Multisensory Response Enhancement of Semantically Congruent and Incongruent Audio-Visual Stimuli

Ruben Jurna
Utrecht University

Bachelor Thesis
Liberal Arts & Sciences – Cognitive Neurobiological Psychology
15ECTS

Supervisor: Dr. Maarten van der Smagt
Second supervisor: Nathan van der Stoep
Student number: 3964809
Date: 01-08-2016

Abstract

This research explores whether multimodal semantic congruency facilitates larger multisensory response enhancement (MRE) compared to MRE facilitated by the principle of congruent effectiveness. Semantically matched audio-visual stimuli pairs with dissimilar response times in their unisensory condition, and semantically non-matched stimuli pairs with similar response times in their unisensory condition, were used. The results are in accordance with the principle of congruent effectiveness; unisensory conditions with similar RT's elicit larger MRE in their respective multimodal condition. No significant differences in MRE were found between matched and non-matched conditions. However, this does not necessarily imply multimodal semantic congruency does not influence MRE. Due to the fact the manipulation in the experiment was unsuccessful, the current results cannot distinguish between the effects of multimodal semantic congruency- and the principle of congruent effectiveness on MRE.

Table of Contents

Preface

Liberal Arts & Sciences is an education that adopts an interdisciplinary approach to tackle scientific issues. The idea behind this approach is that one can combine theories, insights and methods from different scientific disciplines. This creates the possibility to fully explain a complex scientific problem, which cannot be explained with the insights from one scientific field alone. Especially contrasting views between fields of study provide a challenging opportunity to look beyond the scope of one's disciplinary worldview. When interdisciplinary research is performed correctly the result is always more than the sum of its individual disciplinary parts (Repko, 2012). The brain is a highly interdisciplinary entity, regulating bodily functions, lower- and higher order cognitions simultaneously. It seems appropriate researchers with an interdisciplinary background are exploring the functionalities of the brain.

Interdisciplinary research builds on the insights of different kinds of scientific fields. Therefore it is also necessary one specializes in a particular disciplinary field. Because of my predilection to study human behaviour I specialized in cognitive neurobiological psychology. This particular field attempts to explain behaviour by examining underlying cognitive processes and brain functions.

The field of artificial intelligence (AI) aspires to create intelligent systems, which can successfully act on their perceived environment. For instance, AI-enabled products such as smart advisors, personal robots and autonomous cars are currently being developed (Zhang, 2016). Human perception depends largely on the correct integration of multiple sensory signals. Multiple signals from different sensory modalities have to be correctly processed in order to perceive the world as a flowing actuality. Intelligent systems such as personal robots need integrative features resembling the human brain to be able to perceive their environment on the same level humans do. In the current study an experimental research is conducted concerning the dynamics governing integration of signals from different sensory modalities.

The purpose of this research is to investigate dynamics concerning multisensory enhancement. Specifically, the aim is to assess whether semantically matched audio-visual stimuli pairs will elicit larger multisensory enhancement compared to non-matched pairs. A better understanding of the dynamics governing multisensory behaviour could help achieve AI's goal to create intelligent systems with perceptual integrative features resembling the human brain.

Introduction

We often receive information about an object or event through different sensory modalities. It is essential this information is integrated to create an accurate and comprehensive understanding of the world. For example, when we see a car sounding its horn at a duck crossing the road, our brain correctly integrates the sound of the horn with car and the sound the duck makes when it is flying away, with the duck. This process of combining information from multiple sensory modalities is called multisensory integration (Meredith & Stein, 1986). The first neurophysiological studies of multisensory integration used single cell recordings in the superior colliculus (Meredith & Stein, 1983; Meredith & Stein 1986; Meredith, Nemitz & Stein 1987). The researchers discovered neurons with a more frequent spiking rate to multisensory stimuli compared to unisensory stimuli. This shows that on a neurological level integration of different sensory modalities takes place. This increase in neuronal response to multisensory stimuli, compared to unisensory stimuli, is called the *multisensory response enhancement* (MRE). In behavioural studies MRE becomes apparent when a shorter response time is found in the multimodal condition compared to the shortest response time from the respective unimodal conditions (Stevenson et al., 2014).

Although we often receive multiple signals from multiple sensory modalities simultaneously, the process of integrating these signals seems to work effortlessly. How signals from different sensory modalities are integrated remains unknown. This problem is known as the cross modal binding problem and has been extensively investigated by the scientific fields of psychology as well as philosophy (Kant, 1965; Mashour, 2004; Reynolds & Desimone 1999; Riesenhuber & Poggio, 1999). While a definitive explanation for this problem has yet to be discovered, the research above shows neuroscience is leading the way in this endeavour.

Three neural principles governing multisensory integration on a neural level have been found. Two of these principles can be extended to behavioural measures. Other principles, which seem to be of influence in multisensory behaviour, will also be discussed.

Three rules or principles have been found governing optimal neuronal responses to multisensory stimuli: the *spatial rule*, the *temporal rule* and the *principle of inverse effectiveness* (Stein & Meredith, 1993). The first and second principles are similar in nature. The spatial rule states that the multisensory integration is more likely to occur when multisensory stimuli originate at the same location (Meredith & Stein, 1986). The second principle, the temporal rule, is almost identical to the spatial rule; applying to temporal onset of stimuli (Meredith & Stein 1987). The third rule is called the principle of inverse effectiveness. It states that unisensory stimuli, which elicit relatively weak neuronal responses, will elicit larger MRE's in their multimodal conditions, compared to stimuli that elicit strong neuronal responses in their unimodal conditions (Meredith & Stein, 1983; Senkowski, Saint-Amour, Hofle & Fox, 2011; Senkowski, Stein & Meredith, 1993).

An important question is when the integration of different sensory modalities takes place. Is each signal processed separately and integrated later, or are multisensory signals processed collectively? Miller (1982) proposed a co-activation model where signals from different modalities contribute to the same activation pool that initiates a response. As opposed to a separate activation model where signals from different modalities are processed separately. Later research showed Miller's evidence to support his claims was rather weak and the design of his experiment was potentially flawed (der Heijden, Schreuder, Maris & Neerincx, 1982).

When unisensory signals are being processed, activation from one sensory modality is present. When multisensory signals are being processed activation from two sensory modalities is present. Otto and Mamassian (2012) found that when evidence is accumulated for multisensory signals there is increased activity within separate activation pools of the respective modalities, which produces additional noise compared to when only one modality is present. Otto and Mamassian (2012) show multisensory enhancement can be partially explained by the additional noise from an extra sensory modality that is present, when processing multisensory signals. This should always be taken into account when analysing data from multisensory research.

Response time (RT) is time elapsed between presentation of a sensory stimulus and the corresponding behavioural response. Miller (1982) found a decrease in RT to multimodal stimuli, compared to stimuli from a single modality. This is called the *redundant signal effect* (RSE). RSE is often used to measure multisensory enhancement and is based on probability summation. Multisensory integration research uses the so-called race model (Miller, 1982; Raab, 1962) to distinguish multisensory enhancement due to probability summation from multisensory integration. This model derives its name from the apparent race that occurs between different sensory modalities to finish processing a signal first. Only when performance on a multisensory task exceeds the race model, multisensory enhancement can be attributed to integration of different sensory signals. An important note here is that due to a time constraint all results from this research will be regarding MRE. No claims will be made regarding multisensory integration.

It has been found that, at least partially, neurological principles governing multisensory enhancement apply to behavioural measures as well. Especially the spatial and temporal rules seem to hold well in behavioural studies (Bolognini, Frassinetti, Serino & Ladavas, 2005; Hairston, Laurienti, Mishra, Burdette & Wallace, 2003). Bolognini (2005) found that when a visual and an auditory stimulus are spatially aligned, the number of correct responses on a visual detection task increased. This increase was not found when the visual and auditory stimuli were presented at different locations. Bolognini also found that the improvement, due to spatial alignment, was only present if the stimuli were presented at the same time. Moreover, Stevenson (2012) confirmed that there is an interaction effect between the spatial location and the temporal onset of audio-visual stimuli. This shows the spatial and temporal rules do not act independently.

Behavioural measures show conflicting findings with respect to the principle of inverse effectiveness (Holmes, 2007; Ross, Saint-Amour, Leavitt & Javitt & Foxe, 2007; van der Smagt, Buijing & van der Stoep, 2014; Stevenson, 2012). Ross (2007) found that when absolute MRE's were analysed, the principle of inverse effectiveness was not present; contrary to when relative MRE's were used. This shows that the occurrence of the principle of inverse effectiveness depends on which statistical method is used. Combined, this is strong evidence the principle of inverse effectiveness does not always hold in behavioural measures. Holmes (2009) even suggested this principle does not exist at all in behavioural measures.

