Research practical report

Representational similarity analysis of pairs of visual stimuli

Introduction

Perception can be described as a sensory experience of the world. This process results in enhanced recognition of stimuli and depends on the processing of a sensory stimulus (Badde et al., 2020). Humans are able to perceive the world with all of the five senses, including through sight and hearing, and the integration of information from different sensory modalities, forming cohesive perception (Ernst & Bü, 2004).

The notion that mental representations are formed of words and images alone is widely accepted in the literature (Nelson & Castaño, 1984). As visual and auditory perception are involved in the representation of information within the mind; it is hypothesised that memory and learning of verbal or visual stimuli can be enhanced when the corresponding verbal associations or visual imagery is also presented. This premise, derived from the dual coding theory of multimedia learning, suggests that an individual can build connections between the visual and auditory stimuli through the presentation of both verbal and visual stimuli (Mayer & Sims, 1994). It is known that the presentation of multimedia stimuli requires both spatial and working memory (Brunyé et al., 2008), supporting this assumption that audio-visual stimuli must be dually coded. Through understanding whether perception of visual images can be enhanced through an accompanying auditory stimulus, this theory can be supported. This is because examining how similar pairs of audio-visual stimuli are perceived to be related to one another, it can be determined whether an individual is likely to have stored a particular item through coding it in one (either visual or verbal) or two (visual and verbal) ways (Shen, 2010).

This report examines the relationship between pairs of visual stimuli, identifying whether this hypothesis can be supported using the computational technique of representational similarity analysis (RSA). Understanding of the perception of visual stimuli alone can determine how visual stimuli are perceived in relation to one another and can be compared with the findings of how the same visual stimuli are perceived when presented alongside an auditory stimulus.

Methods

The data was obtained from 5 individual participants, as this was the provided dataset. A multiple arrangements task was used to arrange the stimuli with distance representing how similar two stimuli were; stimuli perceived as being similar were placed close together whereas dissimilar pairs of stimuli were placed apart, as shown in the figure below.

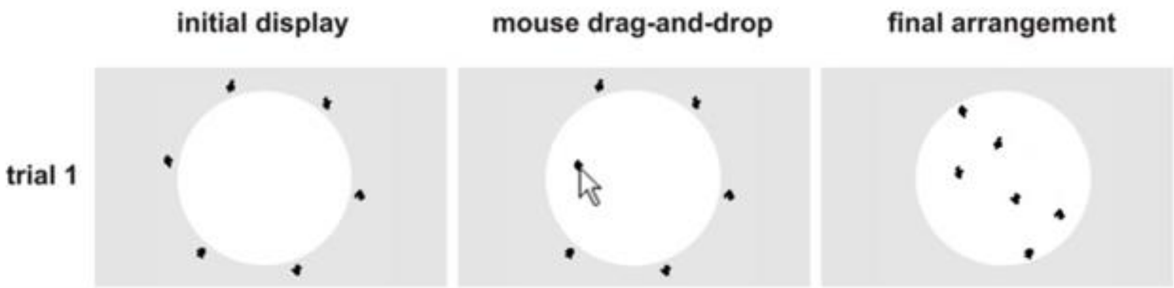

Figure 1. Illustration of a multiple arrangements task. Adapted from (Kriegeskorte & Mur, 2012)

In this multiple arrangement task, participants were initially presented with all 50 stimuli under 4 different conditions. Each stimulus was from a range of categories, such as people, animals, and natural scenery and the conditions were: 3 second videos with sound, videos without sound, video's sound only and single representative static frames. Following the initial trial under each condition, objects were presented in groups, allowing stimulus pairs to be positioned representing their similarity within a group. Using the positions of where the objects were placed in relation to each other, a representational dissimilarity matrix (RDM) was created.

