Selective Attention in Augmented Reality

Tomas F. Bøttern, Anders M. N. Hansen, Rasmus Höfer, Claudia E. Pellengahr

Aalborg University Copenhagen

Department of Medialogy

Abstract

The following study investigated selective attention in an augmented reality simulation. The focus of this experiment was on how visual and auditory stimuli influenced reaction time. The visual and auditory stimuli used in the test were high, middle, and low perceptual load and congruent and incongruent distractors.

The experiment is based on the flanker test, however, varies from it as a search task was added to increase complexity. A total of 24 test subjects were exposed to a simulation of an augmented reality setup. The subjects had to react to previously defined unisensory or multisensory cues in different load conditions. The participants were distracted by congruent or incongruent stimuli. Target cues were real objects and distractors were virtual labels. The data was analyzed in a 3(perceptual load) x 2(uni and multimodal) x 2 (congruence) comparison.

Results indicated that reaction time in high perceptual load setups were significantly slower than in low demanding setups. Congruent and incongruent setups did not yield any significantly different reaction time, indicating that the distractors had been ignored. Audio-visual incongruent setups were processed faster than visual incongruent setups, which might indicate that multisensory stimuli increase processing speed; however only in middle perceptual load situations independent from congruence.

Implications of this experiment are that test participants were able to make a clear distinction between the objects and the augmented text labels, thus being able to effortlessly ignore the augmented reality part of the test. Furthermore, audio appears to increase processing speed independent from congruence.

Keywords: augmented reality, perceptual load, task performance, selective attention, flanker, reaction time

Introduction

Over the past years, Augmented Reality (AR) has advanced in display and tracking technologies (Azuma, Behringer, Baillot, Feiner, Julier & MacIntyre, 2001). Currently, AR is found in applications from the e.g. medical- military- entertainment- and educational industry. Within these areas one of ARs goals is to augment the real environment with additional information. The additional information helps in task performance, guides users' senses or helps in decision processes.

Previous studies have primarily focused on the technological aspects of AR, such as the display technologies (for further AR discussion see Dünser, Grasset & Billinghurst, 2008). Besides the focus on how to output and display data, an often discussed issue in AR is the data density. Augmenting the world with large amounts of virtual information may result in cluttered displays overloaded with information (Azuma & Furmanski, 2003). In order to minimize the problems of information overload, Julier, Baillot, Brown & Lanzagorta (2002) developed algorithms for information filtering and restriction (Julier, et al., 2002).

The described issues of information representation in AR might occur due to the human's cognitive limitations. Information processing and decision making in situations where task performance is critical, require a variety of diverse cognitive capabilities (Pavel, Gouping & Kehai, 2003), including:

- A large amount of precise working memory
- Ability to integrate multiple information sources
- Logical reasoning in complex and abstract problem spaces
- Incorporating uncertainties and values
- Enumerating possibilities
- Predicting consequences and discounting future values appropriately
- Appropriate distribution rather than focused attention – for anticipation of problems

Many studies in psychology have investigated the underlying aspects of the cognitive human abilities and limitations. The effects of cognitive limitations are essential for the design of AR, especially in the areas of military and medical AR applications, where performance is critical. Given the current technological advancements in AR these limitations may grow in importance in other areas, such as in entertainment or in commercial applications. One example could be a mobile application which provides additional information about the surroundings, without taking the attentional focus away from reality. Examples can be found in navigation applications, games and many more. It is worth knowing how attention is affected in such applications. In order to prove value in daily life the design of AR needs to avoid attentional overloads.

This study focuses on the limits of attention in AR.

Selective attention

The selective attention processes involved in human cognition is a subject for an ongoing debate in psychology (Lee, 2005). Different approaches describe the process of perception and attention allocation, such as the early vs. late selection, object vs. spatial selection and top-down vs. bottom-up selection (Lee, 2005). The early and late selection describes where the selection process occurs in the information processing stages. Object and spatial based selection describes the issue of whether attention is located in spatial position or in objects. The top-down and bottom-up approaches deal with the direction of information flow that constrains the selection process (Lee, 2005).
Regardless of the debate on which of these approaches are salient to the selective attention process, a common assumption gathers all of them: the assumption of selective attention as a necessity to prevent perceptual overload. Further, in the work of Lavie (1995) the perceptual load has been identified as a necessary condition for selective attention.

Lavie claims that the amount of stimuli which can be processed at a specific time is limited (Lavie, 1995). For limited information, there is a surplus of perceptual capacity, making the processing of irrelevant stimuli possible (Lavie 1995). However, as soon as the overall amount of stimuli exceeds the limit of perceptual capacity selection of relevant vs. irrelevant stimuli appears. In a study on perceptual load Lavie (1995) showed that in high perceptual load situations distractors have less effect than in low perceptual load situation. This study suggests that: "*the ability to ignore irrelevant information is directly linked to the load in the processing of the relevant information*" (Lavie, 1995, p.463).

Stimuli in the receptive field get processed automatically, even if the focus is set on ignoring the stimuli (Galera, Grünau & Panagopoulos, 2005). The flanker compatibility test has been used to study the effect of distractors, or flankers, on response time and accuracy. A typical flanker test setup has a central target that needs to be identified. This target is surrounded by similar distractors, such as arrows or letters for example (Stins, Polderman, Boomsma & de Geus, 2007). The flanker test has been used to investigate how distraction factors affect the identification of targets that are placed in a fixed and known location (Wühr & Mϋsseler, 2005).

The response time to targets depends among others on the congruence of the flankers. Congruent flankers appear to lead to a shorter response time than incongruent flankers (Schomstein & Yantis, 2002; Stins, et al., 2007). Incongruent flankers lead to a higher response time than neutral or congruent flankers (Eriksen & Schultz, 1979). This indicates that incongruent flankers are more distracting than congruent flankers and thus increase reaction time.
Moreover, the proximity of the flanker to the target appears to have an effect on the response time. The closer the flanker is to the task object the more interference the flanker seems to cause (Eriksen & Schultz, 1979; Schomstein & Yantis, 2002). If the flanker is moved away from the target, its influence decreases, in other words, the response time decreases. However, the effect of the flankers will not completely disappear, even if moved farther away from the target (Eriksen & Schultz 1979). It appears that the closer the distraction is placed to the target the more intense distraction is processed.
A simultaneous appearance of flanker and target increased the reaction time (Eriksen & Schultz, 1979). But not only the moment of appearance but also the order of appearance might be an influencing factor on response time. It has been suggested that congruent repetitions lead to more accurate results than a change in congruency (Stins, et al., 2007).

In their study Mozolic, Hugenschmidt, Peiffer & Laurienti (2007) describe the influence of selective attention on multisensory (visual and auditory) stimuli (Mozolic, et al., 2007). The presented hypothesis was that modality-specific selective attention would limit processing of stimuli in the unattended sensory modalities. Although the results did not clearly support their hypothesis, the findings highlight the role of modality specific selective attention in modulating multisensory integration (Mozolic, et al., 2007).
Studies concerning semantic content of stimuli have shown that task performance is enhanced through contextual congruent information (Laurienti, Kraft, Maldjian, Burdette & Wallace, 2004). Findings by Laurienti, et al. (2004) also indicate that non- matching, incongruent stimuli from multiple sensory modalities can impair the ability to process relevant information, and hence, result in a decrease of task performance. When reading a book, the visual information obtained (text) does normally not match the auditory information from the environment. Attending both modalities gives difficulties in understanding the book, thus results in a decrease of task performance. The ability to read the book despite the distracting auditory information originates from the ability to focus attention on a particular feature or stimuli, while ignoring others (Mozolic, et al., 2007).

Looking at the previous paragraphs one could conclude that various cognitive limitations and mechanisms in selective attention might affect human performance. It is likely, that these limitations and mechanisms are pertinent to the human performance in AR. As mentioned before, one of these performance limitations in AR is the data density. It may seem obvious that a display overloaded with information leads to shortcomings in human perception and cognitive capacity, resulting in a decrease of task performance. Traditionally these limitations in AR are addressed through both training and rote learning (long time learning), or the tasks are divided from single operators to a number of individuals (Pavel, et al., 2003). Shortcomings of these approaches are that besides being time and cost consuming, they cannot be generalized and might complicate new or unpredictable situations. One possible solution could be to develop computational methods, as suggested by Pavel, et al. (2003).

The question of interest for this study was to find out how the described cognitive limitations and processes in selective attention affect task performance in AR. Given the theoretical information from the described cognitive studies, the hypotheses for this study were formulated as follows:
(1) Less demanding tasks leave more capacity to process unimodal visual distraction factors.
(2) Less demanding tasks leave more capacity to process multimodal visual and auditory distraction factors.
(3) More demanding tasks leave less capacity to process unimodal visual distraction factors.
(4) More demanding tasks leave less capacity to process multimodal visual and auditory distraction factors.

The underlying aspects in these hypotheses are that in an AR environment a certain amount of cognitive capacity may be used to process the reality. This is a variable which changes and can range from very simple situations, where a low amount of resources are required, to situations where a very high amount of resources are required. In situations with complex tasks the augmentation may counteract its original purpose. When AR is meant to help in decision processes or task performance it might consume attentional capacity which is actually needed to process reality.
The intention of this study is to investigate in which perceptual load conditions the augmentation is distracting from reality.

METHOD

In this study a 3 x 2 x 2 within-subjects test design was used. Factors were three difficulty levels (low-middle-high perceptual load), the congruence of the distractors (congruent-incongruent) and uni vs. multi modal cues (visual- visual & auditory). The task in this experiment was a search task based on the flanker test.

