Eberhard Karls Universität Tübingen
Faculty of Science
Department of Computer Science

Master Thesis in Cognitive Science

Audiovisual Processing in a Spatial Detection Task

Marc Weitz

April 29, 2020

Reviewers

PD Dr. Gregor Hardieß
Cognitive Neuroscience
Institute for Neurobiology
University of Tübingen

Prof. Dr. Jürgen Heller
Research Methods and
Mathematical Psychology
Institute for Psychology
University of Tübingen

Weitz, Marc:
Audiovisual Processing in a Spatial Detection Task
Master Thesis in Cognitive Science
Eberhard Karls Universität Tübingen
Thesis period: Nov. 2019 - Apr. 2020

Abstract

Localisation of objects in the world is achieved by a complex combination of different sensory input. Amongst others, visual and auditory information are combined along the dorsal pathway in the brain to form a unified percept of space. Recent studies on audiovisual object recognition along the ventral pathway reported congruency benefits when auditory and visual information were redundant and impairments when they were conflicting. Whether similar effects also shape spatial perception are unknown. In this thesis we show that a portion of the effects can also be observed in spatial processing. We found remarkably decelerated responses when the position of a task-irrelevant distractor stimulus did not match the actual target's position. Contrary to our expectations, responses were not enhanced when the distractor appeared at the same position as the target under normal sight. Using less visible stimuli seems to provoke an effect as tested in the subsequent experiment. Our results demonstrate some considerable overlap of processing conflicting information along the two different neural pathways. We anticipate our study to be a starting point of bridging the gap between the variety of cross- and unimodal conflict tasks. But understanding how the human resolves conflicts in general but also in particular in spatial processing is also highly relevant for any human-machine interface, such as during driving or other demanding tasks.

Zusammenfassung

Um Objekte in der Welt lokalisieren zu können, muss das Gehirn die Informationen der verschiedenen Sinne verbinden. Visuelle und auditive Informationen werden dazu entlang des dorsalen Pfades des Gehirns zu einer einheitlichen Repräsentation des Raums verarbeitet. Neuere Studien zur audiovisuellen Objekterkennung entlang des ventralen Pfads legen dabei nahe, dass kongruente auditive und visuelle Information die Verarbeitung verbessert während sich die Reaktionen verschlechterten wenn die Informationen im Konflikt zu einander stehen. Ob es die gleichen Effekte auch in der räumlichen Wahrnehmung gibt ist noch unbekannt. In dieser Masterarbeit zeigen wir, dass zumindest ein Teil der Effekte sich auch in der räumlichen Verarbeitung zeigt. Wir konnten zeigen, dass ein irrelevanter zusätzlicher Reiz an einer anderen Position, die korrekte Lokalisierung des eigentlichen Zielreizes signifikant verlangsamt. Im Gegensatz zu unseren Erwartungen wurden die Reaktionen jedoch nicht schneller, wenn der zusätzliche Reiz an der gleichen Position erschien wie der Zielreiz. Die Ergebnisse des Nachfolgeexperiments legen nahe, dass schlechter sichtbare Stimuli jedoch zu einem Effekt räumlicher Kongruenz führen. Unsere Ergebnisse zeigen einige klare Übereinstimmungen in der Verarbeitung sich widersprechender Informationen entlang der beiden neuronalen Verarbeitungspfade. Damit legt unsere Studie einen Grundstein in der Erforschung der Gemeinsamkeiten in der Verarbeitung widersprüchlicher Informationen zwischen aber auch innerhalb der verschiedenen Sinnesmodalitäten. Wie wir Menschen diese Konflikte lösen ist dabei auch höchst relevant für das Design verschiedenster Mensch-Maschine Schnittstellen, wie beispielsweise beim Autofahren oder anderen anspruchsvollen Aufgaben.

Acknowledgements

Many people participated in the success of this thesis and my studies in general; more than I can list here. To begin with, I want to thank all my academic mentors and supervisors who provided me with so many advice, feedback and knowledge throughout my studies. In particular, I want to thank PD Dr. Gregor Hardieß and Prof. Dr. Jürgen Heller who supervised this thesis. Furthermore, I want to thank Konstantin Sering who supported and advised me so many times over the last years. Additionally, I would like to thank everyone in the Cognitive Neuroscience Group, the Mathematical Psychology Group and the Quantative Linguistics Group in Tübingen for their feedback and support and, particularly, Jana Krämer for being the best office buddy I can imagine.

Moreover, I want to thank all my friends who supported me morally over the last month and years. Especially, I would like to thank David-Elias Künstle for the hardest and most challenging critics one could wish for and Annika Striby for all her encouragements and support. Furthermore, I owe many thanks to my flatmates, Robert Geirhos and Frederik Unger, for providing me with excellent food throughout my thesis. I also would like to thank all the proof-readers for their feedback on earlier drafts of this thesis: Claudia Glemser, Jana Krämer, David-Elias Künstle, Annika Striby, and Frederik Unger.

Last but not least, I would like to thank my family for being so supportive in all of my efforts and for having made my studies and this thesis possible.

Contents

List of Figures

Chapter 1

Introduction

Building up a consistent and reliable representation of the space surrounding us is crucial for any kind of goal-directed behaviour. However, spatial localisation of objects relies on a manifold of cues within but also between the sensory modalities. Visuospatial localisation, for instance, relies among others on the position of the stimulus on the retina, the orientation of the eyes and the position of the head. Similarly, auditory localisation relies on several mono- and binaural cues. Most importantly, the interaural time difference (ITD) and interaural level difference (ILD) are evaluated to localise sounds on the horizontal plane (Moore, 2013). The ITD is characterised by the time difference the sound needs to reach the two ears, whereas the ILD is produced by a larger amplitude at the ipsilateral than the contralateral ear. Together with other auditory cues, the human's auditory system reaches a precision up to 1° (Mills, 1958; Perrott & Saberi, 1990).

The question how the inputs of multiple senses are integrated has concerned human scientists for decades. One of the earliest works on multisensory integration might originate from Todd (1912) who measured reaction times to stimuli from two or more sensory modalities which were either presented alone or in combination (as described in Colonius & Diederich, 2020). By doing so, it is usually observed that bimodal presentations provoke faster and more accurate responses than unimodal presentations. Later, this effect has been described as the "redundant signal effect" (cf. Kinchla, 1974; Miller, 1982).

The principles underlying the redundant signal effect have been controversially debated. The reason for that is that faster or more accurate responses are not sufficient to assume an underlying integration process (Stevenson et al., 2014) because such reactions can also be provoked solely by statistical facilitation of independently processed stimuli (Raab, 1962). By assuming the independence of the signals S_1 and S_2, the maximal statistical facilitation provoked by the redundant information of the two signals can be quantified as

$$\begin{aligned} P(RT < t|S_1 \text{ and } S_2) &= P(RT < t|S_1) + P(RT < t|S_2) \\ &\quad -P(RT < t|S_1) \cdot P(RT < t|S_2) \end{aligned} \tag{1.1}$$

with $P(RT < t|S_1 \text{ and } S_2)$ describing the cumulative density function (CDF) of the observed reaction times (RT) within a time t in trials where information of two different sources (S_1 and S_2) are available (cf. Miller, 1982).

A similar model was already presented earlier (Pirenne, 1943). Pirenne (1943) studied the sensitivity of the eyes towards brightness. In his experiment, he either provided stimuli to one eye each or to both eyes simultaneously. By assuming the two eyes as independent detectors he described the binocular detection threshold p_B to be a linear combination of the monocular thresholds p_L and p_R (for the left and the right eye) as

$$p_B = 1-(1-p_L)(1-p_R) = p_R + p_L - p_L p_R. \tag{1.2}$$

This class of models that has been termed race models can explain the results of certain tasks. Nevertheless, it has been shown that other types of information fusion can exceed the amount of statistical facilitation. Miller (1982), for instance, reported data of a bimodal detection task as well as a letter search task which both exceeded the prediction under signal independence assumption. In the bimodal detection task the participants had to respond as fast as possible to an auditory (a 780Hz sinus tone), a visual (an asterisk) or an audiovisual stimulus (a combination of both). Furthermore, no-go trials without a stimulus were included in which the participants must not respond. The results revealed faster reaction times than predicted for the lower quantiles (equivalent to the fastest reaction times). Thus, Miller (1982) concluded that signal integration rather than independent processing must have occurred.

Besides behavioural research, substantial progress has been made understanding the underlying neurobiological and -physiological processes over the last years (Stein, 2012; Stein & Stanford, 2008). Single cell recordings in the superior colliculus (SC) of cats and primates revealed that certain neurons evoke more spikes to audiovisual stimuli than to the unimodal stimuli together (Meredith & Stein, 1983; Wallace, Wilkinson, & Stein, 1996), a mechanism that has been termed "superadditivity". In mammals, the SC is a structure lying in the dorsal region of the midbrain and is mainly associated with spatial processing. The neurons in the SC are topographically organised and neuronal activation is associated with responses directed toward the corresponding point in space. Besides audiovisual integration, the SC in cats has been shown to further integrate somatosensory information (Meredith & Stein, 1986), promoting it to be a central region of multisensory integration.

The integration of different sensory modalities has also been shown in the human brain (see e.g. Stein, 2012). Moreover, already the cortices associated with early sensory processing (see Schroeder & Foxe, 2005, for an introduction) such as the primary visual cortex (Watkins, Shams, Josephs, & Rees, 2007; Watkins, Shams, Tanaka, Haynes, & Rees, 2006) or the auditory cortices (Lakatos, Chen, O'Connell, Mills, & Schroeder, 2007) appear to integrate input from other senses to a certain extend. The reason for this early level integration is still unclear, but might be in use for cross-modal pre-activations to enhance processing in the other modality.

A quantitative mathematical description of the observed weighting between the different sensory signals has been suggested by using Kalman filters (Ghahramani, Wolptrt, & Jordan, 1997; Wolpert, Ghahramani, & Jordan, 1995). The Kalman filter (Kalman & Bucy, 1961) can be described as a weighted average

$$\hat{s} = \sum_i w_i \hat{s}_i \text{ with } \sum_i w_i = 1 \tag{1.3}$$

of the maximum likelihood estimates of the different sensory inputs $\hat{s}_i$ that are weighted by their respective relative reliability w_i of the sensory input, with

$$w_i = \frac{\frac{1}{\sigma_i^2}}{\sum_j \frac{1}{\sigma_j^2}}. \tag{1.4}$$

For the bimodal audiovisual case with signals s_A and s_V and their respective variances σ_A^2 and σ_V^2, this results in a weighting for the auditory signal of

$$w_A = \frac{\frac{1}{\sigma_A^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_V^2}} = \frac{\sigma_V^2}{\sigma_A^2 + \sigma_V^2}. \tag{1.5}$$

As can be inferred from Equation (1.5), the relative weight of the auditory signal w_A increases with increasing variance of the visual signal and approaches 1 if the visual variance is sufficiently high.

The most important prediction of this model is that if the dominant of the two senses gets obstructed during bimodal processing, the non-dominant sense becomes integrated more. It is assumed that for most species, including the human, vision dominates the other senses (Colavita, 1974; Eimer, 2004; E. I. Knudsen & Knudsen, 1985, 1989a, 1989b; Pick, Warren, & Hay, 1969). Besides empirical evidence of this prediction (Ernst & Banks, 2002; Ernst & Bülthoff, 2004), Werner and Noppeney (2009) also provided neuroimaging evidence of this hypothesis. In their study, participants had to categorise sounds or pictures of tools and instruments. Both stimulus sets were systematically degraded at three levels each (intact, degraded and noise). In the subsequent categorisation task, the participants brain activity was measured. The brain activity was primarily subadditive for intact bimodal stimuli but the additivity increased with degradation. Concurrently, the categorisation accuracy increased as a function of degradation for bimodal stimuli relative to the best unimodal stimulus.

This ranking becomes particularly important when the information of different senses are in conflict. In the case of audiovisual processing, this has become known as the "ventriloquism" effect (see Pick et al., 1969, for an early work on cross-modal conflicts). Originally, ventriloquism described a method of projecting one's voice to somewhere else. The probably most famous examples of this technique can be found in performers making their puppet appear to talk. But nowadays ventriloquism can be also found while watching television or Netflix. All this situations have in common that the physical auditory and visual sound sources differ, but are perceived together with usually the visual source emitting the sound. The mechanisms and factors influencing whether the two sensory information are bound by the brain are still not fully understood. A growing body of evidence indicates that spatiotemporal coincidence plays a crucial role in multisensory integration (Colonius & Diederich, 2004; Meredith, 2002; Stein, 2012). As that, the time window as well as the spatial proximity in which the sensory signals have to occur is rather broad. For spatial coincidence, the literature is even more ambiguous. Even though spatial proximity is particularly suggested by the neuroscientific literature, other studies also reported significant influences of non-spatial sounds (Iordanescu, Grabowecky, Franconeri, Theeuwes, & Suzuki, 2010; Iordanescu, Grabowecky, &

Suzuki, 2011; Meyerhoff & Suzuki, 2018; Sekuler, Sekuler, & Lau, 1997; Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008). Therefore, it is still unclear to which extend spatial proximity is a requirement for multisensory integration.

But also higher level factors, such as semantic congruency (Y.-C. Chen & Spence, 2010; Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004), cross-modal correspondences (Spence, 2011), or the "unity assumption" (Y.-C. Chen & Spence, 2017; Vatakis & Spence, 2007), have been suggested to influence cross-modal binding and, thus, the ventriloquism effect. Y.-C. Chen and Spence (2010), for instance, presented pictures together with natural sounds that could either be semantically congruent (e.g. the sound of a dog barking matched with the picture of a dog) or incongruent (e.g. the sound of a tool was matched with the picture of an animal) or neutral (e.g. white noise). In a series of five experiments, the authors demonstrated significantly better accuracies for congruent semantic stimulus matching than for unimodal or incongruent pairings in a picture identification task. Simultaneously, the picture identification was impaired when the stimulus pairing was incongruent in comparison to the unimodal or congruent condition.

Contrarily to spatial processing, these factors are more concerned with semantic processing. However, a growing body of neuroscientific research suggests that spatial and semantic processing are implemented differently in the brain (Goodale & Milner, 1992). According to the dual stream hypothesis, spatial information is processed along the dorsal pathway whereas semantic processing is mainly associated with the ventral pathway. Possible shared operations might be implemented at the level of higher and more executive cognitive processes. Attention (Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010), working memory (Quak, London, & Talsma, 2015) and long-term memory (Matusz, Wallace, & Murray, 2017; Meyerhoff & Huff, 2016) have also been suggested to interact with multimodal processing in a top-down fashion. But whether and how the cognitive and neurobiological processes operating along the pathways are comparable is still unknown.

The goal of the present thesis was, therefore, to examine the similarities of information processing along the two processing pathways. Similar to the semantic congruency experiment of Y.-C. Chen and Spence (2010), we developed a spatial conflict task to test the observed enhancements and impairments of congruent and incongruent presentation respectively. Instead of semantic congruency and incongruency, we used spatial visual and auditory stimuli that were either presented from the same (congruent) or a different (incongruent) position. The latter setting has already intensely been examined within the ventriloquism literature. In compliance with the observed impairments caused by semantic incongruency, we therefore expected significantly slower reaction times in the incongruent condition compared to the unimodal or congruent condition. On the other hand, some studies also reported significantly enhanced responses for non-spatial sounds as well, enabling also expectations in the other direction. In this case, both congruent and incongruent condition should be faster than the respective unimodal condition if spatial information is neglected. Spatial congruency is expected to enhance processing compared to the unimodal condition under both assumptions.

To test these hypotheses, a series of three experiments was conducted. In a first study reported in Chapter 2, we created virtually displaced auditory stimuli and tested whether the participants localised the sounds as expected. In the subsequent two experiments (see Chapters 3 and 4), the participants had to detect the position of a target stimulus in a four alternative forced choice (4AFC) task. In both experiments the target's modality as well as the presentation condition (unimodal, congruent, incongruent) were manipulated in a 2×3 factorial design. In contrast to the second experiment (reported in Chapter 3), the visual stimulus' contrast was reduced in the final experiment (see Chapter 4 for details) resulting in an additional factor of contrast with two levels.

Chapter 2

Auditory Stimuli Creation and Validation

2.1 Introduction

To localise objects in the world by their sounds, the human's auditory system incorporates a variety of cues. Most importantly, the interaural time difference (ITD), the difference in time of the signal reaching the ipsilateral ear in contrast to the contralateral ear, and the interaural level difference (ILD), the difference in loudness between the left and the right ear, are evaluated (Moore, 2013). For low frequency sounds (< 1500 Hz) the ITD usually dominates the ILD (Goupell & Stakhovskaya, 2018; Wightman & Kistler, 1992). However, larger influences of ILDs have been observed in highly reverberant environments (Rakerd & Hartmann, 2010). Furthermore, it is worth to note that ITD and ILD processing mainly enables sound localisation on the horizontal plane, but not vertically. Further cues, mainly based on the reverberation on the pinnae, have to be incorporated to localise the altitude of a sound. Taken together, these cues can concede a precision up to 1° of the human's auditory system (Mills, 1958; Perrott & Saberi, 1990).