To do this, the data points were loaded in python to form two arrays, each array holding the data associated with one of the elements in the pair of stimuli. The Euclidean distance between each element of the two arrays was calculated, using the cdist method in the SciPy spatial module; the distance represented the similarity between pairs of stimuli with 0 representing most similar. To form an RDM, as RDM's are symmetrical about the diagonal of zeros, the matrix constituted of the lower triangle, which is equivalent to the lower triangle of the matrix. The diagonal was then subtracted, representing the pairs of stimuli where both elements of the pair were the same item, as being most similar, 0. The resulting matrix represented similarity between a pair as being dark blue and similarity as being yellow from a scale of 0 to approx. 0.05. An individual RDM was constructed using the data points of each participant's multiple arrangements task.

To determine the existence of a linear relationship between pairs of stimuli, the Pearson Correlation Coefficient was calculated using each participants RDM. This was computed using the pandas DataFrame.corr function, which computed the pairwise correlation of the dataset using the Pearson method of correlation. A value of 0 represented no linear relationship between a pair, a value of -1 represented a perfect negative relationship and a value of 1 indicated a perfect positive linear relationship. In this project, based on the interpretation of the Pearson Coefficient by Rea and Parker (2014), the following understandings of the r-value were used:

$\|r\|$	Interpretation
0.00 < 0.10	Negligible
0.10 < 0.20	Weak
0.20 < 0.40	Moderate
0.40 < 0.60	Relatively strong
0.60 < 0.80	Strong
0.80 <= 1.00	Very strong

In this study, the independent variables are the stimuli, and the dependant variable is the distance between each pair of stimuli. Each stimulus is presented with each of the other 50 stimuli, from which the distance between a pair of stimuli is represented in the circular area, representing the similarity between the stimuli. In this project, the trials conducted under the single representative static frames condition was used.

Results

The sample consisted of all five participants which completed the multiple arrangements task, the RDM's of all five participants is shown in figures 2 to 6 below. For the purpose of this report, the findings and analysis of participant 1 alone will be discussed in detail.