Reaction time was used to calculate the effects of the different test factors. Participants were exposed to a search and reaction task. The task was to hit one of two possible buttons as soon as the corresponding target appeared on screen. Inspired by the flanker test, each target would have a distractor assigned, which could be either congruent or incongruent. However, unlike the flanker test the objects would appear in random formations and locations in the same vertical line. This extra task was included to increase the task difficulty and complexity.

Previous tests have investigated the flanker effect with 2D virtual shapes, such as squares and diamonds (Green & Bavelier, 2003), birds, cows, cats and dogs (Wells & Hamm 2009) or letters (Miller 1991). In order to simulate an AR setup this study chose real basic geometric-shaped objects as targets. To mix real and virtual elements the targets were labelled with virtual text elements, which represented an augmented reality setup. A total amount of six basic geometric objects were chosen: cube, arrow, cone, cylinder, octagon, pyramid and triangle. Of these objects two target shapes were selected; cube and arrow. These two objects could be considered well known and easily identifiable shapes and they have been used as targets in previous flanker tests (Green & Bavelier, 2003; Shomstein & Yantis, 2005). All objects were coloured in white in order to avoid any distracting impact of colour. Good visibility was guaranteed by placing them in front of a dark background (Figure I, Appendix III and IV).

Only one of the two target objects appeared in a single setup, either an arrow or a cube appeared on screen. The visual load levels (low, medium and high), were defined by the number of objects that were displayed on screen. It was considered a low perceptual load when only one object was visible on screen. Three objects represented a medium and six objects a high perceptual load. Previous studies reported that the proximity of elements in the receptive field matters (Duncan, 1984; Wells & Hamm, 2009), hence, all objects were positioned with the same spacing of 3,5-4cm.

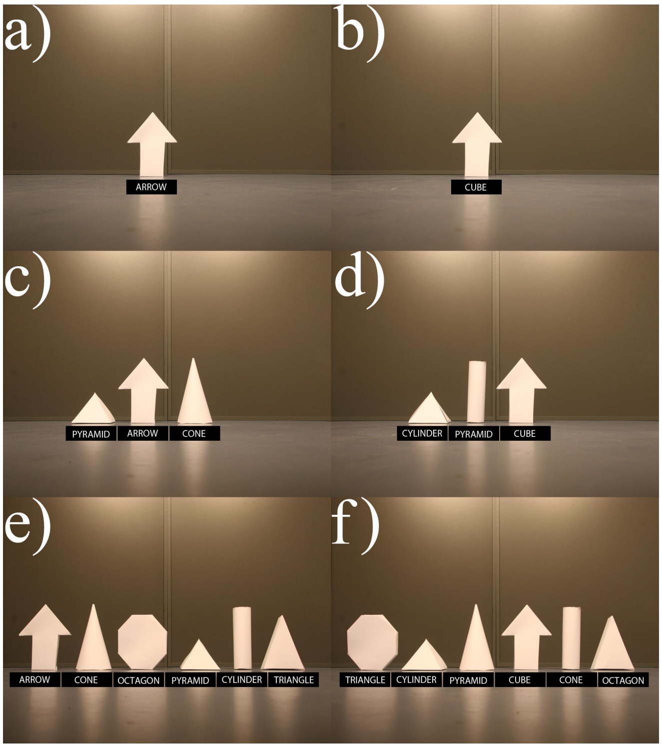


Figure I – Six setups depicting the different perceptual load situations of the arrow. a) Low Congruent b) Low Incongruent c) Medium Congruent d) Medium Incongruent e) High Congruent f) High Incongruent

The virtual text labels were added below each of the objects. When the distractors were congruent, the shapes were matched with the correct text labels. In incongruent setups the target object (cube or arrow) was labelled with the name of the target object that was not displayed (Figure I, Appendix III and IV). Thus, in incongruent cases when the target is a cube the text 'arrow' would appear at the bottom and vice versa. The text appeared as white text in a black box. The font format was Myriad Pro Regular 30pt sharp. The text labels filled approximately one third of the objects space. Figure II shows how the size of objects and text in the test setup related to the screen size and the position of the participant.

Figure II – Experimental setup example with participant.

It has been suggested that attention resources are not distributed equally over the visual field (Shomstein & Yantis, 2005). Distractors presented in a horizontal meridian seem to have a higher impact, as they appear to be more difficult to ignore (Shomstein & Yantis, 2005). In order to ensure that the distractors were processed the text labels were added beneath the objects. Duncan (1984) stated that: "*identifying several stimuli in a complex display is improved if they are made to form a 'strong' perceptual group*" (Duncan 1984 p.502). In the current experiment a visual group of the distractors and the objects was achieved by adding the text as close as possible to the according object.

After 24 setups were completed congruent or incongruent auditory cues were added to the visual task. The additional auditory distractors were used to enhance the effect of the congruent or incongruent visual text. This means that in case of an incongruent text there would also be an incongruent sound. The words 'cube' and 'arrow' were pronounced and used as auditory cues they were activated simultaneously to the visual distractors.

Apart from object based processing location based processing should be mentioned as well.

One example is the spatial conflict task which deals with spatially congruent stimuli. If the object that needs to be identified is located in the left part of the screen, reaction times of the left hand response will be quicker (Stins, et al., 2007). In order to avoid this effect in the present task setup the button labels for 'cube' and 'arrow' were changed after half of the test subjects.

In low perceptual load setups only one shape was displayed on screen, either cube or arrow. In higher load setups the non-target objects were randomly added. The order of the presented objects was varied in every setup. Furthermore, perceptual load and congruent-incongruent situations were mixed so that the order of the setups was random. Overall, setups were in the exact same random order for every participant (Appendix II).

The test started with five training setups. After each setup a black screen was displayed for two seconds. A countdown from three initiated the test. The maximum response time for one setup was set to six seconds.

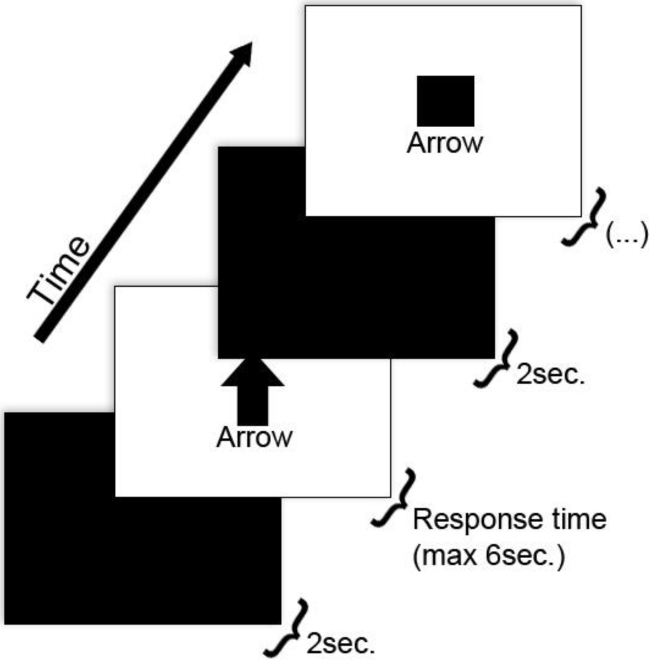


Figure III - Sequence of setups displayed during the test procedure. Between each search task the black screen was presented for two seconds. The maximum time for the search task was set to six seconds. In each task either the cube or the arrow had to be identified and the corresponding button pressed. A button press resulted in an immediate switch to the black screen.

Procedure

The user base consisted of 24 Medialogy students in the age span of 18-30 years. Recruitment of the participants was conducted by addressing the students in their semester assigned rooms. A total of nineteen men and five women voluntarily participated. Participants were given a short introduction to the procedure of the test. One of the tests had to be discarded, due to technical error.

At the beginning of the test a brief oral questionnaire was conducted which required the test subjects to state their age, level of video game usage (none, casual, heavy) and their handedness. Distraction factors from the environment were limited by enclosing the test subjects with the help of partitions.

In order to ensure that all participants had an equal understanding of the objects and their labels the real objects were presented prior to the test. Participants were instructed that reaction time and accuracy were of importance for the test. Two large buttons were used to record reaction times. One button represented the cube object and the other the arrow object. The participants were asked to position their hands above the buttons but not to lean onto them, in order to ensure that all participants had the same initial hand position (Figure II).

Previous to the final experiment a pilot test was conducted. The pilot test revealed that visual feedback on button presses was necessary. Feedback was provided by switching to a black screen as soon as a button was pressed.

Furthermore, to ensure that participants had a correct understanding of the task a training sequence needed to be added before the actual test started.

RESULTS

Results were analyzed in seconds. Tables are provided in the appendix (Appendix VI and VII).

Manipulation checks

To assure the validity of the test setup some manipulation checks were conducted. First, the selection of target shapes was verified. The mean response time of all cube setups (*M* = .7672s) was compared with the mean response time of all arrow setups (*M* = .7426s) and no significant difference was found (*n* = 11; *t* = 1,005, *p* = 0,339). This allows comparing of the target shapes, as they were processed equally in time.

As target objects appeared in the left, middle or right part of the screen, the location of the target objects were checked for an effect on response time (RT). Reaction times of objects located in the left (*M* = .77s), middle (*M* = .74s) and right (*M* = .77s) part of the screen showed no significant difference. Thus, it can be assumed that the location off the target objects on screen did not have a significant influence on the results.

The sample was checked for influences of gender, handedness or age on response times to perceptual load, visual/auditory stimuli and incongruent/congruent situations. There were no significant influences on response time.

Hypothesis results

To test the hypotheses the three perceptual load conditions were compared. The perceptual load levels (1 object, 3 objects, 6 objects) were compared by paired t-tests. Between the low and middle perceptual load condition a clear significant difference in reaction time was calculated (*n* = 11, *t* = -7,340, *p* < .001). The middle and high perceptual load setups were significantly different in reaction time (*n* = 13, *t* = -2,928, *p* = .013). The low and high perceptual condition had a clear significant difference (*n* = 13, *t* = -8,425, *p* < .001).

