MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„The Influence of Continuous, Lateralized Auditory Stimulation on Visuospatial Attention"

verfasst von / submitted by

Rebecca Rosa Schmid, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2019 / Vienna 2019

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 840
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Psychologie
Betreut von / Supervisor:	Univ.-Prof. Dr. Ulrich Ansorge

Acknowledgements

I want to thank my supervisor Prof. Dr. Ulrich Ansorge for his kind and profound support toward completing my thesis in view of the short amount of time. Many sincere thanks I owe to Dr. Ulrich Pomper for letting me participate in his project(s), for his support and advice at any time during the whole process of completing the master thesis.

For proof-reading in English and countless moments of fun during my master studies in and outside the University, I would also thank my fellow student and friend Christina Krumpholz alias Gabi. To my family I owe a debt of gratitude for many years of support in difficult situations, even though they did not always agree on my decisions. Finally, I would like to express my gratitude to my fiancée Pascal for standing behind my back with patience and love during the last seven years.

Rebecca Schmid, Juli 2019

Table of Contents

The Influence of Continuous, Lateralized Auditory Stimulation on Visuospatial Attention

Attention has been – and is still – one of the most intensely studied topics within the field of psychology and cognitive neurosciences. Although our senses constantly receive a vast amount of sensory input, our brain does not have the capacity to process all of them (Marois & Ivanoff, 2005). Nor would it be efficient in terms of behavior to do so. Thus, to avoid sensory overload, attention functions as a cognitive mechanism of *selectively processing* a *limited amount* of incoming sensory information from our environment, while other information is neglected (Broadbent, 1958; Deutsch & Deutsch, 1963; Duncan, 1984; Treisman, 1986). Attended stimuli are processed more thoroughly and more rapidly than non-attended ones, or even potentially distracting stimuli (Posner, 1978; Spence & Parise, 2010).

At the level of selection, a further distinction must be made between its different modi operandi, because attention can be administered in a spatial or in a non-spatial manner (Duncan, 1984; Egly, Driver, & Rafal, 1994; Serences, Schwarzbach, Courtney, Golay, & Yantis, 2004). Thus, it can be deployed to a particular location in space, analogously to a *beam of light*. This metaphor goes back to Posner, Snyder and Davidson (1980), who compared (visual) selective attention to a “spotlight that enhances the efficiency of the detection of events within its beam” (p. 172). Moreover, it is also possible to direct attention to a perceptual object representation as a unit, or just its particular feature properties, such as color, shape or even motion. That is of importance in our everyday life, because we often know what we are looking for but not exactly where it is. For example, when somebody cannot find his or her car on a parking deck and is scanning the rows for a Coupé in grey color.

In the abovementioned case, attention is voluntarily shifted from one location of an object to another, implying that the observer is controlling the focus of attention in a so called top-down manner towards a goal (locating the car). This is often referred to as *endogenous* attention. By contrast, however, a sudden onset of an unexpected event in the environment may capture attention *exogenously* or stimulus-driven, indicating a reflexive, involuntary shift towards the location in which the event occurs (Spence, 2010). For instance, the sudden appearance of a person, previously hidden by a tree, could draw the observer’s attention in a bottom-up fashion while he or she is looking out for the own car. Despite intensive research over the past decades, it is still controversial to what extent bottom-up and top-down attention processes are involved in various tasks (Connor, Egeth, & Yantis, 2004; Gaspelin, Ruthruf, & Lien, 2016).

Many literature on attention research is settled in the visual domain (Desimone & Duncan, 1995; Knudsen, 2007; Shinn-Cunningham, 2008), most likely because vision is regarded as our dominant sense, and there is indeed much evidence that humans preferably direct their attentional resources towards their visual modality (Bertelson & Aschersleben, 1998; Botvinick & Cohen, 1998, Colavita, 1974; Posner, Nissen, & Klein, 1976;).

However, similar principles govern perception in other modalities as well (Busse, Roberts, Crist, Weissmann, & Woldorff, 2005; Marinato & Baldauf, 2019; Serences, Shomstein, Leber, Golay, Egeth, & Yantis, 2005; Shomstein &Yantis, 2006). For instance, evidence suggests that attention mechanisms act on auditory objects in a similar way as on visual objects (Alain & Arnott, 2000; Best, Ozmeral, & Shinn-Cunningham, 2007; Scholl 2001). Consistent with the biased-competition model of visual attention (Desimone, 1999), listeners are apparently only able to attend actively to one object at a time in the auditory modality as well (Cusack, Carlyon, & Robertson, 2000; Vogel & Luck, 2002). To stay in line with top-down and bottom-up attention control processes, we can applicate a concurrent example within the auditory modality: In the well-known *cocktail party effect* (Cherry, 1953), listeners orient their attention endogenously to a certain person speaking to them at a noisy cocktail party, while ignoring all the other voices around them. Given the sudden onset of an unexpected event, such as another person mentioning our name in a different conversation, our attention is captured reflexively by exogenous orienting towards that person's voice.