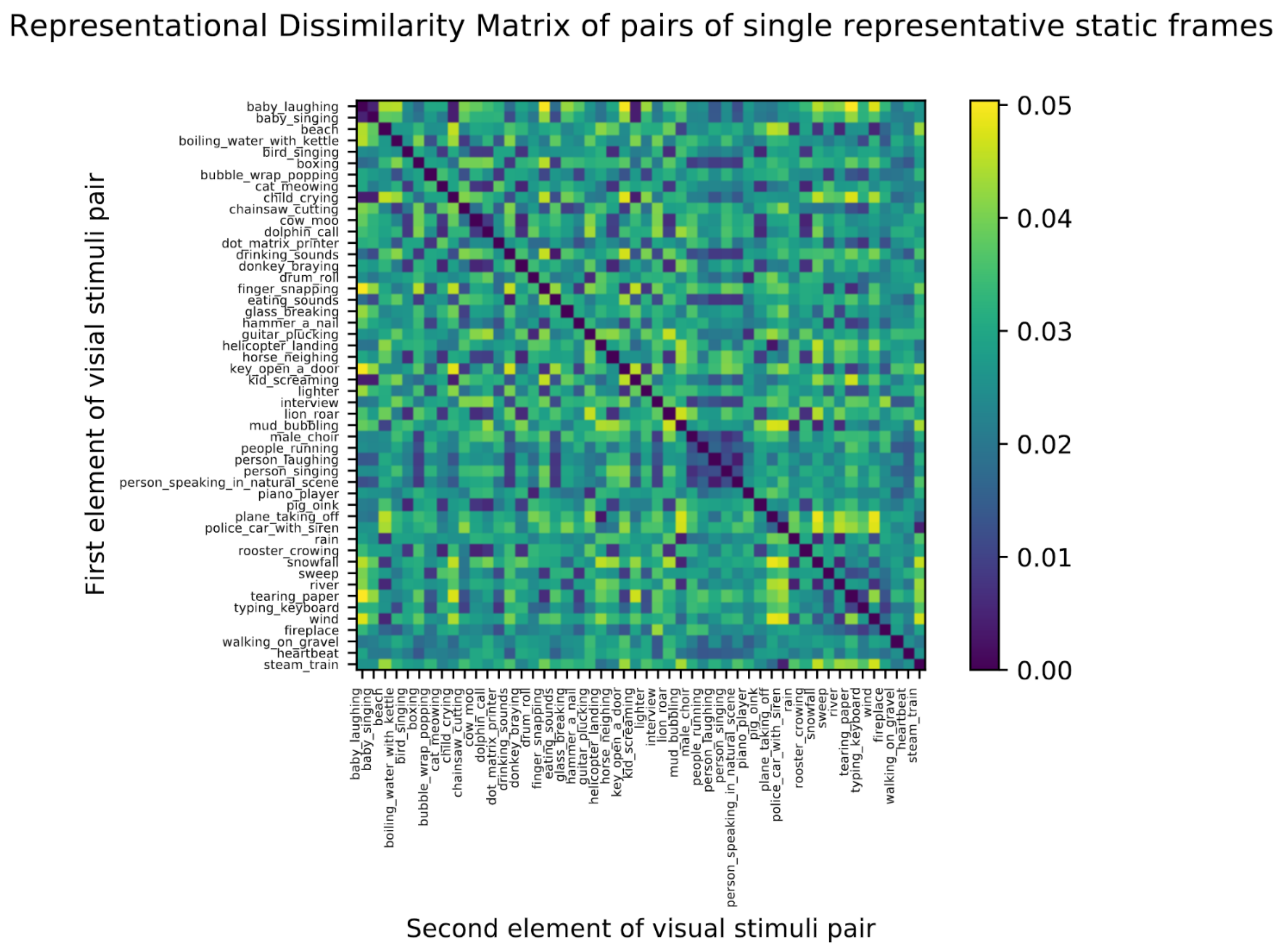

Figure 2. Representational dissimilarity matrix for participant 1

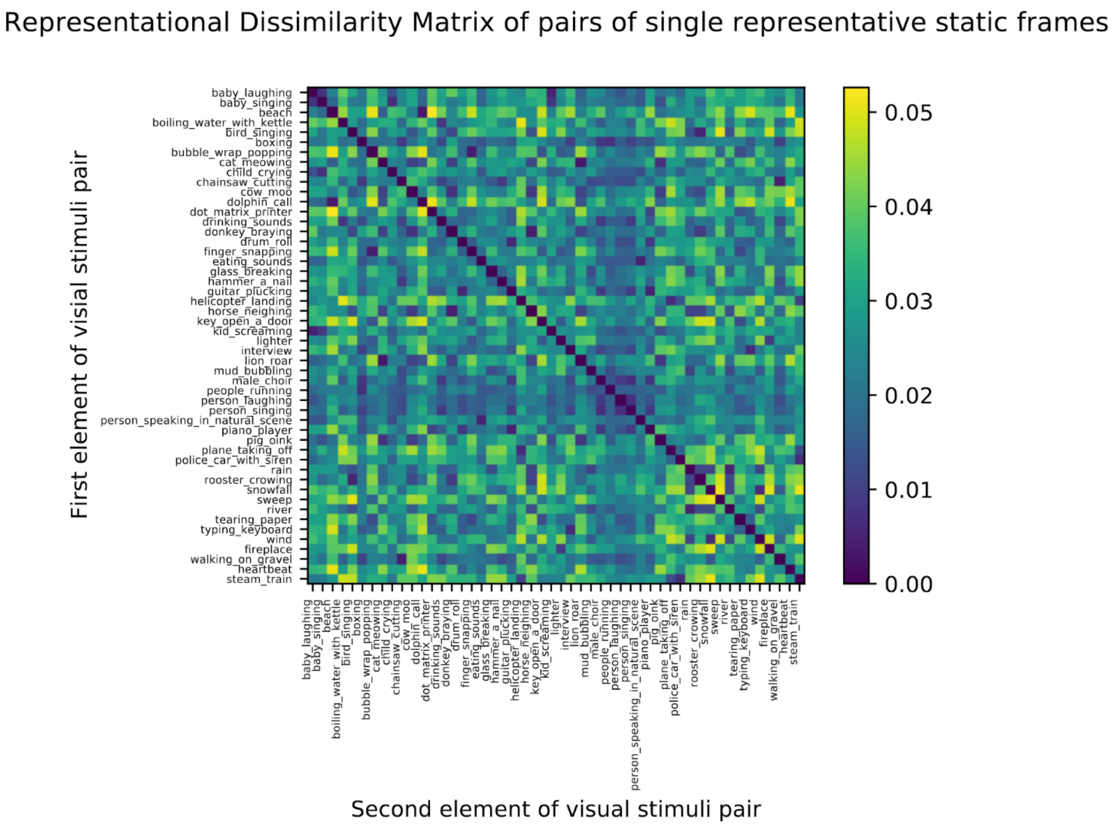

Figure 3. Representational dissimilarity matrix for participant 2

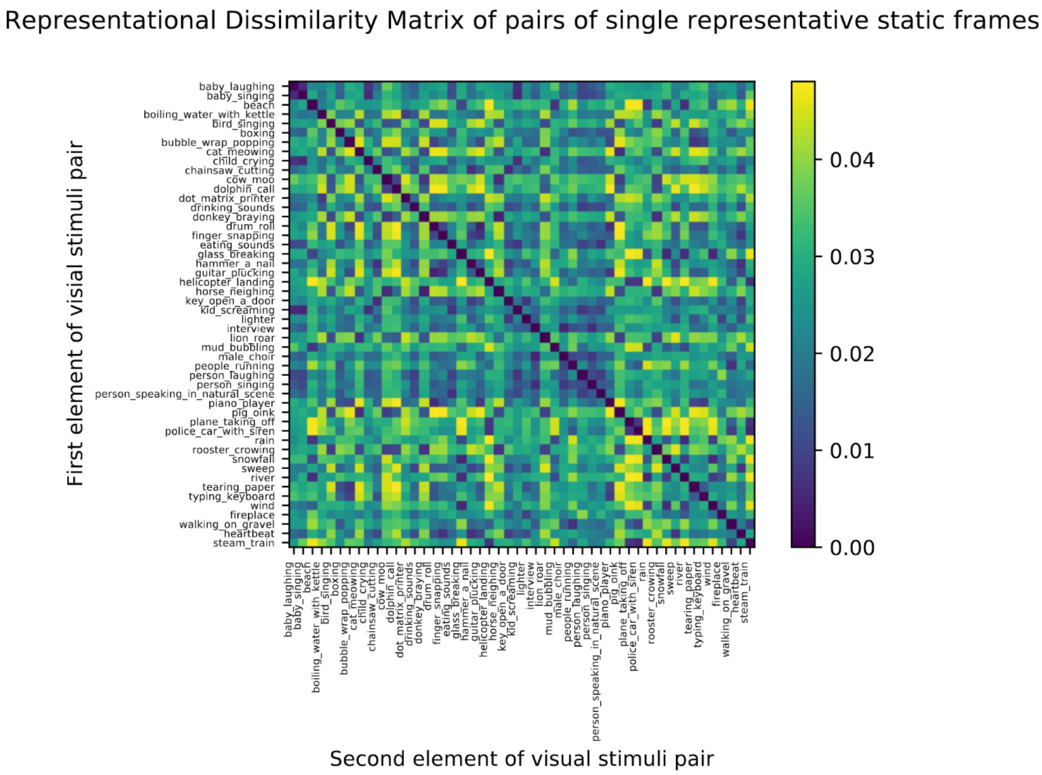

Figure 4. Representational dissimilarity matrix for participant 3

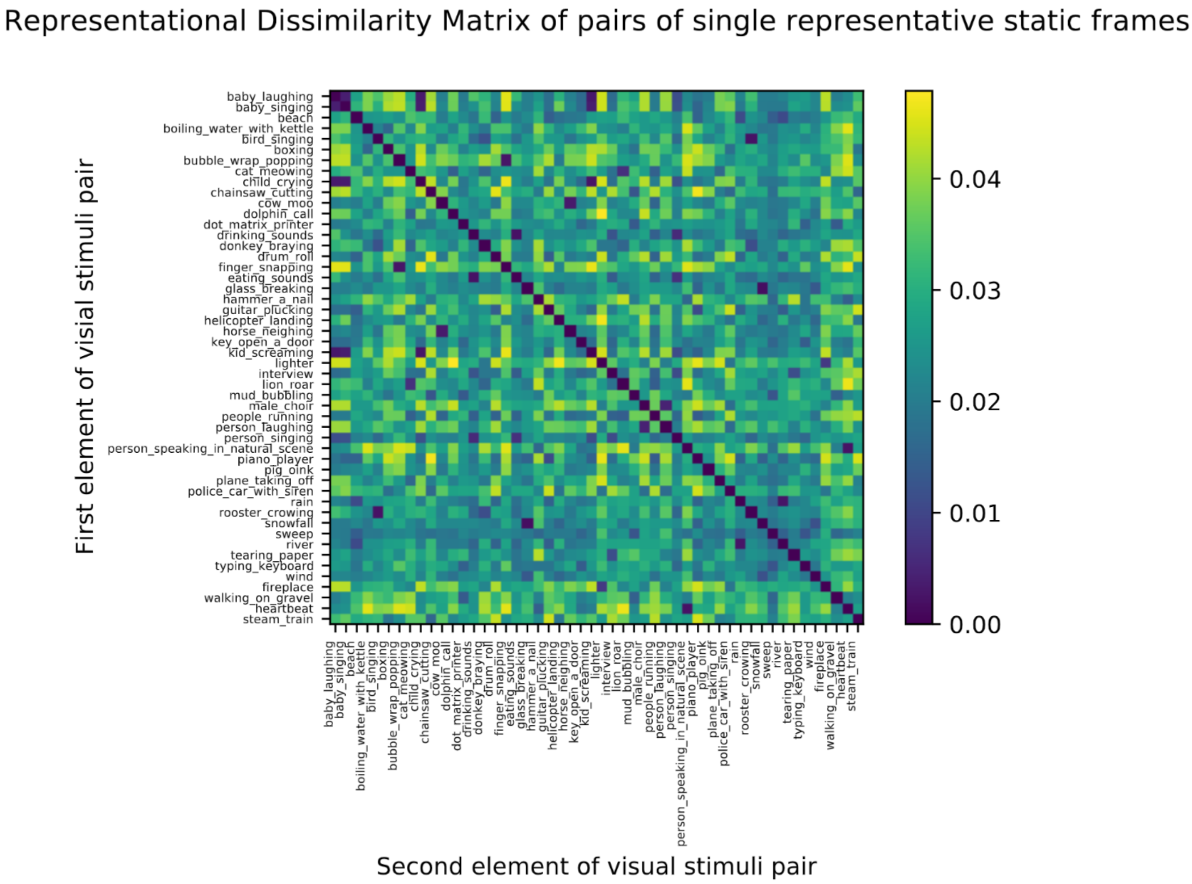

Figure 5. Representational dissimilarity matrix for participant 4

