources. In the text group, the sounds were described textually and appeared in addition to the monologue or dialogue subtitles in the scenes in which the sound occurred at that moment. In the visual group, the GIFs were faded in as soon as the corresponding sound occurred in the respective scene (seen in Fig. 1). Both videos were muted so that neither the monologues or dialogues nor the sound could be heard. The Clips were edited with the help of ShotCut. The Simpsons scenes were accessed via the Disney+ streaming platform."

Fig 1: The Simpsons Season 9 episode 21

A total of 64 people participated in the experiment. The survey lasted 11 days. Participation took place via Zoom or Discord. Some of the experiments took place in smaller groups of 2 to 4 persons, but for the most part they took place in individual settings. The experimental procedure was explained to all participants. Previously, they confirmed the consent form. After the explanation of the procedure, the participants were given the opportunity to answer their emerging questions. Once participants gave the signal to be ready, the video was screened once and participants viewed the content. After the video, they completed the two tasks included in the online questionnaire. Finally, participants entered their sociodemographic data.

V. METHOD

Two online questionnaires were created. In addition to demographic data, these included explanations and instructions for the two tasks to be completed. For demographic data, the following information was elicited: gender, age, degree program, and current semester. The participants were asked whether they had a visual impairment. If they answered in the affirmative, the follow-up question resulted as to whether they corrected them appropriately during the experiment. The two tasks were set up identically for both questionnaires. The first task consisted of 30 input fields, which were assigned to the 30 scenes. To check the serial recall, the participants should fill in the numbered fields with the correct designations according to the learned sequence of the contents. If subjects could not match a content due to lack of accessibility, they were asked to type it into any free input field. For fields where no content came to mind, they should fill it with any letter. The reason for this was that the input fields were mandatory. The second task was performed directly after the first and involved recognition of the items. It also consisted of 30 items.

In this case, the subjects did not have to fill in any fields. Comparable to a single-choice task, they could freely choose between four answer options that served as clue stimuli. Accordingly, there was only one correct answer. The other answer choices served to deceive participants because they had strong similarities to the correct answer. It should be mentioned that the subjects did not have the possibility to switch between the first and the second task during the processing. This ensured that the first task could not be manipulated by the second. In addition, subjects were instructed not to take written notes or recordings of the content presentation during the learning phase.

VI. RESULTS AND ANALYSIS

Statistical procedure

For statistical analysis, Matlab Statistics and Machine Learning Toolbox was used. An evaluation of descriptive statistics was performed. The age and gender distribution as well as the number of subjects were taken into account. Afterwards, the mean values of the two groups were calculated with respect to the retrieval types (free retrieval, serial retrieval, recognition) of the content and compared using t-tests for independent samples in order to draw conclusions about the differences between the two stimulus types with respect to the presentation types. To test the hypotheses, the t-test was calculated. The significance level was set at $p < 0.05$.

Descriptive statistics

A total of 64 people participated in the experiment. From the GIF group, the data sets of a total of four participants were discarded. The reason given was that the participants' responses did not come close to matching the responses of the other participants. In some cases, participants had misunderstood the tasks or completed the questionnaire too quickly compared to the other participants. To exclude outliers, a threshold of 5 correctly answered items was set. Subjects who scored less than or equal to 4 points were removed from the sample. The threshold is based on the assumption that subjects would have to recall at least 5 correct items from short-term memory alone. This consideration is consistent with the fact that 7 ± 2 items can be well reproduced by short-term memory [3]. The sample to be evaluated thus consisted of a total of 60 participants. Of these, 34 were students (57%) and the remaining 25 were working adults.

One youth, age 12, was in the text group. 41 of the subjects were male (68%). The mean age of subjects in the GIF group was 26.83 years (SD=6.29), in the text group 26.77 years (SD=5.87). The mean age of the total participants was 26.8 years (SD=2.97). The number of participants with visual impairment was identical in both groups. Among the 24 participants who had a visual impairment, 21 compensated for it during the test. The remaining three participants who did not compensate for their visual impairment were all in the text group. The number of participants in the two groups was not identical after sorting out the four data sets. In the GIF group, 33 people originally participated. However, only the records of 29 participants were analyzed. A total of 31 people attended the text group. Here the data sets were without outliers, so all data were retained.