Next the reaction times to congruent and incongruent setups were compared. All congruent setups (*M* = .7367s) were compared with all incongruent setups (*M* = .7731s). There was no significant difference between these two conditions (*t* = -1,559, *p* = .150). This is an unexpected result as it was assumed that incongruent setups would need more time to be processed than congruent setups. Consequently, congruent and incongruent setups were compared in the different perceptual load levels. In the low load situation no significant difference between congruent (*M* = .6206s) and incongruent (*M* = .6408s) setups were found (*n* = 16, *t* = -1.247, *p* = .231). In the middle perceptual load level congruent (*M* = .7893s) and incongruent (*M* = .7650s) setups were not significantly different (*n* = 14, *t* = .817, *p* = .428). In the high load setups congruent (*M* = .8915s) and incongruent (*M* = .8359s) stimuli did not lead to significantly different reaction times (*n* = 18, *t* = 1.224, *p* = .238). (Table I)

Compared variables	Mean (s)	*n*	*t*	*p*
Cube	0.7671	11	1,005	.339
Arrow	0.7426			
Low load	0.6097	11	-7,340	< .001
Medium load	0.7886			
Medium load	0.7880	13	2,928	.013
High load	0.8825			
Congruent	0.7367	11	-1,559	.150
Incongruent	0.7731			
Audio congruent	0.7458	16	-0.225	.825
Audio incongruent	0.7515			
High visual load	0.8599	18	-0.408	.689
High audio load	0.8714			
Medium visual load	0.8065	14	4,998	< .001
Medium audio load	0.7186			
Low visual load	0.6199	16	-0.368	.718
Low audio load	0.6272			

Table I – Variables with mean times (s), calculated t-tests and their corresponding p-values

Twenty-four setups consisted of visual stimuli and 12 setups of audio and visual stimuli. By enforcing the distractor with auditory stimuli the expected result was that the response time would increase. Comparing congruent (*M* = 7458s) and incongruent (*M* = .7515s) response times in the multi modal setup did not yield any significant results (*n* = 16, *t* = .225, *p* = .825). This was an unexpected result since nearly no difference between the response times was measured.

When investigating the difference between the congruent visual (*M* = .7483s) and congruent auditory (*M* = .7328s) setups no significant difference appeared (*n* = 15, *t* = 1.026, *p* = .323). Against expectations the congruent sound did not help the processing of the stimuli. Next, the incongruent trials were compared in visual (*M* = 7786s) and auditory (*M* = 7390s) setups, the difference of the incongruent trials was not significant

($n = 12$, $t = 1.823$, $p = .095$). When looking at the means this result indicates that the incongruent auditory stimuli did not distract the identification of the target but decreased the response time.

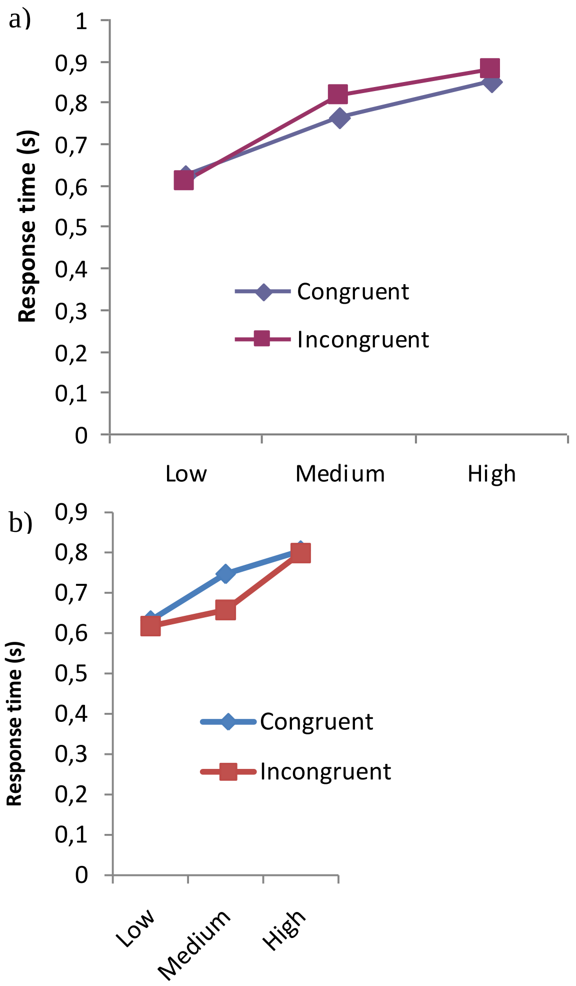


Figure IV - Mean reaction times in seconds for the three different load conditions with: a) unimodal distractors b) multimodal distractors

Due to the unexpected results the perceptual load was crosschecked with the uni- and multimodal setups. High visual load (*M* = .8599s) and high visual-auditory load (*M* = 8714s) showed no significant difference in response time ($n = 18$, $t = -.408$, $p = .689$), the visual setups led to slightly faster responses. The low visual load (*M* = 6199s) compared to the low visual-auditory load (*M* = 6272s) did not yield any significant results. However, the low visual-auditory load was slightly quicker. In the medium load setups a significant difference between the medium uni- (*M* = 8065s) and the medium multimodal (*M* = 7186s) setups was found ($n = 14$, $t = 4.998$, $p < .001$). This indicates that three objects on screen and auditory stimuli significantly enhanced the processing of the target.
Investigating the reason for this unexpected difference shows that incongruent unimodal stimuli (*M* = .8188s) are processed significantly slower than incongruent visual-auditory stimuli (*M* = .6725s). The audio distractor enhanced the processing of the target object instead of slowing it down ($n = 17$, $t = -5{,}381$, $p < .001$).

Calculating the 3x(2x2) ANOVA supported the results obtained from the paired t-test. In the setups with low visual load the F-Value (0.2079) was significantly lower than the F-Crit (2.7081), meaning that congruent and incongruent distractors did not affect response time. This case is repeated in the high visual load were the F-Value (0.7512) was significantly lower than the F-Crit (2.7081).
Only in the medium visual load setups the F-Value (5.2374) was higher than the F-Crit value (2.7081).

Confusion errors (3,6%) were not included in the analysis. In a total of 828 trials 30 invalid answers were recorded. In the unimodal setups 2,6% of the responses were invalid. In the multimodal setups 5% were invalid. Nine confusion errors occurred in the low perceptual load, fourteen in the medium load and seven in the high load setup.

DISCUSSION

When perceptual capacity is left, due to a low amount of stimuli, irrelevant stimuli are still processed. When the overall amount of stimuli exceeds the limit of perceptual capacity, selective attention filters relevant vs. irrelevant stimuli (Galera, et al., 2005; Lavie, 1995). It is assumed that task performance is enhanced through congruent information, whereas incongruent information can result in a decrease of task performance (Laurienti, et al., 2004). Thus it might be hypothesized that the same effects can be observed when transferred to a more complex situation, e.g. an Augmented Reality environment.

Expected effects were that in tasks requiring a low amount of perceptual load (1 object on screen) semantic incongruent distractors would delay the response time. Furthermore, it was expected that in situations with a high amount of perceptual load (6 objects on screen) the response times in congruent and incongruent situations would have no significant difference. This refers to the theory that selective attention filters irrelevant information as soon as the overall amount of stimuli exceeds the limit (Lavie, 1995).

The data from the test results demonstrates that response times increased significantly in the different perceptual load situations (1, 3 or 6 objects on screen). This shows that the change of perceptual load was successful and the difficulty of the task increased. It was hypothesized that less demanding tasks leave more capacity to distraction factors. Considering that it is significantly faster to process one than three or six objects it could be concluded that the task was significantly easier.

In this study the data of congruent and incongruent distractors in the three difficulty levels showed no significant difference. Here the expected outcome of a significant difference in low perceptual load situations was not found; the same applied for the multimodal situations. Moreover, no significant difference between congruent and incongruent distractors was measurable, meaning that the distractors most likely where disregarded by the test subjects. Further, the response time in some of the incongruent situations were faster than in their congruent counter parts. This led us to disregard the congruency of the distractors. It was assumed that test participants did not process the distractors, since the findings in previous experiments indicate the theory that semantically incongruent stimuli degrades performance (Eriksen & Schultz, 1979; Galera, et al., 2005; Schomstein & Yantis, 2002).

Taking the previous paragraphs into consideration it can be stated that although the low perceptual task was significantly faster to process, no effect of the distractors congruent or incongruent was identified.

One explanation for the results might be that the choice of distractors (text) might have influenced the outcome. It could be that the text was ignored because it was not semantically processed. In other flanker experiments (Green & Bavilier, 2003; Wells & Hamm 2009, Miller 1991), distractors were either congruent matching shapes or incongruent non-matching shapes. Thus, it can be argued that the mixing of text and shapes may not represent a contextual equality and thereby be easier to disregard when processing visual stimuli.
The initial reason for choosing text was to use an element that is likely to be a part of an Augmented Reality environment.

Although the results of the positioning of target objects (cube or arrow – left middle or right on the screen) did not show a significant difference in response time, the horizontal alignment may have had an influence on the processing of distractors. Galera, et al. (2005) suggested that distractors have higher effects on the horizontal meridian. The distractors spatial position in a horizontal line beneath the target objects might have influenced the disregarding of the distractors. Additionally, perceptual units might be formed according to colour, intensity or orientation (Lee, 2005). In complex displays it is easier to identify these perceptual groups (Duncan, 1984). The unit of textual distractors at the bottom of the objects might have been identified as one perceptual group. The textual distractors always had the same colour and same location. Thus the location and resemblance of the distractors could have made it easier to disregard them.