However, much of the past research on attention has focused on one sensory modality at a time. This strategy was backed by the idea that attention has traditionally been considered as a selection process among the competing inputs within just a single modality (Driver, 2001). But over the last few decades, there has been a growing interest in the existence and nature of *multisensory attention processes*, involving the parallel processing of one or more modality-specific stimuli in different sensory modalities (Alais, Newell, & Mamassian, 2010; Keil & Senkowski, 2018). This new approach also has its merits as it is a more realistic representation of our daily life. Just imagine a car trip where the driver is concentrating on what is happening on the road, while operating the vehicle and also having an ear to what his or her co-driver is telling him or her *simultaneously*. Thereto, a fundamental question arose about the true nature of internal processes concerning the ability to divide attentional resources between tasks in different modalities. Despite several decades of extensive research, it remains still not entirely clarified if – or, to what extent – these processes act in a supramodal or in a modality-specific manner. In other words, are attentional resources independent of the sensory modality and allocated according to current tasks and goals, or does each sensory modality have its own exclusive resources (Spence, 2010; Wahn & König, 2017)?

There is evidence that dividing attention between tasks presented in different sensory modalities is easier than between tasks presented within the same modality (Duncan, Martens, & Ward, 1997; Hancock, Oron-Gilad, & Szalma, 2007; Lavie, 2005; Sarter, 2007; Wickens, 2008), which supports the modality-specific view. According to this, there should be no reduction in attention in the previous example about the car trip, which means having a conversation should not impact the detection of critical road traffic events by the driver, because two separate and independent attention mechanisms control the processing of the visual and auditory information. However, some behavioral and even neurophysiological data have challenged the solely modality-specific view of the allocation of attentional resources by demonstrating the existence of *cross-modal links in spatial attention* in a large number of studies (Driver & Spence, 1998; Eimer & Schröger, 1998; Teder-Sälejärvi, Münte, Sperlich, & Hillyard, 1999). A frequently observed cross-modal effect is that divided attention to concurrent events in two modalities is generally more effective, when those stimuli are spatially close to each other. In other words, when a stimulus is presented in one modality, it can facilitate the detectability or the speed of responding to a subsequent, spatially coincident stimulus in another modality (Eimer, 1999; Spence & Driver, 1996, 1997; McDonald, Teder-Sälejärvi, Russo, & Hillyard, 2003). Thus, the definition of cross-modal interaction includes any process involving the impact from one sensory modality on the perception or processing of another sensory modality. This impact could therefore both enhance and limit attentional processing in a dual-task situation as described above (Keil & Senkowski, 2018). Such effects have been demonstrated in various combinations of visual, auditory, olfactory, gustatory, tactile, and even painful stimuli until today (Calvert, Spence, & Stein, 2004). Regarding the aim of the present study, we will further focus on interactions between vision and audition only.

A relevant study on this topic, which is worth mentioning here, was conducted by Spence and Driver (1997). Spatially uninformative cues in form of peripherally presented abrupt sounds or flashes were shown prior to visual or auditory targets that required a location discrimination. Brief sounds attracted visual attention, reflected by faster and more accurate visual discriminations on the cued side. Surprisingly, abrupt flashes failed to attract auditory attention, whereby Spence and Driver concluded an asymmetrical link between the visual and the auditory modality with respect to exogenous spatial attention. Spence and Driver (1996) also found links between endogenous auditory and visuospatial attention in an earlier experiment, where a centered arrow cue indicated the likely location of target stimuli in one modality. Target stimuli in the other modality were presented less frequently (about 25% of all trials) and were more likely to be shown at the uncued side. When the cue indicated the location of auditory targets, visual discrimination was faster when visual targets were presented at cued location. Likewise, when the cue indicated the location of visual targets, auditory discrimination was faster at cued than at uncued locations. Overall, these results lead to the conclusion, that endogenous spatial attention, unlike exogenous spatial attention, follows a symmetrical link between the visual and auditory modality.

In previous research, as exemplified in the two studies described above, attention is usually manipulated via brief, abrupt onset stimuli, resulting in a change of the processing of immediately following targets. Furthermore, if attention is induced exogenously or endogenously was most often realized by whether the preceding cue has an informative value for the prediction of the target locus (Spence & Driver, 1997). Let us think about the exemplary driving situation above again, where a driver must pay attention on what is happening on the road while listening to his or her co-driver's words. First, voluntary listening to what the co-driver is saying does not necessarily have to be related to the detection of visual events on the road. Second, attention has to be spatially oriented towards a *continuous* auditory stream towards the co-driver's voice, rather shortly presented sounds. The question now arises, whether a voluntary orientation towards continuous, lateralized auditory stimulation could also have a cross-modal effect on visuospatial attention in dual-task scenarios?

In fact, considerably less published research can be found here. Driver and Spence (1994) first conducted a study in which participants were instructed to shadow one of two streams of speech, from an equidistant location either in their right or left hemifield, while simultaneously monitoring a quickly changing stream of unrelated visual stimuli for a target stimulus (also called RSVP task). In the 'same-side condition', the visual stream was presented on the same side as the auditory stream, whereas in the 'opposite-sides condition' the attended auditory and visual streams where presented in different hemifields. As a result, participants performed significantly worse in the speech-shadowing task, when visual and auditory sources were presented from a different rather than the same spatial location (Experiment 2). The finding that people tend to divide their attention between audition and vision more easily when information in both modalities are presented spatially close to each other leads to the conclusion that also in dual-task situations cross-modal spatial processes influence task performance.