The interaural time difference can be easily derived from the spatial position of the sound origin if the frontal and lateral distance are known. Therefore, the difference between the distances to the left and the right ear can be calculated by

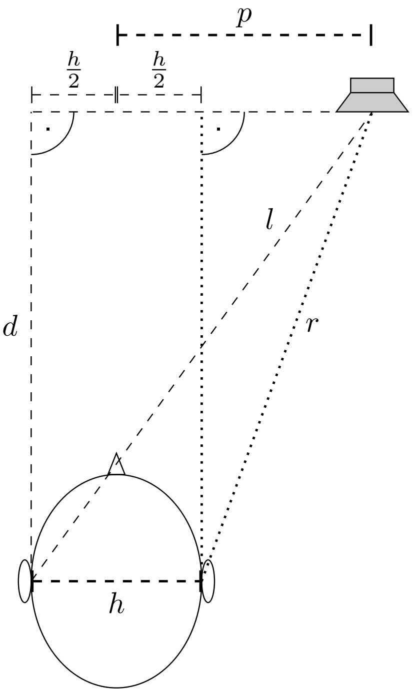


Figure 2.1: Geometry of spatial sound detection. The sound has to travel a longer way (l) to the left than to the right ear (r). These distances can be derived using the Pythagorean theorem if the frontal (d) as well as lateral distance (p) and the head's diameter (h) are known. By dividing the difference of the signals by the speed of sound (c_{air}), the interaural time difference can be calculated.

the Pythagorean theorem

$$l, r = \sqrt{\left(p \pm \frac{h}{2}\right)^2 + d^2} \tag{2.1}$$

with d being the distance in viewing direction. h describes the distance between the two ears of the participant and p denotes the lateral distance of the sound source. This relation is also visualised in Figure 2.1.

To calculate the time difference between the left and right ear, the obtained distances of Equation (2.1) have to be divided by the speed of sound which in air is approximately $c_{air} = 343\frac{m}{s}$. Thus, the ITD can be calculated as

$$\text{ITD} = \frac{r - l}{c_{air}}. \tag{2.2}$$

To virtually move the signal of a frequency f, the created single mono channel has to be shifted in phase for the ipsilateral channel of the stereo signal. Therefore, assuming the signal's origin is lateralised right, the right channel has to be shifted in phase by the interaural time difference, whereas the left channel does not need to be shifted:

$$L_{\text{amp}}(t) = \sin(2\pi f t) \tag{2.3}$$

$$R_{\text{amp}}(t) = \sin(2\pi f t + \text{ITD}) \tag{2.4}$$

However, this panning procedure yields some disadvantages. First, stimulus presentation using this procedure has to be done using headphones as stereo speaker presentation would again undergo the same process. This would result in signals that reach the ear with different ITDs themselves. Stereo headphone presentation, on the other hand, has the disadvantage that the ITD panned signal is only perceived on the axis between the ears, but not frontally as one would prefer.

A frontal representation of the signal using headphones could be obtained using head-related transfer functions (HRTFs; see for instance Blauert, 1997; Brungart & Rabinowitz, 1999; Moore, 2013; Wightman & Kistler, 1989). HRTFs describe the physical change of the signal from its origin to the inner ear. They have to be measured individually by comparing the signal reaching the inner ear and the emitted signal from a certain position. A significant problem with head-related transfer functions is that they vary significantly between individuals due to anatomical variations of the head, the pinnae, the ear canal and other anatomical structures involved in sound processing. Approximations with standardised HRTFs are possible but result in a significant loss of quality.

Other panning procedures for stereo speakers focus on the perceived loudness for sound lateralisation. In the easiest way of loudness related panning, the signal's amplitude can be linearly increased on the ipsilateral and linearly decreased on the contralateral channel. The summed amplitude of left and right stays constant, which is beneficial if one, for instance, wants to combine a stereo signal to mono. But as the amplitudes do not physically sum up, the sound is not perceived equally over the whole range.

Another approach is to adjust the amplitudes non-linearly. Multiplying the signal with the sine function of the panning angle θ allows to shift the signal roughly by this angle. To pan the signal to the left, this influence is positive. To pan the signal to the right, it is negative, respectively (see Equations (2.5) and (2.6)).

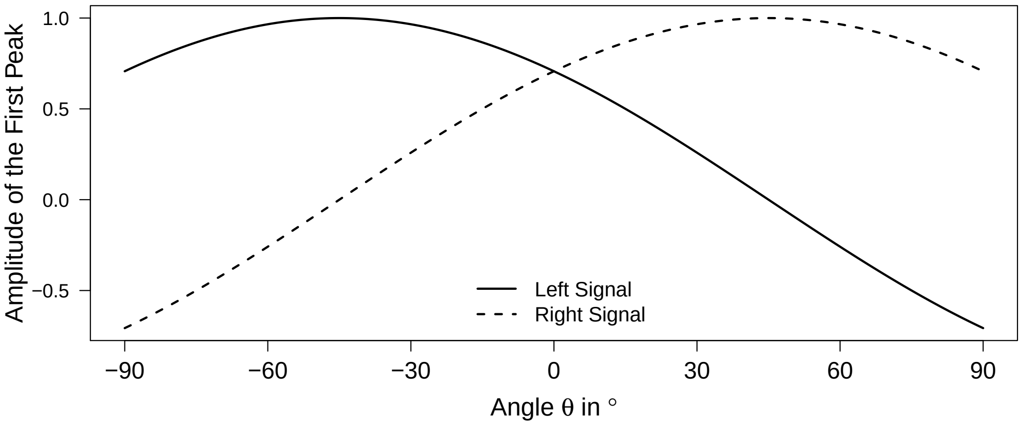


Figure 2.2: Peak amplitudes of the left and right channel as produced using power panning as a function of the panning angle θ. Left and right channel signals are panned from a mono sound signal with a peak amplitude of 1.

However, as can be inferred from Figure 2.2, the summed amplitude is no longer constant. A reduction from stereo to mono is therefore more difficult, but the perceived pitch is more equalised.

$$L_{\mathrm{amp}} = \frac{\sqrt{2}}{2}(\cos\theta + \sin\theta) \cdot \sin(2\pi f t) \tag{2.5}$$

$$R_{\mathrm{amp}} = \frac{\sqrt{2}}{2}(\cos\theta - \sin\theta) \cdot \sin(2\pi f t) \tag{2.6}$$

In summary, both, stereo headphone as well as stereo speaker, have certain advantages and disadvantages. Headphones, for instance, allow for a very high level of control over the auditory stimulus that reaches the ear, as the transmission distance is short. Furthermore, headphones isolate the participants ears from external noise. On the other hand, this also leads to an omission of several cues the auditory system usually expects such as missing head shadows, room acoustics or the reflections of the pinnae. Contrarily, all of this information is kept using stereo speaker, but the control over the incoming signal is reduced. To achieve the best results, the physical and the virtual sound origin should match. However, achieving this is difficult. The best way how this can be accomplished might be using separate speakers for each position.

However, as no such system was available in the lab, we manipulated the signals

ILD as described in Equations (2.5) and (2.6) to create four stimuli with a sampling rate of 44.100 Hz associated with the far left, left, right and far right positions of the visual stimuli. These stimuli were presented using stereo speakers placed at the left and the right side of the display. Stereo speakers were preferred over headphones to achieve that the sound is perceived originating from the axis between the two speakers. As the theoretical panning angle only matches the visual angle under perfect conditions, we decided to validate the created stimuli prior to the other experiments. Therefore, a 4AFC task was performed with the four spatial sounds. As dependent variables we measured the relative response frequencies for all four positions as well as the participants' sensitivity for each position. As a satisfactory criterion, we expected overall high response frequencies for the correct position and only few to none incorrect responses. Simultaneously, the sensitivity of all positions was expected to be different from zero and not to vary between the positions.

2.2 Method

2.2.1 Participants

Twenty-four subjects (19 - 27 years , 12 females) participated in the experiment. The participants received either course credit or a monetary compensation of 10€ per hour. All participants had normal or corrected-to-normal vision. Informed consent was provided by all participants prior to testing. The sample size was determined to conduct all possible block combinations of the subsequent experiment (see Chapter 3) which had four blocks (yielding $4! = 24$ combinations).

2.2.2 Materials

All stimuli were programmed using the PsychoPy library (Peirce, 2007) in Python 3 (Van Rossum & Drake, 2009). The experiments were conducted on a 24 inch LED screen (60 Hz, 1920 x 1080 pixels) with a viewing distance of 50 cm. The screen was controlled by a standard PC running Windows 10 and the PsychoPy Standalone Application (Version 3.2.4). To control the viewing distance and to avoid fatigue, the participant's head was placed on a chin rest. Auditory stimuli (780 Hz pure sine wave tones without onset ramps; approx. 50-55 dB at the listeners' ear; sample

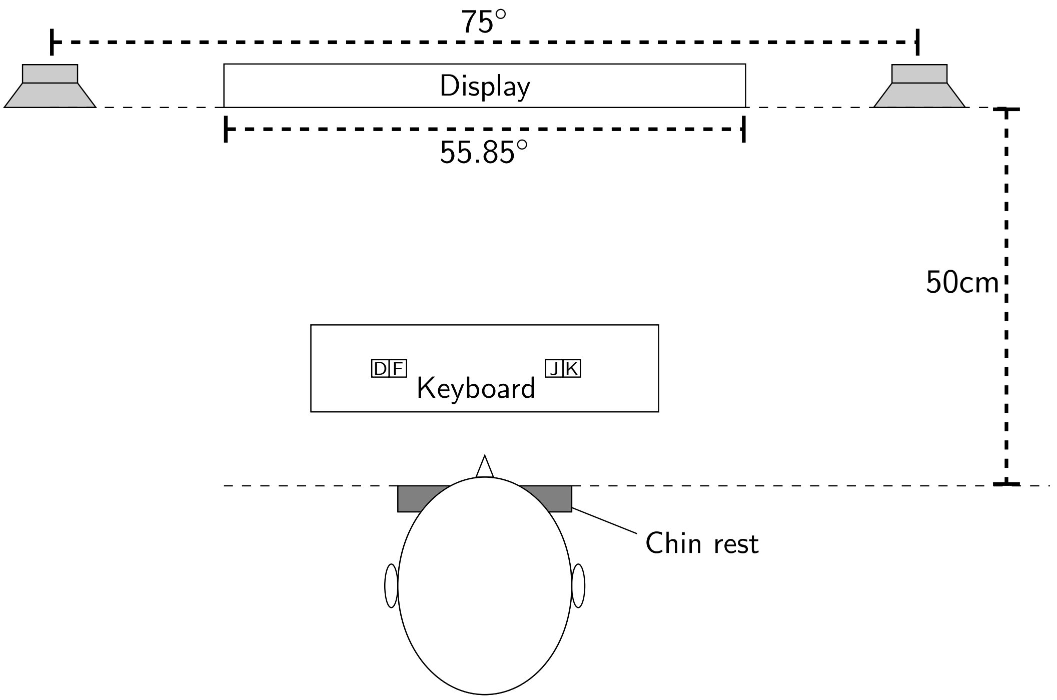


Figure 2.3: General setup of all experiments. The participant was placed roughly 50 cm apart from the computer screen. Depending on the experimental condition, either a visual stimulus in form of a white dot was shown on the computer screen, a spatial sound was presented using the stereo speakers, or both were presented. The stereo speakers were placed on the left and right side of the computer screen 75° apart from each other at the same height as where the visual stimuli were presented.

rate = 44.100 Hz) were presented using Creative MF1680 stereo speakers, placed on the left and the right side of the Fujitsu B24T-7 LED proGREEN computer screen. On the computer screen, visual stimuli were presented at four positions at -21°, -7°, 7° and 21° relative to the center of the screen. Each of the visual stimuli consisted of a white dot with a diameter of 0.5°. For response submission the D, F, J and K button on a standard keyboard were used, which was placed in front of the participant (see Figure 2.3 for an overview of the experimental setup).

2.2.3 Procedure

A 4AFC task was conducted to test for the participants ability to correctly localise the sound origins of the four created auditory stimuli. Each stimulus was presented

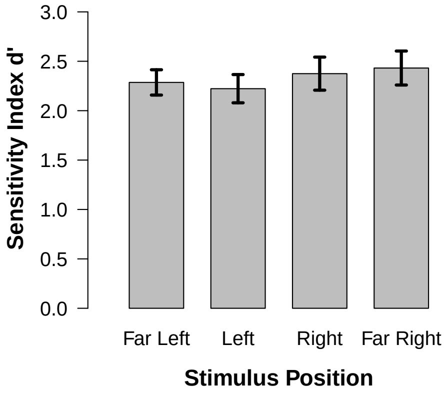


Figure 2.4: Average sensitivity indices (d') for each position of the auditory stimulus. Error bars indicate the within-subject standard error. All four positions were well detectable and discriminable by the participants resulting in average d' larger than two and did not differ significantly from each other.

20 times for a duration of 10s or until the participant responded. Simultaneously, the four visual stimuli were shown with the letters D, F, J and K on the lower side of the screen indicating the respective response key.

In each trial, a fixation cross accompanied by the instruction to specify the position of the sound was presented. The response keys were shown in spatially congruent order on the lower part of the screen and remained there over the course of the experiment. After one second, the four spatial positions were presented in form of white points with a diameter of 0.5°. Simultaneously, the sound was played without a rammed onset on the two stereo speaker. Each trial lasted until the participant responded with one of the four keys.

2.3 Results

On average, the participants responded correctly in 69 of 80 trials with a standard deviation of 5.5 trials. To test that the stimuli are well distinguishable, we applied Signal Detection Theory (see Wickens, 2001, for an introduction). For that purpose, we calculated the sensitivity index d' for each of the four positions by

$$d' = z(\mathsf{HR}) - z(\mathsf{FAR}) \tag{2.7}$$

with HR denoting the hit rate, FAR the false alarm rate and $z(p)$ the inverse of the cumulative density function of the standard normal distribution, yielding the quantile for a certain probability $p \in [0, 1]$. The hit rate, the probability of detecting the signal at a certain position given that this position was presented, was calculated by dividing the hits by the number of trials the stimulus was presented:

$$\text{HR} = \frac{\#\text{ hits}}{\#\text{ hits} + \#\text{ misses}}. \tag{2.8}$$

The false alarm rate, the probability of detecting the signal at a certain position given that another position was presented, was calculated by dividing the false alarms by the number of trials where the position was not presented:

$$\text{FAR} = \frac{\#\text{ false alarms}}{\#\text{ false alarms} + \#\text{ correct rejections}}. \tag{2.9}$$

Hit rates and false alarm rates of zero or one were adjusted by $\frac{0.5}{20} = 0.025$ to avoid edge cases (Stanislaw & Todorov, 1999). The d' were calculated for each position and participant separately and aggregated later to calculate the within-subject variability. Figure 2.4 shows the resulting d' for each position averaged over all 24 participants. The error bars indicated the within-subject standard error of the mean. The overall discriminability and the variability between the positions was tested using linear mixed-effect models (Baayen, 2008; Bates, Mächler, Bolker, & Walker, 2015) with a random effect for the participants. Hierarchical nested model testing of an intercept model against a model with an additional fixed effect for position indicated no variation of the d' between the positions $(\chi^2(3) = 2.24, p = 0.52)$. If responses were mixed up, this confusion was usually within the same side but only rarely between the sides (see Figure 2.5). Testing the intercept model against a model without an intercept, revealed that the d' differed significantly from zero $(\chi^2(1) = 68.01, p < .001)$.

2.4 Discussion

The goal of this experiment was to demonstrate that the created stimuli are distinguishable and spatially localisable. Furthermore, a spatial correspondence to the visuospatial position was expected to be observed. To test these hypothesis, each

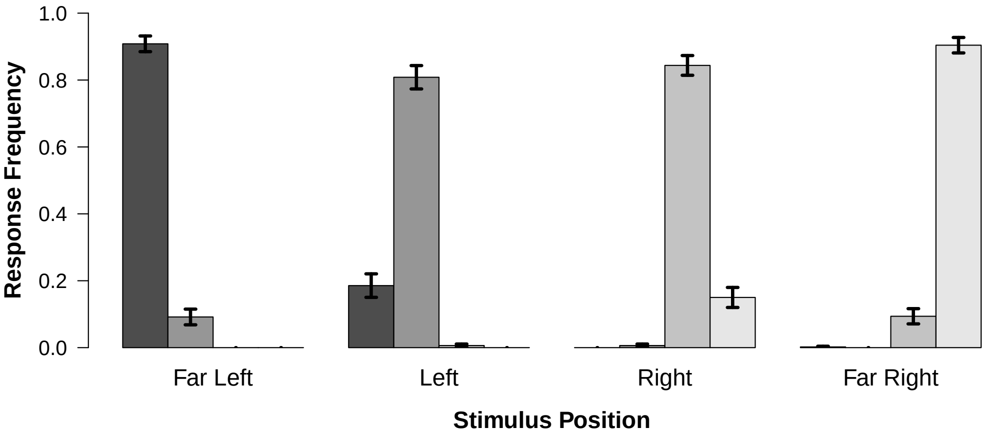


Figure 2.5: Mean relative response frequency as a function of the position of the auditory stimuli. The grouped bars indicate the responded position from far left to right. The error bars indicate one within-subject standard error. All participants showed good detection rates for all positions. It is worth to note that confounded responses were mainly within the same hemisphere (e.g. the left stimulus position was mainly confounded by the far left position, etc.). The discrimination between left- and right-sided stimuli was overall very good as not many left-sided stimuli were detected as right-sided and vice versa.

participant had to respond 80 times to a 4AFC task with 20 repetitions of each of the four stimuli. We applied signal detection theory to calculate a measurement of detectability, the sensitivity index d'. The detectability of the stimuli per position as measured as d' is depicted in Figure 2.4. The linear mixed model analysis confirmed both of our hypotheses: The d' differed significantly from zero but not between the positions.