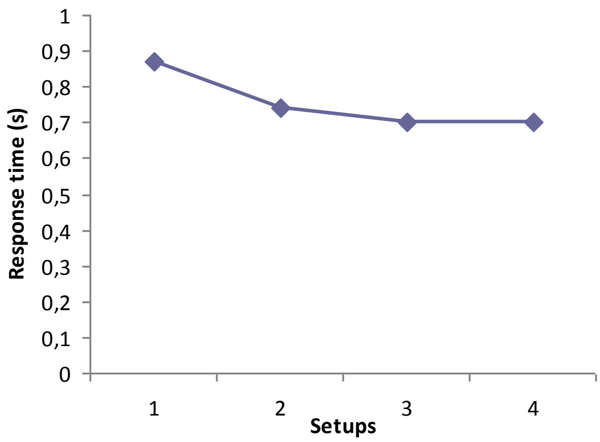


Figure V - Learning curve of the unimodal medium congruent setup. For all learning curves see Appendix V.

In total 36 setups were tested. A short training sequence at the beginning of the test assured that the participants were prepared for the task. During the test the participants improved their overall reaction time; it seems that a learning effect occurred. Comparing the first setups to the last setups the overall response times decreased ~0.3 seconds. It seems that participants were able to learn from the early setups and thus processed latter setups quicker. This learning effect could additionally have had an effect on the congruent and incongruent setups. After several setups the participants might have figured out which stimuli in the receptive field were task relevant and which were task irrelevant.

In a normal flanker test incongruent distractors have an effect on response times. In the current study a search task was added to the flanker setup. However, a difference between congruent and incongruent distractors was unexpectedly not found. A possible explanation might be that the search task was too complex and thus did not leave capacity to process additional stimuli in the receptive field. In order to determine the difficulty of the task the error rate had to be considered. Of 828 responses only 30 confusion errors were made (9 congruent / 21 incongruent). Only 3,6% of all the answers were wrong. This low rate of confusion errors indicates that the search task was not too difficult.

When looking at the results obtained from the multimodal distractors the response times indicate that the distractors have been ignored. The visual distractors may have been ignored for the same reasons as in the unimodal setups. However, the missing effects of the auditory distractors cannot be explained by the same reasons. The most plausible explanation would be that selective attention to the visual modality (the search task) prevented the integration of the auditory modality. Thus, the results indicate that modality-specific selective attention attenuates multisensory integration as suggested by Mozolic, et al.(2007).

CONCLUSION

This study dealt with the effects of selective attention in augmented reality.

The purpose was to investigate how perceptual load and distractors influenced the task performance of an AR simulation. A total of 36 different setups were devised; the setups consisted of different difficulty levels with low, medium or high load, congruent or incongruent cues either visual or visual and auditory. The results showed that perceptual load influenced reaction times. Auditory stimuli were found to affect the semantic comprehension in the medium visual load setups. The difference between congruent and incongruent setups did not yield the expected outcome.

In our AR simulation these results might indicate that the processing of real objects is not harmed by the augmented text labels. Due to the different nature of text labels, as opposed to real objects, the task to distinguish among them appears not to hinder initial processing. Furthermore, our findings indicate that focusing on the visual modality impairs the integration of the auditory stimuli. This would imply that important notifications need to be multimodal in order to make sure they are processed, provided that the attention is focused on one modality.

Lavie (1995) and Galera, et al. (2005) suggest that all stimuli in the receptive field get processed, even if the participants try to ignore them. Contradicting to this theory our results indicate that distractors can be deliberately ignored by the participants. This applies only under the conditions that: attention is focused on a very specific task, the distractors are clearly distinguishable from the targets, the distractors are located in the same location and they appear simultaneously.

In order to verify the actual causes of the unexpected outcome of this study future research has to be conducted.

References

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *Computer Graphics and Applications, IEEE*, *21*(6), 33-47.

Azuma, R., & Furmanski, C. (2003). Evaluating label placement for augmented reality view management. *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality, Malibu*, 66-75.

Duncan, J. (1984). Selective Attention and the Organization of Visual Information. *Journal of Experimental Psychology: General*, *113*(4), 501-517.

Dünser, R., Grasset, R., & Billinghurst, M. (2008). A survey of evaluation techniques used in augmented reality studies. *International Conference on Computer Graphics and Interactive Techniques, Singapore*

Eriksen, C., & Murph, T. D. (1987). Movement of attentional focus across the visual field: A critical look at the evidence. *Perception & Psychophysics*, *42*(3), 299-305.

Eriksen, C. W., & Schultz, E. W. (1979). Information processing in visual search: A continuous flow conception and experimental results. *Perception & Psychophysics*, *25*(4), 249-263.

Galera, C., Grünau, M. v., & Panagopoulos, A. (2005). Automatic focusing of attention on object size and shape. *Psicológica*, *25*, 147-160.

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*(29), 534 - 537.

Julier, S., Baillot, Y., Brown, D., & Lanzagorta, M. (2002). Information filtering for mobile augmented reality. *IEEE Computer Graphics and Applications*, *22*(5)

Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research, Springer*, *158*(4)

Lavie, N. (1995). Perceptual Load as Necessary Condition for Selective Attention. *Journal of Experimental Psychology*, *21*(3), 451-468.

Lee, K. (2005). Cooperative Aspects of Selective Attention. *Advances in Natural Computation - Daejeon, 3610*

Miller, J. (1991). The flanker compatibility effect as a function of visual angle, attentional focus, visual transients, and perceptual load: A search for boundary conditions. *Perception & Psychophysics*, *49*(3), 270-288.

Mozolic, J. L., Hugenschmidt, C. E., Peiffer, A. M., & Laurienti, P. J. (2007). Modality-specific selective attention attenuates multisensory integration. *Experimental Brain Research*, *184*(1), 39 - 52.

Pavel, M., Gouping, W., & Kehai, L. (2003). Augmented cognition: allocation of attention. *Proceedings of the 36th Annual Hawaii International Conference on System Sciences (HICSS'03)*, *5*(5)

Shomstein, S., & Yantis, S. (2002). Object-based attention: Sensory modulation or priority setting? *Perception & Psychophysics*, *64*(1), 41-51.

Stins, J. F., Polderman, J., Boomsma, D. I., & de Geus, E. J. (2007). Conditional accuracy in response interference tasks: Evidence fromthe Eriksen flanker task and the spatial conflict task. *Advances in Cognittive Psychology*, *3*(3), 389-396.

Wells, I. C., & Hamm J. P. (2009). The Effects of Inverting Natural Stimuli in a Flanker Task. *Canadian Journal of Experimental Psychology*, *63*(1), 33-39.

Wühr, P., & Müsseler, J. (2005). When do irrelevant visual stimuli impair processing of identical targets? *Perception & Psychophysics*, *67*(5), 897-909.

APPENDIX

I) Test Manuscript

The test manuscript was created to ensure that each test participant was given the exact same introduction and ensuring the same premises for the entire user base. The actual description can be interpreted as a workflow requiring two test supervisors along with one test participant at any given time.
Before conducting the actual test setup, participants had to answer a few questions regarding their abilities.

- Handedness (left/right)
- Computer Video experience (Non/Casual/Heavy Gamer)
- Age and Semester
- Their gender

After the brief verbal questionnaire, the test subjects were given a short explanation to the premises of the test.

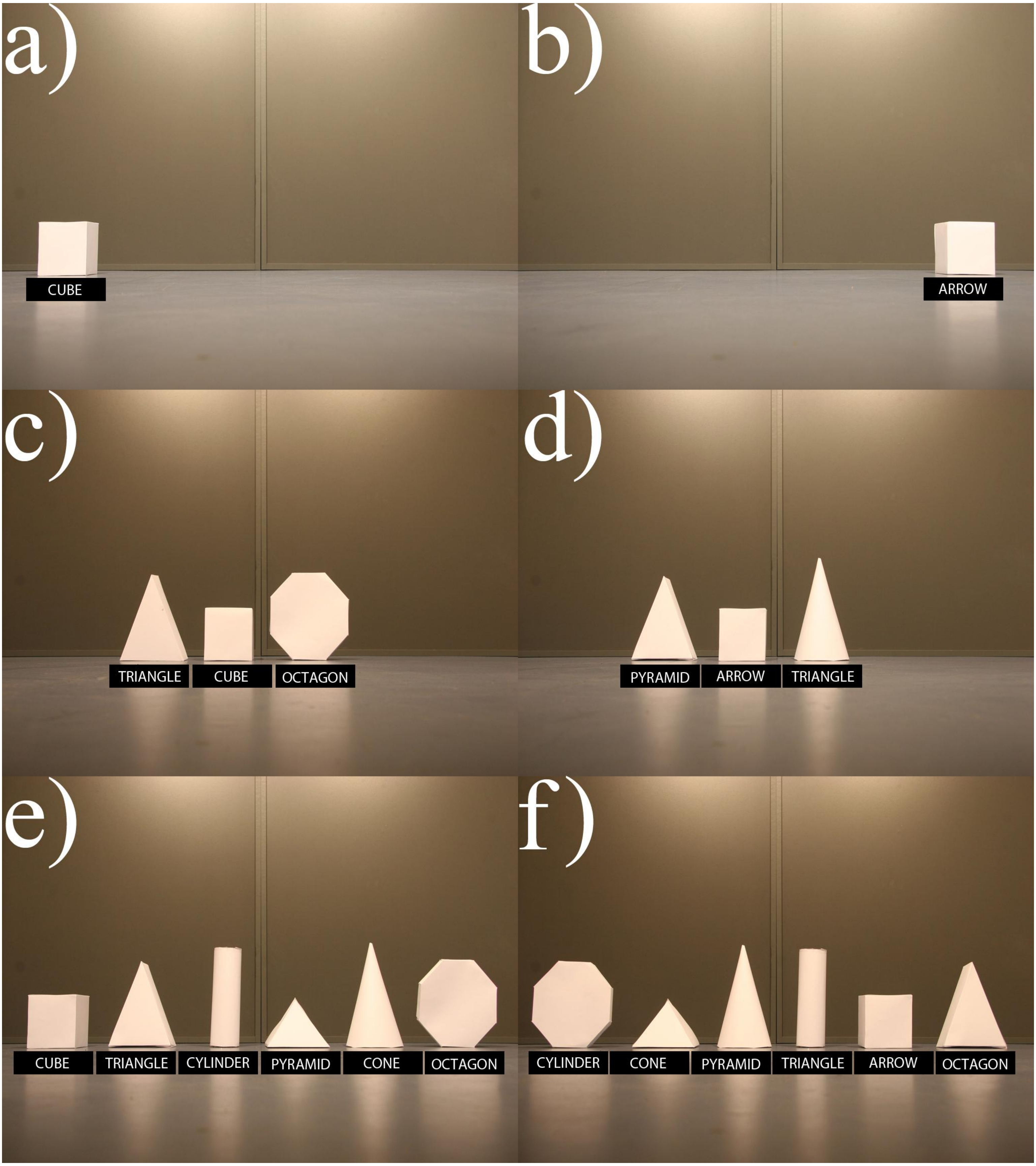
- A physical representation of the real objects
- Emphasis on the Cube and Arrow
- Introduction to the physical setup (Buttons)