Besides three principles governing multisensory enhancement on a neural level, behavioural studies also found other principles governing multisensory behaviour. The principle of *congruent effectiveness* states that when the distribution of response times between unimodal conditions is similar, MRE in the respective multimodal condition will be larger (Otto, Dassy & Mamassian, 2013). Recent research confirms the effect of the principle of congruent effectiveness in behavioural measures (van der Smagt & van der Stoep, 2015).

Another important factor influencing multisensory enhancement is the *unity assumption*, i.e. the degree to which a person infers two sensory stimuli belong to the same source or event (Welch & Warren, 1980). As discussed, the spatial and temporal rules show the assumption of unity is strongest when the spatial and temporal onset of a stimulus is similar. Other forms of multimodal similarities also seem to influence this assumption. Stimuli that are semantically or synaesthetically congruent are more likely to be bound together (Vatakis, Ghazanfar & Spence, 2008; Vatakis & Spence 2007). Synaestical congruence refers to more basic stimulus properties like pitch, size and brightness (Spence, 2011). For instance, Melara (1989) showed that congruent pitch-colour combinations elicit larger multisensory enhancement than incongruent combinations. Vatakis and Spence (2007) found that subjects were better in distinguishing which stimulus came first when looking at a face of a talking male combined with a female voice (incongruent stimuli) than when the stimuli were of congruent nature. Moreover, Chen and Spence (2010) found that a semantically congruent sound improved performance on a picture identification task whereas an incongruent sound impaired performance. Taken together this evidence shows stimuli with congruent features concerning location, temporal onset, identity and basic stimuli properties are more often correctly bound together and perceived as coming from the same event.

Recently, van der Smagt et al., (2014) conducted a research were subjectively matched audio-visual stimuli were created. Participants matched the auditory intensity of a white noise stimulus to the brightness of a visual stimulus. They found larger MRE's in the subjectively matched condition compared to the +5 dB and -5 dB conditions. The results also confirmed the presence of the principle of congruent effectiveness. The smallest difference in RT's between unimodal conditions was found in the subjectively matched condition (van der Smagt & van der Stoep, 2015). This shows the principle of congruent effectiveness applies strongest in this condition. Hence, the results of this research cannot confirm if the increase in MRE found was due to the effect of subjective matching or the principle of congruent effectiveness. The design of the current experiment circumvents this problem.

The research discussed implicates similarities of any kind between unisensory stimuli would facilitate larger MRE in their respective multimodal condition. Surprisingly, there has not yet been much research concerning the influence of semantic congruency on MRE. This is remarkable because multimodal congruencies concerning identity or meaning are very relevant in multisensory perception (Vatakis et al., 2008; Vatakis & Spence 2007).

The focus of the current research will be to assess the influence of multimodal semantic congruency- compared to the influence of the principle of congruent effectiveness, on MRE. Because multimodal congruencies seem to enhance MRE it is deemed likely that semantic congruency will influence MRE in similar ways. To be able to distinguish the effect of semantic congruency from the effect of congruent effectiveness, it was necessary semantically congruent audio-visual stimuli pairs with dissimilar RT's were created.

Two manipulations were performed in course of the experiment. First, two visual stimuli, a car and a duck, which yielded dissimilar RT's, were created. Secondly, staircase procedures were used to equalize the RT's yielded by the auditory stimuli to the RT's yielded by the semantically incongruent visual stimuli. Thus, two semantically incongruent audio-visual stimulus pairs with similar RT's in their respective unimodal conditions were created. This would also result in two semantically congruent audio-visual stimulus pairs with unequal RT's in their respective unimodal conditions. This manipulation ensures the multisensory enhancement in the semantically congruent conditions cannot be ascribed to the principle of congruent effectiveness. The improvement in MRE in the semantically incongruent conditions can only be ascribed to the principle of congruent effectiveness. Figure 1 shows the expected MRE if multimodal semantic congruency influences MRE more strongly than the principle of congruent effectiveness. Figure 2 shows the expected MRE when the principle of congruent effectiveness influences MRE more strongly than multimodal semantic congruency. To ensure optimal multisensory enhancement, the spatial and temporal rule were followed, i.e. the audio-visual stimulus pairs were presented at the same time and location.

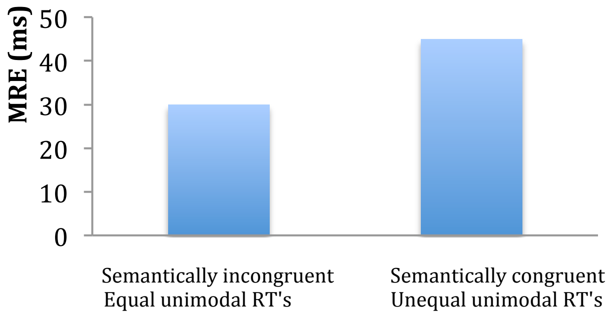


Figure 1. Expected MRE if multimodal semantic congruency facilitates larger MRE compared to the principle of congruent effectiveness.

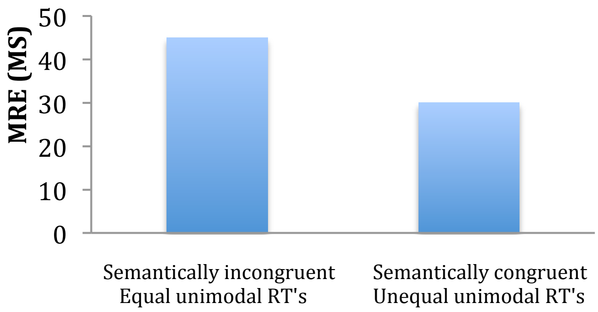


Figure 2. Expected MRE if the principle of congruent effectiveness facilitates larger MRE compared to multimodal semantic congruency.

Methods

Participants

Thirty-two healthy volunteers participated in the experiment. All subjects reported normal hearing and had normal or corrected-to-normal visual acuity. Nineteen were included in the analysis. This group consisted of thirteen males with a mean age of 26,3 (range = 22-34, SD = 4.0) and six females with a mean age of 31.6 (range = 23-46, SD = 8.6). All subjects signed an informed consent form prior to the experiment.

Apparatus & Stimuli

The subjects were placed at a table 90 cm from the projection screen. On the table a white fabric projection screen, mouse, keyboard and response box were placed. The response box was used to record the RT for each trial. A chinrest was used to maintain a fixed distance from the screen. The experiments were projected with an Acer 1261P beamer with a refresh rate of 60HZ. The beamer was located on a gantry above the subjects' head. The beamer was the only light source in the room. The auditory stimuli were presented with Harman/Kardon HK206 speaker boxes, which were placed behind the fabric screen at the exact location where the visual stimuli were presented. The code of the experiment was developed in Matlab 2010a using psychtoolbox version 3 (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007) and executed by a Compac dc5750 HK206 computer.

A red fixation cross (visual angle 0.39°) was presented in the middle of the screen on a grey background (intensity 58.2 cd/m2). In a pilot study the contrasts of two visual stimuli, a duck (3.59° diameter, intensity 457.2 cd/m2, figure 3) and a car (3.27° diameter, intensity 69.6 cd/m2, figure 4), were manipulated to create two visual stimuli with unequal RT's to stimulus onset. The contrast (Michelson) between the background and visual duck stimulus was 0.77 and between the background and visual car 0.09. The auditory stimuli were the sound of a duck and the horn of a car. The intensity of the auditory stimuli was different for each subject, the values where extracted from the completed staircases. The stimuli were presented on the same height as the fixation cross on either the left or right side. The distance between the centres of the fixation cross and the visual stimuli was 15 cm or 9.44° visual angle.

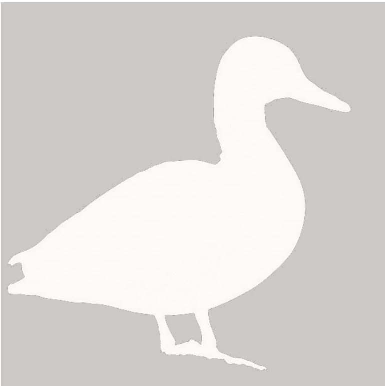


Figure 3. Visual duck stimulus

Figure 4. Visual car stimulus

Procedure

The entire experiment consisted of three parts: a pilot study, four staircase procedures and the main experiment. The participants were instructed to place their head on the chinrest and place their hands on the button of the response box before the start of each part of the experiment. They were instructed to respond as fast as possible when a stimulus was presented and to refrain from responding when there was no stimulus present. Before each part of the experiment began, the subjects were given an opportunity to ask questions.