However, this experimental design does not consider how a continuous, lateralized auditory stream could affect *covert* visuospatial attention in a real dual-task situation like in the former mentioned example, in which the driver would not merely concentrate on occurring traffic events in either the left or the right visual hemisphere only. Thus, Spence and Read (2003) extended these results some years later in a more realistic scenario involving a car-driving simulator. Participants had to shadow one of two speech-sounds presented from loudspeakers that were located either to their side in one condition (as with a co-driver's talking next to them) or in front of them in another condition, while sitting in a static car driving along busy roads on a screen. Results showed better performance in the auditory task when participants had to shadow the loudspeakers in front of them rather than from the side. Once again, this shows the benefits of spatially congruent visual and auditory sources of information, even when the streams from different modalities were unrelated. What is more, that Spence and Read failed to demonstrate a significant cross-modal bias on driving performance, no matter of which auditory stream participants had to shadow (and even not in their follow-up study, where the driving task was more demanding). They concluded, that their results therefore reflect the fact, that participants would prioritize the driving task over the shadowing task, because the consequences of making a mistake while driving would be more far-reaching.

As a consequence, we asked ourselves, if we could find a cross-modal bias from endogenously oriented audiospatial attention towards a continuous, lateralized auditory stimulation on visuospatial attention in a more systematic and controlled experimental setting. Since neglecting performance in one task over the other should not lead to more serious consequences than in the more realistic dual-task situation by Spence and Read (2003), no task should be prioritized for reasons of survival. Previous studies also found that spatial cross-modal effects become even more pronounced when task difficulty or perceptual load increases (Driver & Spence, 1994; Spence, Ranson, & Driver, 2000). Thus, we additionally wanted to examine if a variation of semantic complexity of auditory stimulation could also affect this bias in a similar way. To our knowledge, no previous studies on cross-modal effects have investigated this research question so far.

We set up a dual-task paradigm within the visual and the auditory modality. Participants had to perform a speeded visual discrimination task, in which they had to report the orientation of peripherally presented white triangles. In parallel, they had to perform an auditory task, in which a continuous stream of acoustic signals was presented bilaterally via loudspeakers. Participants always had to actively listen to just one of them at a time. In the first condition, the content of auditory stimulation was of low semantic complexity (LSC), implemented as random sequences of tonepips with omissions of sound as targets. In the second condition, the content of auditory stimulation was of high semantic complexity (HSC), implemented as two short stories with subsequent comprehension questionnaire.

We expected to observe a cross-modal spatial congruity effect, which means faster and more accurate responses in congruent trials (i.e., visual targets presented on the side currently attended within the auditory modality). Furthermore, if this bias is affected by the semantic complexity of auditory stimulation, we expected to find a significant interaction between the conditions of different auditory semantic complexity (i.e., LSC and HSC) and the spatial congruity manipulation.

Eye tracking was used for eye fixation control purposes only. Auditory task performance was also evaluated, but only for controlling whether the participants were actually attending the relevant auditory stream according to the instruction.

Method

Participants

An a-priori power analysis with G*Power (version 3.1.9.4; Faul, Erdfelder, Lang, & Buchner, 2007) suggested a total sample size of 20 participants, $f = .35$, $\alpha = .05$, $1 - \beta = .80$. Twenty-two undergraduate Psychology students (16 female, $M_{age} = 23.36$ years, $SD_{age} = 3.33$ years) from the University of Vienna took part in the experiment. They received course credit for their participation. Their vision was normal or corrected to normal. German was either mother tongue or at least C1 level. Participants were treated in line with established ethical standards: Prior to testing, they were thoroughly instructed and gave their informed consent in written form, knowing that they could abort the experiment at any time without negative consequences. Their well-being was closely monitored throughout the testing. At the end, all participants were debriefed about the theoretical background, study design and hypotheses both in orally and written form. Data from three participants were excluded from further analysis, because they performed poorly on the auditory task in the LSC condition (i.e., exceeded the criterion in more than half of the blocks). A potential reason for this could be that these participants did not correctly remember the instruction to which loudspeaker they should pay attention to, or they had a more basic problem differentiating the spatial location of relevant auditory targets between both sides. Finally, data from 19 participants remained for further analysis.

Apparatus and Stimuli

The experiment was programmed and executed in MATLAB (version 8.3) using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) and the Eyelink toolbox (Cornelissen, Peters, & Palmer, 2002) running on a desktop computer with Windows 7 Professional (Microsoft, Inc.). Visual stimuli consisted of white triangles (CIE L*a*b*, 17.5/22.6/-29.1) which participants viewed at an eccentricity of 6.1° either left or right of a central fixation cross (0.4° × 0.4°). The triangles had an initial width of 0.7° and height of 0.4°. Depending on the individual performance, task difficulty was adapted after every fourth trial. Thus, if accuracy had exceeded 75%, size and brightness of triangles were reduced by factor 0.5. Conversely, if accuracy had fell below 75%, size and brightness of triangles were increased by factor 0.5. Visual stimuli were presented against a black background (8.2/22.5/-29.1) on a 19-inch CRT monitor (Sony Multiscan G400; resolution 1,280 × 1,024 pixels, refresh rate 85 Hz) with a Nvidia GeForce (GT 220, color 32 bit/96 DPI) graphics card. Viewing distance was held constant at 64 cm with a chin and forehead rest. Participants responded by key press with their right index finger on the number pad of a standard keyboard.