After familiarization with the objects and the button functionality, the goal of the test (response time) was conveyed to the test participant. The test participant was issued headphones but if and when a sound would occur was not revealed. The first five screens of the test were introductory to the buttons and cube/arrow functionality, after which the real test did commence.

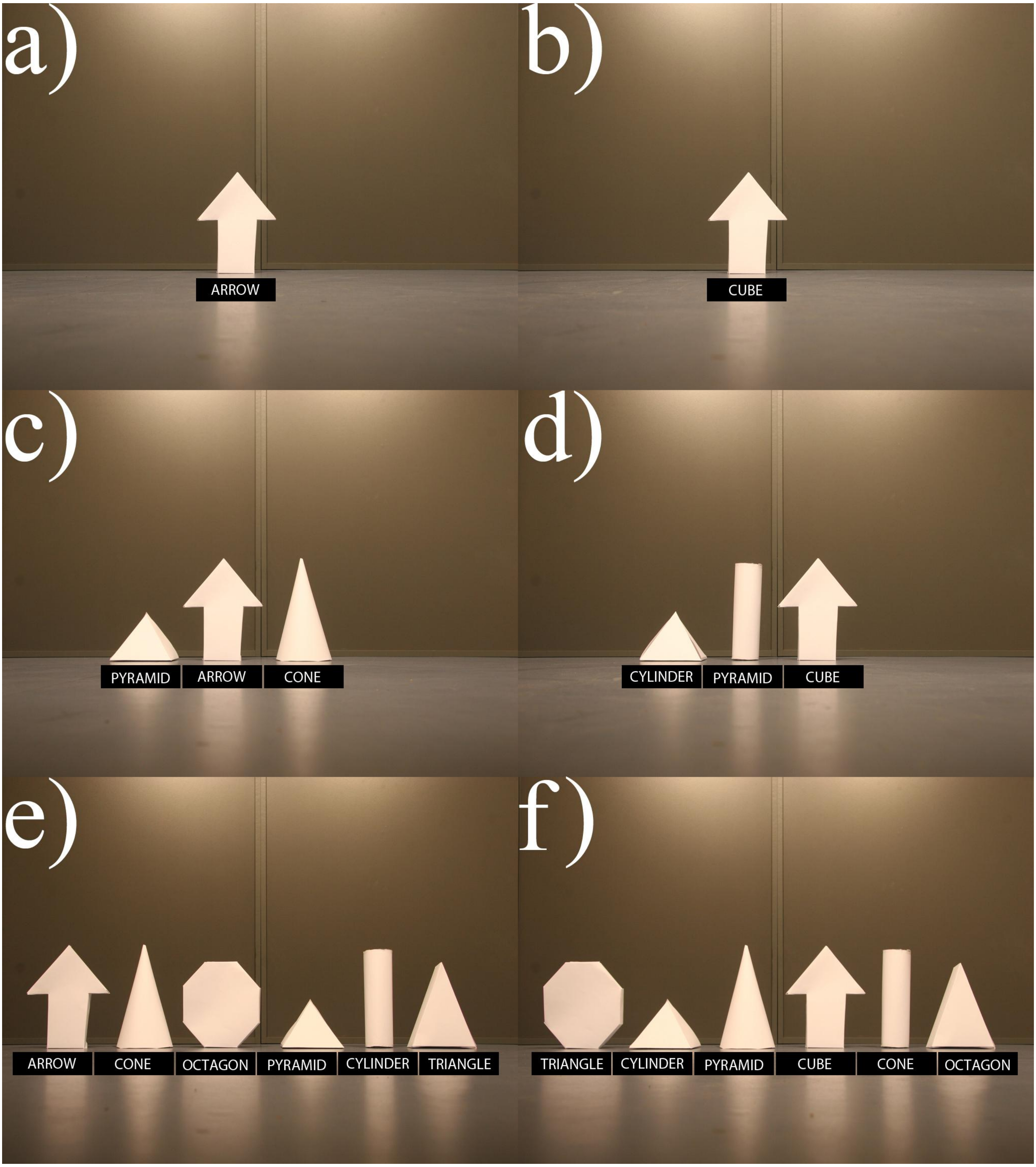
II)Test setup sequence

Setupnr.	Cube / Arrow	Load	Objects	Congruence
VISUAL SETUPS				
1	C	MIDDLE	triangle, cube, octagon	Congruent
2	A	LOW	arrow	Incongruent
3	A	MIDDLE	pyramid, cylinder, arrow	Incongruent
4	A	MIDDLE	arrow, cylinder, cone	Congruent
5	C	HIGH	octagon, pyramid, cone, cylinder, cube triangle	Incongruent
6	A	LOW	arrow	Congruent
7	C	HIGH	cube, triangle, cylinder, pyramid, cone octagon	Congruent
8	C	MIDDLE	pyramid, cube, cone	Incongruent
9	C	LOW	cube	Congruent
10	A	LOW	arrow	Congruent
11	C	HIGH	octagon, cylinder, pyramid, triangle, cube, cone	Incongruent
12	A	MIDDLE	triangle, arrow, octagon	Incongruent
13	A	HIGH	octagon, pyramid, cone, arrow, cylinder, triangle	Incongruent
14	C	LOW	cube	Congruent
15	C	MIDDLE	cube, cylinder, cone	Congruent
16	A	HIGH	arrow, pyramid, octagon, triangle, cone, cylinder	Incongruent
17	A	LOW	arrow	Incongruent
18	C	HIGH	cylinder, triangle, cube, pyramid, cone, octagon	Congruent
19	C	LOW	cube	Incongruent
20	A	HIGH	arrow, cone, octagon, pyramid, cylinder, triangle	Congruent
21	A	MIDDLE	pyramid, arrow, cone	Congruent
22	C	MIDDLE	cube, pyramid, cylinder	Incongruent
23	A	HIGH	cone, arrow, octagon, cylinder, triangle, pyramid	Congruent
24	C	LOW	cube	Incongruent
AUDIO / VISUAL SETUPS				
25	C	MIDDLE	cube, cylinder, cone	Congruent
26	A	HIGH	arrow, cone, octagon, pyramid, cylinder, triangle	Congruent
27	C	LOW	cube	Incongruent
28	C	MIDDLE	pyramid, cube, cone	Incongruent
29	A	MIDDLE	pyramid, arrow, cone	Congruent
30	A	HIGH	octagon, pyramid, cone, arrow, cylinder, triangle	Incongruent
31	A	LOW	arrow	Congruent
32	C	HIGH	octagon, cylinder, pyramid, triangle, cube, cone	Incongruent
33	C	LOW	cube	Congruent
34	A	MIDDLE	triangle, arrow, octagon	Incongruent
35	C	HIGH	cube, triangle, cylinder, pyramid, cone octagon	Congruent
36	A	LOW	arrow	Incongruent

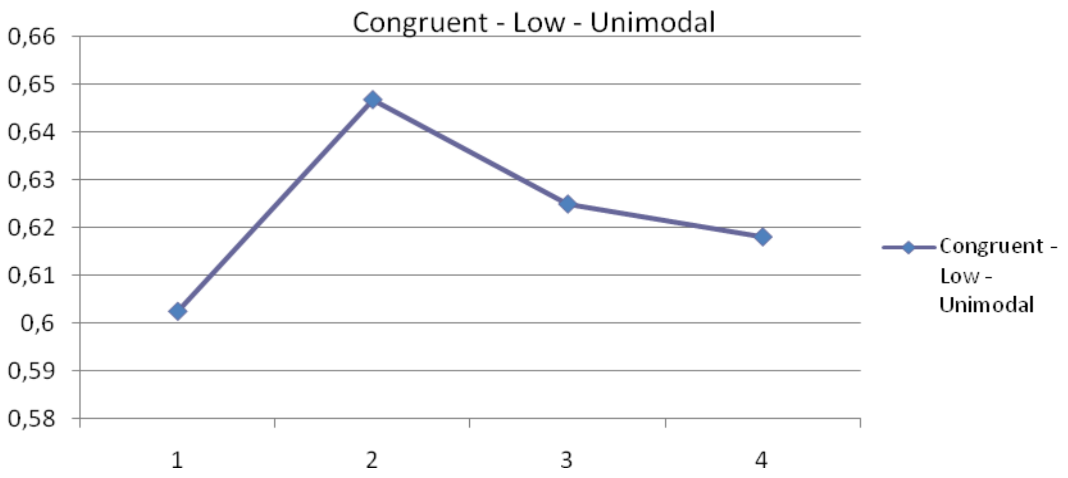
III) Image of Congruent Incongruent Cube Setups