In all experiments a trial consisted of four phases: fixation cross, blank screen, presentation of a target and an inter-trial interval (see figure 5). The fixation cross was presented for a random duration of 750-1250 ms before it was removed. A blank screen, background colour, was projected for a random duration of 200-250 ms before the onset of a target. Targets consisted of visual (V), auditory (A) or audio-visual stimuli (AV). The target modality and location (left/right from fixation cross) was randomized. The response window was 1500 ms from target onset. After a response was given an inter-trial interval of 2000 ms followed. If no response was given the trial continued to the inter-trial interval. Trials without a target (catch trials) were included in the pilot study and the main experiment to assess if subjects were behaving as instructed in the course of the experiment.

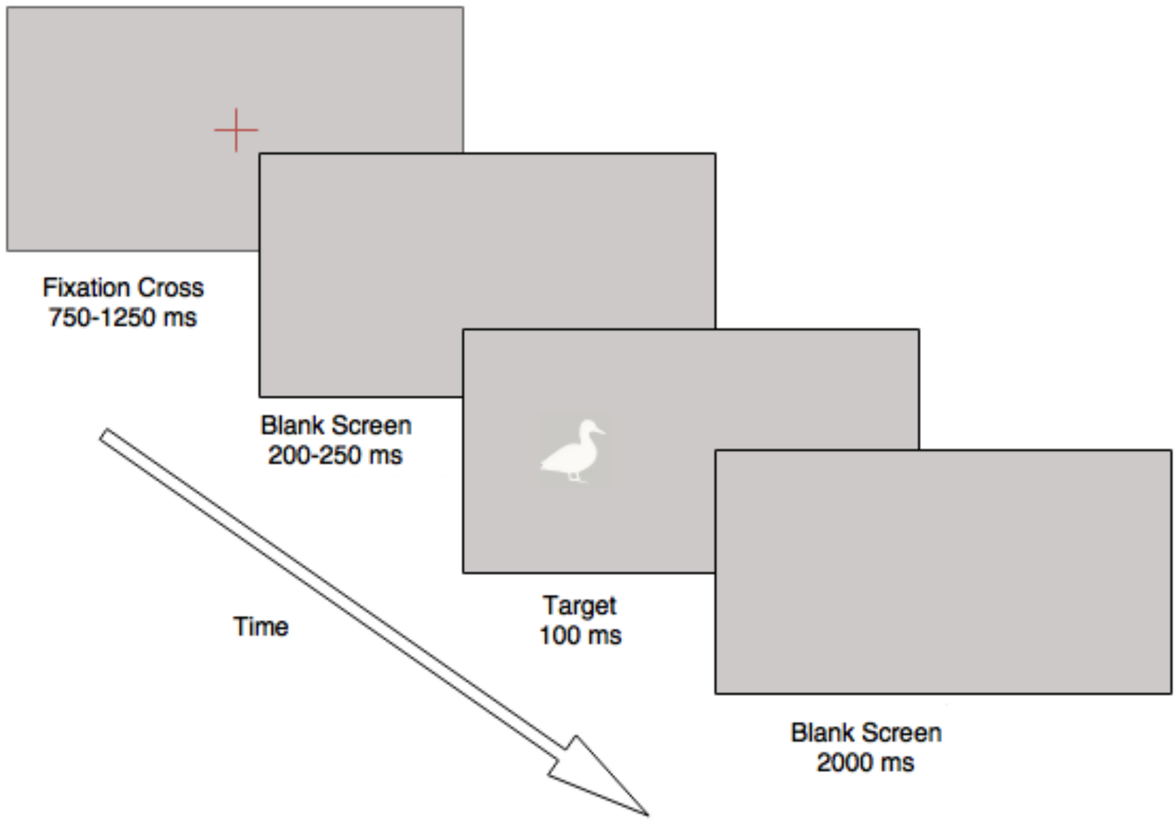


Figure 5. Illustration of the different phases of a trial. This was kept equal for the pilot study, staircases and main experiment. Only the type of target differed in the three parts of the experiment: pilot study (visual), staircase (auditory) and main experiment (visual, auditory and audio-visual targets). Note that the audio-visual targets were always presented at the same time and location.

Pilot

In the pilot study subjects performed a speeded detection task to the visual targets only. The subjects were instructed to respond as fast as possible when they saw either a car or a duck and to refrain from responding when no stimulus was presented. The pilot study consisted of forty trials for each visual stimulus and eight catch trials were added. The median RT's to the visual stimuli were used as input for the staircase procedure.

Staircases

A double staircase (starting level staircase up: 37.72 dB starting level staircase down: 51.70 dB) was used to equalize the RT yielded by the auditory car stimulus to the RT yielded visual duck stimulus. This was also done for auditory duck and visual car stimuli. The staircases were programmed to take the average RT[*] of three trials and compare this to the median RT of the respective visual stimulus given as input. If the average RT yielded by the auditory stimulus was larger than the median RT yielded by the visual stimulus, the intensity was increased, and vice versa. If the RT to the auditory stimulus fell within a range of 10 ms of the given input, the volume remained at the current level. The RT's of both stimuli were considered equal when the difference in RT was 10 ms or less. The average values of the up- and down staircases were used as auditory intensity values for the auditory stimuli in the main

[*] The median RT of three numbers is always the middlemost number. Using the average here results in a more precise estimation of the actual RT.

experiment (Appendix table 1). The staircase procedures stopped after twelve steps (see Appendix figure 1 for staircase graphs for each participant). The order of the staircases (car/duck) was randomized across subjects.

After ten participants it became apparent that the RT's, yielded by the auditory stimuli with the lowest possible intensity setting from the staircase, were often still shorter than the RT's yielded by the visual stimuli. An attempt was made to increase the RT to the auditory stimuli by introducing earplugs in the experiment. The remaining participants were given earplugs before the beginning of the staircase procedure. The earplugs were to be held in for the remainder of the experiment.

Main experiment

The main experiment consisted of nine conditions with a total of 370 trials. Seventeen practice trials were added at the beginning of the experiment (every condition was presented left and right and one catch trial was added). The unimodal conditions consisted of: visual duck (VD), visual car (VC), auditory duck (AD) and auditory car (AC). The multimodal conditions consisted of each possible combination of the unimodal conditions where an auditory and visual target was present: VDAC, VCAD, VDAD, VCAC. Every condition consisted of 38 trials except for the VCAC condition, which contained 36 trials. This discrepancy came to be because there were 68 instead of 50 catch trials added in the main experiment. These eighteen extra trials accidently replaced trials from other conditions. A break was optional after completion of 1/3 and 2/3 of the experiment.

Data analysis

Due to an error in the staircase script, at first it was possible the staircases reached negative values. This would be practically impossible. The staircases of participants 1-7 contained at least one negative value; they were excluded from the analysis. Another five participants were excluded because at least one of their staircases did not converge to an auditory intensity. This is indication that the manipulation (equalisation of RT's between incongruent audio-visual stimuli) had failed. One subject was excluded from analysis because his response rate in the catch trial condition exceeded the 10% limit. The RT data of the remaining nineteen participants were further analysed.

Trials with RT's between 100 ms and 1000 ms and catch trials without a response were considered correct responses. RT's shorter than 100 ms indicate anticipatory responses. RT's longer than 1000 ms indicate the participant was not paying attention during that trial. This led to removal of 0.14% of all trials containing a stimulus. It was also recorded when there was no response given in a trial containing a stimulus. This was the case in 0.50% of all trials. Responses during catch trials were also registered. Overall there was a response of 1.93% in the catch condition from the main experiment. Only correct responses were further analysed.

Median RT's of each subject were used in the analysis because RT data is often skewed; median RT's are less influenced by the presence of outliers. The absolute and relative MRE's were calculated for all subjects using median RT's from their respective conditions (see figure 6 for illustration):

Absolute MRE (aMRE) = min (RT auditory, RT visual) – RT audio-visual

$$\text{Relative MRE (rMRE)} = \frac{\text{min (RT auditory, RT visual)} - \text{RT audio-visual}}{\text{min (RT auditory, RT visual)}} \times 100\%$$

To assess if the manipulations from the pilot and staircase succeeded, two-tailed paired *t* tests were performed in order to compare the RT's of the unimodal conditions. MRE's of semantically congruent conditions were compared to MRE's of semantic incongruent conditions. Pearson correlation tests were performed for each subject to assess if differences between unimodal conditions affect the MRE of their respective multimodal condition. This would be in correspondence to the principle of congruent effectiveness. The analysis was done in excel 2010 and SPSS version 23.