In the LSC condition, auditory stimulation was realized by two bilaterally presented streams of tonepips. These had varying pitches within a frequency range of 310 – 1000 Hz. Initial tonepips started at 440 Hz, and subsequent tonepips were randomly increased or decreased by 5%. One stream was presented at a rate of five Hz and the other one at a rate of six Hz. The auditory target consisted of an omission of one tonepip (i.e., no sound) from a prespecified side. In the HSC condition, auditory stimulation was realized by two bilaterally presented short stories from Greek mythology and were both told by the same narrator. Both stories were taken from a publicly available online source. The target story was *Inachos und Eris* (Köhlmeier, 2011) and the non-target story was *Aigisthos* (Köhlmeier, 2013). Auditory stimulation was presented in both conditions via two loudspeakers (Logitech Z150), placed at medium height, directly left and right next to the monitor, with an external USB sound card (Behringer U-Control UCA222). The volume was individually adjusted to each participant.

Gaze data were recorded monocularly using an EyeLink 1000 Desktop Mount (SR Research, Mississauga, Ontario, Canada) video-based eye tracker sampling at 1000 Hz. Prior to the beginning of each condition, the signal was calibrated on participants' right eye using a nine-point calibration and validation sequence. If participants had left their position in the chinrest before all blocks of the respective condition were completed, a recalibration of the eye tracker was performed.

Task and Design

The experiment used a repeated measures design with the two within-subject variables Auditory Semantic Complexity (LSC vs. HSC) and Spatial Congruity (visual targets presented spatially congruent or incongruent with the attended auditory stream). In the LSC condition, each participant completed eight blocks, in which they had to attend either to the left or right auditory stream (randomized blockwise). Each block consisted of twelve trials, during which a visual target stimulus was shown either left or right from a central fixation cross (randomized trialwise; see Figure 1 for trial setup). Visual target stimuli appeared after 2 to 4 s for a duration of 250 ms. Participants had to press 8 (arrow pointing upwards) with the right index finger on the number pad of the keyboard as fast as possible when the triangle was pointing upwards. Conversely, if the triangle was pointing downwards, participants were instructed to press key number 2 (arrow pointing downwards) as fast as possible. If no response was given within a time interval of 1 s after onset of the visual target, the trial was evaluated as a ‘miss’. During the trials, the index finger remained on the key number 5 to ensure that the distance to both response keys was equal.

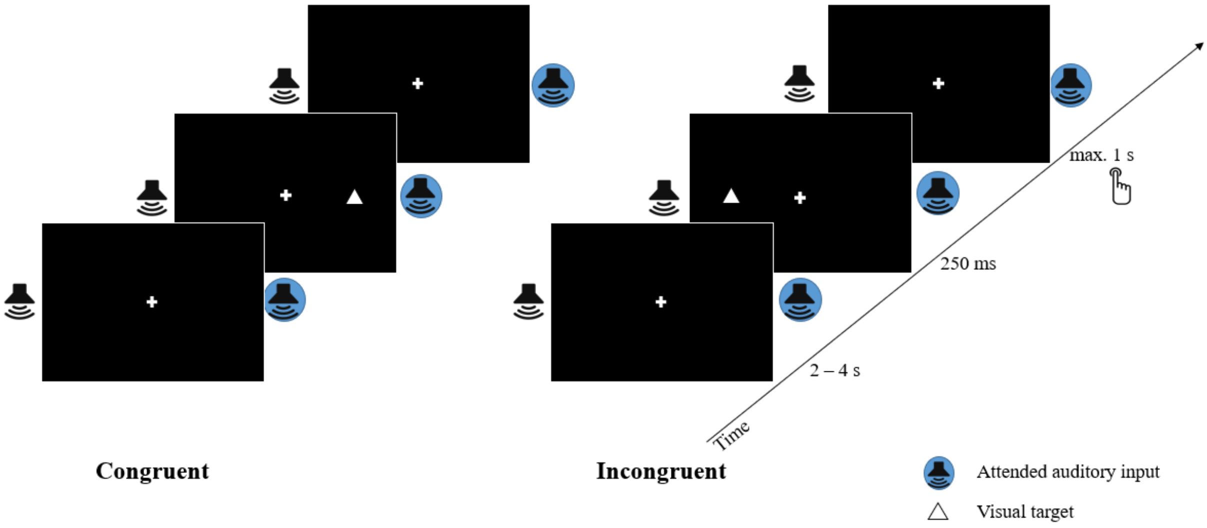

Figure 1. Sample trial sequence. Here, participants needed to actively listen to the right loudspeaker. After an interval of 2 to 4 s, a visual target (white triangle) appeared for 250 ms. Within a maximum interval of 1 s after target onset, participants had to press key number 8, as the triangle is pointing upwards. On the left, an exemplar congruent trial is shown, where auditory stimulation and the visual target stimulus were presented on the same side. On the right, an exemplar incongruent trial is shown, where auditory stimulation and the visual target stimulus were presented on opposite sides.