Mean values in free recall

To investigate the first hypothesis, the means of the free recall items of the text and GIF groups were calculated and compared using a t-test for independent samples. The evaluation results showed that the GIF group (M=15.07, SD=4.26, n=29) achieved a higher mean score in free recall than the text group (M=13.29, SD=4.31, n=31) (Fig. 2). Comparison of means revealed no significant difference between the two groups, t-test(58)=1.58, p=0.060.

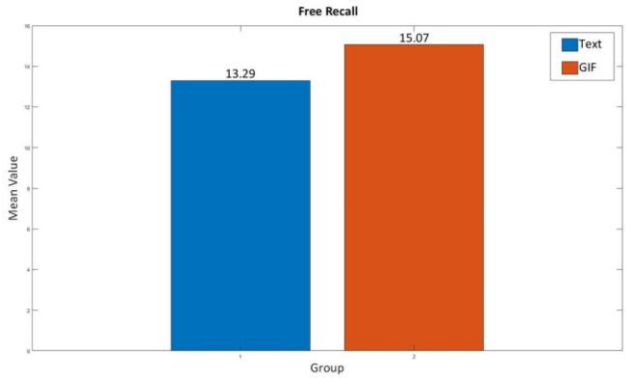

Fig 2: Mean values of the text and GIF group at free retrieval

Mean values in serial recall

To test the second hypothesis, the first step was to check whether the participants' correct answers matched the items. For example, if the text “cough” or the GIF of a person coughing was shown at the fourteenth position during the experiment, then the retrieval evaluation made sure that the response at the fourteenth position corresponded to the given label. In the evaluation, the means of the text and GIF groups were calculated and compared using a t-test. The mean for the text group (M=6.32, SD=4.14, n=31) was smaller than that for the GIF group (M=7.14, SD=4.18, n=29) (Fig. 3). Variance inequality cannot be assumed because Levene's test showed no significant result, F(1.58)=0.001, p=0.972. The comparison revealed no significant difference between the text and GIF groups in serial recall, t-test(58)=0.747, p=0.229.

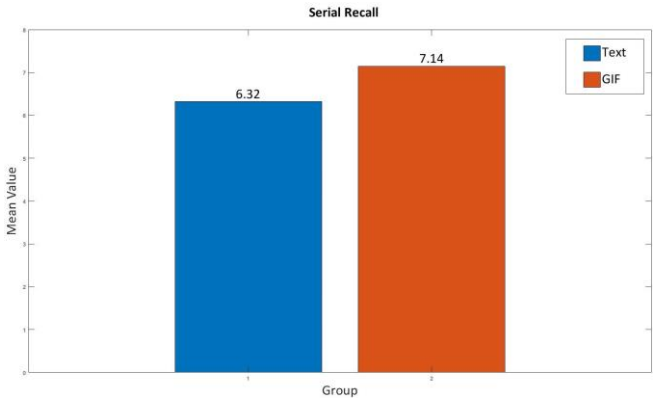

Fig 3: Mean values of the text and GIF group during serial recall

Mean values for recognition

To test the third hypothesis, the mean values of the items in terms of recognition were obtained for the text and GIF groups and compared using the t-test. The results of the evaluation indicate, as shown in Fig. 4, that the text group (M=15.74, SD=4.88, n=31) had a lower mean score than the GIF group (M=16.14, SD=4.40, n=29). The homogeneity assumption of variance was satisfied because Levene's test did not yield a significant result, F(1.58)=0.074, p=0.787. In the comparison performed, no significant difference was found between the means, t-test(58)=0.324, p=0.374.