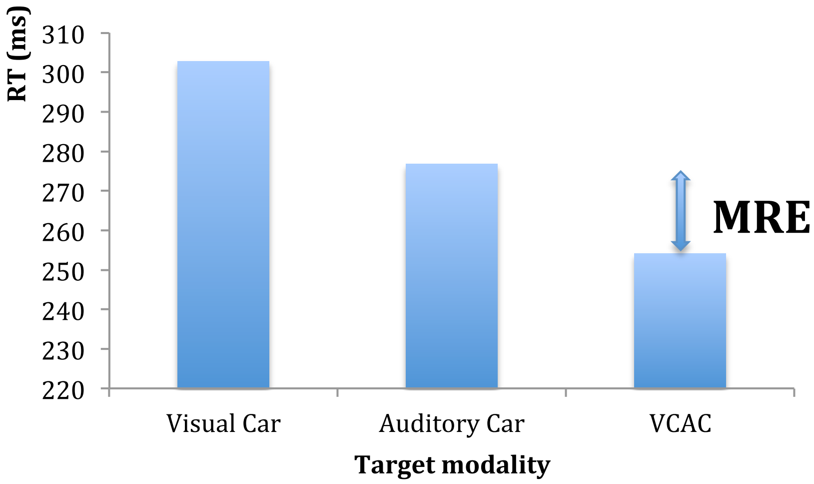


Figure 6. Illustration of the calculation of MRE. Values were used from subject 8 in the condition VCAC.

Results

Manipulation & response times

To obtain an overview of the data the median RT's of each unimodal and multimodal condition are shown in table 1. Note that median RT's from each subject are used in this calculation. A two-tailed paired *t* test showed a significant difference between the RT's to VD (M = 289, SE = 14) and VC (M = 332, SE = 15, $t(18) = -10.95$, $p < .001$, $r = 0.97$). This shows that the manipulation from the pilot was successful. The average differences in RT between the unimodal conditions are shown in table 2.

The manipulation from the staircase would have been successful if the difference in RT's between the conditions VD and AC would be zero and this was also true for VC and AD. Overall, differences seem to be quite large (30-55 ms), indicating the manipulation of the staircases was not successful.

Two-tailed paired-samples *t* tests were conducted to compare these conditions. There was a significant difference in RT found between the VC (M = 332, SE = 16) and AD conditions (M = 305, SE = 14, $t(18) = 3.34$, $p = .004$, $r = 0.86$). This shows the equalisation of the RT's to VC and AD failed. No significant difference was found between the conditions VD (M = 289, SE = 14) and AC (M = 280, SE = 17, $t(18) = 1.17$, $p = .257$, $r = 0.89$). The difference in RT between VD and AC was not significant, indicating this manipulation could have been successful.

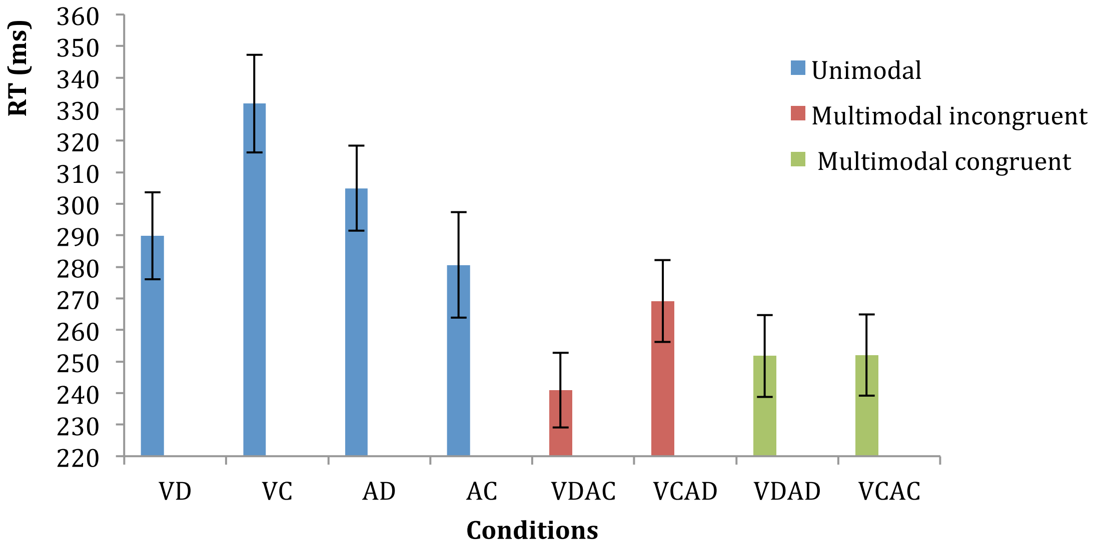


Table 1. Average of median RT's for each condition across participants. The error bars represent standard error of the mean.

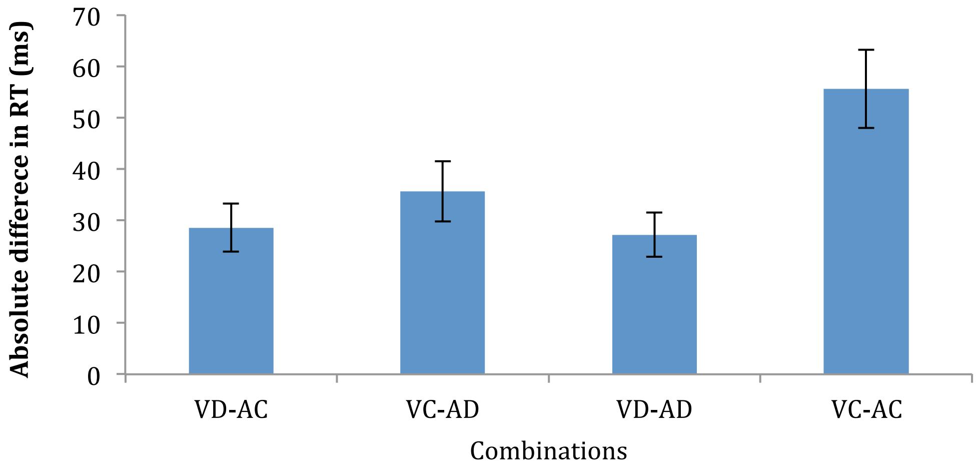


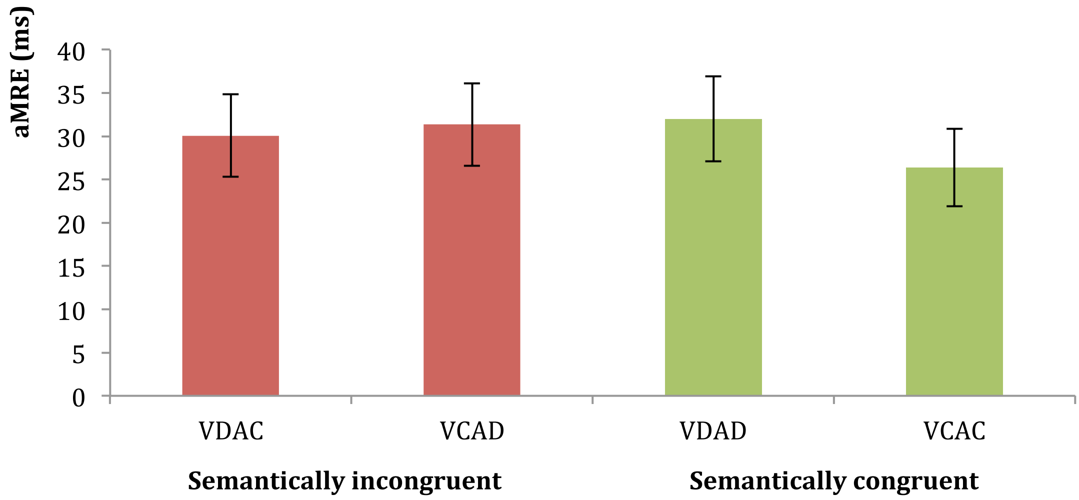
Table 2. Averages of absolute differences in RT in ms between unisensory conditions. Error bars represent standard error of the mean.

Multisensory response enhancement

To investigate if the absolute and relative MRE's were significant for each condition responses to multisensory conditions were compared to the respective unisensory conditions. Two-tailed paired sample *t* tests (all *t* 's > 5.1, *p*'s < .01) show a significant improvement in RT for all multimodal conditions (*p* values were corrected using the Bonferroni method). Table 3 shows the average absolute and relative MRE for each condition. The largest aMRE, stems from the VDAD condition, which is also the condition where the principle of congruent effectiveness applies the most; this condition has the smallest unimodal difference across all conditions (table 2).

To assess if the aMRE's and/or rMRE's of semantically congruent conditions are larger than incongruent conditions two-tailed paired *t* test were performed on all possible combinations. No significant differences were found between these groups, indicating no difference between MRE's in semantically congruent and incongruent conditions.