Furthermore, in each block, between zero and three auditory targets were presented in both the attended and the non-attended stream. There were at least ten tonepips distance between two auditory targets, regardless from which side they were presented, and a minimum distance of 1500 ms between visual and auditory targets. The laterality of auditory targets had no predictive power for subsequent visual targets. The participants' auditory task was to count the targets (tonepip omissions without any sound) only from one prespecified side, simultaneous to the visual speeded response time task. This was to ensure that participants permanently turned their attention to the required hemisphere in the auditory modality. The counted number should then be entered via number pad of the keyboard at the end of each block. The participants were also allowed to count the targets with their left hand during the blocks, as the task was intended to be purely perceptual and not requiring any working memory.

In the HSC condition, each participant completed six blocks with one block consisting of 34 trials. The visual task was identical to the LSC condition. Again, participants had to attend the auditory stream either from the left or the right side only (randomized blockwise). Between the blocks, they had to complete a subsequent comprehension questionnaire about the semantic content of auditory input in the previous block. Up to three short questions were asked by the experimenter to ensure that participants had permanently turned their auditory attention to the correct loudspeaker according to the instruction.

Procedure

All participants were tested individually in the eye-tracking laboratory at the Institute of Psychology in Vienna, Liebiggasse 5. Participants were welcomed at the beginning and asked to carefully read and sign the agreement for voluntary participation in the study. Demographic information (date of birth and gender) was also requested. The concrete hypotheses were not explained at the beginning in order to avoid biases caused by the participants' expectations. Therefore, it was only explained that this was a study on visual attention.

The study consisted of a total of three conditions with varying levels of semantic complexity of auditory stimulation. The visual task was the same throughout (with exception of the adaptive difficulty). In order to slowly accommodate participants to the complex dual task, the experimental conditions were always run through in the same order (LSC, HSC). In between, another condition with an auditory n-back task was carried out, which, however, was not relevant regarding the research question of this paper and is therefore only mentioned here for the sake of completeness. Each of the three conditions lasted for about 15 minutes. Including practice trials and self-paced breaks this resulted in a total study duration of about 75 minutes.

During the practice trials, participants were introduced to the auditory stimulation stepwise. For a start, omissions of sound (auditory targets) were presented in monotone tonepips sequences unilaterally. Afterwards, targets were presented in a randomly generated tonepips sequence unilaterally in the first place and then once again bilaterally. Subsequently, participants were instructed to undergo a practice trial counting auditory targets (between zero and three) from one side only and responding via keypress. Afterwards, the visual task alone was practiced in three trials with increasing difficulty and eventually both tasks should be carried out together in parallel.

Before starting with the LSC condition, participants were informed that their eye movements were recorded to verify that they kept their gaze on the central fixation cross and responded without making eye movements. Short breaks could be made between the blocks as needed. However, the eye tracker had to be recalibrated if participants had left the chin rest. They were also instructed to respond to visual stimuli as quickly and accurately as possible. Prior to the beginning of each block, participants received written instruction via the screen regarding which loudspeaker to attend to. Following each block, feedback on the individual performance (hits, misses, and mean RTs) was displayed on screen.

Before starting with the HSC condition, participants were briefly instructed that the visual task remained unchanged, but additionally, they would hear two short stories from Greek mythology (one from each speaker), whereby they only actively listened to one of them consecutively. Between the blocks, up to three short comprehension questions were asked by the experimenter to ensure that participants attended the relevant side. An exemplary question was „Mit wem hat Inachos dort gelebt?“ [„With whom did Inachos live there?“]. Answers could be given mostly with just a few words, in this case the right answer would have been “Mit seiner Frau und Tochter” [“With his daughter and wife”]. The criterion for excluding participants from further analysis remained the same as in the LSC condition (> 50% of questions answered either incorrectly or not at all). However, no participant exceeded this criterion. After completion of all three conditions, participants were debriefed about our hypotheses orally and in written form.

Data Analysis

Analysis of auditory task performance. Since we were interested in the influence of auditory stimulation on the detection of visual stimuli, visual task performance was of primary importance for us. However, we additionally analyzed the auditory task performance to ensure that participants followed the instructions and shifted their audiospatial attention accordingly. Finding a fair (exclusion-) criterion for auditory performance, which could be applied equally to the LSC and HSC condition, was not trivial, as the HSC condition allowed far more potential errors in replying to auditory targets. We have therefore decided on a criterion of blocks with more than one error. Participants who exceeded the criterion in more than half of all blocks were excluded from further statistical analyses.

