Impacts of incidental collaborative bimodal presentation on list recall

Thi Le[1] and Zeth duBois[2]

[1] Department of Psychology, University of Idaho

Abstract

In this study we explore the effects of using bimodal format encoding on free recall performance in intermediate term memory. To control for variables and reduce complexity, this study focuses purely on the encoding task of a sequential list under two conditions of presentation. In both conditions, the items to be recalled are presented as clear typological English words presented with a background speaker narration, in audio only. In one condition, the words from the memory list incidentally coincide with the narrator's speech, while in the second condition, the list words have zero incidence in the transcription. We used a within-subjects design and ran repeated-measures ANOVA statistical test to check for ordered effects. The study found no significant difference between the number of words correctly recalled and the audio conditions. The environment is a limitation in the effective control of the study conditions. The study was completed via Qualtrics survey, disabling the researchers' ability to control for improper procedures or errors in remote participation. Furthermore, the statistical power was quite small. Future studies should have a larger size sample, and should include one other control condition to control for confounds.

Keywords: working memory, bimodal, recall, words

Impacts of incidental collaborative bimodal presentation on list recall

There are possibly as many methods to construct a study to test recall as there are permutations in presentation methods, creating an opportunity for junior researchers to hone skills in both study design and conduct, without duplicating prior published studies exactly. These distinctions are presented in two main categories; conditions for the encoding exposure, and type of recall, as either free, cued, or serial recall. We conducted a literature review to learn from prior work conducting similar studies.

Studies have shown the impact of using the bimodal format on working memory for the purpose of understanding the information transmitted in learning and teaching. How does bimodal presentation encode list items in intermediate term memory, without comprehensive abstraction? We review a fews studies using different methods of bimodal presentation found to impact either raw list recall or comprehension.

Thompson and Paivio (1994) found that bimodal encoding using picture-sound pairs increased recall performance not by encoding a combined memory icon, but rather by retaining the perceptual qualities of each stimulus separately. Their conclusion supported the idea of better recall performance in dual-modality encoding.

With an interest in contemporary use of multimedia for learning tasks, Zheng et al. (2021) showed that adding subtitles to audio lessons facilitated comprehension. Although the study was configured to score the benefit on comprehension, not simply raw list recall, the method of matching text on screen to information in the audio track is inline with our intended methods.

In Japan, Wen and Michihiko from Tohoku University (2005) conducted a study to investigate the effect of listening to either familiar or unfamiliar music and volume on working

memory. As a result, they found no difference between the participants who listened to the familiar music and those who listened to the song that was unfamiliar. This study is similar to our study the most since it asked participants to remember a list of words within a period of time and compare the free recall performance. However, this study uses sound and does not present the words one at a time, rather allows participants to spend six minutes with the recall list.

Daud and Suirman (2016) conduct a series of memory recall tests with subjects treated with either Mozart music or white noise. This is a within-subject design with the order of treatment randomly assigned. Additionally, the first memory test is designed to be ‘easy’, while the second is ‘hard’. With both easy and hard tasks, the groups given white noise treatment scored significantly higher than during a silent control condition. The report shows the Mozart treatment with a $p = 0.06$ in the hard test, nearly meeting a 0.05 significance criteria. The report does not include an order effects analysis, which we find to be useful in a between-subjects 2x2. Daud and Suirmann include an EGG measurement, with intent to discover how the memory task under audio conditions may be measured in brainwave recordings, but this is not relevant to our study as we are not seeking to research how treatment causes neurological effects, merely how they impact practical outcome.

Cautionary studies point to issues of divided attention. Craik et al. (1996) find that subjects given concurrent reaction-time tasks during the time allotted to memory encoding perform significantly poorer during retrieval. We highlight this study with the awareness that participant attention spent on processing audio while reading words may create a divided attention confound-in both the matching incidence condition, and the mismatching condition.

In the study of Bower and Grilligan (1979), they found that the free recall performance depended a lot on individual differences. In other words, the recall performance of each

individual will be influenced by how much of the information that they needed to remember related to them. This might seem not to be a concern for our study, but there is a possible chance that some participants found some words are easier to remember than other participants.

Aside from professional studies cited, our study is primarily based on one of our team member's undergraduate research. That study was conducted with a similar idea of having a bimodal format of the information input. As a result, the data supported the hypothesis that auditory input helped to increase the number of words correctly recalled. However, the study used a list of related terms generated from the same schematic topic for the treatment group, which might help the encoding and recall of terms by relational grouping, confounding the hypothesis. It would be easier for people to remember a list of words with a common theme.