Therefore, it was possible to pool the raw RT data of multimodal semantically incongruent conditions; the same was done for the multimodal semantically congruent conditions. Pooling data here facilitates a more precise estimation of the investigated effects. The RT's yielded by unimodal conditions (VD + VC and AD + AC) were pooled to calculate the aMRE's in the semantically congruent- and incongruent condition. Although the mean of the pooled aMRE in the semantically congruent condition (M = 34, SE = 4) was larger than in the semantically incongruent condition (M = 38, SE = 4), a two-tailed paired-sample *t* test showed this difference was not significant.



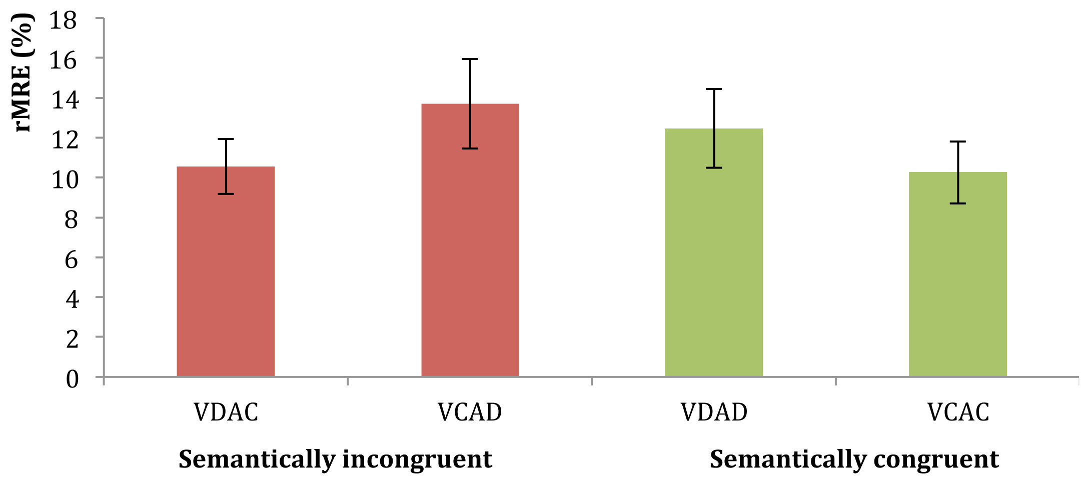


Table 3. Averages of absolute and relative MRE's in each condition. Error bars represent standard error of the mean. Red bars represent semantically incongruent conditions; green bars represent semantically congruent conditions.

Congruent effectiveness
The principle of congruent effectiveness states unimodal conditions with similar RT's elicit larger MRE's in their respective multimodal condition. The largest absolute difference is between VC-AC. Looking back at table 3; the smallest MRE is also from the VCAC condition. This is what one would expect following the principle of congruent effectiveness. In order to investigate this relationship for every participant, differences in RT's between unimodal conditions and their respective MRE's were compared using Pearson correlation

tests. The correlation was not significant for any of the participants individually. However, across all participants this correlation was significant. A Pearson test showed a relationship between differences in RT from unimodal conditions and aMRE (M = -7, SE = 4.56, $t(75) = -1.49$, $p = .001$, $r = -0.38$). This shows that smaller unimodal differences elicit larger MRE's in their respective multimodal condition (figure 7). The same analysis was done for unimodal differences and rMRE's (M = -26, SE = 3.51, $t(75) =-7.5$, $p = <.001$, $r = -0.45$). It shows rMRE's follow a similar pattern as aMRE's when set out against their unimodal differences (figure 8).

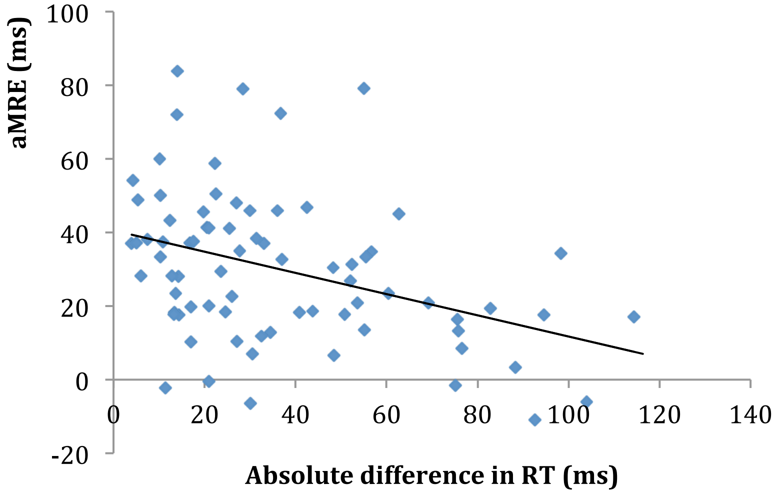


Figure 7. Absolute differences in RT between unimodal conditions plotted against their respective MRE's. An overall negative correlation (*r*= -0.38) is found.

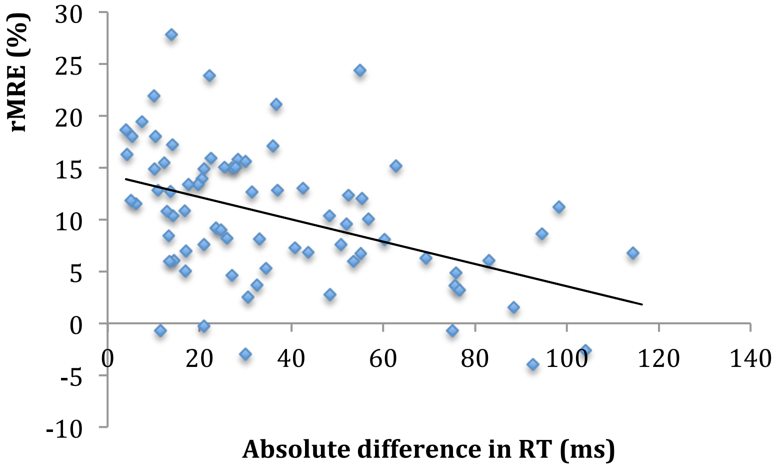


Figure 8. Differences in RT between unimodal conditions plotted against their respective aMRE's. An overall negative correlation (*r*= -0.45) is found.

Discussion

The main purpose of the current research was to examine whether multimodal semantic congruency facilitates larger MRE compared to the effect of the principle of congruent effectiveness. Consistent with the results of van der Smagt and van der Stoep (2015), the current data show a negative relationship between differences in unimodal RT's and MRE i.e., the principle of congruent effectiveness. Across all participants, the data show larger MRE when the difference in RT between the respective unimodal conditions was small. These findings are in line with the model proposed by Otto (2013) and the findings of van der Smagt and van der Stoep (2015).

The results showed a significant difference in RT between the visual stimuli, indicating the first manipulation was a success. It turned out to be difficult to equalize RT's from visual stimuli to auditory stimuli using staircase procedures. The results showed a significant difference in RT between the unimodal conditions VC and AD, indicating the manipulation in this condition could not have been successful. The difficulty in this experiment was the AD stimulus. Table 1 and 2 show a large difference between the conditions VC and AD, which should have similar RT's. The RT yielded by AD should have been at least 20-30 ms shorter to declare their respective RT's were equal. On average the RT yielded by AD was too short to equalize to the RT yielded by the VC stimulus. The difference in RT between the unimodal conditions VC-AD was large, causing the difference in RT of the unimodal VD-AD condition to be relatively small. Therefore, MRE in this condition is most influenced by the principle of congruent effectiveness. Therefore, the current design of this experiment could not confirm whether multimodal semantic congruency has a larger effect on MRE compared to the principle of congruent effectiveness. No significant differences in aMRE or rMRE were found between the semantically congruent- and incongruent conditions.

However, this does not necessary imply multimodal semantic congruency has no effect on MRE at all. Table 2 shows the largest difference in RT between unimodal conditions is between VC and AC. This implies MRE's in this condition cannot be ascribed to the principle of congruent effectiveness. Following the principle of congruent effectiveness one would expect smaller MRE in this condition. MRE in this condition is not significantly smaller than MRE's in the semantically incongruent conditions. This could be seen as weak evidence multimodal semantic congruency between audio-visual stimuli elicits larger MRE.

Because the staircase procedures did not produce the desired results, attempts were made to bring the RT's of the visual and auditory stimuli closer together. By enlarging the visual stimuli an attempt was made to decrease the RT's to the visual stimuli (subject 15-16). Increasing the size of the visual stimuli did result in shorter RT's. Unfortunately this also resulted in more similar RT's between VD and VC, which was clearly undesirable. Subsequently, an attempt was made to decrease the RT's to the auditory stimuli by introducing earplugs in the experiment (subjects 11-32).