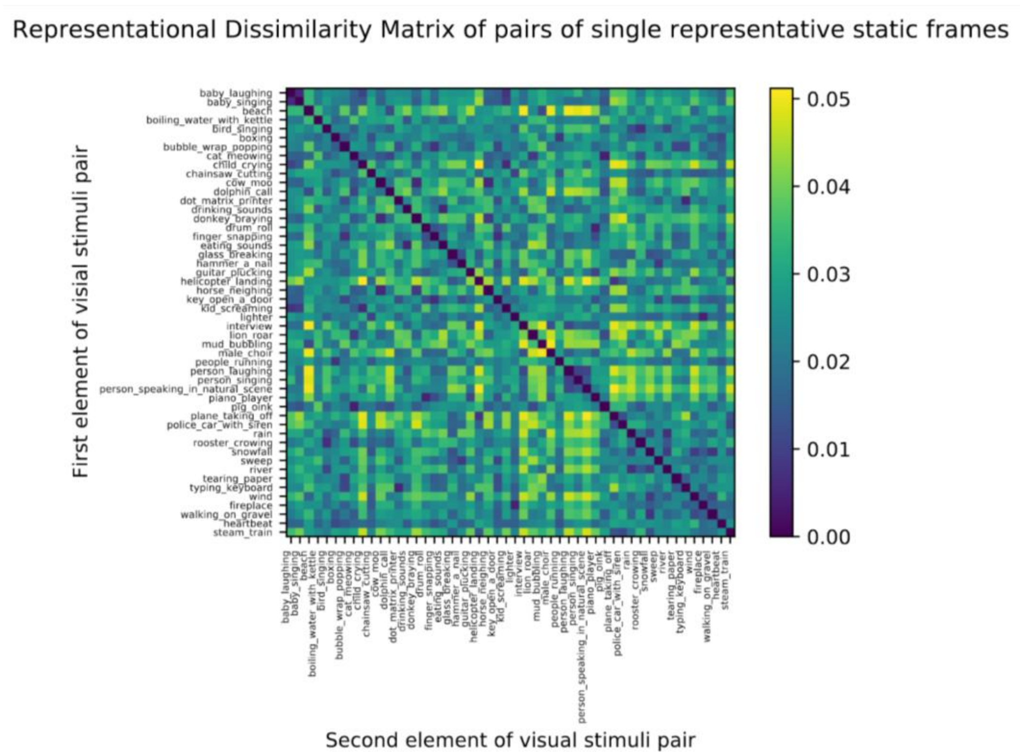

Figure 6. Representational dissimilarity matrix for participant 5

The Pearson Correlation Coefficient calculated for participant 1 indicated that there was a perfect positive linear correlation between a pair of the same stimulus, representing these stimuli to be perceived by participant 1 as ‘identical’. Particular ‘groups’ were formed using the results of the Pearson Correlation, identifying the following seven ‘groups’:

1. Children
2. Nature
3. Music
4. Animals
5. People
6. Stationary/tools
7. Vehicles

Within each of these groups, participant 1 perceived item pairs to be similar, passing the r-value threshold of 0.8 for a strong positive correlation. The Pearson Correlation Coefficient for each of these grouped pairs is shown in the tables below.

Table 1.

R value of items within the children group

	Baby Laughing	Baby Singing	Child Crying	Kid Screaming
Baby Laughing	1	-	-	
Baby Singing	0.965	1	-	
Child Crying	0.971	0.958	1	
Drinking Sounds	0.801	0.809	0.856	
Eating Sounds		0.801	0.844	
Kid Screaming	0.979	0.967	0.960	1

Table 2.

R value of items within the nature group

	Beach	Mud Bubbling	Rain	Snowfall	River	Wind
Beach	1	-	-	-		-
Mud Bubbling	0.864	1	-	-	-	
Rain	0.837	0.912	1	-		
Snowfall	0.938	0.916	0.876	1	-	-
River	0.941	0.910	0.876	0.943	1	-
Wind	0.922			0.806	0.879	1

Table 3.