Therefore, the average d' larger than two suggest that all positions were easily detectable. Given the negligible subject-specific variations as depicted in the error bars in Figure 2.4, the detectability seems additionally to be very stable across participants. However, it is worth to note that participants had to reach a threshold of 60 out of 80 trials. 7 out of the 24 participants failed to reach this threshold in the first try and had to repeat the procedure either due to an insufficiently low number of correctly answered trials or due to technical malfunction. Because of this repetition procedure, the average response frequencies as well as the d' might be slightly biased. However, as the aim was to test for the participants overall ability to detect and match the auditory stimuli, this should not confound the results.

Another argument that could be raised against the consistency of the results is that the participants could have learned an arbitrary mapping based on the qualitative differences of the sound signals instead of localising the sound sources directly. If such a learning process occurred, this would systematically confound the results of our subsequent experiments. This idea might gain some support from the observation that there is some substantial and perceivable confusion in the responses (see Figure 2.5). However, as the participants did not receive feedback during this experiment and, thus, the mapping between a sound signal and a position should be at chance level under this assumption. Even with a repetition, it seems unlikely that the participants guessed the mapping correctly. Thus, it is more reasonable that the participants used their existing mapping than that they have learned a new mapping in the course of the experiment.

Another interesting observation that can be drawn from the response frequencies is the larger extend of intra- compared to inter-hemispheric confusion as shown in Figure 2.5. Inter-hemispheric confusion barely occurred at all and were so seldom that they likely resulted from involuntary random responses. The inter-hemispheric confusion, on the other hand, appears to be substantial and persistent across participants. An explanation of this observation might be the better spatial resolution frontally than laterally (Mills, 1958). Therefore, the more centred presented stimuli show a better discriminablility as the minimum audible angle of source discrimination is here significantly smaller than it is with increasing azimuth. One way to encounter for this artefact might be to use a stimulus spacing based on auditory discriminablility rather than equidistance. However, using scaled positions would, on the other hand, significantly increase the visual search times in the subsequent experiments that are already supposed to be higher for peripheral than for central cues.

The difference between intra- and interhemispheric confusions might also be systematically induced by the used panning procedure. As mentioned, the human auditory system uses a variety of cues to calculate the spatial origin of a sound. On the horizontal plane, these are mainly the interaural time difference (ITD) and the interaural level difference (ILD). However, the applied panning procedure only modified the sounds amplitude and, thus, the ILD but not the ITD. An ITD, on the other hand, is still induced by the two stereo speakers, making the signal reaching the ear not perfectly traceable. Nevertheless, loudness panning seems more reasonable on stereo speakers than ITD panning as the panned signal would again be superimposed by the ITD produced by the speakers. Headphone presentation, as would be reasonable when using ITD manipulation, would, however, drop the useful spatial information that is provided by the environment and is expected by the listener.

Another possible approach would have been to use head-related transfer functions (HRTFs) to manipulate the perceived origin of the auditory signal. Contrarily to the ITD manipulation, HRTFs keep further spatial cues and could, thus, let the spatial sound origin appear more realistic. However, several concerns have to be raised against this approach. First, HRTFs describe the sounds alteration during transmission. Besides ITD and ILD, the sound is further significantly altered by the shape of the pinnae that is highly variable across participants. Therefore, one would have to measure HRTFs for each participant separately. Standardised HRTFs that can be obtained from several databases could also be used but would lead to a significant loss of quality. Moreover, manipulating single sine waves does not seem to produce satisfactory results as the combination of HRTFs with the signal are performed as a multiplication in the frequency domain. Therefore, different stimuli such as sound bursts over different frequencies might be better, but would prevent traditionally used stimuli of multisensory integration experiments (e.g. Miller, 1982).

Given the good overall detection and discriminability results observed in both, sensitivity (see Figure 2.4) as well as response frequencies (see Figure 2.5), we conclude that the used stimuli fit their purpose. Nevertheless, it is worth to note, that for even better and more reliable results a hardware implementation of sound and light sources should be considered. However, the implementation of such an apparatus would have exceeded the time constraints of this thesis and was, therefore, only virtually modelled. Constructing such an apparatus would still be beneficial to improve the signal's quality.

Chapter 3

Audiovisual Spatial Congruency Does Not Improve Localisation in Conflict Situations

3.1 Introduction

Objects in the world usually stimulate more than one sense during perception. Nevertheless, we perceive events like the bouncing of a ball or the bark of a dog as one instead of two events even though the spatial information derives from visual as well as auditory input. Therefore, at some point along the neural information processing pathway, the information from different modalities has to be combined in order to produce a unified percept. Even though research over the last hundred years yielded a substantial understanding on how humans perceive objects in space in each modality separately, our understanding on how the brain combines these information is still limited.

For a long time, vision was thought to be the dominant, and from other senses almost impenetrable, sense of perception. However, although vision might dominate the input from other senses (e.g. Colavita, 1974; Eimer, 2004; Ernst & Banks, 2002; E. I. Knudsen & Knudsen, 1985, 1989a, 1989b; Rock & Victor, 1964; Warren, Welch, & McCarthy, 1981), other sensory input is able to alter our perception as well. For instance, in the classical McGurk effect the participant sees a person saying repeatedly /ga/ while hearing the person saying /ba/, resulting in very consistently perceiving /da/ (McGurk & MacDonald, 1976). Another example is the rubber hand illusion (Botvinick & Cohen, 1998) in which the congruent visual and tactile stimulation leads to a perceived transfer of ownership of the rubber hand.

In more spatially related tasks, a systematic bias towards an irrelevant visual stimulus has been reported (Bertelson & Radeau, 1981; Pick et al., 1969). In these studies, participants have to report the origin of a sound. Simultaneously, a visual stimulus is presented at a different position. Under these circumstances, the participants' reactions become systematically biased towards the position of the distracting stimulus. Besides sound localisation, the same effect has also been observed in the other direction, but is usually more subtle (Bertelson & Radeau, 1981; L. Chen & Vroomen, 2013).

The mechanisms underlying these observations are still unclear but might stem from the neuronal underpinnings of multisensory integration. Audiovisual integration has been intensively studied on the level of single neurons, particularly in the superior colliculus (SC) of cats and other mammals (Meredith, 2002; Meredith & Stein, 1983; Stein, 2012; Stein & Stanford, 2008). The SC is a central structure of the midbrain and is arranged topographically representing the external space. It has been demonstrated that at least some neurons in the SC receive input from more than one sensory modality (see Stein & Stanford, 2008, for an overview). Furthermore, separate but overlapping auditory and visual receptive fields have been found that are assumed to enhance spatial processing of input from the same location if it arrives in close temporal proximity (Meredith, 2002; Meredith & Stein, 1996). It is worth to note that these cells do not function exclusively multimodal and can also elicit responses to unimodal input. However, it has been found that the firing rate increases "superadditively" for spatiotemporally coinciding multimodal input (King & Palmer, 1985; Meredith & Stein, 1983; Stein & Stanford, 2008).

Nevertheless, sounds that fall in temporal but not in spatial proximity have also been reported to alter human behaviour (Iordanescu et al., 2010; Iordanescu et al., 2011; Meyerhoff & Suzuki, 2018; Sekuler et al., 1997; Van der Burg et al., 2008). Van der Burg et al. (2008), for example, showed that search times in a visual search task significantly decrease if a non-spatial sound is presented. In their study, participants had to find and report the orientation (horizontal or vertical) of a target stimulus shown between many other distracting stimuli with random orientation between 0° and 90°. All stimuli changed their color randomly. In the trials in which a non-spatial sound occurred in temporal proximity to the color change of the target stimulus, the search times significantly decreased and became independent of the number of distractors. In another line of research, it was further shown that a non-spatial sound at the time of overlap disambiguates whether participants perceive two disks moving in opposite directions as bouncing or streaming (Meyerhoff & Suzuki, 2018; Sekuler et al., 1997). Thus, although spatial proximity is assumed to be a requirement for multisensory integration, non-spatial auditory stimuli are also capable of triggering integration.

What multisensory integration is in this context is rather unspecific. As Stevenson et al. (2014) pointed out, significantly faster or better reactions are not sufficient to be interpreted as integration. Reaction times can also exceed the respective unimodal condition just by independently processing a second stimulus based on statistical facilitation (Raab, 1962). Defining a baseline model under the assumption of independent processing is therefore crucial to interpret responses on multimodal presentations. For simple detection tasks, such a model has been proposed by Miller (1982). However, contrarily to detection tasks in which the participant only has to decide to respond or not, most studies incorporate multiple possible responses in n alternative forced choice tasks (nAFC).

In summary, the sensory input is integrated in the brain in a complex mechanism to form a representation of space. The question whether spatial concordance or just the temporal co-occurrence of the stimuli leads to multisensory integration is still unresolved. Research on the ventriloquism effect and spatial discordance suggests impaired reactions to spatially incongruent audiovisual stimulus combinations. In these combinations, the positions of the target in one and the distractor in a different modality differ. As pointed out, this assumption of spatial relevance is also supported by neurophysiological research on spatial processing in other animals. Contrarily, other research on perceptual disambiguation as well as on visual search also suggests integration mechanisms without a spatial component.

Simultaneously, recent studies on audiovisual semantic processing showed significant enhancement of processing redundant auditory and visual information in comparison to processing the sensory input in isolation (Y.-C. Chen & Spence, 2010). Moreover, in their experiment, responses were impaired when the two sensory inputs were in conflict. Even though semantic content is commonly assumed to be processed along the ventral path (Goodale & Milner, 1992), the results of Y.-C. Chen and Spence (2010) are in great compliance with the findings on spatial multisensory processing. This is particularly interesting as spatial informa-

tion is processed along the dorsal path and whether the same mechanisms apply along both paths is still unknown. Finding comparable effects in both processing domains would therefore decisively shape our understanding of the neuronal information processing underlying human perception.
To examine this relationship, we designed a spatial 4AFC task with a 2×3 factorial design. As independent variables, we manipulated the modality the participant had to respond to (auditory vs visual) as well as the properties of the audiovisual stimulus pairing (unimodal vs congruent vs incongruent). We expected faster responses for the congruent conditions in comparison to the respective unimodal and incongruent conditions. Based on the ambiguity in the literature, two competing hypotheses arose for the processing of spatially incongruent audiovisual information. Under the assumption of spatial relevance, incongruent responses should be impaired in comparison to the unimodal and congruent conditions. If mostly temporal information is processed and spatial proximity is to a great extend neglected, incongruent conditions might also provoke better performance than the respective unimodal condition. Therefore, if spatial information is relevant, a similar result pattern as for semantic congruency is expected. Such a similarity would suggest a much closer relationship between processing along the ventral and dorsal path than expected.

3.2 Method

3.2.1 Participants

Twenty-four subjects (19-27 years, 12 females) participated in the experiment. The participants received either course credit or a monetary compensation of 10€ per hour. All participants had normal or corrected-to-normal vision and were tested to hear spatial sounds prior to the experiment. Informed consent was provided by all participants prior to testing. The sample size was determined to conduct all possible block combinations of the four blocks ($4! = 24$ combinations) to avoid ordering effects without a power analysis, as no prior effect size was known.

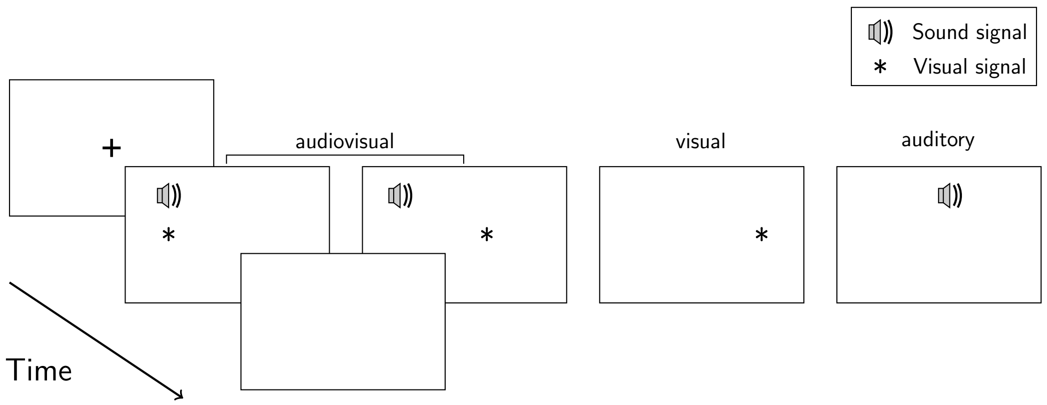


Figure 3.1: Schematic procedure of the second and third experiment. In each trial an auditory stimulus, a visual stimulus or a combination of both was presented at one of four positions. In half of the audiovisual trials the unimodal stimuli matched in position (congruent condition) whereas in the other half the distracting stimulus originated from a different position (incongruent condition). All trials were tested twice; once with each modality as the target stimulus.

3.2.2 Materials

As in the previous experiment, all stimuli were programmed using the PsychoPy library (Peirce, 2007) in Python 3 (Van Rossum & Drake, 2009). The experiments were conducted on a 24 inch LED screen (60 Hz, 1920 x 1080 pixels) with a viewing distance of 50 cm. The screen was controlled by a standard PC running Windows 10 and the Psychopy Standalone Application (Version 3.2.4). To control the viewing distance and to avoid fatigue, the participant's head was placed on a chin rest. Auditory stimuli (780 Hz pure sine wave tones without onset ramps; approx. 50-55 dB at the listeners' ear; sample rate = 44100 Hz) were presented using Creative MF1680 stereo speakers, placed on the left and the right side of the Fujitsu B24T-7 LED proGREEN computer screen. On the computer screen visual stimuli were presented at four positions at -21°, -7°, 7° and 21° relative to the center of the screen. Each of the visual stimuli consisted of a white dot with a diameter of 0.5°. For response submission the D, F, J and K buttons on a standard keyboard were used, which was placed in front of the participant (see Figure 2.3 for an overview of the experimental setup).

3.2.3 Procedure

Prior to the experiment, participants were tested for their ability to hear spatial sounds. Therefore, the spatial auditory stimuli of the experiment were presented for ten seconds or until response in randomised order. Each of the four stimuli was repeated 20 times, summing up to 80 trials in total. In this manipulation check, the participant had to indicate the perceived source by clicking the corresponding key on the keyboard (see Chapter 2 for details). After the manipulation check, the participants were instructed for the main experiment.

Before each block, the participant was additionally instructed in which modalities stimuli were presented and to which of the modalities he or she had to respond. Each block instruction was followed by a sequence of 24 test trials to check whether the participant understood the task. During the test trials, which followed the same procedure as the main trials, the participants received feedback on their correctness after each trial and their overall performance after completion of all test trials.

In each of the main trials, a white circle, a spatial sound or both were presented for 1000ms or until the participant responded (cf. Figure 3.1). The participant's task was to respond as fast but also as correct as possible to the position of the target stimulus. Target as well as presentation conditions were balanced. In each target condition, a third of all trials consisted just of the target stimulus (unimodal condition) whereas in the other two third a distractor stimulus of the other modality was additionally presented. The position of the distractor matched the position of the target stimulus in half of the audiovisual trials (congruent condition) and was presented at a different position in the other half (incongruent condition). Congruent and incongruent presentation made thus also a third of all trials each. The position of the target, the distractor as well as their combinations were uniformly distributed. Incorrect trials were repeated at the end of the block until all trials were answered correctly. After each block, the participant had to leave the room to make a small break to avoid fatigue.

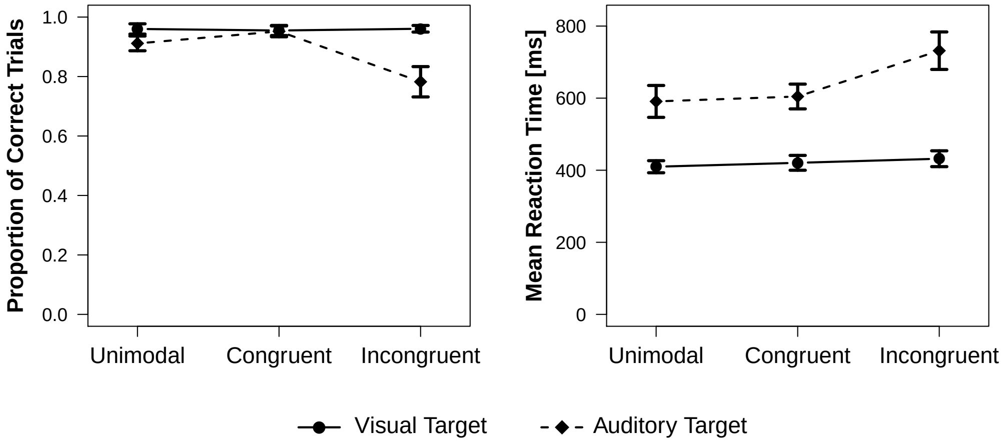


Figure 3.2: Correctness and reaction times are plotted for all six conditions. The error bars indicate the 95% within-subject confidence intervals. The *left plot* shows the proportion of correct trials per condition. Participants detected visual and auditory targets at a comparable level in the congruent condition, but differed in the unimodal and incongruent conditions. In particular, the expected response conflict was only observed for auditory but not for visual targets. The *right plot* shows the mean reaction time for each condition. Participants were faster detecting a visual compared to an auditory target with a response conflict observed only for auditory targets.