In the setup of the experiment an important issue was not taken into consideration. The RT's yielded by auditory stimuli are usually much shorter than RT's yielded by visual stimuli (Shelton & Kumar, 2010; Pain & Hibbs, 2007). The conversion of a photon to a bioelectric signal takes longer than the conversion of a pressure wave to a bioelectric signal (Goldstein, 2013). A simplified explanation here is that visual stimuli travel a more complicated neural pathway in order to be fully processed.

This experiment tried to equalize the RT to auditory targets to the RT of visual targets. However, if RT's to auditory targets are always shorter than visual targets it is theoretically possible that for every possible intensity value the RT's to the auditory targets remained shorter than the RT of their respective visual counterparts.

Future research

The current research shows equalizing RT's of a sensory modality, which yields shorter RT's, to a sensory modality that yields larger RT's, is difficult. First and foremost, the sensory modality, which yields shorter RT's, should be equalised to the modality that yields larger RT's. In the current research, this would have meant two auditory stimuli with different intensity settings (and thus dissimilar RT's) would have been constructed. Subsequently, the median RT yielded by these auditory stimuli would have been measured. Afterward, the RT yielded by the visual stimuli should have been equalized to the RT yielded by their semantically incongruent auditory counterparts.

A second improvement would be to adapt a dynamical step size in the staircase procedures. This would have resulted in better equalisation of RT's between the semantically incongruent audio-visual stimuli.

Conclusion

The current results confirm the effect of the principle of congruent effectiveness in behavioural measures. No significant differences between the effects of multimodal semantic congruency- and the principle of congruent effectiveness on MRE have been found. In the current design there are no conclusions to be made regarding the effect of semantic congruency-, compared to the effect of the principle of congruent effectiveness on MRE. The reason for not finding any differences between the MRE's of the multimodal conditions lies in the fact the manipulation was unsuccessful. Due to the failed manipulation the effects of the principle of congruent effectiveness and multimodal semantic congruency could not be assessed separately.

Aknowledgmements

I would like to thank Dr. Maarten van der Smagt for his supervision during this research and a special thanks goes out to Nathan van der Stoep for constructing the staircase procedure while he was abroad for a seminar.

References

Bolognini, N., Frassinetti, F., Serino, A., & Làdavas, E. (2005). "Acoustical vision" of below threshold stimuli: Interaction among spatially converging audiovisual inputs. *Experimental Brain Research*, *160*(3), 273–282. http://doi.org/10.1007/s00221-004-2005-z

Chen, Y. C., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition*, *114*(3), 389–404. http://doi.org/10.1016/j.cognition.2009.10.012

Goldstein, E. B. (2013). *Sensation and Perception, 9th ed.* Belmont, CA: Wadsworth Cengage Learning.

der Heijden, A. H., Schreuder, R., Maris, L., & Neerincx, M. (1984). Some evidence for correlated separate activation in a simple letter-detection task. *Perception and Psychophysics*, *36*(6), 577–585. http://doi.org/10.3758/BF03207519

Hairston, W. D., Laurienti, P. J., Mishra, G., Burdette, J. H., & Wallace, M. T. (2003). Multisensory enhancement of localization under conditions of induced myopia. Experimental brain research, 152(3), 404-408.

Holmes, N. P. (2007). The law of inverse effectiveness in neurons and behaviour: multisensory integration versus normal variability. *Neuropsychologia*, 45(14), 3340-3345.

Holmes, N. P. (2009). The principle of inverse effectiveness in multisensory integration: some statistical considerations. Brain topography, 21(3-4), 168-176. Kleiner M, Brainard D, Pelli D, 2007, "What's new in Psychtoolbox-3?" *Perception,* 36 ECVP Abstract Supplement.

Mashour, G. (2004). The cognitive binding problem: from Kant to quantum neurodynamics. *NeuroQuantology*, *2*(1), 29–38. http://doi: 10.1016/j.neubiorev.2013.09.009

Meredith M. A., Stein B. E. (1983) Interactions among converging sensory inputs in the superior colliculus. *Science* 221(4608):389–391

Meredith, M. A., & Stein, B. E. (1986). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Research, 365*(2), 350-354.

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *The Journal of neuroscience, 7*(10), 3215-3229.

Melara, R. D. (1989). Dimensional interaction between color and pitch. *Journal of Experimental Psychology. Human Perception and Performance*, *15*(I), 69–79. http://doi.org/10.1037/0096-1523.15.1.69

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*(2), 247–279. http://doi.org/10.1016/0010-0285(82)90010-X

Miller, J., & Ulrich, R. (2003). Simple reaction time and statistical facilitation: A parallel grains model. *Cognitive Psychology*, *46*(2), 101–151. http://doi.org/10.1016/S0010-0285(02)00517-0

Otto, T. U., & Mamassian, P. (2012). Noise and correlations in parallel perceptual decision making. *Current Biology*, *22*(15), 1391–1396. http://doi.org/10.1016/j.cub.2012.05.031

Otto, T. U., Dassy, B., & Mamassian, P. (2013). Principles of multisensory behavior. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(17), 7463–74. http://doi.org/10.1523/JNEUROSCI.4678-12.2013

Pain, M. T. G., & Hibbs, A. (2007). Sprint starts and the minimum auditory reaction time. *Journal of Sports Sciences*, *25*(1), 79–86. http://doi.org/10.1080/02640410600718004

Repko, A.F. (2012). *Interdisciplinary Research: Process and Theory.* Los Angeles/London/New Delhi/Singapore/Washington DC: Sage.

Reynolds, J. H., & Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron*, *24*(1), 19–29. http://doi.org/10.1016/S0896-6273(00)80819-3

Riesenhuber, M., & Poggio, T. (1999). Are cortical models really bound by the "binding problem"? *Neuron*, *24*(1), 87–93. http://doi.org/10.1016/S0896-6273(00)80824-7

Raab, D. H. (1962). Statistical facilitation of simple reaction times. Transactions of the New York Academy of Sciences, 24, 574-590. doi:10.1111/j.2164-0947.1962.tb01433.x

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, *17*(5), 1147–1153. http://doi.org/10.1093/cercor/bhl024

Senkowski, D., Saint-Amour, D., Hoffle, M., & Foxe, J. J. (2011). Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *NeuroImage*, *56*(4), 2200–2208. http://doi.org/10.1016/j.neuroimage.2011.03.075

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971–995. http://doi.org/10.3758/s13414-010-0073-7

Shelton, J., & Kumar, G. P. (2010). Comparison between Auditory and Visual Simple Reaction Times. *Neuroscience & Medicine*, *01*(September), 30–32. http://doi.org/10.4236/nm.2010.11004

Smagt, M. J., Buijing, I. F., & van der Stoep, N. (2014, May). Subjective crossmodal correspondence and audiovisual integration. Poster presented at the annual meeting of 15th Vision Sciences Society Conference, St. Pete Beach, Florida. Abstract retrieved from http://www.researchgate.net/publication/266154198_Subjective_crossmodal_correspondence_and_audiovisual_integration

Smagt, M. J. & van der Stoep, N. (2015, May). Inter-individual response differences predict multisensory response enhancement. Poster presented at the annual meeting of 16th

Vision Sciences Society Conference, St. Pete Beach, Florida. Abstract retrieved from https://www.researchgate.net/publication/281519927_Inter-individual_response_differences_predict_multisensory_response_enhancement

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA, US: The MIT Press.

Stevenson, R. A., Bushmakin, M., Kim, S., Wallace, M. T., Puce, A., & James, T. W. (2012). Inverse effectiveness and multisensory interactions in visual event-related potentials with audiovisual speech. *Brain Topography*, *25*(3), 308–326. http://doi.org/10.1007/s10548-012-0220-7

Stevenson, R. A., Ghose, D., Fister, J. K., Sarko, D. K., Altieri, N. A., Nidiffer, A. R., ... & Wallace, M. T. (2014). Identifying and quantifying multisensory integration: a tutorial review. *Brain topography*, *27*(6), 707-730. doi:10.1007/s10548-014-0365-7

Van der Stoep, N., Van der Stigchel, S., Nijboer, T. C. W., & Van der Smagt, M. J. (2016). Audiovisual integration in near and far space: effects of changes in distance and stimulus effectiveness. *Experimental Brain Research*, *234*(5), 1175–1188. http://doi.org/10.1007/s00221-015-4248-2

Vatakis, A., & Spence, C. (2007). Crossmodal binding: evaluating the “unity assumption” using audiovisual speech stimuli. *Perception & Psychophysics*, *69*(5), 744–756. http://doi.org/10.3758/BF03193776

Vatakis, A., Ghazanfar, A. A., & Spence, C. (2008). Facilitation of multisensory integration by the “unity effect” reveals that speech is special. *Journal of Vision*, *8*(9), 1–11. http://doi.org/10.1167/8.9.14

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638–667. http://doi.org/10.1037/0033-2909.88.3.638

Brainard, D. H. (1997). The Psychophysics Toolbox, *Spatial Vision 10*:433-436. [PDF] Pelli, D. G. (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies, *Spatial Vision 10*:437-442.