Analysis of visual task performance. To investigate whether we could observe a cross-modal spatial congruity effect in participants’ sensitivity towards the discrimination of a visual target, we used signal detection theory (SDT). When participants were required to report the presence of a stimulus, they could give one of four different types of responses according to SDT: Hits, misses, false alarms and correct rejections (Koppen, Levitan, & Spence, 2009).

We defined a triangle pointing upwards as signal. Accordingly, pressing the key number 8 (i.e., arrow pointing upwards) was a correct signal response. The trial was evaluated as a hit, in case of a triangle pointing upwards and as a false alarm, in case of a triangle pointing downwards. Conversely, we defined a triangle pointing downwards as noise. Accordingly, pressing the key number 2 (i.e., arrow pointing downwards) was a correct noise response. The trial was evaluated as a correct rejection in case of a triangle pointing downwards and as a miss, in case of a triangle pointing upwards. Further on, we calculated hit rates and false alarm rates as follows: The number of hits divided by the number of all signal trials and the number of false alarms divided by the number of all noise trials, respectively.

Thus, the hit rate and the false alarm rate are fully capable to describe a participants' individual performance. These two indicators additionally reflect the response bias plus the degree of overlap between signal and noise distributions (also called *sensitivity*). Both factors are confounded by most performance measures: If the hit rate varies between conditions, it will not be clear whether they differ in response bias, sensitivity, or even both. Noting that sensitivity is linked to the distance between the mean of noise distribution and the mean of signal distribution (i.e., overlap between the two distributions will decrease, if this distance increases, respectively, if the means maintain a constant separation but the standard deviations decrease), a possible solution to this matter is to quantify sensitivity by determining the distance between the means through hit and false alarm rates, relative to their standard deviations by z- transformed scores (under the assumption of a cumulative normal distribution) of the proportion of hits minus z- transformed scores of the proportion of false alarms. This measure of SDT is also called *d'* (Stanislaw & Todorov, 1999). Therefore, we calculated participants' *d'* scores to analyze their relative sensitivity towards visual targets between congruent and incongruent trials in both conditions.

Furthermore, the luminance and size of visual stimuli were increased for frequently incorrect responses and decreased for frequently correct responses in order to adjust the task difficulty individually. This adaptive procedure is commonly used to counteract floor and ceiling effects as well as fatigue of participants. In the present study, the numerical frequency of adaptations made per condition (LSC, HSC) could also be an indicator of whether the auditory semantic complexity is reflected in a lower or higher task difficulty, in addition to the analysis of d' values.

Analysis of response times. As outliers we excluded all trials in which RTs deviate more than two *SD*s from the mean RT within the conditions (i.e., separately for both levels of Spatial Congruity and Auditory Semantic Complexity). The analysis of mean response times was based on trials with correct answers only.

To test if our hypotheses reached level of statistical significance ($p < .05$), sensitivity (*d'*) and RT data were further analysed by conducting a repeated measures analysis of variance (ANOVA) with the variables of Spatial Congruity and Auditory Semantic Complexity. Furthermore, ruling out a t-test should reveal if there is any significant difference in the frequency of adapted luminance and size values of visual stimuli. All analyses were conducted with MATLAB (version 8.3) or IBM SPSS Statistics (version 25).

Results

Auditory Task Performance

In the LSC condition, participants exceeded the criterion of more than one error per block in 5.9 % (*SE* = 1.5 %) of all blocks on average. In the HSC condition, more than one of the participants' answers per block were incorrect, or they did not answer at all, in 5.2 % (*SE* = 2.1%) of all blocks on average.

Visual Task Performance

To investigate whether we could observe a cross-modal response bias in participants' sensitivity towards the discrimination of visual targets, and whether this bias was affected by the complexity of auditory semantic content, we calculated mean *d'* scores for congruent and incongruent trials separately for each condition (see Table 1 for descriptive statistics).

Table 1

Means and standard deviations of d' scores in congruent and incongruent trials for the LSC and the HSC condition

	Congruent			Incongruent		
Levels of semantic complexity	*M*	*SD*	95% CI	*M*	*SD*	95% CI
Low	2.88	1.36	[2.22, 3.53]	2.76	1.43	[2.07, 3.45]
High	1.84	0.53	[1.58, 2.09]	1.80	0.39	[1.61, 1.99]

Note: CI = confidence interval.

Test of assumptions. Prior to statistical analysis, distribution of sample means was tested using the Shapiro-Wilk procedure. Although results indicate non-normality in two factorial levels of mean *d'* values ($p_s < .015$), ANOVA is considered as a very robust *F*-statistic. Thus, this violation can be neglected (Blanca, Alarcón, Arnau, Bono, & Bendayan, 2017; Lix, Keselman, & Keselman, 1996; Schmider, Ziegler, Danay, Beyer, & Bühner, 2010).

Analysis of variance. We found a significant main effect for Auditory Semantic Complexity, $F(1, 18) = 18.41$, $p < .001$, $\eta^2 = .51$, attributable to higher *d'* scores in the LSC compared to the HSC condition (see Table 1). No main effect was found for Spatial Congruity, $F(1, 18) = .46$, $p = .507$, $\eta^2 = .03$. Furthermore, there was no interaction between both variables, $F(1, 18) = .108$, $p = .746$, $\eta^2 = .01$.