3.3 Results

A two-factorial repeated measurements ANOVA of the reaction times revealed faster responses to visual than to auditory targets ($F(1,23) = 177.1, p < .001, \hat{\eta}_p^2 = 0.89$). Moreover, reaction times varied significantly between the presentation conditions ($F(2,46) = 33.98, p < .001, \hat{\eta}_p^2 = 0.60$). The interaction between presentation condition and target modality was also significant ($F(2,46) = 23.02, p < .001, \hat{\eta}_p^2 = 0.50$). The results are shown in the right plot of Figure 3.2. Reaction times were only analysed for correctly answered trials. As all incorrectly answered trials were repeated, 72 observations per condition and participant were measured. Response accuracies were good in all conditions except of the incongruent condition with auditory target. Bonferroni corrected posthoc paired t-tests were applied between the different presentation conditions revealing slower responses to incongruent than unimodal ($t(23) = 3.22, p < .05, d = 0.43$) as well as to congruent ($t(23) = 4.26, p < .01, d = 0.21$) trials with vi-

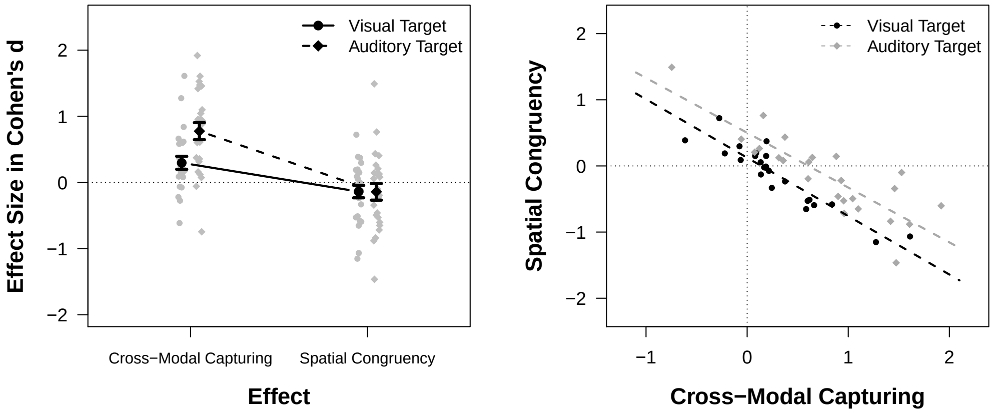
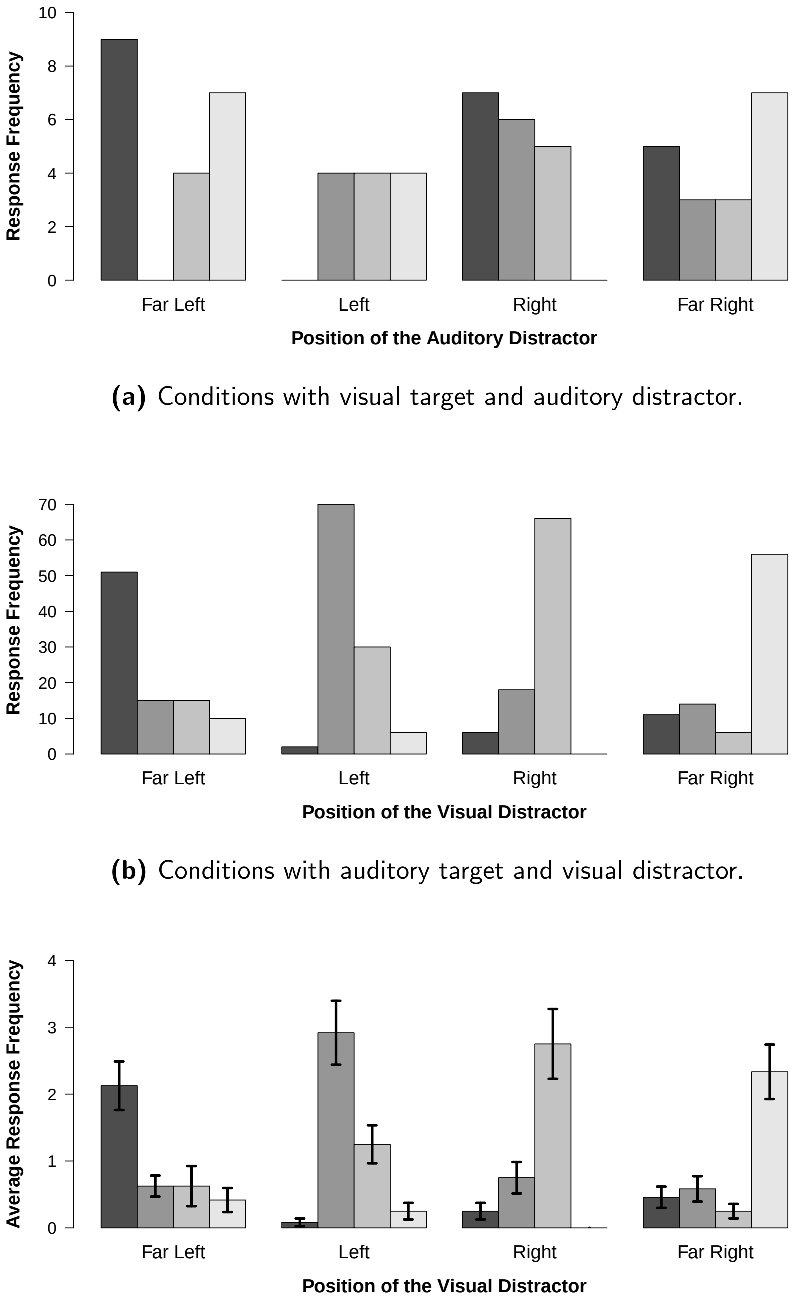


Figure 3.3: Effect sizes and individual differences of the two effects are shown. The *left plot* shows the effect sizes for cross-modal capturing and spatial congruency for all participants (grey symbols) as well as the average effect sizes (black symbols). Error bars indicate one standard error of the mean. In the *right plot* the individual cross-modal capturing effects are plotted against the spatial congruency effects revealing a negative relationship between them.

sual target. No difference between the reaction times of the congruent and the unimodal condition with visual target was observed. The same result pattern was observed for auditory targets, with reactions to incongruent trials being slower than to unimodal ($t(23) = 5.73, p < .001, d = 1.16$) or congruent ($t(23) = 7.46, p < .001, d = 1.02$) trials. Again, the comparison between the congruent and the unimodal condition was not significant.

Additionally, an exploratory analysis was conducted on the reaction time data. Effect sizes as difference in mean between the conditions in units of the pooled standard deviation were calculated for each participant for the effects of cross-modal capturing (the difference between unimodal and incongruent condition) as well as for spatial congruency (the difference between unimodal and congruent condition). Effect sizes per participant as well as the respective averages are shown in the left plot of Figure 3.3. A linear relationship between the two effects was observed as shown in the right plot of Figure 3.3 ($F(1, 44) = 120.96, p < .001$). Furthermore, there seems to be an effect of the target's modality ($F(1, 44) = 22.55, p < .001$). An interaction term was included, but was not significant. The model fitted the data well ($F(3, 44) = 47.88, p < .001$) with an overall explained variance of $R^2 = .77$.



(a) Conditions with visual target and auditory distractor.

(b) Conditions with auditory target and visual distractor.

(c) Average response frequency per participant.

Figure 3.4: Number of incorrectly responded positions for each position of the distactor. Each group belongs to one distractor position with the responded position from far left to far right within each group. Figures 3.4a and 3.4b show the absolute number a position was responded for the respective distractor's position. Figure 3.4c shows the average response frequency per participant for the auditory target condition. For auditory targets the incorrect responses in the incongruent conditions were mainly influenced by the irrelevant visual distractor.

3.4 Discussion

Goal-directed behaviour originates in perceiving the world around us. For that reason spatial information ("Where is something?") as well as semantic information ("What is something?") have to be processed. Physiologically, both processing pathways are structurally separated and whether the same mechanisms apply to both is still unknown. In this experiment, we conducted a spatial detection task comparable to previous studies on semantic congruency in order to compare how the behavioural responses are related to each other. This is of particular importance as the literature on multisensory integration is ambiguous in this aspect: Spatial congruency was argued to be a requirement besides temporal coincidence (Meredith, 2002; Meredith & Stein, 1996). However, the necessity of spatial proximity has been questioned by several studies reporting multisensory integration based on solely non-spatial sounds (Iordanescu et al., 2010; Iordanescu et al., 2011; Meyerhoff & Suzuki, 2018; Sekuler et al., 1997; Van der Burg et al., 2008). Comparable processing along the dorsal "where" and the ventral "what" path is only possible if spatial information of both modalities is relevant. Otherwise, no redundancy of information would exist.

The present study demonstrates strong agreement with previous studies showing that spatially displaced but irrelevant distractors impair processing. This phenomenon is particularly well studied for auditory targets that are accompanied with spatially separated visual distractors, a phenomenon that has become known as ventriloquism effect. During ventriloquism, the human brain binds audiovisual spatial information together in a way that the sound is experienced as originating from the position of the visual stimulus. Besides the significantly decelerated response times, this is also reflected in the lower accuracy of this condition compared to the rest (cf. Figure 3.2). Moreover, when a participant has responded incorrectly, the responded position matched almost always the position of the visual distractor (cf. Figures 3.4b and 3.4c), whereas no clear trend is observable for auditory distractors (cf. Figure 3.4a). The mismatch for auditory targets seems to be independent of the distance to the distractor.

Whether this reflects real ventriloquism is unclear as ventriloquism refers to the persistent perception of both stimuli having the same source. This conclusion cannot be drawn as the participants' task was to make speeded responses to the target. Moreover, in most of the trials did the participants correctly report the target's position and only in roughly a fifth of the trials the position was not stated correctly. This suggests rather an involuntary shift of attention towards the visual stimulus which given the speeded response instructions led to more responses of the distractor position. An attentional capturing by the accompanying distractor is therefore reasonable.

As similar result pattern was also observed for the visual targets, but was more subtle. This difference between the target modalities is also depicted in the left plot of Figure 3.3. This is in great compliance with the current literature that also describes that auditory capturing of visual stimuli is less intrusive than vice versa (L. Chen & Vroomen, 2013). However, given the low number of incorrect responses in the latter condition, it remains unclear whether the distractor really captures the visual target or just involuntarily also attracts attention. In the latter case, the saliency of the distractor might not be high enough to provoke immediate responses but additional time would be required to resolve the conflict as observed.

Together, these to effects highlight the relevance of spatial information in multisensory integration. But how can it be that so many studies demonstrated multisensory integration without a spatial component of the sound? As pointed out earlier, attention seems to play an integral part during this process (Talsma et al., 2010). In the studies reporting effects of non-spatial sound, spatial information was usually of low importance. For instance in the study of Van der Burg et al. (2008), the additional auditory information was only relevant on the temporal domain as it indicated the time of relevant event. Additional spatial information might be beneficial in this case but the temporal information is already sufficient to pay attention to the change of this moment.

The two directions of cross-modal capturing as discussed so far are also in line with the observations in processing semantically incongruent stimuli (e.g. Y.-C. Chen & Spence, 2010). In their study, responses to stimuli accompanied with a semantically incongruent stimulus of the other modality were significantly impaired. However, contrary to their results, spatial congruent presentation of audiovisual stimuli did not lead to faster processing. A possible reason for that is that responses to the visual stimuli were already very fast. Moreover, even though the response times varied significantly between the unimodal and incongruent condition, the absolute time difference is rather small between all visual conditions.

The significantly faster reactions to visual than to auditory targets suggests furthermore that vision is dominating the processing in this task as frequently reported in the literature (Colavita, 1974; Eimer, 2004; Ernst & Banks, 2002; Ernst & Bülthoff, 2004; Rock & Victor, 1964). A strongly dominating visual sense would imply that the relative weight of the auditory sense would diminish and, therefore, not be observable any longer. Our results would, thus, be in compliance with the inverse effectiveness principle of multisensory integration stating that the relative enhancements become larger the lower the dominant signal is (Stanford & Stein, 2007). Vice versa, if one sense dominates the other, the benefit produced by multisensory integration decreases. This hypothesis can be easily tested by reducing the dominance of the visual sense.

Such a reduction can be obtained in many ways. For low level stimuli like dots, the brightness, contrast or overall noise can be manipulated. All of these manipulation have in common that the salience of the target is reduced and response times increase for that reason. Simultaneously, it is likely that with decreasing saliency the ability of attracting involuntary attention is also reduced. Given the idea that attentional processing caused the observed effects as pointed out earlier, these effects should be smaller under this manipulation.

This idea is further supported by the exploratory analysis that we have conducted on the data. In great compliance with the idea of involuntary bottom-up driven shifts of attention, the individuals' effects formed a linear relationship (see right plot of Figure 3.3). In this plot the differences between the means in the unit of pooled standard deviations are plotted against each other. As illustrated by the two regression lines, participants showing stronger effects of cross-modal capturing showed smaller effects of spatial congruency and vice versa. The reason for that might be that some participants were better in focusing on one modality whereas other are easier distractable by the other modality. If involuntary attentional shifts are reduced by lowering the contrast, the average should be shifted along this lines in direction of spatial congruency.

Taken all together, the results of this experiment clearly support the relevance of spatial information in multisensory processing. However, this relevance might be relative to the respective task and in consequence to the respective informativeness of the spatial information. Furthermore, our results are in great compliance with the idea that attention promotes multisensory integration to a large extent. Nevertheless, not all of our hypotheses could have been confirmed as spatial redundancy provoked neither in the visual nor in the auditory target conditions any benefits. We attribute this observation to the high saliency of the visual stimulus which, given the inverse effectiveness principle of multisensory integration, likely cancels out the positive influences of the additional stimulus. These hypotheses are tested in the next chapter.

Chapter 4

Spatial Congruency Effects Might Increase Under Visual Uncertainty

4.1 Introduction

The previous experiment reported in Chapter 3 revealed significant cross-modal influences of both visual as well as auditory distractors on the respective target. Contrarily to our expectations, spatial congruency of the target and the distractor did not enhance processing. Even though the reasons for that are manifold, the likeliest explanation seems to be the excessive salience of the visual stimulus as observed in the very fast reaction times and the low within-subject variance (cf. Figure 3.2) in the visual target conditions. Moreover, this argument is supported by neurophysiological underpinnings of multisensory integration (Stanford & Stein, 2007; Stein & Stanford, 2008). Stanford and Stein (2007) argued that the occurrence of "superadditive" spiking behaviour, in which the number of emitted action potentials exceeds the sum of incoming spikes, increases with decreasing stimulus saliency. In other words, a second signal is unlikely to provide additional information when the first signal's informativeness is already high. This principle has become known as the reverse effectiveness principle of multisensory integration.

Behavioural correlates of this observation have also been reported (Ernst & Banks, 2002; Ernst & Bülthoff, 2004). Perceptual misjudgements in visuohaptic processing, for instance, depend on the quality of the visual signal (Ernst & Banks, 2002). In these experiments, participants had to judge the size of an object by sight or haptics. The authors systematically varied the distortion of the visual sense resulting in an increasing bias towards the visual domain, the better the visual signal's quality was. Contrarily, the correct haptic information gets reported more with decreasing visual reliability.

This observation has also been quantified using a Kalman filter (Alais & Burr, 2004; Ernst & Banks, 2002; Kalman & Bucy, 1961; Wolpert et al., 1995):

$$\hat{s} = \sum_i w_i \hat{s}_i \text{ with } \sum_i w_i = 1. \tag{4.1}$$

In its basic definition, a Kalman filter is defined as a weighted average of the incoming signal estimates $\hat{s}_i$ that are weighted by a factor w_i. Additionally, the weights of the respective signals are normalised so that the sum over all weights equals one. With respect to multisensory integration this corresponds to the assumption of relative weighting of the sensory signals. This can be easily demonstrated by assuming two signals s_A and s_V for the auditory and visual signal as well as their corresponding weights w_A and w_V:

$$\hat{s} = w_A \hat{s}_A + w_V \hat{s}_V \text{ with } w_A + w_V = 1. \tag{4.2}$$

By inserting the normalisation restriction (second term in Equation (4.2)) into the first, one obtains

$$\hat{s} = w_A \hat{s}_A + (1 - w_A)\hat{s}_V, \tag{4.3}$$

which shows that under these assumptions, the second signal is weighted relative to the first signal's reliability and vice versa.

In the context of our experimental setup, this predicts a stronger audiovisual processing when the visual signal's quality is lower. To follow this idea, the visual signal's reliability needs to be reduced. For low level stimuli like the white dots in our experiment, brightness, contrast or overall noise manipulations are commonly used to reduce visibility and to decelerate reactions. Alais and Burr (2004), for example, used extensive blurring to decrease the reliability of visual blobs.

In our experiment, we decided to decrease the visual stimulus' quality by decreasing its luminance and, thus, its contrast compared to the background. As contrast measurement, we used Weber contrasts

$$C = \frac{L_{Stimulus} - L_{Background}}{L_{Background}} \tag{4.4}$$

with $L_{Stimulus}$ as the luminance of the stimulus and $L_{Background}$ as the background's luminance. As stimulus' contrasts for this experiment, we chose $C_1 = 0.25$ in the low contrast condition and $C_2 = 0.05$ in the very low contrast condition. As a comparison, the visual stimulus' contrast of the second experiment reported in Chapter 3 was approximately $C_0 = 3.98$. The two contrast levels were based on pretests to obtain similar reaction times in comparison to the auditory target condition.