Zhang, B. T. (2016). Humans and Machines in the Evoluton of AI in Korea. *AI magazine,* 37(2), 108-112. http://dx.doi.org/10.1609/aimag.v37i2.2656

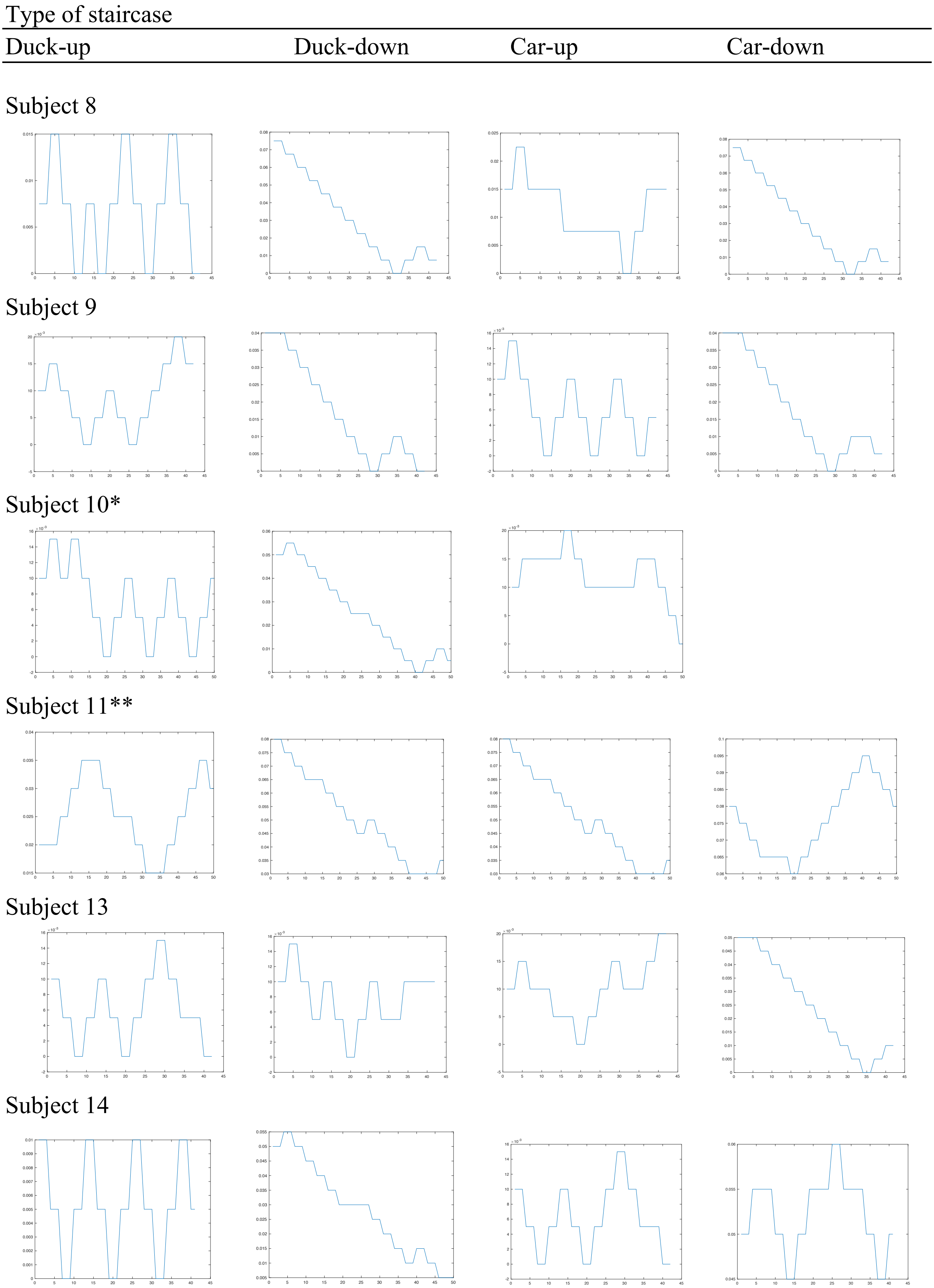
Appendix

Table 1. Intensity levels used for the auditory stimuli in the main experiment for each subject included in the analysis.

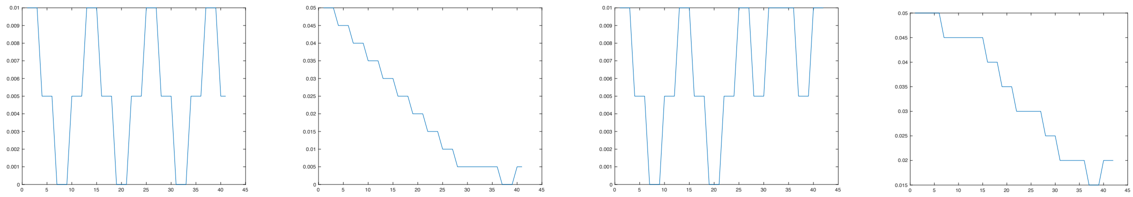
Subject	Intensity Duck	Intensity Car
8	47.82 dB (0.0325)	35.08 dB (0.075)
9	35.08 dB (0.0075)	31.56 dB (0.005)
10	35.08 dB (0.0075)	31.56 dB (0.005)
11	47.82 dB (0.0325)	35.08 dB (0.075)
13	31.56 dB (0.005)	41.10 dB (0.015)
14	31.56 dB (0.005)	31.56 dB (0.005)
16	31.56 dB (0.005)	44.62 dB (0.0225)
19	31.56 dB (0.005)	46.37 dB (0.0275)
21	31.56 dB (0.005)	46.37 dB (0.0275)
22	37.58 dB (0.01)	44.62 dB (0.0025)
23	41.10 dB (0.015)	37.58 dB (0.01)
24	35.08 dB (0.0075)	42.44 dB (0.0175)
25	31.56 dB (0.005)	35.08 dB (0.0075)
26	37.58 dB (0.01)	44.62 dB (0.0225)
27	31.56 dB (0.005)	35.08 dB (0.0075)
28	37.58 dB (0.01)	47.82 dB (0.0325)
29	31.56 dB (0.005)	44.62 dB (0.025)
31	31.56 dB (0.005)	44.62 dB (0.025)
32	31.56 dB (0.005)	47.82 dB (0.0325)

Note. Auditory intensities displayed in decibel (dB). The numbers in the brackets are the auditory values used in Matlab.

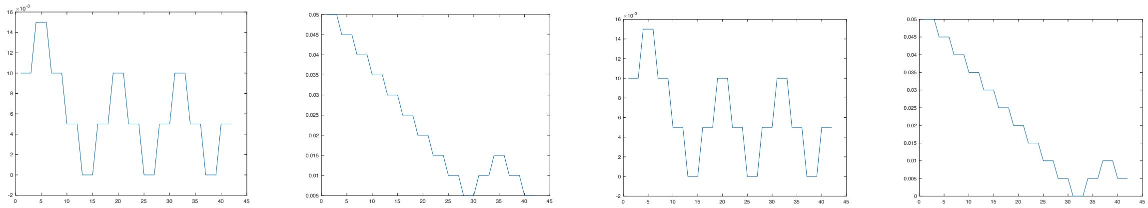
Figure 1. Graphs of the staircases for every subject that was included in the analysis.