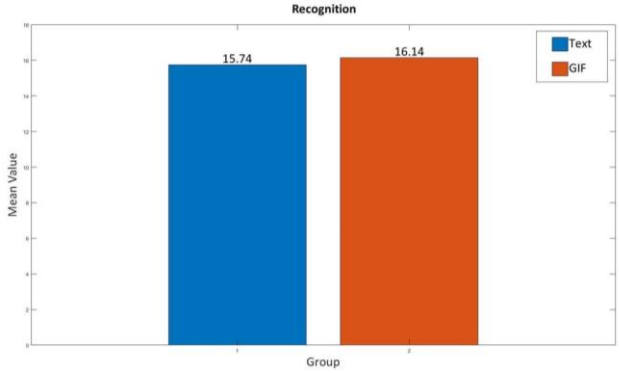

Fig 4: Mean values of the text and GIF group on recognition

Position effects

To analyze the primacy and recency effect in the two groups, the participants' correct answers to the respective items of the questionnaire were evaluated from free recall.

Position effects on text during free recall

Fig. 5 shows the position curve of the text group. It can be seen from the curve that the initial positions (primacy effect) flatten slightly as the number of positions increases. Exceptions are positions five and eight, where the sounds “incoming call” and “someone going down the stairs” are positioned. While the five was reproduced correctly more often, the recall values for the eight were significantly lower. In contrast, it can be seen that the curve rises again from the 20th to the last position (recency effect). Peak values occur repeatedly in the interval from position 11 to 26, but no maximum values are reached in this range as at the start and end positions. The maximum values are produced by sounds such as “fart sounds” (10), “bell

ringing" (15), "truck horns" (21), "saxophone sounds" (23), and "gunshots" (26).

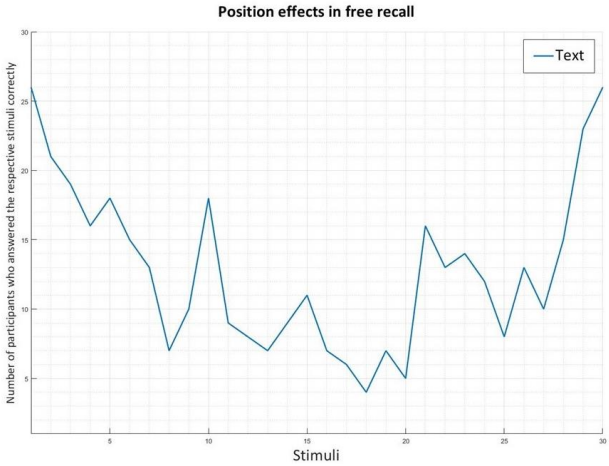

Fig 5: Position effects in text

Position effects on GIF during free recall

The GIF group also exhibited the primacy and recency effect. The initial positions show that the subjects found it easy to reproduce the content to be memorized without error. As the number of items increased, the scores of the individual items decreased. The lowest value was reached at the twenty-first position. In the interval of [1 - 21], individual peaks occur again and again. This suggests that certain sounds were better stored in memory regardless of position. Two much stronger deviations can also be seen in the interval]21,27[. These are caused by the sounds "police siren" (22) and "shots being fired" (26). From position 27, a rapid increase begins until the last position. However, the last element does not contain the same retrieval value as the start element (Fig. 6).

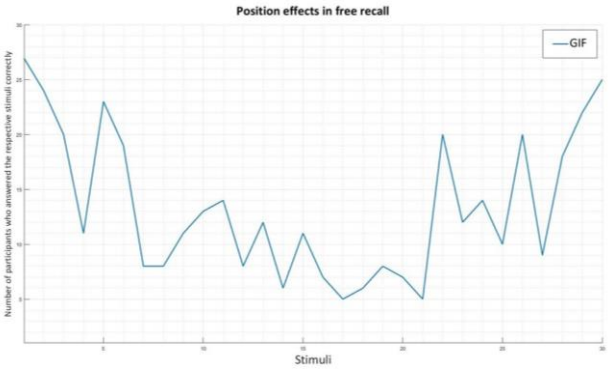

Fig 6: Position effects in GIF

VII. CONCLUSION

The study examined the comparison of nonverbal stimuli in the form of moving images (GIFs) with verbal stimuli in the form of written words. Short scenes from the cult series Simpsons were enriched with these stimuli to find out which stimulus form outperformed the other in terms of memory performance.