By doing so, the goal of this experiment was to test again the hypotheses of the previous chapter. Spatial congruency, provoked by the redundancy of the target's and distractor's position, was expected in form of faster reaction times in the congruent conditions compared to the respective unimodal condition. The observed cross-modal capturing effects of the previous experiment were expected to be replicable. However, in comparison to the previous experiment, this effect might decline due to the decreased saliency of the visual distractor but, in contrast, might increase for auditory distractors. Finally, as reducing the visual stimuli's contrast aimed to decelerate the responses in these conditions, the reaction times were expected to not vary between the two target modalities.

4.2 Method

4.2.1 Participants

Nine subjects (19 - 26 years, 4 females) participated in the experiment. The participants received either course credit or a monetary compensation of 10€ per hour. One of the participants was the author of this thesis. All participants had normal or corrected-to-normal vision and were tested to hear spatial sounds prior to the experiment. Informed consent was provided by all participants prior to testing. The sample size was limited to nine after the outbreak of the COVID-19 pandemic. As in the previous experiment, a number of 24 participants was targeted.

4.2.2 Materials

The experimental setup was mainly the same as in Chapter 3. Instead of white dots, the contrast of the visual stimulus was reduced and presented at two different levels (low contrast with $C_1 = 0.25$ and very low contrast with $C_2 = 0.05$).

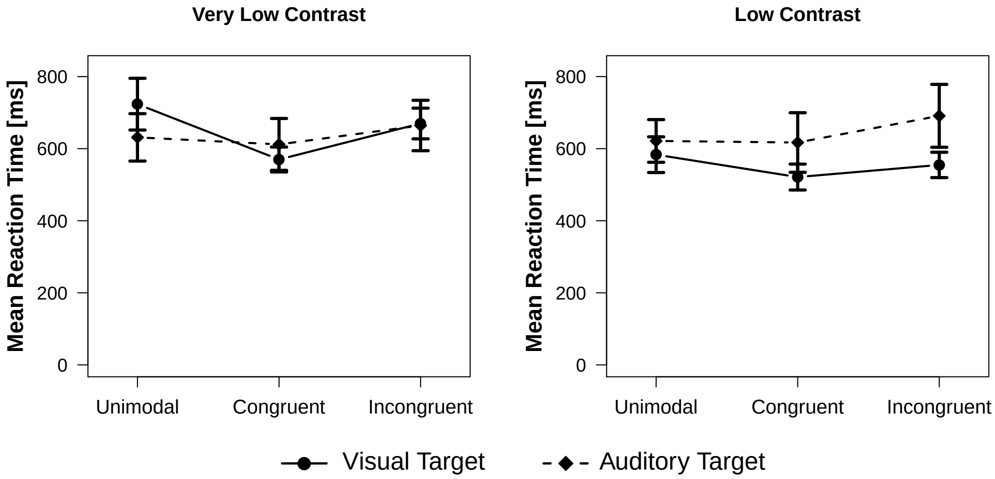


Figure 4.1: Mean reaction times are plotted for all six conditions at two different contrast levels. The error bars indicate one within-subject standard error of the mean. The conditions with lower contrast stimuli are shown on the *left* and with the higher contrast level on the *right*. For visual targets, faster reaction times were observed in the congruent condition than in the incongruent or unimodal condition. This was particularly the case for lower contrast. Reaction times for auditory stimuli did not vary much between the conditions.

4.2.3 Procedure

The procedure of this experiment was the same as in the previous chapter, except that in half of the test and main trials a visual stimulus of low contrast and in the other half a visual stimulus of very low contrast was used. The manipulation check that was performed prior to the experiment was the same, but the results are not reported.

4.3 Results

A three-factorial repeated measurements ANOVA of all correctly answered trials revealed that the reaction times significantly differed between the presentation conditions ($F(2, 16) = 22.58, p < .001, \hat{\eta}_p^2 = .74$). Neither the main effect of target modality nor their interaction was significant. As the third factor, we analysed the influence of contrast (as a measure of visibility). Overall, lower contrast seems to lead to slower reaction times ($F(1, 8) = 88.01, p < .001, \hat{\eta}_p^2 = .92$). Furthermore, the level of contrast seems to interact with the target modalities

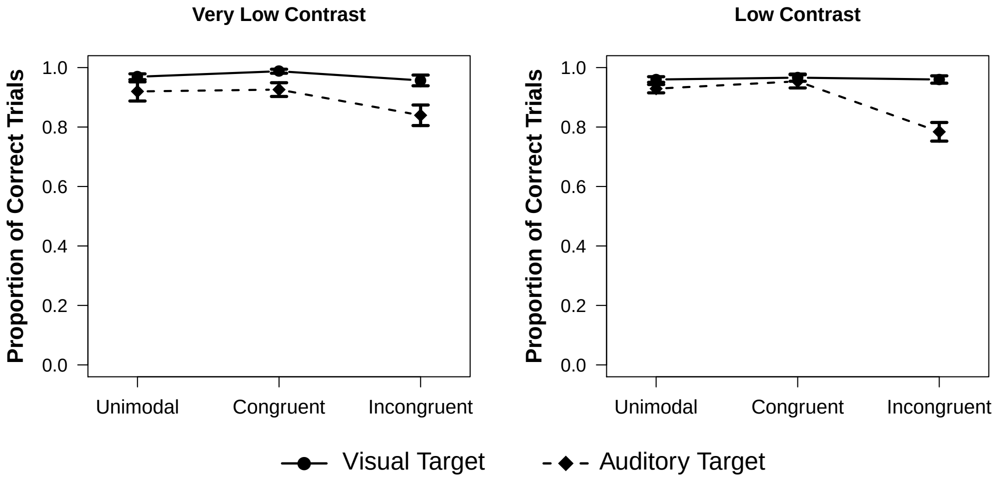


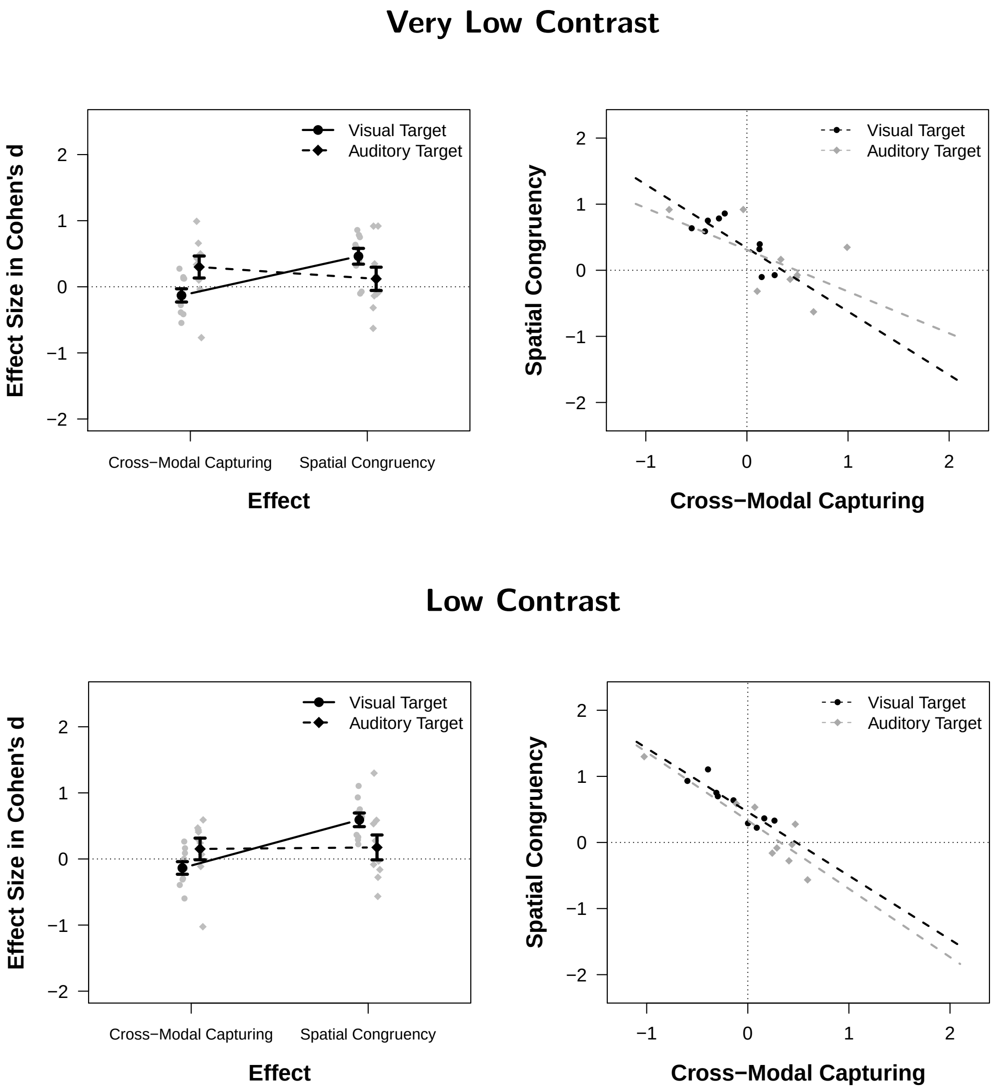
Figure 4.2: Average response correctness is shown for all six conditions at two different contrast levels. The error bars indicate one within-subject standard error of the mean. Participants responded well in all conditions. However, response correctness decreased in the incongruent conditions.

$(F(1,8) = 41.8, p < .001, \hat{\eta}_p^2 = .84)$ but not with the presentation conditions. Finally, the data also supports the three-way interaction of presentation condition, target modality and contrast $(F(2,16) = 11.96, p < .001, \hat{\eta}_p^2 = .60)$.

Bonferroni-corrected posthoc t-tests only indicated significant differences between the congruent and incongruent condition of the very low contrast condition with auditory $(t(8) = 5.49, p < .01, d = .24)$ as well as visual target $(t(8) = 6.29, p < .01, d = .75)$ and of the low contrast condition with visual target $(t(8) = 5.06, p < .05, d = .31)$.

Again, we calculated the two effects for each participant individually as depicted in Figure 4.3. Only slight differences can be observed between the two contrast levels. Plotting the two effects against each other as in the previous experiment, revealed a similar linear relationship between them, but the difference between the target modalities decreased (see right plots in Figure 4.3). Regression lines for the two linear models (one for each contrast level) are plotted for visual guidance.

As the outbreak of the COVID-19 pandemic abruptly stopped the data collection, this section rather sketches the planned analysis of the data than it reports useful results.



;

Figure 4.3: Effect sizes and individual differences of the two effects are shown for the two contrast levels. The *left plots* show the effect sizes for cross-modal capturing and spatial congruency for all participants (grey dots) as well as the average effect sizes (black dots). Error bars indicate one standard error of the mean. In the *right plots* the individual cross-modal capturing effects are plotted against the spatial congruency effects revealing negative relationships between them.

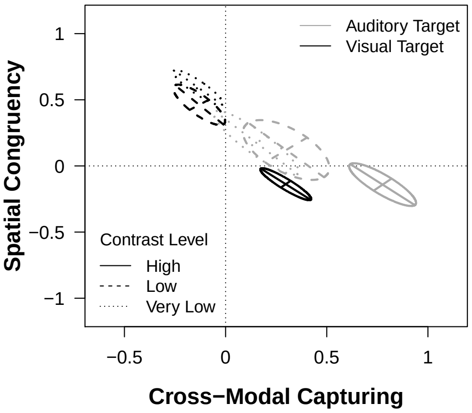


Figure 4.4: The negative relationship between spatial congruency and cross-modal capturing is shown with centroids and 95% confidence ellipses for the three contrast and two target levels. Lower contrast levels seem to show more spatial congruency and simultaneously less cross-visual capturing.

4.4 Discussion

Hearing the horn of a lighthouse becomes particularly important when it is foggy and you cannot see the coast. In more general terms, auditory spatial information becomes of specific importance when the visual information is not reliable. To demonstrate that was, amongst other things, the purpose of this experiment. However, as the outbreak of the COVID-19 pandemic abruptly stopped the data collection, the results of this experiment are not inferentially interpretable. For that reason, we will discuss the data only on a descriptive level and highlight the possible implications they would have if the statistical analysis confirms these trends after data completion.

As in the previous experiment, we were interested in the relation between semantic and spatial processing. Even though some correspondence along the processing pathways appears to exist, anatomically they are processed along different pathways (Goodale & Milner, 1992). Whether the same mechanisms of multisensory integration operate along both paths is still unknown. If they do, this should be reflected in comparable result patterns for both types of information processing. Recently, Y.-C. Chen and Spence (2010) reported semantic congruency effects as well as response impairments based on semantic incongruency. The latter expectation has already been confirmed within the scope of spatial processing in the previous experiment. The so far collected data is ambiguous in this respect. Even though the means of the incongruent condition reveal slower reaction times than the means of the unimodal condition for auditory targets, responses to visual targets seem to be faster (see Figure 4.1).

If this trend holds, this would be in line with the studies reporting multisensory integration with non-spatial sounds (Iordanescu et al., 2010; Van der Burg et al., 2008). Therefore, spatial information would not be of the same relevance as reported frequently. This is particularly interesting as for other species it is well studied that multisensory spatial integration is implemented as overlapping unisensory receptive fields in the superior colliculus (Meredith, 2002; Meredith & Stein, 1996). As the neuronal underpinnings of multisensory integration are well studied and in great compliance with behavioural findings this trend might also be simply an artefact of the uncompleted data collection.

Another argument against the irrelevance of spatial information is depicted in Figure 4.2. As already reported in Chapter 3, the response accuracy for auditory targets with an accompanying visual distractor are impaired at both contrast levels. We did not analyse whether the incorrect responses in these conditions were again dominantly the position of distractor due to the limited amount of data. However, given the results of the previous experiment, this hypothesis seems to be very likely. If spatial information would not be processed there should be no difference between the conditions, but descriptively such a difference is likely.

Spatial congruency, on the other hand, might be observable for visual but not for auditory targets. At both contrast levels, the response times of the congruent conditions were faster than the respective unimodal condition. Contrarily, the congruent and unimodal conditions with auditory target did not seem to vary on average independent of the contrast level. This is partially in line with our expectations as we expected congruency effect independent of the target modality. If this trend holds it would raise the question why redundancy is beneficial for visual but not for auditory targets. In the previous chapter we discussed the role of attention as an important process underlying multisensory integration. Following this line of thought, the auditory stimuli would have to provide significantly more spatial information than the visual stimuli. In this case, the spatially coinciding informative auditory stimulus would dominantly provide the relevant information that lead to a faster reaction time. However, this can not explain why responses to incongruent presentations are also faster for visual targets.

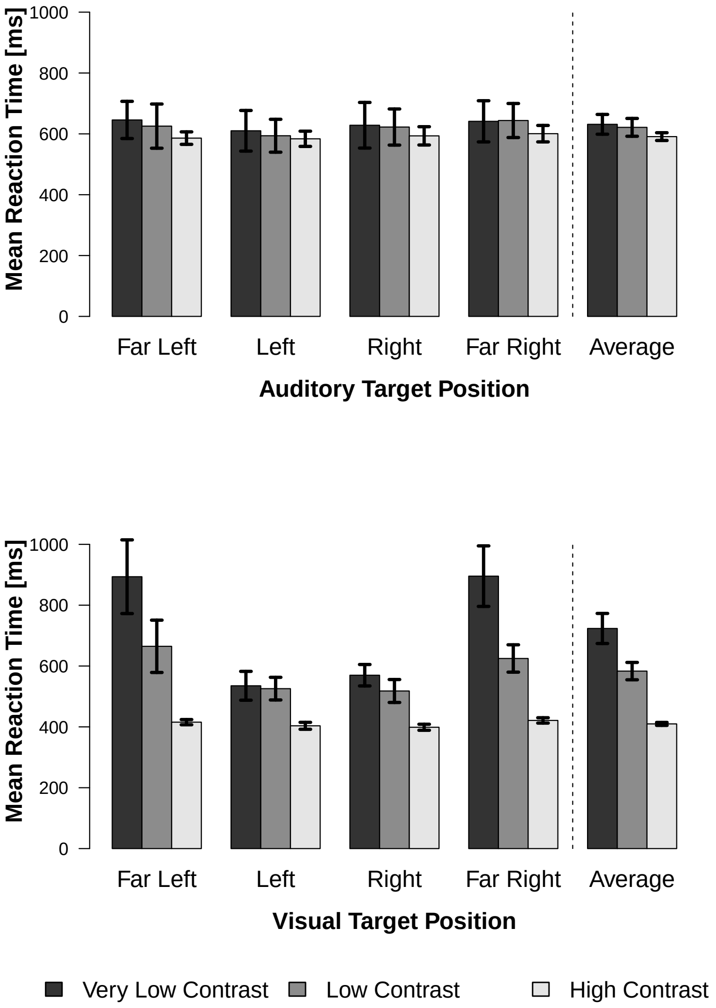


Figure 4.5: Average reaction times by target position and contrast level for the two unimodal conditions. Error bars indicate the standard error of the mean for $n = 24$ (high contrast) and $n = 9$ (very low and low contrast) observations. As expected, reaction times for auditory targets do not vary between positions or contrast levels. In contrast, the reactions to visual targets slowed down as a function of contrast level. Moreover, reactions to peripheral targets (on the far left and right) were peculiarly impaired.

Nevertheless, the role of attention is partially supported. In the previous chapter, we reported a linear relationship between spatial congruency and cross-modal capturing and concluded that this is likely an artefact of dividing the attention across two modalities with some participants being better in focusing on one modality whereas others are easier distractable. The same relationship seems to be existent for the previous study even though the amount of data points makes fitting a regression model difficult (see Figure 4.3). We assumed that if attention is the underlying mechanism of this relation, the centroids of the conditions should be affected by contrast manipulation as attention is as well. This hypothesis might hold when the centroids of the six conditions are compared (see Figure 4.4). As expected the centroids seem to be shifted along the linear relation to evoke stronger spatial congruency effects when the visual stimulus quality gets decreased. Simultaneously, cross-modal capturing decreases as well.