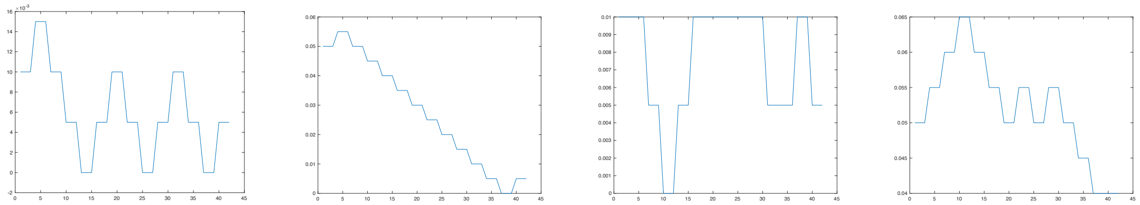
Subject 16



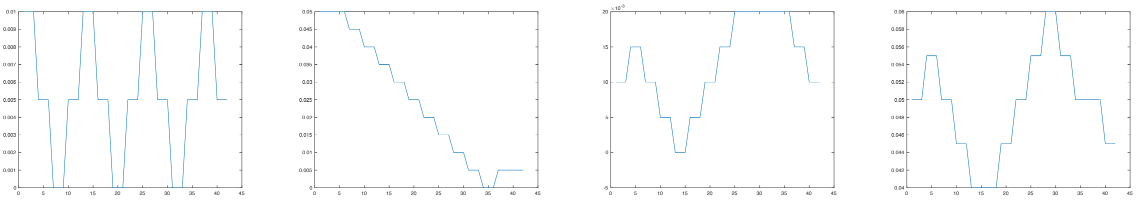
Subject 19



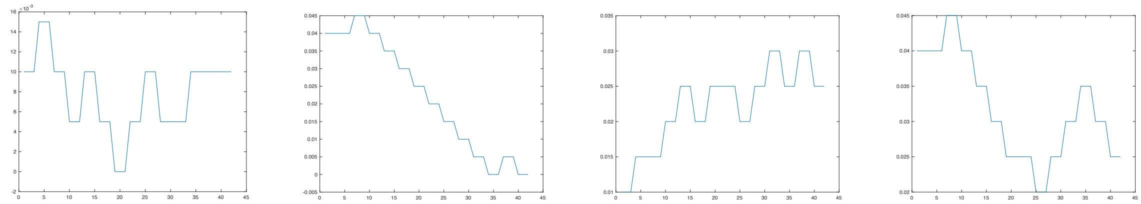
Subject 21



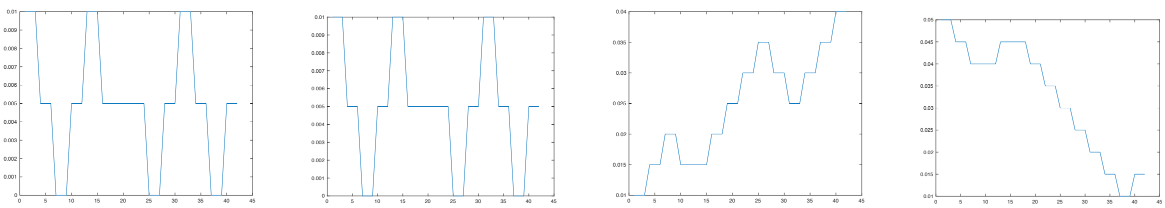
Subject 22



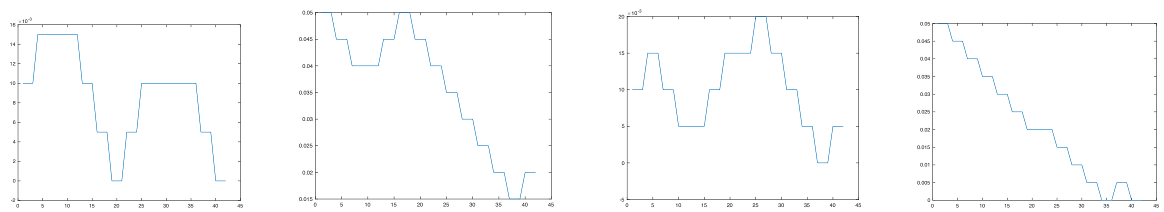
Subject 23



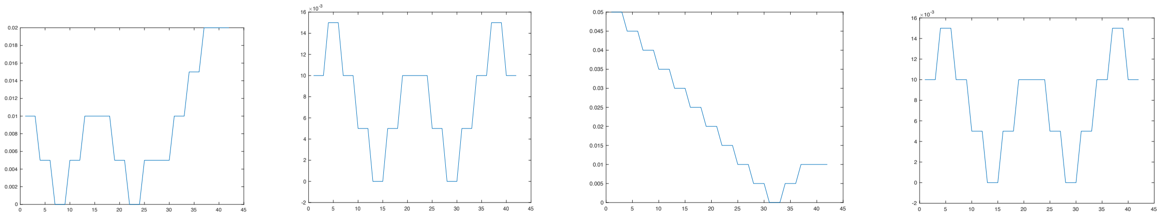
Subject 24



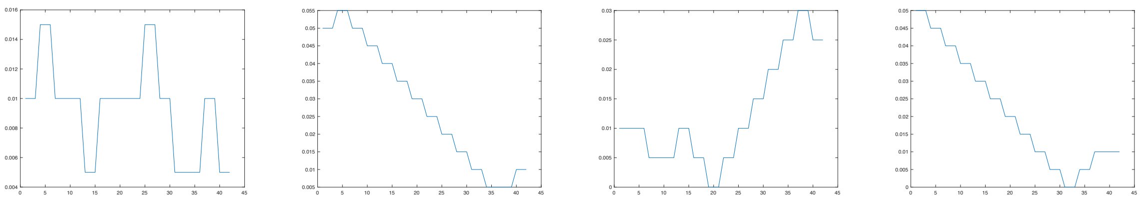
Subject 25



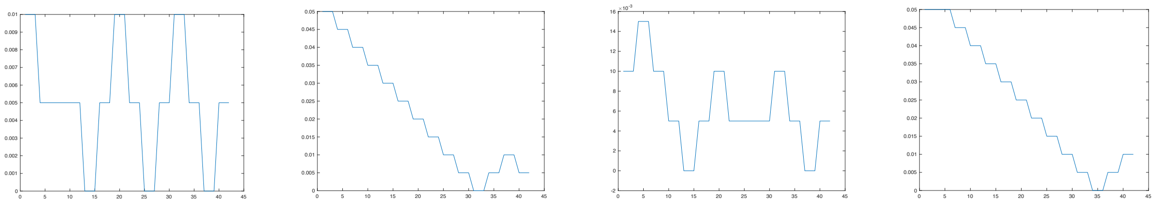
Subject 26



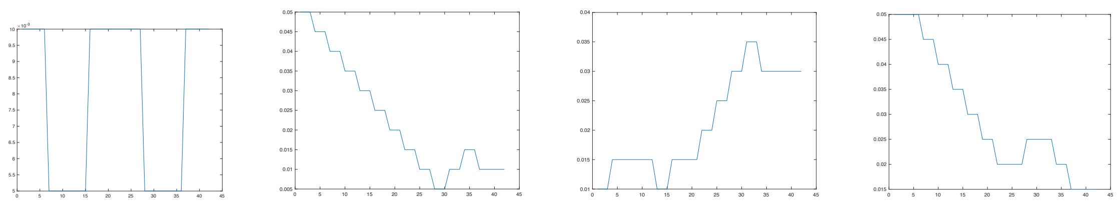
Subject 27



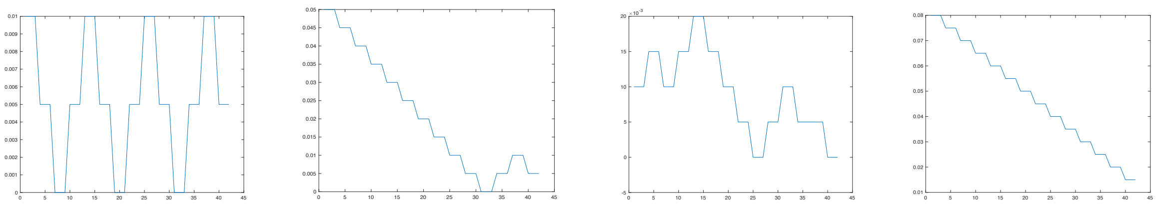
Subject 28



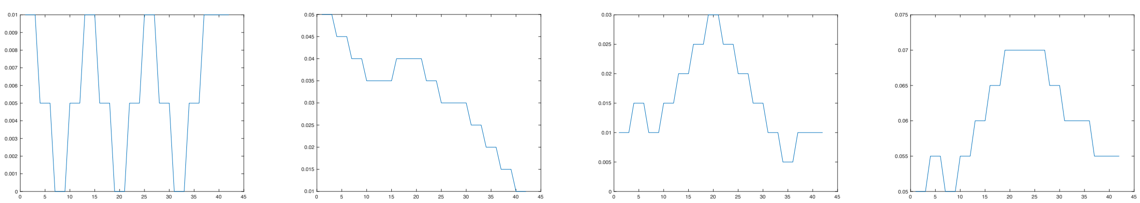
Subject 29



Subject 31**



Subject 32



*After completion of the two car staircases the intensity settings for this subject reached 0, effectively removing the auditory car stimulus from all trials. It was decided to assign this person the lowest possible auditory value for the auditory car stimulus.

**Due to an error the starting values of the car-down staircase were set at 0.08 instead of 0.05.