The assumption that GIFs have higher recall values in free recall compared to words could not be substantiated. When calculating the mean values, a difference was found. Although the mean value of the GIFs was higher than that of the words, the t-test did not show statistical significance. Therefore, hypothesis 1 had to be rejected. Hypotheses 2 and 3 could not be confirmed either. The calculation of the t-test did not show statistical significance. Especially

since in both serial recall and recognition, the mean scores of the GIFs were higher than those of the verbal stimuli. One reason for this could be the small number of subjects and the associated low test strength. Calculation of the t-test for hypothesis 1 yielded a p value of p=0.060, showing a trend near the significance level. It can be assumed that the size of the sample also had an influence on the results of the second and third hypotheses.

Another possible reason could be that the individual scenes also had an impact on the memorability of the nonverbal stimuli. Since both the scenes were moving images and the visualized sounds were GIFs that had to be memorized. In the case of verbal stimuli in the form of written words, the influence of the individual scenes did not exist since it was a different form of representation of stimuli. An aspect that attracted attention when coding the responses of the GIF group underlines this assumption again. In the GIF group, terms that described the content of the scenes rather than the GIFs were used more frequently. For example, the words “explode”, “burn”, “fireworks”, and “police” were reproduced in response. This indicates that participants' attention was focused on the content of the scenes and the stimuli they should have remembered faded into the background. This behavior was barely or only slightly evident in the verbal stimuli.

Another assumption is that subjects wanted to understand what was happening in the scenes, so participants' attention switched from the GIFs to the transcribed monologues or dialogues. Because the position of the GIFs was not the same as that of the verbal version of these stimuli, gaze switching occurred more frequently. This was not the case with the text, because the subtitles of the monologues or dialogues were in the same visual area as the verbal stimuli.

From the list in the last paragraph, one might conclude that these terms could be associated with emotions. Some of the 30 items that had to be memorized also had to do with emotions. Thus, it can be seen from the position curve for both the GIF group and the text group that the individual peaks contain terms that could be associated with emotions. These include, for example, “crying baby”, “fart sounds”, “thunder rumbling”, “truck horn”, “police siren”, and “gunshots falling”. The fact that these items in particular show high values can be explained by the fact that the subjects' attention was focused more on the verbal and nonverbal stimuli. The study, entitled “Emotion drives Attention”, indicated that emotionally charged stimuli received significantly more attention than neutral stimuli [27].

What could be considered from the analysis for future work would primarily be to look at both stimuli once with and once without movie scenes to compare how much the scenes distract the viewers' attention from the target stimulus and whether there is indeed a significant difference between verbal and nonverbal stimuli. In addition, it could also be investigated how strongly attention is drawn to emotion-linked stimuli in the verbal presentation form. In the case of the GIF group, the content was inserted at the lower left edge. As mentioned above, it is assumed that this would affect the results of the subjects' memory performance. For future work, it is a good idea to place the content at the same position coordinates where the words are inserted into the text group. Nevertheless, to avoid obscuring the content of the scenes too much, the GIFs should not have a background, but be transparent. However, it is important to choose a larger sample.