R value of items within the music group

	Drum Roll	Guitar Plucking	Piano Player
Drumroll	1	-	-
Guitar Plucking	0.910	1	-
Pianoplayer	0.847		1

Table 4.

R value of items within the animals group

	Bird Singing	Cow Moo	Dolphin Call	Donkey Braying	Horse Neighing	Lion Roar	Pig Oink	Rooster Crowing
Bird Singing	1	-	-	-	-	-	-	-
Cat Meowing	0.811						-	-
Cow Moo	0.872	1	-	-	-	-	-	-
Dolphin Call	0.838	0.903	1	-	-	-	-	-
Donkey Braying	0.876	0.920	0.898	1			-	-
Horse Neighing	0.836	0.856	0.896	0.937	1			-
Lion Roar	0.892	0.927	0.913	0.959	0.948	1		-
Pig Oink	0.869	0.842	0.869	0.883	0.910	0.908	1	-
Rooster Crowing	0.912	0.964	0.907	0.897	0.859	0.923	0.858	1

Table 5.

R value of items within the people group

	Boxing	Drinking Sounds	Eating Sounds	Interview	Male Choir	People Laughing	People Singing	Person Speaking In Natural Scene
Boxing	1	-						
Drinking Sounds	0.927	1						
Eating Sounds	0.920	0.986	1					
Kid Screaming		0.814	0.804					
Interview	0.879	0.877	0.865	1				
Male Choir	0.890	0.832	0.831	0.907	1			
Person Laughing	0.808	0.897	0.885		0.896	1		
Person Singing	0.890	0.888	0.898	0.872			1	
Person Speaking In Natural Scene	0.853	0.927	0.911	0.807		0.958	0.831	1

Table 6.

R value of items within the stationary/tools group

	Bubble wrap popping	Dot matrix printer	finger snapping	Hammer nail	Key opening door	lighter	sweep	Tearing paper	Typing keyboard	Boiling water with kettle	Chainsaw cutting
Bubble wrap popping	1	-									
Dot matrix printer	0.872	1									
finger snapping	0.831		1								0.854
Hammer a nail	0.815		0.911	1							
Key open a door				0.893	1						0.834
lighter	0.811		0.93	0.895	0.929	1					
sweep	0.864	0.823	0.922	0.895	0.889	0.892	1				
tearing paper	0.801		0.900	0.895	0.906	0.864	0.865	1			0.850
typing keyboard	0.870	0.900	0.929		1	1	0.804	0.836	1		
Boiling water with kettle										1	
Chainsaw cutting										0.851	1

Table 7.

R value of items within the vehicles group

	Helicopter Landing	Plane Taking Off	Police Car With Sirens	Steam Train
Helicopter Landing	1			
Plane Taking Off	0.955	1		
Police Car With Sirens		0.820	1	
Steam Train			0.976	1

Participant 1 did not perceive the following stimuli as 'similar' with a r-value of above 0.8 for the following items:

1. Glass breaking.

2. People running.

3. Fireplace.

4. Walking on gravel

5. Heartbeat

For these stimuli, the following table identifies what the item was perceived to be most similar and least similar to, with its corresponding r value.

Table 8.

Items perceived to be most and least similar to the items participant 1 did not perceive to be strongly similar to other stimuli, and the corresponding R value

Stimuli	Most similar to		Least similar to	
Glass breaking.	River	0.694	Baby singing	-0.545
People running.	Person singing	0.777	Hammer a nail	-0.617
Fireplace.	Typing Keyboard	0.768	People running	-0.556
Walking on gravel	Interview	0.636	Fireplace	-0.463
Heartbeat	Person singing	0.723	Donkey braying	-0.432

For these five stimuli, though there appears to be no correlation between these stimuli and a 'group', there is a clear difference between the group of the item it is most similar to and the group of the item it is least similar to.