Task difficulty. The luminance and size values decreased in a total of 5,7368 times (SE = 1,1892) in the LSC condition and 6,3889 times (SE = 1,1995) in the HSC condition. The difference was not significant, $t(18) = 0.52$, $p < 0.613$.

Response Times

To investigate whether we could observe a cross-modal response bias in participants' speeded discrimination of visual targets, and whether this bias was affected by the complexity of auditory semantic content, we calculated mean response times (RTs) for congruent and incongruent trials separately for each condition (see Table 2 for descriptive statistics).

Table 2

Means and standard deviations of RTs in congruent and incongruent trials for the LSC and the HSC condition

	Congruent			Incongruent		
Levels of semantic complexity	*M*	*SD*	95% CI	*M*	*SD*	95% CI
Low	727	70	[693, 760]	741	70	[708, 773]
High	714	62	[684, 744]	719	58	[691, 747]

Note: CI = confidence interval. All results are displayed in ms.

Test of assumptions. Sample means of RT data were normally distributed, as assessed by the Shapiro-Wilk test (ps > .070).

Analysis of variance. We found a significant main effect for Spatial Congruity, $F(1, 18) = 7.38$, $p = .014$, $\eta^2 = .29$, attributable to faster RTs in spatially congruent compared to incongruent trials (see Table 2). No main effect was found for Auditory Semantic Complexity, $F(1, 18) = 3.01$, $p = .100$, $\eta^2 = .14$. Furthermore, there was again no interaction between both variables, $F(1, 18) = 1.36$, $p = .259$, $\eta^2 = .07$.

Discussion

Previous studies on divided attention across different senses mainly investigated cross-modal links of spatially localized, brief sounds on concurrent visuospatial attention. However, little is known about the impact of a continuous, lateralized auditory stimulation. Based on this discontinuity, we applied a dual-task paradigm to reveal a possible cross-modal bias from an auditory task with stimuli of either low or high semantic complexity on a simultaneously performed visual speeded discrimination task. In the light of our results, the expected cross-modal effects in visual task performance and speed could only be partially confirmed. Although participants responded significantly faster in spatially congruent trials, their sensitivity towards the discrimination of visual targets only showed a small tendency in higher *d'* values, when presented on the same side as the relevant auditory stream. However, this tendency did not reach the level of statistical significance.

A possible explanation for our mixed results could derive from a different nature of required attention control in both tasks. Compared to the auditory task which aimed strongly at top-down attentional control due to a continuous binaural presentation of target and non-target stimuli, the visual task was less straightforward by demanding only a simple binary speeded discrimination of two different types of stimuli. Thus, due to the lack of presenting a continuous stream of stimuli, visual spatial attention may not necessarily have been applied in a *sustained* way, what is described as a defining feature of top-down attention control (Müller & Rabbitt, 1998). Therefore, we conclude that the task within the visual modality possibly aimed not predominantly at top-down controlled spatial attention, but rather on bottom-up control processes.

There is a growing body of evidence that these are two distinct forms of spatial attention arising from separate neuronal mechanisms (Carrasco, 2011; Cheal & Lyon, 1991; Corbetta & Shulman, 2002; Kurtz, Shapcott, Kaiser, Schmiedt, & Schmid, 2017; Lu & Dosher, 2000; Roque, Wright, & Boots, 2016). Based on a series of experiments, Prinzmetal and colleagues (2005a; 2005b; 2009) recently argued that endogenous orienting of spatial attention could enhance the perceptual representation of an attended stimulus, while exogenous orienting may simply increase participants' tendency to respond first to the stimulus presented in one location rather than another. Endogenous orienting of attention is therefore more likely to affect the accuracy of responses, while exogenous orienting of attention is more likely to influence the speed of responses instead (Spence, 2010). This view is consistent with our actual findings of significantly faster response times, but not a higher sensitivity towards target discrimination in spatially congruent trials, given that bottom-up control is actually more involved within the visual task.

These considerations are supported by previously used experimental designs in dual-task studies which succeeded to provide evidence for cross-modal links in endogenous spatial attention in task performance. For example, Driver and Spence (1994) firstly ruled out a *continuous dual-task performance*, where participants had to monitor a rapid serial visual presentation (RSVP) for occasionally appearing specified target shapes at a fixed location, while simultaneously shadowing (i.e., repeat it aloud without deeper processing of semantic content of attended message) one of two auditory speech streams. They found that participants performed significantly better in blocks in which the visual and the auditory target stream appeared on the same side, rather than on the opposite side. A few years later, Spence and Read (2003) could replicate Driver's and Spence's results in a more realistic scenario. Therein, participants had to repeat again semantically unrelated/meaningless syllables from one of two loudspeakers while simultaneously performing a demanding task in a car-driving simulator. Results indicated a significantly better auditory task performance when participants had to shadow the loudspeaker in front of them rather than from the side. Thus, Spence and Read could reveal benefits in multisensory dual-task performance when the attended sources of information are located spatially close to each other.