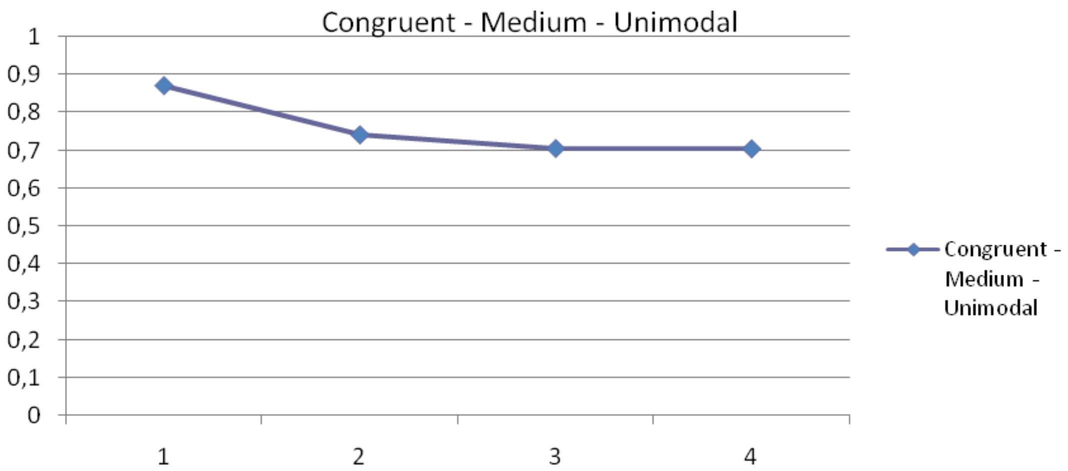


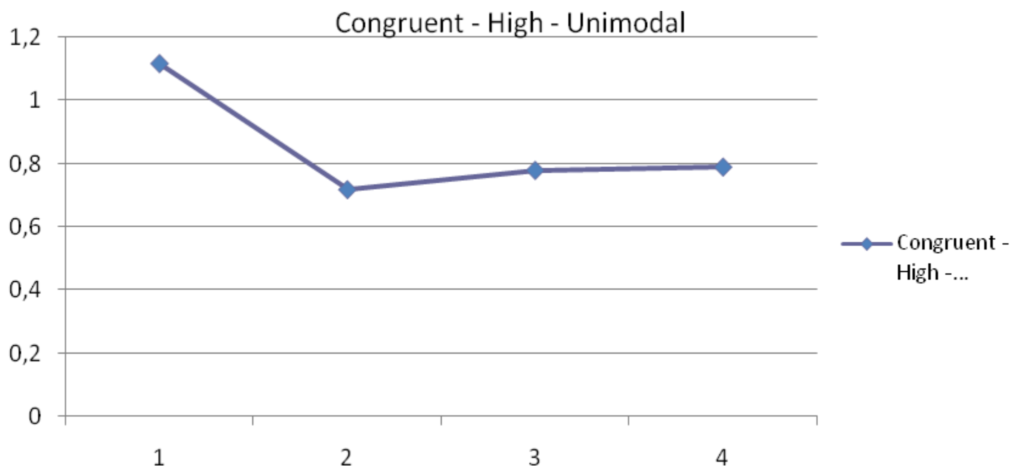
IV) Image of Congruent Incongruent Arrow Setups

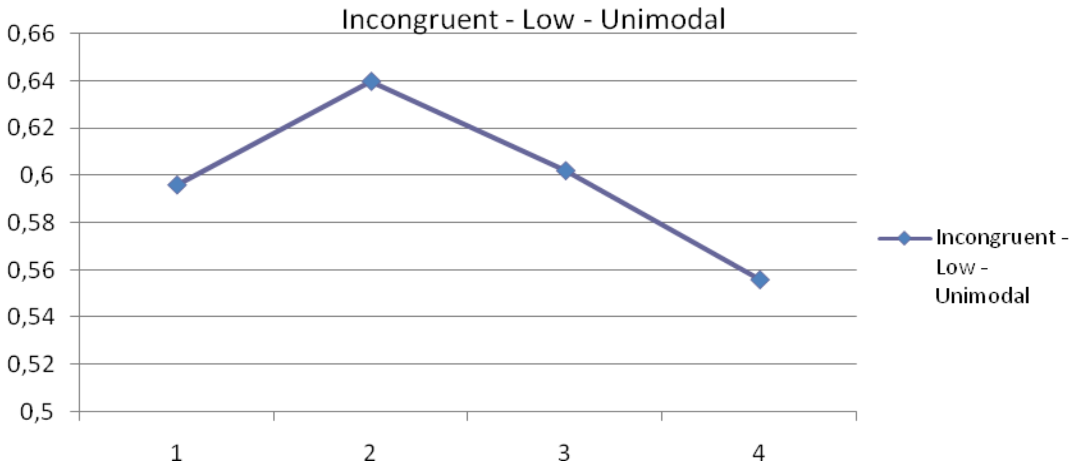


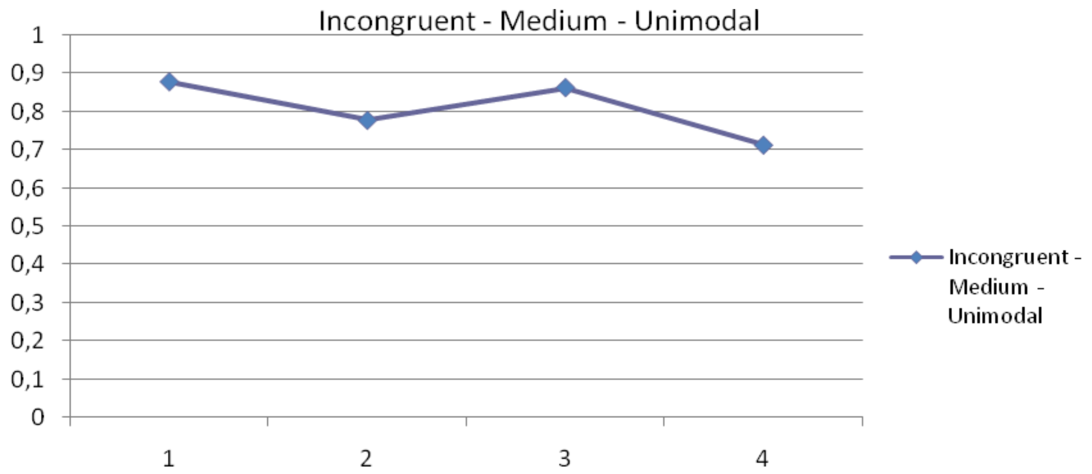
V) Learning Curves for each Respective Setup

