Advances in Social Science, Education and Humanities Research, volume 615
Proceedings of the 2021 4th International Conference on Humanities Education and Social Sciences (ICHESS 2021)

Selecting Audio and Visual Stimuli: What Do We Prefer If They Happened Simultaneously?

Yutong Ding[1, †], Yandi Wu[2, †] and Taojun Zhang[3, *, †]

[1] *Malvern College Qingdao, Qingdao, China*
[2] *Rabun Gap Nacoochee School, Georgia, United States*
[3] *Syracuse University, New York, United States*
[*] *Corresponding author. Email: tzhang52@syr.edu*
[†] *These authors contributed equally.*

ABSTRACT
Visual and audio stimuli become increasingly important in human lives as technology advances. To consider how humans receive, process, and react to stimuli, we investigate attention, specifically selective attention. Selective attention describes how humans focus on a particular stimulus in the environment for a certain period. The present study aimed to examine how do individuals make choices between auditory and visual stimuli. Ten participants with the age range from 15 to 22 years participated in this study. They were asked to choose between pairs of visual and auditory stimuli. The reaction time and congruence were recorded. The result showed that, compared to auditory stimuli, participants preferred responding to visual stimuli rather than auditory stimuli. The time it took to answer visual stimuli was longer than auditory stimuli, and the results were consistent. According to the result, we can conclude that individuals were more sensitive to visual stimuli than auditory stimuli.

Keywords: *Attention, Visual stimuli, Auditory stimuli*

1. INTRODUCTION

Attention has been the center of psychology research for decades. Attention is a person's mental processing capability of objects [1]. Attention has numerous forms, including selective attention, divided attention, and automaticity. While they often accompany each other in understanding the human mind's process, there's a distinction between attention and perception. Attention often takes place before perception and sensation. It involves focusing mental processing capacity on something to exclude other possible objects or thoughts [1]; focalization and concentration of consciousness are of their essence [2]. However, when multiple stimuli attract one's attention simultaneously, it would be difficult for them to focus specifically on a particular stimulus. Just as the cocktail party effect, coping with the scenario, people have attended an unattended channel to pay attention to selected stimulus while filtering out the other stimulus. These two channels are when people consciously or unconsciously pay attention to specific subjects while selectively neglect background information. They are also part of a larger concept: selective attention.

Selective attention is the process of focusing on a particular object in the environment for a certain period. The selective attention researched in this study also involves attentional change, which is often closely related to conscious perception changes. A new stimulus attracts attention and populates consciousness [3].

Additionally, a phenomenon known as attentional capture is also one of the main focuses of this study. With the increasingly rapid development of technologies, both visual and audio stimuli are in every aspect of our lives. They often attract people's attention quickly, prompting them to divide their attention onto the new alternative [4].

Recognizing the increasingly important role that attention plays in people's daily lives, more researchers have studied the topic, specifically focusing on visual and audio stimuli. There has been a finding regarding the physical processing of objects in the unattended auditory channel [5], suggesting that even in the absence of attention, the unattended background information is still being extensively processed implicitly. Another research further strengthens the idea that even unattended stimuli may be semantically processed and that people consciously attend to stimuli that they deem as task-

relevant [6]. This idea can also be perceived in a more recent study that people's sensory systems rely on common principles for extracting relevant sensory events [7].

In addition, one research regarding the visual cortex found that correspondence across attention state between neuronal and behavioral performance on a given task is restricted to specific regions of the visual cortex [8]. Attention is allocated to visual search targets to resolve ambiguities in neural coding that arise when multiple objects are processed simultaneously [9].

However, not many present types of research have combined visual and auditory processing and study their impact on the human responsive system together. Most of them have focused exclusively on how human brains process either auditory or visual stimuli one at a time, but not together.

This study is almost the first study combining both visual stimuli and auditory stimuli together. We aim to measure whether the users are more sensitive to auditory stimuli or visual stimuli. The hypothesis is that individuals are more sensitive to visual stimuli than auditory stimuli when facing a computer screen in a quiet situation. Specifically, we recorded the participants' frequency, reaction time, and the congruence of the two different kinds of stimuli. We predict to get higher frequency, less reaction time, and more congruent answers due to the visual stimuli.

2. METHODS

2.1. Participants

Ten participants were recruited for this experiment voluntarily. Among the participants, the age ranged from 15 to 22 years of age. All the participants were familiar with the experiment after the directions were given, and they were all right-handed and had normal or corrected-to-normal visions. All the participants reported neither significant history of sensory disorders nor psychotherapy. Furthermore, this study was approved by the local ethics committee.