However, the implications of this experiment are also limited as the responses were not equally decelerated over the four positions. As depicted in Figure 4.5, response times were stronger impaired for peripheral visual stimuli even though the reaction times increased equally on average. Preferably, reaction times should be slowed down equally for all positions. This, however, is difficult to achieve as the visual and auditory system are less sensitive in the periphery (Mills, 1958; Wilson & Sherman, 1976). Increasing the visual stimulus' size at the peripheral positions, for instance, would be an option to counteract the decelerated peripheral response times and to account for the larger receptive fields in this area. On the other hand, this might again differ between the participants and, therefore, introduce new noise. Ideally, further research should incorporate a trade-off to obtain more equally decelerated response times without adding to much additional noise.

In summary, whether spatial and semantic congruency produce similar effects remain unknown due to the limited amount of data. The question whether they do is still of high relevance to understand how semantic information is processed as for these type of tasks no sufficient animal model exists. Understanding the similarities and differences between dorsal and ventral path might help to relate also other aspects of neuronal spatial processing that is known from animal studies. As an important factor of multisensory integration, we suggested involuntary bottom-up driven attention to explain our findings as it has already been discussed in the literature (Talsma et al., 2010). The relation to semantic processing as well as the differences between individuals highlight the importance of further investigations on the interplay of multisensory integration and attention. In particular, it would be of interest whether benefits can also be provoked by top-down driven attention or whether multisensory integration only operates on the perceptual side.

Chapter 5

General Discussion

Does hearing the bark not only help to identify but also to localise the dog? The goal of this thesis was to relate recent findings on cross-modal semantic congruency to the processing of spatial information. The results of the three experiments reported in this thesis demonstrate severely impaired performance when processing spatially incongruent stimulus pairings. On the other hand, performance was neither impaired nor enhanced when the position of the distractor matched with the target in the second experiment. The data of the third experiment is not inferentially interpretable as data collection had to be stopped too early. Descriptive comparisons of the means in this experiment indicate no clear pattern. Spatially congruency might enhance performance for visual but not for auditory targets when the visibility of the stimuli is impaired. Spatially incongruent stimulus pairings might impair performance for auditory but not for visual targets. To the contrary, performance might be even enhanced for visual targets.

This is partially in compliance with recent findings on cross-modal semantic congruency. Y.-C. Chen and Spence (2010), for instance, reported impaired performance when a semantically-incongruent sound was presented together with the target picture, as compared to when a white noise burst was presented instead. Our results confirm these findings and demonstrate that this incongruency effect seems to be impenetrable by the visual quality of the signal (see Chapter 4). However, as discussed in Chapter 3, our findings are also well related to the literature on cross-modal capturing. During cross-modal capturing one stimulus captures a spatially displaced stimulus of a different modality forming one unified percept of both stimuli. This phenomenon is well known for natural speech as ventriloquism during which the spoken words of an artist are perceived as originating from its puppet. Whether the semantic incongruency effects reported by Y.-C. Chen and Spence (2010) can be related to ventriloquism is unclear. Semantic based explanations of the ventriloquism effect have been discussed (Spence, 2011), but still most research has focused on the spatial and temporal factors of multisensory integration (Stein, 2012; Stein & Stanford, 2008). Our findings support the idea that spatial factors play the main role during ventriloquism. Particularly, that the effects of cross-modal capturing were well observable under all circumstances and with stimuli that carried way less semantic information than natural language strongly endorse this idea.

Spatially redundant stimuli pairings, on the other hand, do not seem to enhance processing. This seems to hold at least for settings under common circumstances with well detectable visual stimuli (see Chapter 3). Whether the same applies for less well visible stimuli is still unclear. A possible explanation why spatial congruency enhances processing just under impaired view can be found in the inverse effectiveness principle of multisensory integration (Stanford & Stein, 2007). Broadly speaking, additional signals only enhance performance based on their relative informativeness in the situation. Under clear view, vision already provides so much spatial information that redundant auditory information is irrelevant, but when visibility decreases the relative informativeness of audition increases. Important in this context is that not all senses are equally informative under all circumstances with vision usually dominating the other senses in the most situations (Colavita, 1974). Under visual impairment the performance seems to be at least descriptively enhanced for visual stimuli (see Figure 4.1). Whether this trend holds in general is unclear as the available data is limited.

Although the so far collected data also suggest speeded responses for spatially incongruent stimulus pairings, the reliability of the data can be questioned. Nevertheless, our findings are partially in compliance with several recent studies suggesting multisensory integration on the behavioural (Iordanescu et al., 2010; Iordanescu et al., 2011) as well as in the psychophysical level (Meyerhoff & Suzuki, 2018; Sekuler et al., 1997) without a spatial property of the sound. However, in none of these experiments, the spatial component of the sound would have been of use as solely the temporal information was relevant. For our results, this would also suggest the irrelevance of the spatial information of the auditory signal and therefore questions the potential effects of spatial congruency in general.

The spatial informativeness of the auditory stimuli, however, has been discussed thoroughly and has been tested experimentally in Chapter 2. Even though we constrained the signal on small amount of physical cues, namely the interaural level difference, the participants were able to detect and report the positions of the virtual sound origins well. Nevertheless, better performances might be observable for real-life experimental settings in which physical sound sources are used. Besides the interaural level difference, the interaural time difference as well as monoaural cues can contribute to the overall perception. Whether this would change the observed result pattern of Chapter 3 is unclear. On the one hand, the additional cues might enrich the auditory signal in a way that its relative weight is sufficient that its integration provokes significantly faster reactions in the congruent condition as compared to the unimodal condition. On the other hand, given the strong dominance of the visual sense, it is doubtful whether the additional cues are sufficient to provoke spatial congruency in everyday-like situations.

An important factor whether stimuli can provoke spatial congruency might be their ability to capture attention. As recently discussed, attention might be a key mechanism modulating multisensory integration (Talsma et al., 2010). The results of our experiments are in compliance with this idea as reducing contrast can be related to a lower capability of involuntary bottom-up attention shifts towards the visual stimulus. To examine this relationship, we calculated the individuals' effects of cross-modal capturing and spatial congruency. If attention modulates multisensory integration, both effects should depend on each other based on the idea that attention is divided to both modalities, but limited in general. These assumptions explain the observed linear relationship as illustrated in Figure 3.3 well and suggest that the variations between the participants are individual trade-offs of spreading the limited attention across two modalities.

This hypothesis has just been exploratorily examined in the first experiment and its exact reason is still unknown. Nevertheless, with the limited data of the second experiment, this relationship seems to be replicable. Calculating the centroids for all conditions as depicted in Figure 4.4 indicated that these are also contrast dependent as lower contrast levels seem to show more spatial congruency and simultaneously less cross-visual capturing. Following the idea of attention as the underlying mechanism, a reduction of contrast reduces the capability of the visual signal to capture attention. In consequence, cross-modal capturing reduces whereas spatial congruency increases as discussed. However, as the two experiments were conducted with different participants, it is unclear whether the centroids' shift has to be attributed to contrast alone or to which extend sampling bias influenced the results. A replication of the experiments with contrast as a within factor would thus be advisable to preclude sampling bias as a possible source.

Whether this method of analysis is beneficial for other conflict tasks as well is still unknown but likely. Applying this method to other cross-modal conflict studies (Y.-C. Chen & Spence, 2010; Cowan & Barron, 1987; White & Prescott, 2007) might shed further light on the question whether attention is the mediator of this relationship and how it interacts with multisensory processing. The interplay between attention and multisensory integration is a major current focus of multisensory integration research (Lunn, Sjoblom, Ward, Soto-Faraco, & Forster, 2019; Talsma et al., 2010) and is still under debate. The proposed method might help to clarify the partially contradictory findings and might further help to relate multimodal to unimodal attention as it can be applied to unimodal conflict tasks (Eriksen & Eriksen, 1974; Navon, 1977; Stroop, 1935) as well. Similar response patterns for multimodal and unimodal tasks might suggest that attention is rather cue than modality organised.

Structural similarities between spatial and semantic attention would also explain the similarity in processing of semantically and spatially incongruent stimulus pairings. The discrepancy between spatial and semantic congruency might be, following this line of argument, attributable to the different a priori abilities of the stimuli to capture attention. For real-life application this would imply some critical limitations. In most everyday situations the visual quality is good, but improved responses by additional stimuli would still be beneficial. Particularly in emergency situations, a faster detection of potentially dangerous events can save life providing additional multisensory information to speed up reactions.

Our results provide first insights to the similarities and differences between rather semantic processing like object recognition and spatial processing. In compliance with recent findings on semantic congruency, responses to spatially incongruent presentations were impaired. Contrarily to Y.-C. Chen and Spence (2010), spatial congruency does not enhance responses at least under normal sight. Whether it improves reactions under degraded vision was tested in the last experiment. As data collection had to be stopped early, the question remains unknown. Positional differences in the increase of reaction times with decreasing contrast would have further lowered the reliability of the results. With regard to further research, this implies that peripheral stimuli have to be adjusted to avoid excessively decelerated reaction times at these positions.

Beyond the scope of perception, our results provide further evidence for a multifaceted interplay between attention and multisensory integration. In this thesis, we also proposed a new analysis method to examine the relationship between spatial congruency and cross-modal capturing. As pointed out in Chapter 3, the observed linear relationship between the two effects might be caused by differently a priori distributed attention. Whether this relationship is cross-modal by nature or does also apply to conflict tasks within a modality is still unknown. Future research should, therefore, apply the method also to other conflict tasks and modality combinations.

Besides the field of cognitive science our results have also practical implications for clinical psychology. As the relationship between spatial congruency and cross-modal capturing appears to be very stable across healthy participants, the proposed method might also be used as a diagnostic tool to detect disabilities to distribute attention across modalities. This might be of particular interest for the assessment of driving ability during neurorehabilitation. However, how participants suffering from neglect or other neurological disorders react in this setting has to be examined first.

Bibliography

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257–262. doi:10.1016/j.cub.2004.01.029

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using r*. doi:10.1017/CBO9780511801686

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics*, *29*(6), 578–584. doi:10.3758/bf03207374

Blauert, J. (1997). *Spatial hearing: The psychophysics of human sound localization*. Cambridge, MA: MIT press.

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature*, *391*(6669), 756–756. doi:10.1038/35784

Brungart, D. S., & Rabinowitz, W. M. (1999). Auditory localization of nearby sources. head-related transfer functions. *The Journal of the Acoustical Society of America*, *106*(3), 1465–1479. doi:10.1121/1.427180

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. doi:10.1016/j.cognition.2009.10.012

Chen, Y.-C., & Spence, C. (2017). Assessing the role of the 'unity assumption' on multisensory integration: A review. *Frontiers in Psychology*, *8*. doi:10.3389/fpsyg.2017.00445

Chen, L., & Vroomen, J. (2013). Intersensory binding across space and time: A tutorial review. *Attention, Perception, & Psychophysics*, *75*(5), 790–811. doi:10.3758/s13414-013-0475-4

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, *16*(2), 409–412. doi:10.3758/bf03203962

Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: A time-window-of-integration model. *Journal of Cognitive Neuroscience*, *16*(6), 1000–1009. doi:10.1162/0898929041502733

Colonius, H., & Diederich, A. (2020). Formal models and quantitative measures of multisensory integration: A selective overview. *European Journal of Neuroscience*, *51*(5), 1161–1178. doi:10.1111/ejn.13813

Cowan, N., & Barron, A. (1987). Cross-modal, auditory-visual stroop interference and possible implications for speech memory. *Perception & Psychophysics*, *41*(5), 393–401. doi:10.3758/bf03203031

Eimer, M. (2004). Multisensory integration: How visual experience shapes spatial perception. *Current Biology*, *14*(3), R115–R117. doi:10.1016/j.cub.2004.01.018

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. doi:10.3758/bf03203267

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. doi:10.1038/415429a

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–169. doi:10.1016/j.tics.2004.02.002

Ghahramani, Z., Wolptrt, D. M., & Jordan, M. I. (1997). Computational models of sensorimotor integration. In *Advances in psychology* (pp. 117–147). doi:10.1016/s0166-4115(97)80006-4

Goodale, M. A., & Milner, A. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25. doi:10.1016/0166-2236(92)90344-8

Goupell, M. J., & Stakhovskaya, O. A. (2018). Across-frequency processing of interaural time and level differences in perceived lateralization. *Acta Acustica united with Acustica*, *104*(5), 758–761. doi:10.3813/aaa.919217

Iordanescu, L., Grabowecky, M., Franconeri, S., Theeuwes, J., & Suzuki, S. (2010). Characteristic sounds make you look at target objects more quickly. *Attention, Perception, & Psychophysics*, *72*(7), 1736–1741. doi:10.3758/APP.72.7.1736

Iordanescu, L., Grabowecky, M., & Suzuki, S. (2011). Object-based auditory facilitation of visual search for pictures and words with frequent and rare targets. *Acta Psychologica*, *137*(2), 252–259. doi:10.1016/j.actpsy.2010.07.017

Kalman, R. E., & Bucy, R. S. (1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering*, *83*(1), 95–108. doi:10.1115/1.3658902

Kinchla, R. A. (1974). Detecting target elements in multielement arrays: A confusability model. *Perception & Psychophysics*, *15*(1), 149–158. doi:10.3758/bf03205843

King, A., & Palmer, A. (1985). Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. *Experimental Brain Research*, *60*(3). doi:10.1007/bf00236934

Knudsen, E. I., & Knudsen, P. F. (1985). Vision guides the adjustment of auditory localization in young barn owls. *Science*, *230*(4725), 545–548. Publisher: American Association for the Advancement of Science Section: Reports. doi:10.1126/science.4048948

Knudsen, E. I., & Knudsen, P. F. (1989a). Vision calibrates sound localization in developing barn owls. *Journal of Neuroscience*, *9*(9), 3306–3313. Publisher: Society for Neuroscience Section: Articles. doi:10.1523/JNEUROSCI.09-09-03306.1989

Knudsen, E. I., & Knudsen, P. F. (1989b). Visuomotor adaptation to displacing prisms by adult and baby barn owls. *Journal of Neuroscience*, *9*(9), 3297–3305. Publisher: Society for Neuroscience Section: Articles. doi:10.1523/JNEUROSCI.09-09-03297.1989

Lakatos, P., Chen, C.-M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron*, *53*(2), 279–292. doi:10.1016/j.neuron.2006.12.011

Laurienti, P., Kraft, R., Maldjian, J., Burdette, J., & Wallace, M. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research*, *158*(4). doi:10.1007/s00221-004-1913-2

Lunn, J., Sjoblom, A., Ward, J., Soto-Faraco, S., & Forster, S. (2019). Multisensory enhancement of attention depends on whether you are already paying attention. *Cognition*, *187*, 38–49. doi:10.1016/j.cognition.2019.02.008

Matusz, P. J., Wallace, M. T., & Murray, M. M. (2017). A multisensory perspective on object memory. *Neuropsychologia*, *105*, 243–252. doi:10.1016/j.neuropsychologia.2017.04.008

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. doi:10.1038/264746a0

Meredith, M. A. (2002). On the neuronal basis for multisensory convergence: A brief overview. *Cognitive Brain Research*, *14*(1), 31–40. doi:10.1016/s0926-6410(02)00059-9

Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, *221*(4608), 389–391. doi:10.1126/science.6867718

Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, *56*(3), 640–662. doi:10.1152/jn.1986.56.3.640

Meredith, M. A., & Stein, B. E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. *Journal of Neurophysiology*, *75*(5), 1843–1857. doi:10.1152/jn.1996.75.5.1843

Meyerhoff, H. S., & Huff, M. (2016). Semantic congruency but not temporal synchrony enhances long-term memory performance for audio-visual scenes. *Memory & Cognition*, *44*(3), 390–402. doi:10.3758/s13421-015-0575-6

Meyerhoff, H. S., & Suzuki, S. (2018). Beep, be-, or –ep: The impact of auditory transients on perceived bouncing/streaming. *Journal of Vision*, *18*(10), 1138. doi:10.1167/18.10.1138

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

Mills, A. W. (1958). On the minimum audible angle. *The Journal of the Acoustical Society of America*, *30*(4), 237–246. doi:10.1121/1.1909553

Moore, B. (2013). *An introduction to the psychology of hearing*. Leiden, The Netherlands: Brill.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353–383. doi:10.1016/0010-0285(77)90012-3

Peirce, J. W. (2007). PsychoPy—psychophysics software in python. *Journal of Neuroscience Methods*, *162*(1-2), 8–13. doi:10.1016/j.jneumeth.2006.11.017

Perrott, D. R., & Saberi, K. (1990). Minimum audible angle thresholds for sources varying in both elevation and azimuth. *The Journal of the Acoustical Society of America*, *87*(4), 1728–1731. doi:10.1121/1.399421

Pick, H. L., Warren, D. H., & Hay, J. C. (1969). Sensory conflict in judgments of spatial direction. *Perception & Psychophysics*, *6*(4), 203–205. doi:10.3758/bf03207017

Pirenne, M. H. (1943). Binocular and uniocular threshold of vision. *Nature*, *152*(3867), 698–699. doi:10.1038/152698a0

Quak, M., London, R. E., & Talsma, D. (2015). A multisensory perspective of working memory. *Frontiers in Human Neuroscience*, *9*. doi:10.3389/fnhum.2015.00197

Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, *24*(5 Series II), 574–590. doi:10.1111/j.2164-0947.1962.tb01433.x

Rakerd, B., & Hartmann, W. M. (2010). Localization of sound in rooms. v. binaural coherence and human sensitivity to interaural time differences in noise. *The Journal of the Acoustical Society of America*, *128*(5), 3052–3063. doi:10.1121/1.3493447

Rock, I., & Victor, J. (1964). Vision and touch: An experimentally created conflict between the two senses. *Science*, *143*(3606), 594–596. doi:10.1126/science.143.3606.594

Schroeder, C. E., & Foxe, J. (2005). Multisensory contributions to low-level, 'unisensory' processing. *Current Opinion in Neurobiology*, *15*(4), 454–458. doi:10.1016/j.conb.2005.06.008

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, *385*(6614), 308–308. doi:10.1038/385308a0

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

Stanford, T. R., & Stein, B. E. (2007). Superadditivity in multisensory integration: Putting the computation in context. *NeuroReport*, *18*(8), 787–792. doi:10.1097/wnr.0b013e3280c1e315

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. doi:10.3758/bf03207704

Stein, B. E. (2012). *The new handbook of multisensory processing*. Cambridge, MA: MIT Press.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*(4), 255–266. doi:10.1038/nrn2331

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

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. doi:10.1037/h0054651

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*(9), 400–410. doi:10.1016/j.tics.2010.06.008

Todd, J. W. (1912). *Reaction to multiple stimuli.* New York: The Science Press.