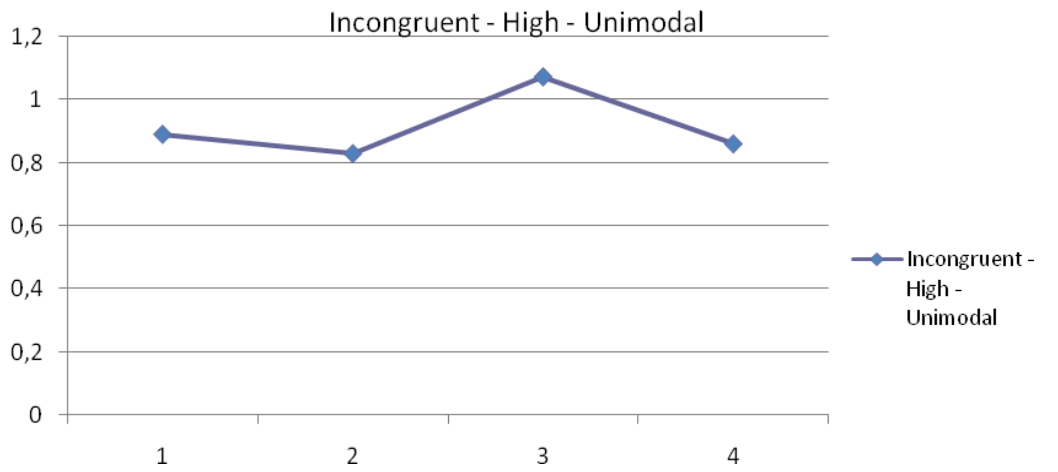


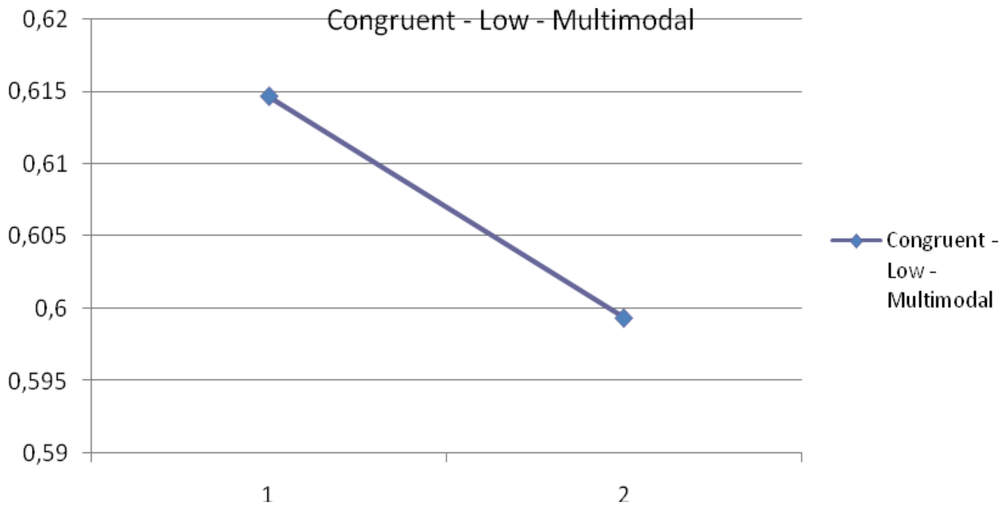


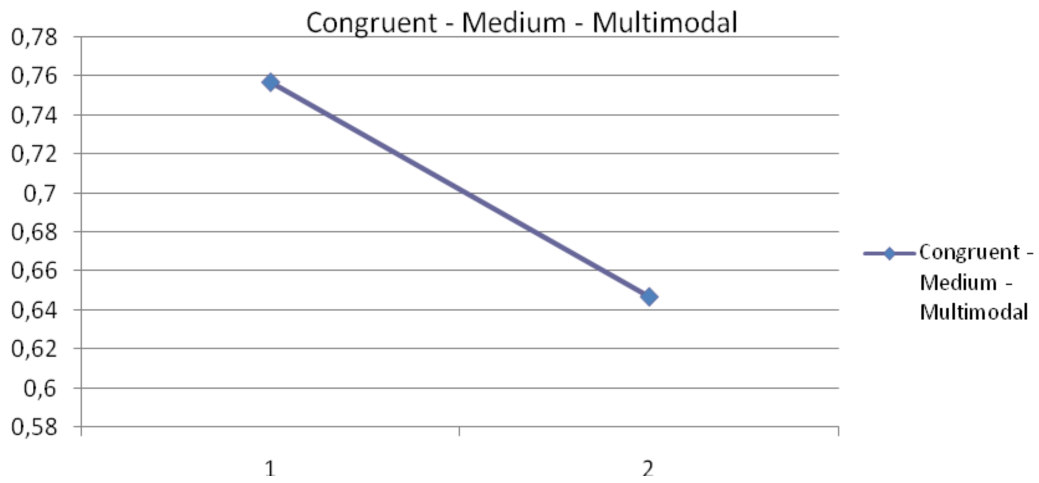


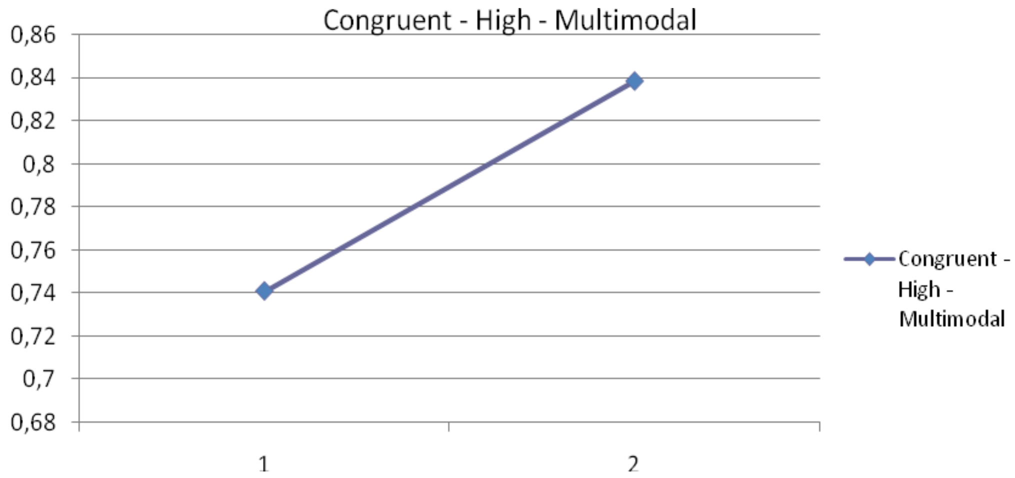

Incongruent - Low - Unimodal
0,66
0,64
0,62
0,6
0,58
0,56
0,54
0,52
0,5
1
2
3
4
Incongruent - Low - Unimodal

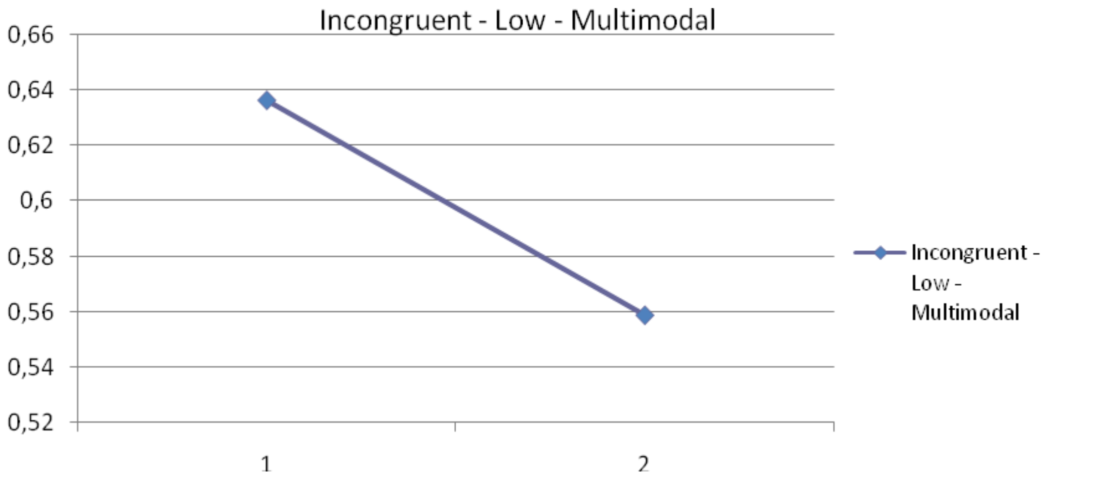

Incongruent - Medium - Unimodal
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
1
2
3
4
Incongruent - Medium - Unimodal

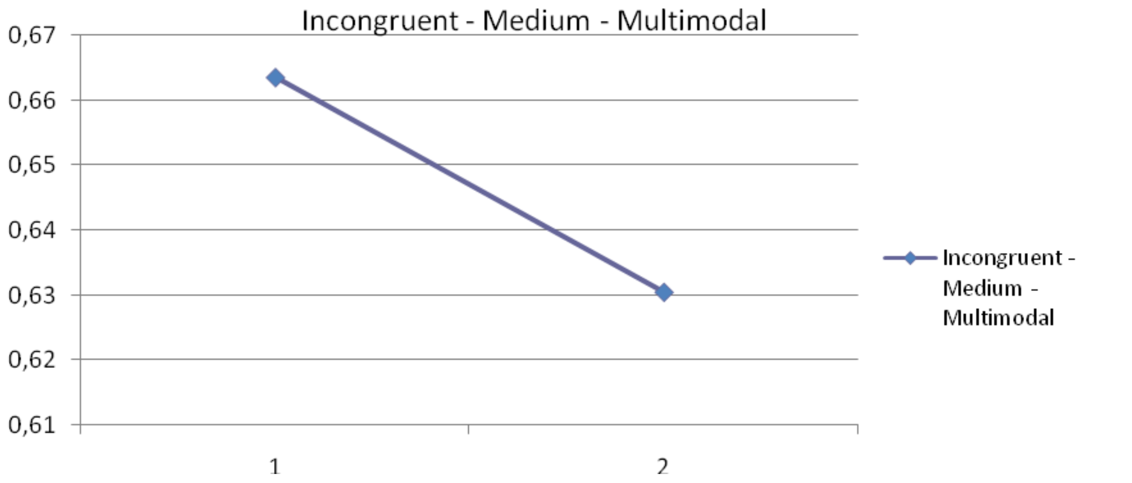

Incongruent - High - Unimodal
1,2
1
0,8
0,6
0,4
0,2
0
1
2
3
4
Incongruent - High - Unimodal

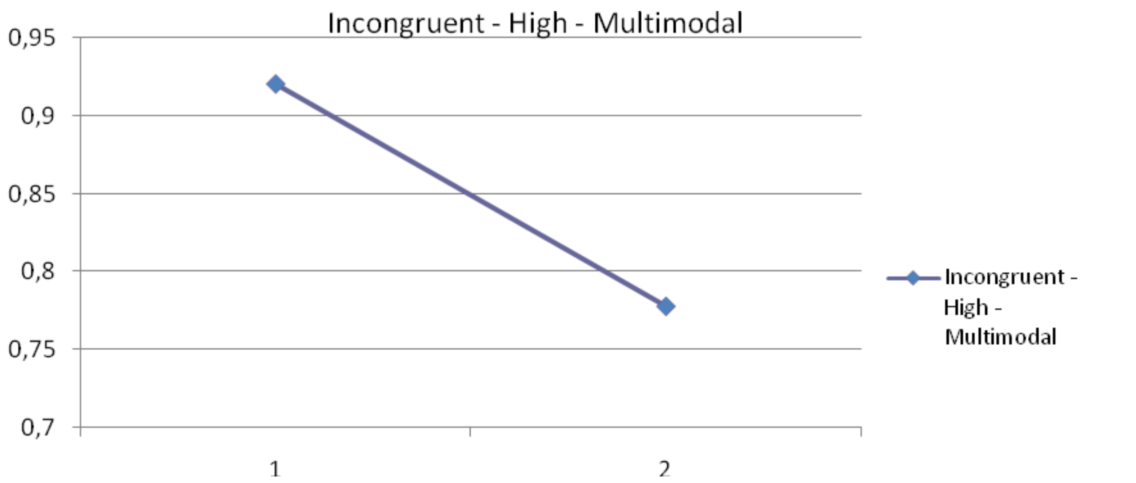

Congruent - Low - Multimodal
0,62
0,615
0,61
0,605
0,6
0,595
0,59
1
2
Congruent - Low - Multimodal