2.2. Materials

One set of 15 pairs of stimulations were utilized in this experiment, in which each pair contains a visual stimulus and an auditory stimulus. Each stimulus contained a single word that is commonly seen (e.g., mouse, month). The participants presented the visual stimuli by merely showing the exact word on the center of the screen. The auditory stimuli – the recording of the target word-- were presented to the participants through the computer speaker, and an automated voice generator generated the voices. Psychology was the platform that organized and presented the stimuli, and the entire experiment should take less than 200 seconds to complete [10].

Figure 1. The flow chart of the experiment.

As shown in figure 1, The introduction message was displayed on the screen, which only appears once for the entire experiment. The first picture from the left displays how a visual stimulus is being presented. Mind that the auditory stimulus is being played from the background as soon as the visual stimulus is shown. The second picture from the left demonstrates the General Reaction phase. The participants are prompted to either press “f” for choosing visual stimulus or “J”; for choosing auditory stimulus. The rightmost picture demonstrates the Content specific phase, in which the participants are also prompted to either press “f” or “j” to confirm the actual content of the stimulus they have seen.

2.3. Procedures

To ensure that the user received the message from both the image and the audio, we distributed simple tasks for the user to complete on Psychopy [10]. The experiment was divided into two phases: the General Reaction phase and the Content-specific phase. Both served the purpose of measuring the preference of the user. In both phases, two stimuli would be presented simultaneously during each iteration, which meant the audio stimuli (recording from the voice generator) would be played in the background. In contrast, the visual stimuli (the actual word) would be displayed on the screen simultaneously. There would be 15 iterations in total.

In the General Reaction phase, the participants would indicate whether they detected the visual or audio stimulus first by pressing “f” for the audio stimulus and “j” for the visual stimulus. This phase served to measure the user's first instinct, as we would want as little time as possible the users spent answering the preferred stimuli [11].

The content-specific phase is, in many aspects, very similar to the General Reaction phase. Still, there are 2 differences: 1) The users were asked to press the “f” or the “j” button to the actual word they see/hear instead of the kind of stimuli. The answers were labeled adjacently: the visual stimulus was still listed on the left, and the auditory stimulus was still listed on the right. And 2) Instead of measuring the reaction speed, we measured congruency as if the users' answers in the Content-specific phase were consistent with those in the General Reaction phase. Incongruent answers could raise concerns in the analysis, but the data would be considered [11].

2.4. *Data analysis*

The software used to analyze the data was JASP [12], in which we performed a t-test, and descriptive statistics test it. All of the data were presented in the results section.

2.4.1.Frequency

After data were collected, the frequencies of answering auditory and visual stimuli were added up separately for each participant. Therefore, each user was assigned to two frequencies: auditory and visual, which both were documented as integers. Before any analysis, data 3 standard deviations away from the mean were removed. Then, a t-test was performed to evaluate the significance of the data. More specifically, the t-test was used to evaluate whether there was a significant difference between the frequencies of the two stimuli being answered.

2.4.2.Reaction Time

Upon completing the data collection, the reaction times of answering both auditory and visual stimuli were analyzed. Unlike the frequency data, we did not calculate the mean reaction time for each kind of stimuli for each user. Rather, we compiled all the reaction time from every participant for each kind of stimuli as a sample; therefore, there would be two samples. Then, like frequency analysis, we removed the data that were 3 standard deviations away from the mean value before any analysis. Finally, a t-test was carried out to evaluate whether there was a significant difference in the means between auditory and visual stimuli reaction time.

2.4.3.Congruence

Once the data collection process had terminated, the numbers of the congruent answers for each individual participant were also added up separately for both stimuli. Therefore, like frequency analysis, each participant was assigned two pieces of data. After removing the data that were 3 standard deviations away from the mean value, the t-test was used to evaluate whether there was a significant difference between the number of congruent answers from the users.

3. RESULTS

Frequency. As noted in Table 1 below, the frequency of participants (*M*=5.400, *SD*=2.675) choosing auditory stimuli was lower than that of visual stimuli (*M*=9.600, *SD*=2.675). In addition, the same standard deviation for the frequency of choosing both stimuli (*SD*=2.675) could imply considerable stability of the data measured. Moreover, the high absolute t-value under the two-sample t-test (*t*=-2.483, *p*=0.035) could also indicate the significant differences between the frequencies of the two kinds of stimuli. Figure 2 documented the mean of the frequency of both stimuli being chosen by the participants, with error bars showing one standard deviation from the mean.