From the RDM of participant 1, it can be seen that item pairs such as key opening door, finger snapping and tearing paper with baby laughing are perceived to be very dissimilar. Through no correlation can be identified between which groups item pairs are perceived to be strongly dissimilar are from, it is clear with the blue squares clustered around 'similar' item pairs, such as people talking, singing and laughing, dissimilarity between pairs is perceived when each item in the pair belongs to a different 'group'.

Discussion

The similarity between the perception of pairs of items was analysed using a multiple arrangements task, from which a representational dissimilarity matrix was performed, and the Pearson Correlation Coefficient was calculated for each participant. The analysis conducted on participant 1 using both these statistical methods. The main aim of this research was to understand how similar two items in a pair were perceived to be, to understand how visual information alone is coded by an individual. This understanding can then be used to better understand how the presentation of verbal information alongside a visual image can aid the representation of the stimulus in the human brain.

The results from participant 1 show that these is a strong relationship between a particular item and other items within its 'group'. The groups in which the participant places each item have a strong Pearson correlation between most items and the other items within the groups. As it would be expected, pairs of stimuli with, for instance, people had a strong linear relationship with other single representative static frames with people in them. It could also be seen that the items which participant 1 perceived to not be similar to any of the other 50 stimuli also represented the dissimilarity between pairs of items which were not in the same group. It is understood that there are many ways humans can represent images. Analogue codes involve central perceptual features to be preserved within the representation of a particular visual image (Sternberg, 2003). Supporting this concept, the results seen in participant 1 support this.

As research into how the human brain codes for and subsequently represents perceived environmental stimuli, future research school focus combining such findings with the brain activity associated with the presentation of these stimuli. The understanding from such research should be used to identify the neural underlying of visual and auditory encoding within the human mind. One of the limitations for this particular project, although only one participants data was analysed for this report, is the small sample size. Inferring how mental representations are coded for with a small sample size with a sample less that 50 would not be sufficient.

Comprehending how the human brain codes and learns new information and environmental stimuli is critical for understanding how humans represent mental images. This would allow for greater knowledge of how humans learn, enabling tools and facilities to enhance the cognitive capacities of the human mind.

References

Badde, S., Ley, P., Rajendran, S. S., Shareef, I., Kekunnaya, R., & Röder, B. (2020). Sensory experience during early sensitive periods shapes cross-modal temporal biases. *ELife*, *9*, 1–14. https://doi.org/10.7554/ELIFE.61238

Brunyé, T. T., Taylor, H. A., & Rapp, D. N. (2008). Repetition and dual coding in procedural multimedia presentations. *Applied Cognitive Psychology*, *22*(7), 877–895. https://doi.org/10.1002/acp.1396

Ernst, M. O., & Bü, H. H. (2004). *Merging the senses into a robust percept*. https://doi.org/10.1016/j.tics.2004.02.002

Hidalgo, A. (2019). Computer science AI - the biological brain. Presentation, University of Birmingham.

Kriegeskorte, N., & Mur, M. (2012). Inverse MDS: Inferring dissimilarity structure from multiple item arrangements. *Frontiers in Psychology*, *3*(JUL). https://doi.org/10.3389/fpsyg.2012.00245

Mayer, R. E., & Sims, V. K. (1994). For Whom Is a Picture Worth a Thousand Words? Extensions of a Dual-Coding Theory of Multimedia Learning. *Journal of Educational Psychology*, *86*(3), 389–401. https://doi.org/10.1037/0022-0663.86.3.389

Nelson, D. L., & Castaño, D. (1984). Mental representations for pictures and words: same or different? *The American Journal of Psychology*, *97*(1), 1–15. https://doi.org/10.2307/1422543

Shen, H. H. (2010). Imagery and verbal coding approaches in Chinese vocabulary instruction. *Language Teaching Research*, *14*(4), 485–499. https://doi.org/10.1177/1362168810375370

Sternberg, R. J. (2003). Wisdom and Education. *Gifted Education International*, *17*(3), 233–248. https://doi.org/10.1177/026142940301700304