To control for the variable of list "internal relativity", we redesigned the study to present participants with arbitrary words extracted from the underlying narrative audio. With relational word sets eliminated, the result of the study has higher internal validity. The purpose of our study is to investigate the effect of audio on working memory; especially when the visual information was presented simultaneously with visual information. In our research, we hypothesized that the number of words recalled in audiovisual with the words repeated by the audio higher than the condition with only the input by seeing the words from the screen.

Method

Participants

Participants are selected from informal network associates, and from associates from the university department. Invitations are sent informally through personal channels of the investigators. The participants were asked to do two conditions which were the experimental group and the control group ($N = 33$) from 19 to 58 years old ($M = 33$, $SD = 12.39$). There was no gender identity in this study.

Materials

All materials were delivered online in the body of a survey administered by Qualtrics. The treatment is composed of multimedia content in the form of two videos and one license free image extracted from the open web. Both videos were created by the investigators using video editing software. The primary source for the video content was extracted from a video on YouTube. Clips were extracted, and then altered by removing all video and keeping the audio track. The participants are recommended to use a device with a full keyboard such as a workstation or laptop to minimize error and delay during recall capture.

In this manner, two video treatments were prepared, called “match” and “mismatch” for internal reference. Figure 1 pictorial represents how the match condition is built by extracting words from the narration on a ten second interval, making a list which is not technically arbitrary, but which the researchers have no direct control of. Perhaps the process could be likened to a random “seed” generated output, similar to how computer code functions use a starting input to produce a deterministic output, yielding a pseudo-random result.

The mismatch list is an ‘arbitrary list’ created in the researchers’ minds. A random seeded function where they shouted out words and wrote them down.

Procedure

The study is administered by Qualtrics in the form of a single survey. Two treatment conditions in the form of videos are prepared as described in materials. Participants were asked to read and sign on the consent form in the page of the survey. The survey logic will randomly present either the match condition or the mismatch condition first. In both conditions, participants were asked to remember the list of words that will be presented on the screen in bimodal format (auditory and visual input) for about 2 minutes, 10 seconds per word.

After that they will be asked to do a free recall memory for the words they visualized. In the treatment condition, participants will watch the video with words on the screen repeated from the audio. Contrast to the treatment group, the audio in the control group did not repeat the words on the screen. Before participants start to do the first condition, they will be asked to test the audio by listening to a 5 seconds video. After completing the first task, participants were asked to do a filler task to refresh their mind as it will help them to perform the second task to minimize the impact from the previous task. The filler task is a finding game in which participants have to find 10 objects hidden in the picture within one minute. Participants are asked to report the number of objects they have successfully found. Next, participants were asked to do the second condition exactly the same as the first condition. There were two questions at the end of the survey that participants needed to answer in order for the researchers to know whether participants performed the tasks correctly. One of the questions asked to identify the speaker's gender. The researcher wanted to include this question to ensure that participants also listened to

the audio while trying to remember the words on the screen. The second question was asking the participant's age.

Results

The results of the study show a slight decrease in performance for the list of words with a matching vocal track, which is the opposite direction proposed by the alternative hypothesis (table 1). However, the mean difference between the two treatments lacks significance (figure 2). Repeated measures anova statistical results including effect size check, partial eta-square, as follows, $F(1,31) = 0.389$, $p = 0.538$, $\eta^2_p = 0.012$. Table 1 and Figure 2 shows the output of statistical analysis provided in software.

Given that the experiment was a time series with two different treatments, it was determined in advance that a within-subjects design would allow the results to be tested for series effects; namely, if running the match treatment before the mismatch treatment would differ from mismatch preceding match. The survey tool we used to administer both the treatment and the post treatment testing, allows for the random assignment of the treatment order, and ensures balanced numbers between the two.

Additionally, within subjects testing allowed for the small sample size to have increased power. This is the reason for running repeated measures anova instead of a simple pairwise t-test.

To check for order effects, between subject effects checks were added, $F(1,31) = 0.499$, $p = 0.485$, $\eta^2_p = 0.016$.

Discussion

Memory tests are prolific fodder for psychology research, especially for academic environments, given the relative ease of instrumentation and a broad accessibility for participants. Given the sheer volume of testing on memory conducted, the authors are unsure if a test with this particular method has been attempted before, but with experience from this study, we can make some confident conclusions about the results.