Table 1. Descriptive statistics of every variable under each kind of stimuli

Variables	Frequency		Reaction-time	
	M	*SD*	*M*	*SD*
Visual	9.600	2.675	1.952	1.397
Auditory	5.400	2.675	1.928	1.952

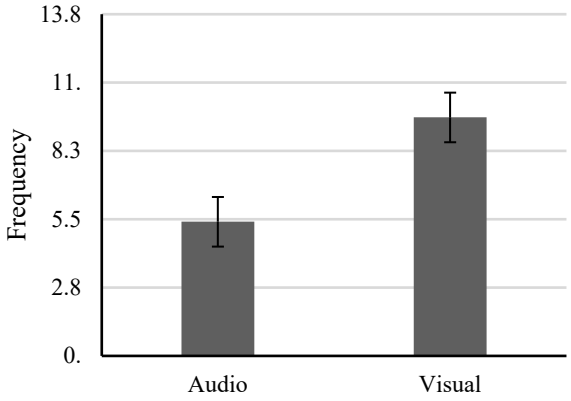

Figure 2. The frequency of both stimuli being chosen by the participants.

Reaction Time. As noted in Table 1 below, the reaction time for the participants for the visual stimuli (M=1.952, SD=1.397) was longer than that of the auditory stimuli (M=1.928, SD=1.952). However, the significant standard deviation difference between the reaction time for auditory stimuli (SD=1.952) and that of the visual stimuli (SD=1.397) could showcase the relative instability of the time needed to answer the auditory stimuli. In addition, the higher absolute t-value for the two-sample t-test (t=-1.343, p=0.186) could also indicate the significant differences between the reaction time when the participants are answering the two kinds of stimuli. However, the significance level was lower than the other two variables (p=0.186 compared to p=0.035). Figure 3 documented the mean of the reaction time of both stimuli being chosen by the participants, with error bars showing one standard deviation from the mean.

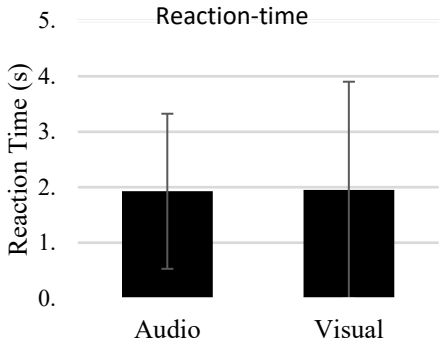

Figure 3. The reaction-time of both stimuli being chosen by the participants.

Congruence. To measure the consistency of the participants' answers, we computed congruence which each piece of datum should ideally match the corresponding ones in the frequency section. In fact, the congruency of the participants and the results for the t-test completely matched the data in the frequency analysis, which implied that the participants made all the choices they intended to. No technological nor operation errors existed.

4. DISCUSSION

The present study investigated the sensitivity of humans to auditory and visual stimuli. We explored how people react to simultaneous visual and auditory stimuli in a quiet environment. Comparing the frequency of the participants' choices (M=5.400, SD=2.675 for the auditory stimuli; M=9.600, SD=2.675 for the visual stimuli; p=0.035 for the two-sample t-test), we could conclude that in a quiet environment, people were more likely to notice visual stimuli compared to auditory stimuli. Comparing the reaction time of the participants (Which as noted in table 1, M=1.928, SD=1.952 for the auditory stimuli; M=1.952, SD=1.397 for the visual stimuli; p=0.186 for the two-sample t-test), we could conclude that there was less logical reasoning occurring when the participants chose the auditory stimuli than the visual stimuli. The datasets for visual stimuli were stabler because of the lower standard deviation. The reason that caused this might be because the multi-stage memory model divided memory into sensory memory, short-term memory, and long-term memory. Short-term memory was formed when sensory memory was attended to. There were two main encoding methods, icon, and echo, in the short-term memory formation stage. In the experiment, participants might subconsciously attend to visual stimuli. In the choice response task, humans might make decisions based on subconscious visual information [13].

The results of this experiment are consistent with a cocktail party effect. When individuals attended to one stimulus first, although they did attend to the later stimulus, they did not clearly remember the details of the next stimulus. Congruent data during the content-specific and general reaction phases provided a clearer picture of which stimuli the participants noticed first compared to incongruent data.

Any inconsistency in the content-specific phase data with the data in the general reaction phase did not prove that the participants' choice in the general reaction phase was invalid. This was because, in the experiment, we asked participants to choose in as short a time as possible. After the participants noticed the first stimulus and formed a short-term memory, they also developed a short-term memory for the second stimulus. This change is closely related to the change in conscious perception. The new stimulus attracts an individual's attention and fills the consciousness [3]. This might have led to inconsistent choices among both phases as participants were influenced by the short-term memory formed for the first stimulus.

5. LIMITATIONS

This study was not free of limitations based on the methods and the demographics of the participants.