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(5), 1053–1065. doi:10.1037/0096-1523.34.5.1053

Van Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual*. Scotts Valley, CA: CreateSpace.

Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Perception & Psychophysics*, *69*(5), 744–756. doi:10.3758/bf03193776

Wallace, M. T., Wilkinson, L. K., & Stein, B. E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of Neurophysiology*, *76*(2), 1246–1266. doi:10.1152/jn.1996.76.2.1246

Warren, D. H., Welch, R. B., & McCarthy, T. J. (1981). The role of visual-auditory "compellingness" in the ventriloquism effect: Implications for transitivity among the spatial senses. *Perception & Psychophysics*, *30*(6), 557–564. doi:10.3758/bf03202010

Watkins, S., Shams, L., Josephs, O., & Rees, G. (2007). Activity in human v1 follows multisensory perception. *NeuroImage*, *37*(2), 572–578. doi:10.1016/j.neuroimage.2007.05.027

Watkins, S., Shams, L., Tanaka, S., Haynes, J.-D., & Rees, G. (2006). Sound alters activity in human v1 in association with illusory visual perception. *NeuroImage*, *31*(3), 1247–1256. doi:10.1016/j.neuroimage.2006.01.016

Werner, S., & Noppeney, U. (2009). Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cerebral Cortex*, *20*(8), 1829–1842. doi:10.1093/cercor/bhp248

White, T. L., & Prescott, J. (2007). Chemosensory cross-modal stroop effects: Congruent odors facilitate taste identification. *Chemical Senses*, *32*(4), 337–341. doi:10.1093/chemse/bjm001

Wickens, T. D. (2001). *Elementary signal detection theory*. doi:10.1093/acprof:oso/9780195092509.001.0001

Wightman, F. L., & Kistler, D. J. (1989). Headphone simulation of free-field listening. i: Stimulus synthesis. *The Journal of the Acoustical Society of America*, *85*(2), 858–867. doi:10.1121/1.397557

Wightman, F. L., & Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America*, *91*(3), 1648–1661. doi:10.1121/1.402445

Wilson, J. R., & Sherman, S. M. (1976). Receptive-field characteristics of neurons in cat striate cortex: Changes with visual field eccentricity. *Journal of Neurophysiology*, *39*(3), 512–533. doi:10.1152/jn.1976.39.3.512

Wolpert, D., Ghahramani, Z., & Jordan, M. (1995). An internal model for sensorimotor integration. *Science*, *269*(5232), 1880–1882. doi:10.1126/science.7569931

Appendix

Additional Tables and Figures

This section of the appendix contains additional figures and tables that were not relevant in the scope of this thesis, but might provide further insights for the interested reader.

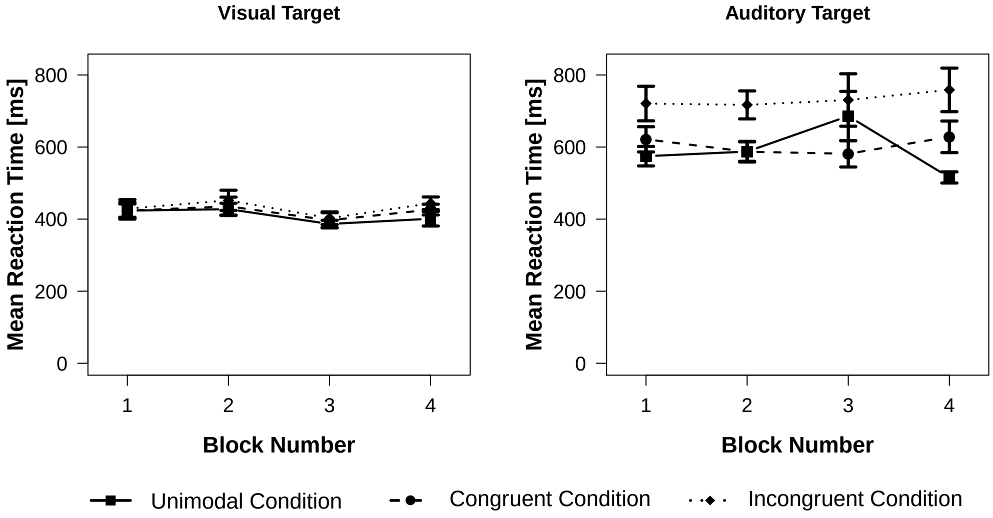


Figure 1: Average reaction times for all six conditions of the second experiment per block. The error bars indicate one standard error of all observations at this block number. The *left plot* shows the visual target conditions, whereas the *right plot* shows the auditory target conditions. Each condition occurred six times at each block position. No notable difference between the block positions, and, thus, no effects of ordering, was observable.

Table 1: Average incorrect response frequencies per distractor position for the incongruent condition with auditory target of the second experiment. Results are averaged over participants ± the standard error of the mean. Also across participants there seems to be a trend to respond to the distractor's position if the response is incorrect as indicated by the higher values on the diagonal compared to the rest (see also Table 2).

	Visual Distractor Position			
Response	far left	left	right	far right
far left	2.13 ± .36	0.06 ± .06	0.25 ± .12	0.46 ± .16
left	0.63 ± .16	2.92 ± .48	0.75 ± .24	0.58 ± .19
right	0.62 ± .30	1.25 ± .28	2.75 ± .52	0.25 ± .11
far right	0.42 ± .18	0.25 ± .12	0.00 ± .00	2.33 ± .41

Table 2: Response frequencies for all incorrectly responded trials of the incongruent conditions as a function of distractor position and target modality of the second experiment. Results are pooled over all participants. Overall, participants responded more correctly when the target was visual as observable in the lower total amount of observations (see upper table). For auditory targets, participants responded more often with the distractor position than other locations (see lower table).

	Auditory Distractor Position (Visual Target)			
Response	far left	left	right	far right
far left	9	0	7	5
left	0	4	6	3
right	4	4	5	3
far right	7	4	0	7

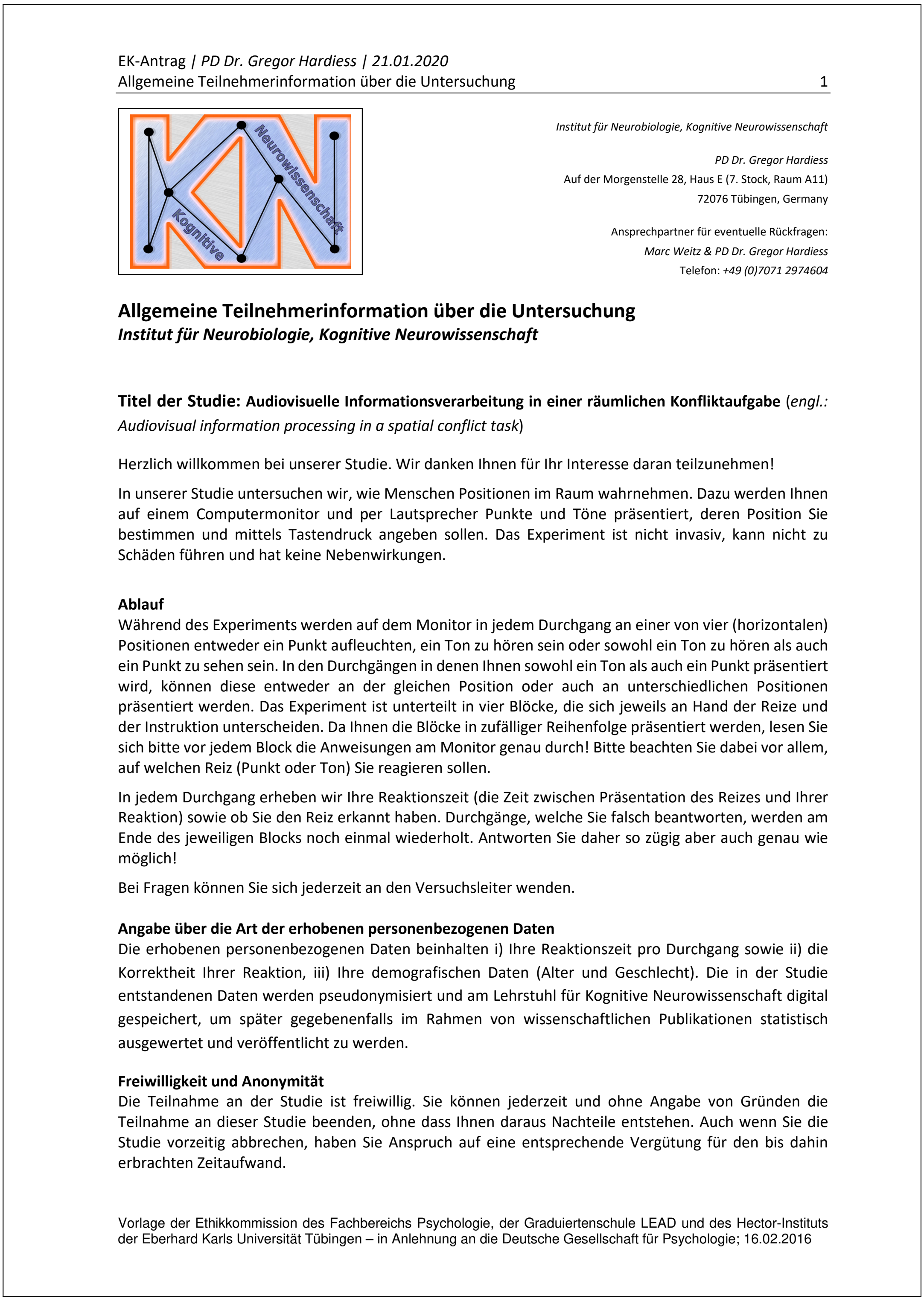
	Visual Distractor Position (Auditory Target)			
Response	far left	left	right	far right
far left	51	2	6	11
left	15	70	18	14
right	15	30	66	6
far right	10	6	0	56

Auditory Stimulus Creation

This section of the appendix contains the code used to create the auditory stimuli.

```
from scipy.io.wavfile import write
import numpy as np


def panner(mono, angle):
    angle = np.radians(angle)

    left = np.sqrt(2)/2.0 * (np.cos(angle)
           - np.sin(angle)) * mono
    right = np.sqrt(2)/2.0 * (np.cos(angle)
            + np.sin(angle)) * mono

    stereo = np.dstack((left, right))[0]

    return stereo


SAMPLERATE = 44100
FREQUENCY = 780
DURATION = 10
POSITIONS = [(-70, "far_left"), (-20, "left"),
             (20, "right"), (70, "far_right")]

TIME_POINTS = np.linspace(0.,
                          1.*DURATION,
                          SAMPLERATE*DURATION)
MONO = np.sin(2. * np.pi * FREQUENCY * TIME_POINTS)

for deg, position in POSITIONS:
    signal = panner(MONO, deg)
    write("tone_{}_{}.wav".format(FREQUENCY, position),
          SAMPLERATE, signal)
```

This code has been heavily influenced by a post on stackexchange (https://dsp.stackexchange.com/questions/21691/algorithm-to-pan-audio/21736#21736)



Institut für Neurobiologie, Kognitive Neurowissenschaft

PD Dr. Gregor Hardiess
Auf der Morgenstelle 28, Haus E (7. Stock, Raum A11)
72076 Tübingen, Germany

Ansprechpartner für eventuelle Rückfragen:
Marc Weitz & PD Dr. Gregor Hardiess
Telefon: *+49 (0)7071 2974604*

Allgemeine Teilnehmerinformation über die Untersuchung

Institut für Neurobiologie, Kognitive Neurowissenschaft

Titel der Studie: Audiovisuelle Informationsverarbeitung in einer räumlichen Konfliktaufgabe (*engl.: Audiovisual information processing in a spatial conflict task*)

Herzlich willkommen bei unserer Studie. Wir danken Ihnen für Ihr Interesse daran teilzunehmen!

In unserer Studie untersuchen wir, wie Menschen Positionen im Raum wahrnehmen. Dazu werden Ihnen auf einem Computermonitor und per Lautsprecher Punkte und Töne präsentiert, deren Position Sie bestimmen und mittels Tastendruck angeben sollen. Das Experiment ist nicht invasiv, kann nicht zu Schäden führen und hat keine Nebenwirkungen.

Ablauf

Während des Experiments werden auf dem Monitor in jedem Durchgang an einer von vier (horizontalen) Positionen entweder ein Punkt aufleuchten, ein Ton zu hören sein oder sowohl ein Ton zu hören als auch ein Punkt zu sehen sein. In den Durchgängen in denen Ihnen sowohl ein Ton als auch ein Punkt präsentiert wird, können diese entweder an der gleichen Position oder auch an unterschiedlichen Positionen präsentiert werden. Das Experiment ist unterteilt in vier Blöcke, die sich jeweils an Hand der Reize und der Instruktion unterscheiden. Da Ihnen die Blöcke in zufälliger Reihenfolge präsentiert werden, lesen Sie sich bitte vor jedem Block die Anweisungen am Monitor genau durch! Bitte beachten Sie dabei vor allem, auf welchen Reiz (Punkt oder Ton) Sie reagieren sollen.

In jedem Durchgang erheben wir Ihre Reaktionszeit (die Zeit zwischen Präsentation des Reizes und Ihrer Reaktion) sowie ob Sie den Reiz erkannt haben. Durchgänge, welche Sie falsch beantworten, werden am Ende des jeweiligen Blocks noch einmal wiederholt. Antworten Sie daher so zügig aber auch genau wie möglich!

Bei Fragen können Sie sich jederzeit an den Versuchsleiter wenden.

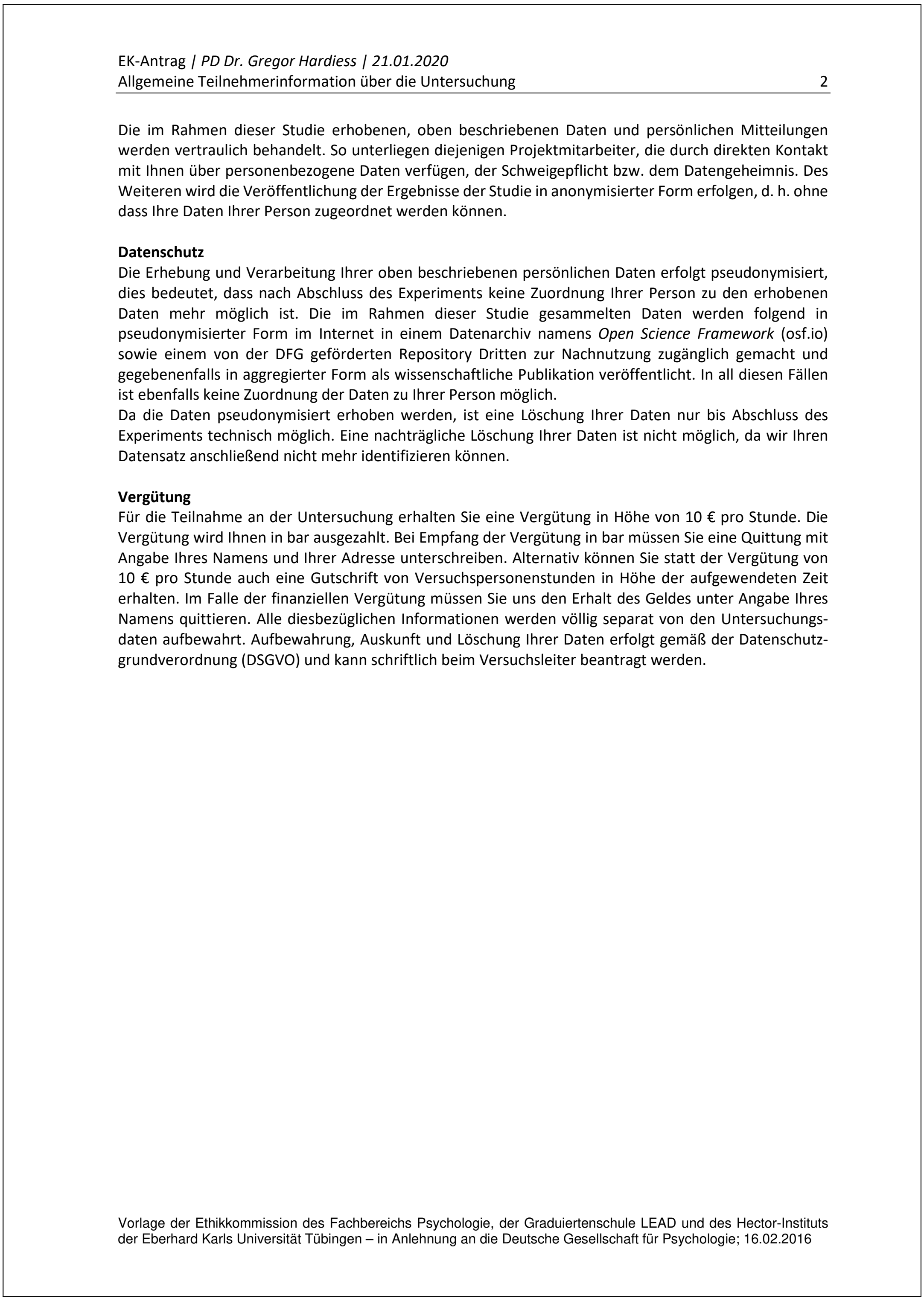
Angabe über die Art der erhobenen personenbezogenen Daten

Die erhobenen personenbezogenen Daten beinhalten i) Ihre Reaktionszeit pro Durchgang sowie ii) die Korrektheit Ihrer Reaktion, iii) Ihre demografischen Daten (Alter und Geschlecht). Die in der Studie entstandenen Daten werden pseudonymisiert und am Lehrstuhl für Kognitive Neurowissenschaft digital gespeichert, um später gegebenenfalls im Rahmen von wissenschaftlichen Publikationen statistisch ausgewertet und veröffentlicht zu werden.