First, it must be strongly noted that working memory capacity, such as found in untrained participants, has a well understood limit of about seven plus or minus 2 (7 +/- 2) items (Miller, 1994). Furthermore, recall tests with the best success allow for the participants to survey the complete set for the experimental duration, enabling the opportunity to create reinforceable patterns that may help retention and recall. Our experimental challenge, without the addition of audio, presents a sequential list of terms over ten second intervals for two minutes, giving the participants no options for casual organization strategies beyond series association.

With Miller's limit in mind, the distribution of scores for a memory test under the best circumstances would score from five to nine (5-9) items. Each subject tested will have an individual score, *only integers,* in and around that narrow band of working memory limits. That is to say, if a subject has a native working memory-to-recall score equivalent to mean population average, namely 7, a minimum 14% change in performance would be required to result in a measurable deviation at all. Mathematically expressed, (7 +1)/ 7 = 1.14.

Beyond this pragmatic observation, considering the experimental design from even a naive perspective, the notion of improving recall under the testing condition of this experiment seems highly improbable. Stated simply, expecting an arbitrary list of words to have more

successful recall merely because they coincided with a periodic extraction from the full context of an oratory accompanying the visual list, seems a long shot indeed.

The authors conceive of a plausible modification to the experiment. Keeping the same methodology of a periodic serial list as the challenge set, instead of a verbose speech with occasional coincidence, the audio could be created to *only* state the word to be memorized--eliminating extraneous verbiage--thus improving "collaborative bimodal recall". Such a trial would still need to overcome the integer limitations mentioned above--14% improvement minimum, on average.

Limitations

Collecting data via Qualtrics, with no human administrator at hand, troubleshooting, eliminating breeches in procedure, assisting in questions, or rectifying misunderstandings during the study was not possible. The researcher cannot control that every participant has the same environment, control for distraction, errors or delays in technology platforms or internet connection or other impacts that could disturb encoding or disrupt testing.

Future studies would also benefit from a larger sample size. Suppose the future studies still collect data via Qualtrics or other online methods. In that case, the researchers should design their instruction carefully to ensure participants take the study seriously without any distractions.

Moreover, it was noticed by the researchers after data collection began that participants could encode and/or recall words from the list with deviations in minor phonology. For example, the list word "number" could be either encoded or recalled as "numbers", the plural. "Communicate" could become "communication". Furthermore, using forms delivered by internet, participants' browsers could make auto-correction changes to entries, or the participants could even simply misspell or mis-key their entries.

This became clear as the data was cleaned for analysis. Confounds discovered of this nature are shown in Table 2, with the terms prone to error boxed in red, and the associated errors submitted listed with yellow highlight at the bottom of the columns. Table 3 shows the changes in dependent variable 1 and dependent variable 2 scores in columnar comparison. Numerical scores highlighted in orange in the corrected scores column are scores that were higher after correction, demonstrating that the match condition scores, corrected for confound, could have been considerably higher.

For clarity, we ran statistical analysis on the corrected scores, and found that our alternative hypothesis was in fact captured--participants had a higher mean recall score with matching audio. However, the difference was still below significance. See Figure 3 for post-hoc corrected descriptive plots.

References

Bower, G. H., & Gilligan, S. G. (1979). Remembering information related to one's self. *Journal of Research in Personality*, *13*(4), 420–432. https://doi.org/10.1016/0092-6566(79)90005-9

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The Effects of Divided Attention on Encoding and Retrieval Processes in Human Memory. *Journal of Experimental Psychology. General*, *125*(2), 159–180. https://doi.org/10.1037/0096-3445.125.2.159

Miller, G. A. (1994). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *101*(2), 343–352. https://doi.org/10.1037//0033-295X.101.2.343

S. S. Daud & R. Sudirman. (2016). Discovering sound effect on visual memory performance based on electroencephalography. *2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, 210–215. https://doi.org/10.1109/IECBES.2016.7843444

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

Wen, W., & Michihiko, K. (2006). The effects of music type and volume on short-term memory. *Tohoku Psychol Folia*, *64*, 68–76.