Diversify the forms of stimuli. First, our experiment only tested participants' responses to visual and auditory stimuli with different words. It did not test people's responses to other kinds of visual and auditory stimuli, such as positive and negative emotions of sound in auditory stimuli. Positive emotions could contain happiness, joyfulness, emotions that could make humans feel joyful. In contrast, negative emotions could contain sadness, anger, anything that would make a human go through a difficult emotional time. In the future, researchers should further investigate whether the participants would prefer auditory or visual stimuli with the content that conveyed positive emotions and the same with negative emotions. Additionally, emotion regulation would be an important factor when performing the analysis [14]. Therefore, in future related experiments, more research is needed to apply and test people's sensitivity to these two stimuli in depth by adding more kinds of auditory and visual stimuli to the participants.

Gender effect. Second, factors that interfere with the experimental results include the gender effect. There were sex-related differences in pitch, brightness, and loudness discrimination, with men performing better than women [15]. Future research should consider the potential effects of sex-related differences more carefully. For example, the researchers should have divided the male and female participants into two groups before conducting the experimental study.

The structure of this experiment. Over the course of this experiment, we found out that the preference of the visual stimuli outnumbered the auditory stimuli. Prior to finalizing the ultimate methods presented in this work, we had chosen to delay the appearance of the visual

stimuli until 0.6s after the auditory stimuli had been presented in our first prototype. The results appeared inverted from the ones presented in this paper: the number of the preference of the auditory stimuli significantly outnumbered that of the visual stimuli, and we had encountered the users that chose the auditory stimuli for each of the 15 iterations. Both prototypes could result in biased responses from the participants, further leading to less accurate response times. We needed to calibrate the prototype to lessen the bias effect before the data collection process in future research.

Despite our study by the above limitations, these findings had two important implications for future research. To begin with, this study provided important practical implications for divided attention. Our experiment was divided into attention-related tasks, which meant the tasks would interfere with each other when the overall demand for resources exceeded the mental resources available. Our experimental results provided data on reaction times for a divided attention experiment on people receiving visual and auditory stimuli simultaneously. This provided detailed experimental data for future related experiments. Moreover, the finding of our experimental research could be applied to video advertisements. That was when watching an advertisement, and people would first notice the visual information. Therefore, advertisements needed to focus more on the visual information that was conveyed to the viewer. At the same time, the producer of the video could convey the key message of the video to the viewer more easily by adding subtitles or other visual information to the video.

6. CONCLUSION

There had been a tremendous amount of studies involved in how humans perceive information from different kinds of stimuli. Overall, this study found that humans preferred simple forms of visual stimuli rather than auditory stimuli, while the reaction time for visual stimuli was longer than that of auditory stimuli. We could further use the result of this study to improve the efficiency of how information could be communicated, such as advertisement.

REFERENCES

[1] McCallum, W. Cheyne (2015). Attention. Encyclopedia Britannica. https://www.britannica.com/science/attention

[2] James, W. (1890). The Principles Of Psychology. Henry Holt and Company.

[3] Hohwy, Jakob (2012). Attention and conscious perception in the hypothesis testing brain. Frontiers.

[4] Allport, D. A., Antonis, B., & Reynolds, P. (1972).On the division of attention: A disproof of the single channel hypothesis. The Quarterly Journal of Experimental Psychology, 24(2), 225–235.

[5] Alho, K. (1992). Selective attention in auditory processing as reflected by event‐related brain potentials. Psychophysiology, 29(3), 247-263.

[6] Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. Behavioral and brain sciences, 13(2), 201-233.

[7] Kayser, C., Petkov, C. I., Lippert, M., & Logothetis, N. K. (2005). Mechanisms for allocating auditory attention: an auditory saliency map. Current biology, 15(21), 1943-1947.

[8] Maunsell, J. H., & Cook, E. P. (2002). The role of attention in visual processing. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 357(1424), 1063-1072.

[9] Luck, S. J., & Ford, M. A. (1998). On the role of selective attention in visual perception. Proceedings of the National Academy of Sciences, 95(3), 825-830.

[10] Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. Behavior Research Methods. 10.3758/s13428-018-01193-y

[11] Saija, J.D., Başkent, D., Andringa, T.C. et al. (2019) Visual and auditory temporal integration in healthy younger and older adults. Psychological Research 83, 951–967.

[12] JASP Team. (2020). JASP (Version 0.14.1) [Computer software]

[13] Leukel, C., Lundbye-Jensen, J., Christensen, M. S., Gollhofer, A., Nielsen, J. B., & Taube, W. (2012). Subconscious visual cues during Movement Execution allow correct Online CHOICE

[14] Nezlek, J. & Kuppens, P. (2008). Regulating Positive and Negative Emotions in Daily Life. Journal of personality. 76. 561-80. 10.1111/j.1467-6494.2008.00496.x.

[15] Rammsayer, T. H., & Troche, S. J. (2011). On sex-related differences in auditory and visual sensory functioning. Archives of Sexual Behavior, 41(3), 583–590.