REFERENCES

1. S. Atkinson, Richard C., and Richard M. Shiffrin. "The control of short-term memory." Scientific american 225.2 (1971): 82-91.
2. Gruber, Thomas. Gedächtnis. Springer-Verlag, 2018.
3. Miller, George A. "The magical number seven, plus or minus two: Some limits on our capacity for processing information." Psychological review 63.2 (1956): 81.
4. Becker-Carus, Christian, and Mike Wendt. Allgemeine Psychologie: Eine Einführung. Springer-Verlag, 2017.
5. Frick-Salzmann, Annemarie. Gedächtnis: Erinnern und Vergessen. Springer Fachmedien Wiesbaden, 2017.
6. Gerrig, Richard J., and Philip G. Zimbardo. Psychologie. Pearson Deutschland GmbH, 2008.
7. Baddeley, Alan. "The episodic buffer: a new component of working memory?." Trends in cognitive sciences 4.11 (2000): 417-423.
8. Tulving, Endel. "Cue-dependent forgetting: When we forget something we once knew, it does not necessarily mean that the memory trace has been lost; it may only be inaccessible." American scientist 62.1 (1974): 74-82.
9. Glanzer, Murray, and Anita R. Cunitz. "Two storage mechanisms in free recall." Journal of verbal learning and verbal behavior 5.4 (1966): 351-360.
10. Ma, Qikun, et al. "InteractiveSubtitle: Subtitle Interaction for Language Learning." Proceedings of the Sixth International Symposium of Chinese CHI. 2018.
11. García, Boni. "Bilingual subtitles for second-language acquisition and application to engineering education as learning pills." Computer applications in engineering education 25.3 (2017): 468-479.
12. Grgurović, Maja, and Volker Hegelheimer. "Help options and multimedia listening: Students' use of subtitles and the transcript." Language learning & technology 11.1 (2007): 45-66.
13. Fang, Fang, Yuqing Zhang, and F. A. N. G. Yuan. "A comparative study of the effect of bilingual subtitles and English subtitles on college English teaching." Revista de Cercetare si Interventie Sociala 66 (2019): 59.
14. Mirzaei, Mohammad Reza, Seyed Ghorshi, and Mohammad Mortazavi. "Combining augmented reality and speech technologies to help deaf and hard of hearing people." 2012 14th Symposium on Virtual and Augmented Reality. IEEE, 2012.
15. Olwal, Alex, et al. "Wearable subtitles: Augmenting spoken communication with lightweight eyewear for all-day captioning." Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology. 2020.
16. Teófilo, Mauro, et al. "Exploring virtual reality to enable deaf or hard of hearing accessibility in live theaters: A case study." International Conference on Universal Access in Human-Computer Interaction. Springer, Cham, 2018.
17. Kushalnagar, Raja, and Kesavan Kushalnagar. "SubtitleFormatter: Making subtitles easier to read for deaf and hard of hearing viewers on personal devices." International Conference on Computers Helping People with Special Needs. Springer, Cham, 2018.
18. Gorman, Benjamin M., Michael Crabb, and Michael Armstrong. "Adaptive Subtitles: Preferences and Trade-Offs in Real-Time Media Adaption." Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021.

19. Agulló, Belén, Mario Montagud, and Isaac Fraile. "Making interaction with virtual reality accessible: rendering and guiding methods for subtitles." AI EDAM 33.4 (2019): 416-428.
20. Rothe, Sylvia, Kim Tran, and Heinrich Hußmann. "Dynamic subtitles in cinematic virtual reality." Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video. 2018.
21. Miltner, Kate M., and Tim Highfield. "Never gonna GIF you up: Analyzing the cultural significance of the animated GIF." Social Media+ Society 3.3 (2017): 2056305117725223.
22. Paivio, Allan, Ronald Philipchalk, and Edward J. Rowe. "Free and serial recall of pictures, sounds, and words." Memory & Cognition 3.6 (1975): 586-590.
23. Paivio, Allan, Timothy B. Rogers, and Padric C. Smythe. "Why are pictures easier to recall than words?." Psychonomic Science 11.4 (1968): 137-138.
24. Shepard, Roger N. "Recognition memory for words, sentences, and pictures." Journal of verbal Learning and verbal Behavior 6.1 (1967): 156-163.
25. Borges, Marilyn A., Mary Ann Stepnowsky, and Leland H. Holt. "Recall and recognition of words and pictures by adults and children." Bulletin of the Psychonomic Society 9.2 (1977): 113-114.
26. Snodgrass, Joan Gay, Rochelle Volvovitz, and Eleanor Radin Walfish. "Recognition memory for words, pictures, and words+ pictures." Psychonomic Science 27.6 (1972): 345-347.
27. Öhman, Arne, Anders Flykt, and Francisco Esteves. "Emotion drives attention: detecting the snake in the grass." Journal of experimental psychology: general 130.3 (2001): 466.
28. C. Y. Lin, M. Wu, J. A. Bloom, I. J. Cox, and M. Miller, “Rotation, scale, and translation resilient public watermarking for images,” IEEE Trans. Image Process., vol. 10, no. 5, pp. 767-782, May 2001.