Zheng, Y., Ye, X., & Hsiao, J. H. (2022a). Does adding video and subtitles to an audio lesson facilitate its comprehension? *Learning and Instruction*, *77*, 101542. https://doi.org/10.1016/j.learninstruc.2021.101542

Table 1

Within Subjects Effects ▼

	Sum of Squares	df	Mean Square	F	p	η^2	η_p^2
Treatment	0.737	1	0.737	0.389	0.538	0.012	0.012
Treatment ✳ order	0.009	1	0.009	0.005	0.944	0.000	0.000
Residual	58.748	31	1.895				

Note. Type III Sum of Squares

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2	η_p^2
order	4.398	1	4.398	0.499	0.485	0.016	0.016
Residual	273.086	31	8.809				

Note. Type III Sum of Squares

Table 2

Ambiguous phonology prone to testing errors

match list!	**unmatch list!**
top	math
spoke	perspective
minute	different
think	people
explain	secret
communicate	ability
thing	number
illustration	happen
useful	image
world	volume
important	understanding
research	letter
minute	numbers
things	happens
illustrate	happening
communi	images
communication	

Table 3

Correcting DV1 and DV2 for confounding ambiguity

dv1 score	dv2 score	dv1! score	dv2! score
4	4	4	4
4	4	5	4
8	9	9	9
5	3	6	3
3	4	3	4
5	7	6	7
5	4	5	4
2	5	3	5
5	8	5	8
7	6	8	6
10	6	10	6
4	6	5	6
6	6	7	6
11	9	12	9
4	5	4	5
1	1	1	1
5	4	7	4
3	3	3	3
6	7	7	7
6	5	6	5
3	5	4	5
7	4	7	4
9	8	10	8
4	5	6	5
2	7	3	7
3	2	3	2
1	3	1	3
5	4	7	5
2	5	2	5
2	0	2	0
6	6	7	6
6	5	6	8
7	8	7	8

Figure 1

Creating the recall list, match vs mismatch

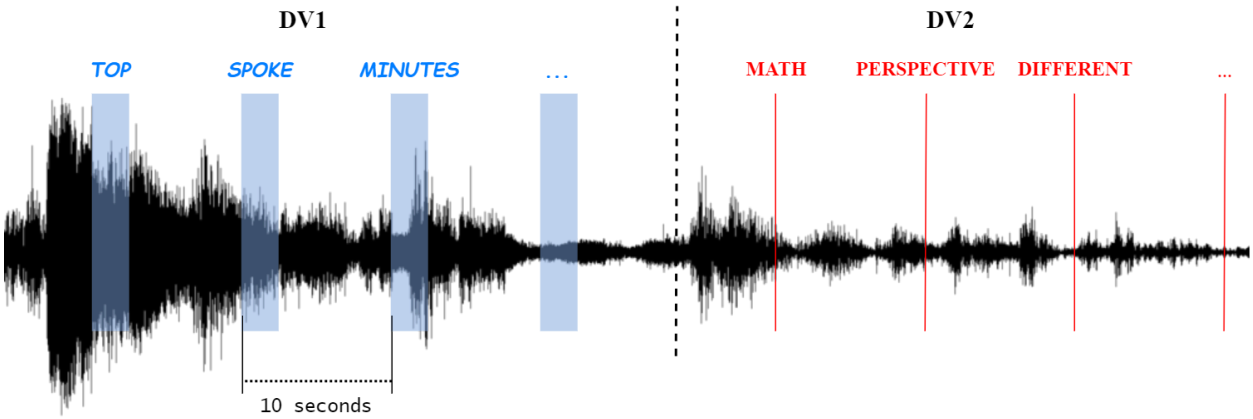

Figure 2

Descriptive plot of results

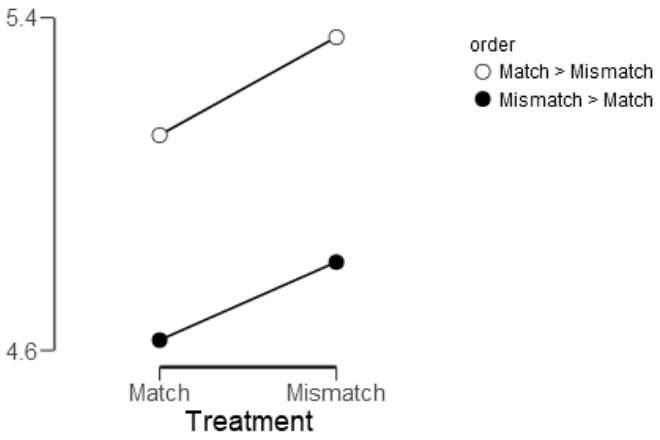

Figure 3

Descriptive plot using hand-corrected data