Freiwilligkeit und Anonymität

Die Teilnahme an der Studie ist freiwillig. Sie können jederzeit und ohne Angabe von Gründen die Teilnahme an dieser Studie beenden, ohne dass Ihnen daraus Nachteile entstehen. Auch wenn Sie die Studie vorzeitig abbrechen, haben Sie Anspruch auf eine entsprechende Vergütung für den bis dahin erbrachten Zeitaufwand.

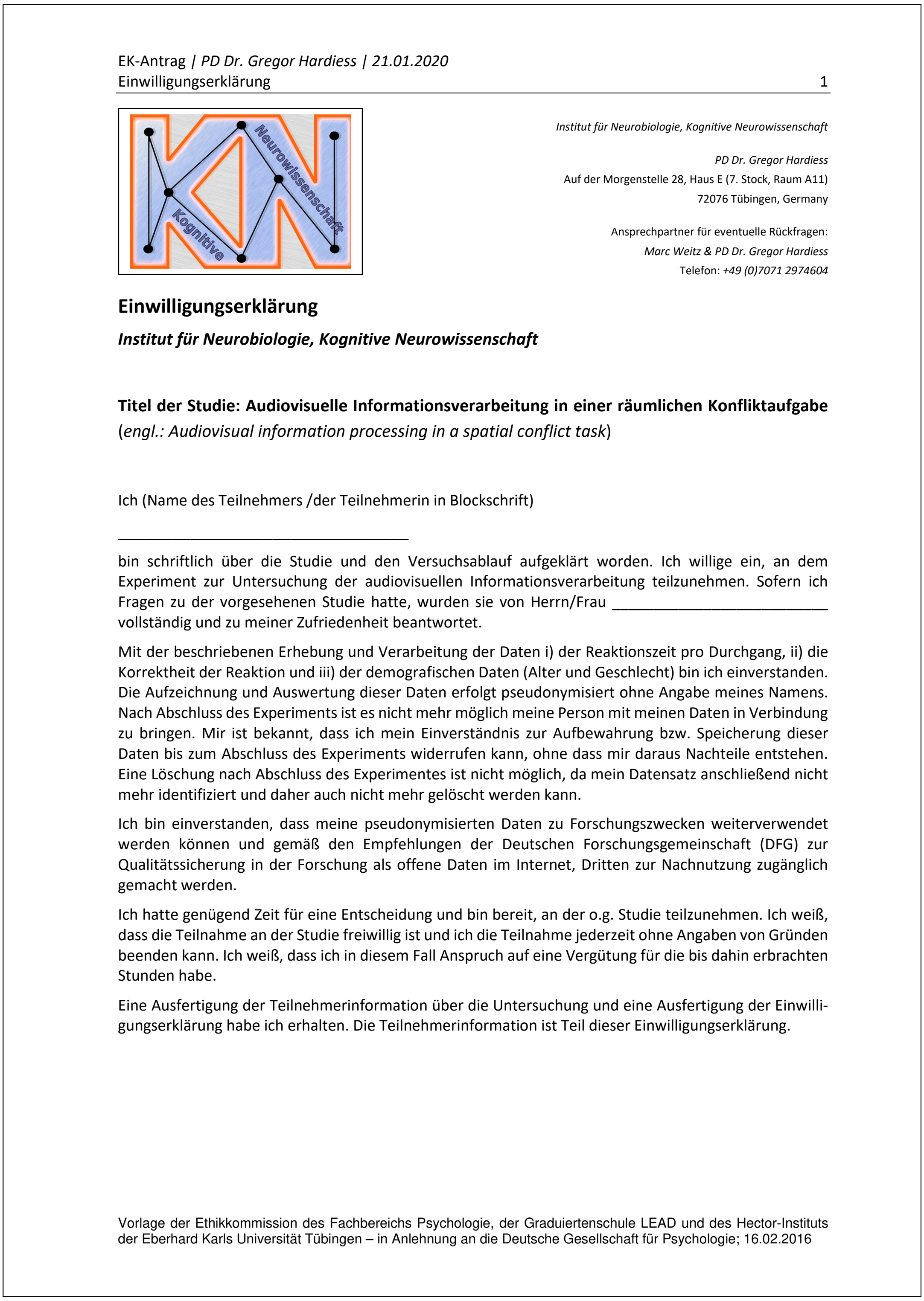
Vorlage der Ethikkommission des Fachbereichs Psychologie, der Graduiertenschule LEAD und des Hector-Instituts der Eberhard Karls Universität Tübingen – in Anlehnung an die Deutsche Gesellschaft für Psychologie; 16.02.2016



Die im Rahmen dieser Studie erhobenen, oben beschriebenen Daten und persönlichen Mitteilungen werden vertraulich behandelt. So unterliegen diejenigen Projektmitarbeiter, die durch direkten Kontakt mit Ihnen über personenbezogene Daten verfügen, der Schweigepflicht bzw. dem Datengeheimnis. Des Weiteren wird die Veröffentlichung der Ergebnisse der Studie in anonymisierter Form erfolgen, d. h. ohne dass Ihre Daten Ihrer Person zugeordnet werden können.

Datenschutz
Die Erhebung und Verarbeitung Ihrer oben beschriebenen persönlichen Daten erfolgt pseudonymisiert, dies bedeutet, dass nach Abschluss des Experiments keine Zuordnung Ihrer Person zu den erhobenen Daten mehr möglich ist. Die im Rahmen dieser Studie gesammelten Daten werden folgend in pseudonymisierter Form im Internet in einem Datenarchiv namens *Open Science Framework* (osf.io) sowie einem von der DFG geförderten Repository Dritten zur Nachnutzung zugänglich gemacht und gegebenenfalls in aggregierter Form als wissenschaftliche Publikation veröffentlicht. In all diesen Fällen ist ebenfalls keine Zuordnung der Daten zu Ihrer Person möglich.
Da die Daten pseudonymisiert erhoben werden, ist eine Löschung Ihrer Daten nur bis Abschluss des Experiments technisch möglich. Eine nachträgliche Löschung Ihrer Daten ist nicht möglich, da wir Ihren Datensatz anschließend nicht mehr identifizieren können.

Vergütung
Für die Teilnahme an der Untersuchung erhalten Sie eine Vergütung in Höhe von 10 € pro Stunde. Die Vergütung wird Ihnen in bar ausgezahlt. Bei Empfang der Vergütung in bar müssen Sie eine Quittung mit Angabe Ihres Namens und Ihrer Adresse unterschreiben. Alternativ können Sie statt der Vergütung von 10 € pro Stunde auch eine Gutschrift von Versuchspersonenstunden in Höhe der aufgewendeten Zeit erhalten. Im Falle der finanziellen Vergütung müssen Sie uns den Erhalt des Geldes unter Angabe Ihres Namens quittieren. Alle diesbezüglichen Informationen werden völlig separat von den Untersuchungsdaten aufbewahrt. Aufbewahrung, Auskunft und Löschung Ihrer Daten erfolgt gemäß der Datenschutzgrundverordnung (DSGVO) und kann schriftlich beim Versuchsleiter beantragt werden.



Institut für Neurobiologie, Kognitive Neurowissenschaft

PD Dr. Gregor Hardiess
Auf der Morgenstelle 28, Haus E (7. Stock, Raum A11)
72076 Tübingen, Germany

Ansprechpartner für eventuelle Rückfragen:
Marc Weitz & PD Dr. Gregor Hardiess
Telefon: *+49 (0)7071 2974604*

Einwilligungserklärung

Institut für Neurobiologie, Kognitive Neurowissenschaft

Titel der Studie: Audiovisuelle Informationsverarbeitung in einer räumlichen Konfliktaufgabe
(*engl.: Audiovisual information processing in a spatial conflict task*)

Ich (Name des Teilnehmers /der Teilnehmerin in Blockschrift)

____

bin schriftlich über die Studie und den Versuchsablauf aufgeklärt worden. Ich willige ein, an dem Experiment zur Untersuchung der audiovisuellen Informationsverarbeitung teilzunehmen. Sofern ich Fragen zu der vorgesehenen Studie hatte, wurden sie von Herrn/Frau ___________________________ vollständig und zu meiner Zufriedenheit beantwortet.

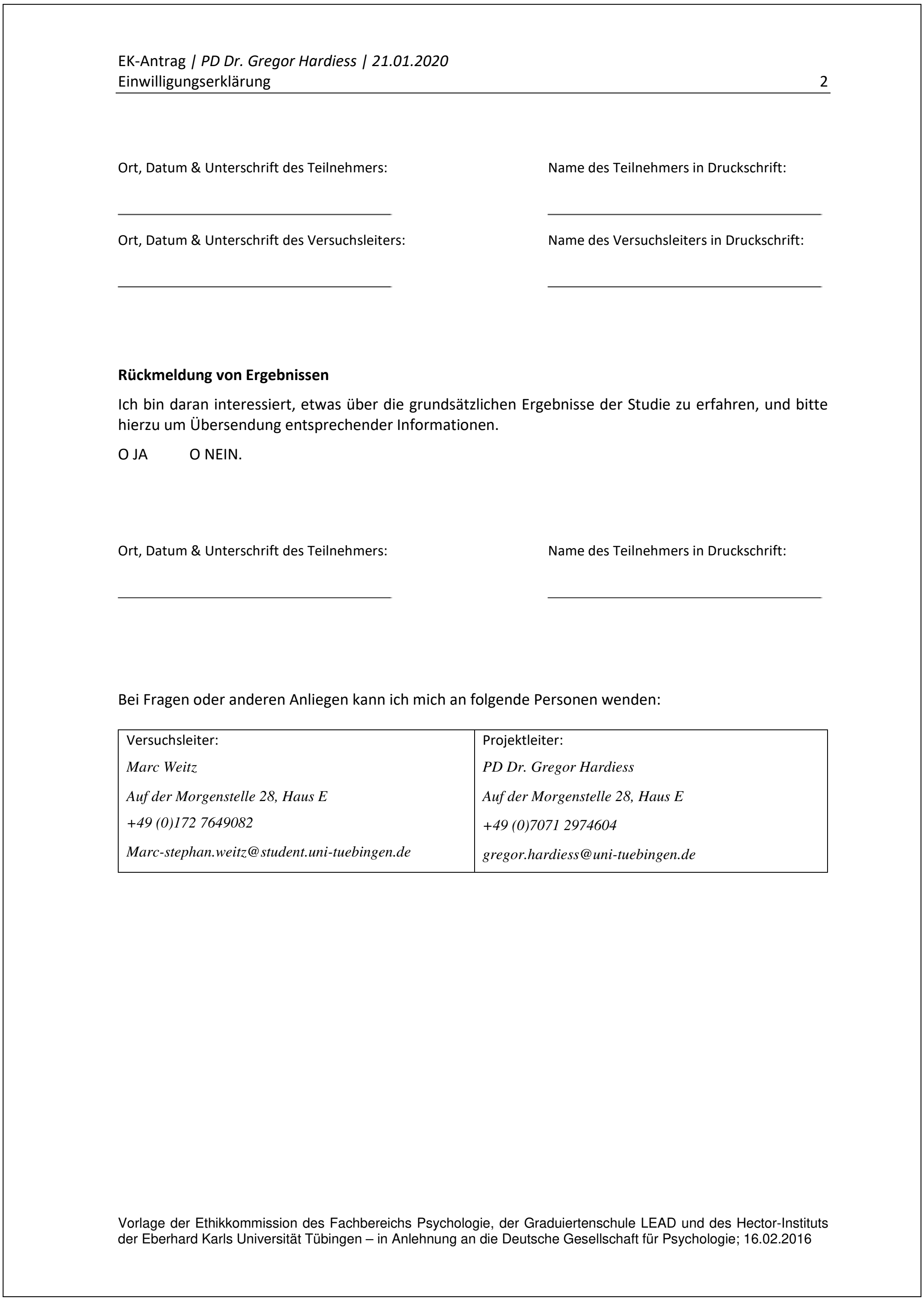
Mit der beschriebenen Erhebung und Verarbeitung der Daten i) der Reaktionszeit pro Durchgang, ii) die Korrektheit der Reaktion und iii) der demografischen Daten (Alter und Geschlecht) bin ich einverstanden. Die Aufzeichnung und Auswertung dieser Daten erfolgt pseudonymisiert ohne Angabe meines Namens. Nach Abschluss des Experiments ist es nicht mehr möglich meine Person mit meinen Daten in Verbindung zu bringen. Mir ist bekannt, dass ich mein Einverständnis zur Aufbewahrung bzw. Speicherung dieser Daten bis zum Abschluss des Experiments widerrufen kann, ohne dass mir daraus Nachteile entstehen. Eine Löschung nach Abschluss des Experimentes ist nicht möglich, da mein Datensatz anschließend nicht mehr identifiziert und daher auch nicht mehr gelöscht werden kann.

Ich bin einverstanden, dass meine pseudonymisierten Daten zu Forschungszwecken weiterverwendet werden können und gemäß den Empfehlungen der Deutschen Forschungsgemeinschaft (DFG) zur Qualitätssicherung in der Forschung als offene Daten im Internet, Dritten zur Nachnutzung zugänglich gemacht werden.

Ich hatte genügend Zeit für eine Entscheidung und bin bereit, an der o.g. Studie teilzunehmen. Ich weiß, dass die Teilnahme an der Studie freiwillig ist und ich die Teilnahme jederzeit ohne Angaben von Gründen beenden kann. Ich weiß, dass ich in diesem Fall Anspruch auf eine Vergütung für die bis dahin erbrachten Stunden habe.

Eine Ausfertigung der Teilnehmerinformation über die Untersuchung und eine Ausfertigung der Einwilligungserklärung habe ich erhalten. Die Teilnehmerinformation ist Teil dieser Einwilligungserklärung.

Vorlage der Ethikkommission des Fachbereichs Psychologie, der Graduiertenschule LEAD und des Hector-Instituts der Eberhard Karls Universität Tübingen – in Anlehnung an die Deutsche Gesellschaft für Psychologie; 16.02.2016



EK-Antrag | *PD Dr. Gregor Hardiess | 21.01.2020*
Einwilligungserklärung 2

Ort, Datum & Unterschrift des Teilnehmers: Name des Teilnehmers in Druckschrift:

______________________________ ______________________________

Ort, Datum & Unterschrift des Versuchsleiters: Name des Versuchsleiters in Druckschrift:

______________________________ ______________________________

Rückmeldung von Ergebnissen

Ich bin daran interessiert, etwas über die grundsätzlichen Ergebnisse der Studie zu erfahren, und bitte hierzu um Übersendung entsprechender Informationen.

O JA O NEIN.

Ort, Datum & Unterschrift des Teilnehmers: Name des Teilnehmers in Druckschrift:

______________________________ ______________________________

Bei Fragen oder anderen Anliegen kann ich mich an folgende Personen wenden:

Versuchsleiter:	Projektleiter:
Marc Weitz	*PD Dr. Gregor Hardiess*
Auf der Morgenstelle 28, Haus E	*Auf der Morgenstelle 28, Haus E*
+49 (0)172 7649082	*+49 (0)7071 2974604*
Marc-stephan.weitz@student.uni-tuebingen.de	*gregor.hardiess@uni-tuebingen.de*

Vorlage der Ethikkommission des Fachbereichs Psychologie, der Graduiertenschule LEAD und des Hector-Instituts der Eberhard Karls Universität Tübingen – in Anlehnung an die Deutsche Gesellschaft für Psychologie; 16.02.2016

Statement of authorship

I hereby certify that this master thesis has been composed by myself, and describes my own work, unless otherwise acknowledged in the text. All references and verbatim extracts have been quoted, and all sources of information have been specifically acknowledged. It has not been accepted in any previous application for a degree. The digital copy attached to this thesis is the same as the printed thesis.

Tübingen, April 29, 2020 Marc Weitz