Congruent - Medium - Multimodal
0,78
0,76
0,74
0,72
0,7
0,68
0,66
0,64
0,62
0,6
0,58
1
2
Congruent - Medium - Multimodal


Congruent - High - Multimodal
0,86
0,84
0,82
0,8
0,78
0,76
0,74
0,72
0,7
0,68
1
2
Congruent - High - Multimodal


Incongruent - Low - Multimodal
0,66
0,64
0,62
0,6
0,58
0,56
0,54
0,52
1
2
Incongruent - Low - Multimodal


Incongruent - Medium - Multimodal
0,67
0,66
0,65
0,64
0,63
0,62
0,61
1
2
Incongruent - Medium - Multimodal


Incongruent - High - Multimodal
0,95
0,9
0,85
0,8
0,75
0,7
1
2
Incongruent - High - Multimodal

VI) Result tables

1. Manipulation checks

1.1 Mean response time to arrow or cube

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	CUBE	,7672	11	,05929	,01788
	ARROW	,7426	11	,07251	,02186

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	CUBE - ARROW	,02460	,08120	,02448	-,02995	,07915	1,005	10	,339

1.2 Location of the target objects on screen

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	LocationLEFT	,7662	17	,12516	,03036
	LocationRIGHT	,7772	17	,10882	,02639
Pair 2	LocationMIDDLE	,7453	13	,08304	,02303
	LocationLEFT	,7583	13	,07957	,02207
Pair 3	LocationMIDDLE	,7505	13	,08003	,02220
	LocationRIGHT	,7570	13	,07636	,02118

Paired Samples Test

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	Df	Sig. (2-tailed)
Pair 1	LocationLEFT - LocationRIGHT	-,01092	,14682	,03561	-,08641	,06457	-,307	16	,763
Pair 2	LocationMIDDLE - LocationLEFT	-,01302	,11182	,03101	-,08059	,05455	-,420	12	,682
Pair 3	LocationMIDDLE - LocationRIGHT	-,00652	,07662	,02125	-,05283	,03978	-,307	12	,764

2. Hypothesis result tables

2.1 Perceptual load comparison low, middle, high

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	LOW	,6097	11	,04504	,01358
	MIDDLE	,7886	11	,09120	,02750
Pair 2	HIGH	,8825	13	,11303	,03135
	MIDDLE	,7880	13	,08658	,02401
Pair 3	LOW	,6204	13	,05652	,01568
	HIGH	,8601	13	,08157	,02262

Paired Samples Test

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	LOW - MIDDLE	-,17886	,08082	,02437	-,23316	-,12457	-7,340	10	,000
Pair 2	HIGH - MIDDLE	,09451	,11636	,03227	,02419	,16482	2,928	12	,013
Pair 3	LOW - HIGH	-,23971	,10259	,02845	-,30171	-,17772	-8,425	12	,000

2.2 Modality specific loads manipulations (visual or auditory & visual)

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	HIGHvisual	,8705	16	,11982	,02995
	MIDDLEvisual	,7991	16	,09734	,02434
Pair 2	MIDDLEvisual	,8026	13	,10005	,02775
	LOWvisual	,6049	13	,05966	,01655
Pair 3	LOWaudio	,6333	18	,11972	,02822
	MIDDLEaudio	,7350	18	,10451	,02463
Pair 4	MIDDLEaudio	,7363	16	,10855	,02714
	HIGHaudio	,8805	16	,11947	,02987

Paired Samples Test

		Paired Differences					T	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	HIGHvisual - MIDDLEvisual	,07137	,12753	,03188	,00341	,13932	2,238	15	,041
Pair 2	MIDDLEvisual - LOWvisual	,19772	,10285	,02852	,13557	,25987	6,932	12	,000
Pair 3	LOWaudio - MIDDLEaudio	-,10169	,10362	,02442	-,15323	-,05016	-4,164	17	,001
Pair 4	MIDDLEaudio – HIGHaudio	-,14419	,12545	,03136	-,21104	-,07734	-4,597	15	,000

2.3 All congruent and all incongruent comparison

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	CONGRUENT	,7367	11	,03662	,01104
	INCONGRUENT	,7731	11	,08448	,02547

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	CONGRUENT - INCONGRUENT	-,03641	,07747	,02336	-,08846	,01564	-1,559	10	,150

2.4 Congruent and incongruent comparison in the different load levels

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	LowInongruent	,6408	16	,07869	,01967
	LowCongruent	,6206	16	,04167	,01042
Pair 2	MiddleInongruent	,7893	14	,13696	,03660
	MiddleCongruent	,7650	14	,06743	,01802
Pair 3	HighCongruent	,8359	18	,16462	,03880
	HighIncongruent	,8915	18	,11165	,02632

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	LowIInongruent - LowCongruent	,02022	,06483	,01621	-,01433	,05476	1,247	15	,231
Pair 2	MiddleInongruent - MiddleCongruent	,02430	,11123	,02973	-,03993	,08852	,817	13	,428
Pair 3	HighCongruent - HighIncongruent	-,05554	,19247	,04537	-,15125	,04018	-1,224	17	,238

2.5 Multimodal congruent and incongruent

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	AudioCongruent	,7458	16	,08798	,02200
	AudioINcongruent	,7515	16	,12602	,03150

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	AudioCongruent - AudioINcongruent	-,00576	,10263	,02566	-,06045	,04893	-,225	15	,825

2.6 Unimodal congruent vs. multi modal congruent and unimodal incongruent vs. multi modal incongruent

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	VISUALcongruent	,7483	15	,08942	,02309
	AUDIOcongruent	,7328	15	,07947	,02052
Pair 2	VISUAincongruent	,7786	12	,08718	,02517
	AUDIOIncongruent	,7390	12	,10074	,02908

Paired Samples Test

		Paired Differences					T	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	VISUALcongruent – AUDIOcongruent	,01544	,05833	,01506	-,01686	,04775	1,026	14	,323
Pair 2	VISUALincongruent – AUDIOincongruent	,03960	,07489	,02162	-,00799	,08718	1,832	11	,094

2.7 Unimodal vs. multi modal in the different load setups

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	HIGHvisual	,8599	18	,11760	,02772
	HIGHaudio	,8714	18	,11607	,02736
Pair 2	MIDDLEvisual	,8065	14	,10152	,02713
	MIDDLEaudio	,7186	14	,08919	,02384
Pair 3	LOWvisual	,6199	16	,06106	,01526
	LOWaudio	,6272	16	,07373	,01843

Paired Samples Test

		Paired Differences					T	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	HIGHvisual - HIGHaudio	-,01151	,11980	,02824	-,07109	,04806	-,408	17	,689
Pair 2	MIDDLEvisual - MIDDLEaudio	,08791	,06581	,01759	,04991	,12591	4,998	13	,000
Pair 3	LOWvisual – LOWaudio	-,00727	,07900	,01975	-,04936	,03483	-,368	15	,718

2.8 Middle perceptual load visual in/congruent and Middle perceptual load audio in/congruent

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	INCONaudioMIDDLE	,6724	17	,11820	,02867
	INCONvisualMIDDLE	,8188	17	,15298	,03710
Pair 2	CONaudioMIDDLE	,7403	18	,07918	,01866
	CONvisualMIDDLE	,7655	18	,07564	,01783

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	INCONaudioMIDDLE - INCONvisualMIDDLE	-,14635	,11214	,02720	-,20401	-,08870	-5,381	16	,000
Pair 2	CONaudioMIDDLE - CONvisualMIDDLE	-,02524	,09515	,02243	-,07255	,02208	-1,125	17	,276

VII) Anova Results

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Dif 1 con	23	14.30908333	0.622134058	0.004742055
Dif 1 incon	23	13.952	0.606608696	0.010808929
dif 1 con sound	23	14.5285	0.631673913	0.013337968
dif 1 incon sound	23	14.2095	0.617804348	0.018929676

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.007456397	3	0.002485466	0.207907732	0.890677822	2.70818651
Within Groups	1.052009829	88	0.011954657			
Total	1.059466226	91				

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Dif 3 con	23	17.57283333	0.764036232	0.006289306
Dif 3 incon	23	18.82475	0.818467391	0.021708634
Dif 3 con sound	23	17.1765	0.746804348	0.007838517
Dif 3 incon sound	23	15.128	0.65773913	0.041098497

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.307398113	3	0.102466038	5.327411373	0.002026402	2.70818651
Within Groups	1.692568993	88	0.019233739			
Total	1.999967105	91				

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
dif 6 con	23	19.56925	0.850836957	0.047472333
dif 6 incon	23	20.51825	0.892097826	0.033265675
Dif 6 con sound	23	18.506	0.804608696	0.021692181
Dif 6 incon sound	23	19.973	0.868391304	0.065263772

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.094485351	3	0.031495117	0.751252265	0.524514482	2.70818651
Within Groups	3.689267125	88	0.04192349			
Total	3.783752476	91				

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Dif 3 con	23	17.57283333	0.764036232	0.006289306
Dif 3 incon	23	18.82475	0.818467391	0.021708634

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.034071638	1	0.034071638	2.433867446	0.125904189	4.061706349
Within Groups	0.615954688	44	0.01399897			
Total	0.650026326	45				

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Dif 3 con sound	23	17.1765	0.746804348	0.007838517
Dif 3 incon sound	23	15.128	0.65773913	0.041098497

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.091225049	1	0.091225049	3.728263814	0.05995809	4.061706349
Within Groups	1.076614304	44	0.024468507			
Total	1.167839353	45				