It is noticeable that in these two studies both tasks predominantly aimed at sustained, top-down attention control. In order to observe stronger cross-modal effects in a future study, a possible implementation could therefore be to adapt the experimental design accordingly, such that more sustained top-down attention is required. One possible way could be an adaption of the applied, ordinary speeded discrimination task by speeding up the presentation rate of stimuli and additionally by including non-targets. Another possibility could be to set up a visual search task which is considered to aim at top-down capture of spatial attention (Roque et al., 2016). Thereto, we would suggest applying an adapted version of the contingent cuing paradigm from Folk, Remington and Johnston (1992).

There is evidence from previously conducted dual-task studies, that cross-modal links in endogenous spatial attention become more pronounced when task difficulty increases (Driver & Spence, 1994; Spence et al., 2000). Hence, a second goal of this study was to determine, if different levels of semantic complexity in auditory stimulation had also an impact on the strength of cross-modal effects in visuospatial attention. As we could not find any interaction between the variables Auditory Semantic Complexity and Spatial Congruity, one may assume that our second hypothesis was disconfirmed. The additional finding that participants performed significantly better in the LSC condition also supports this conclusion, because it can be interpreted in a way that a lower semantic complexity of auditory stimulation was less attention demanding (i.e., lower task difficulty), allowing participants to devote more of their attentional resources towards the visual modality and therefore resulting in a better visual task performance. Staying with previously proposed implementations for a future study, possible underlying interactions may first come to bear, if a more demanding top-down task is also applied within the visual modality (Driver & Spence, 1994; Spence et al., 2000).

A limitation of this study is the comparability of the auditory task performance between the LSC and the HSC condition. Based on the results of Spence and Read (2003), which could show cross-modal effects in shadowing performance, but not in driving performance, we were primarily interested in cross-modal effects of semantically different complex auditory stimulation on visuospatial attention, or in other words, the visual task performance by manipulating the semantic complexity of the auditory content. Nevertheless, it would have been better to standardize the evaluation of the auditory performance in advance, since it is in fact a dual-task paradigm. Hence, we subsequently defined a criterion of blocks with more than one error for the analysis of performance. Strictly taken, however, the calculated scores of both conditions are not sufficiently standardized for a valid comparison, since memorizing the semantically way more complex contents allowed far more error possibilities than omissions of a tonepip. While for the LSC condition, the performance was well standardized and easily countable by the rate of counted auditory targets per block depending on the spatial congruency, this was much more difficult for the HSC condition. Here, an evaluation framework should have been defined in advance in order to be able to quantify the semantic performance systematically and in a controlled manner, since the answers given to the questions asked could gradually coincide with the information provided, which can no longer be determined afterwards. A possible solution to this problem would be to establish a double coding of answers in a follow-up study, which would be checked for inter-rater consistency within a category system.

Another limitation in regard of the comparability of auditory task performance are the differing levels of working memory involvement. In the LSC condition, participants were allowed to count auditory targets with their left hand during the blocks, as it was intended to draw on perceptual resources only. Similar solutions could not be provided in the HSC condition. Therefore, it is possible that some errors in auditory task performance could also be attributed to memory decay.

Conclusion. Over the last few decades, research on multisensory attention could highlight the existence of an interplay between our senses. In fact, there have been relatively few dual-task studies on the impact of endogenous orienting of spatial attention towards a continuous auditory stream. Demonstrating the persuasive nature of cross-modal links in audiovisual spatial attention, our results are inconsistent with popular resource allocation models that claim attention capacities are independent for auditory and visual information processing. Implications for real-world settings outside the laboratory could be, therefore, that endogenous orienting of one's spatial attention towards a lateralized, continuous auditory information source, like a co-driver's voice, could enhance the speed in detection of events within the same visual hemifield, compared to the opposite visual hemifield.

References

Alain, C., & Arnott, S. R. (2000). Selectively attending to auditory objects. *Frontiers in Bioscience, 5*, 202-212. doi:10.2741/

AlainAlais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: Fromphysiology to behaviour. *Seeing and Perceiving, 23*(1), 3-38. doi:10.1163/187847510X488603

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review*, *5*(3), 482-489. doi:10.3758/BF03208826

Best, V., Ozmeral, E. J., & Shinn-Cunningham, B. G. (2007). Visually-guided attention enhances target identification in a complex auditory scene. *Journal for the Association for Research in Otolaryngology*, *8*(2), 294-304. doi: 10.1007/s10162-007-0073-z

Blanca, M., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option?. *Psicothema*, *29*(4), 552-557. doi:10.7334/psicothema2016.383

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

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433-436. doi:10.1163/156856897X00357

Broadbent, D. E. (Eds.). (1958). Perception and communication. New York, NY, US: Oxford University Press

Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H., & Woldorff, M. G. (2005). The spread of attention across modalities and space in a multisensory object. *Proceedings of the National Academy of Sciences*, *102*(51), 18751-18756. doi:10.1073/pnas.0507704102

Calvert, G. A., Spence, C., & Stein, B. E. (Eds.). (2004). The handbook of multisensory processes. Cambridge, MA, US: MIT Press.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484-1525. doi:10.1016/j.visres.2011.04.012

Cheal, M., & Lyon, D. R. (1991). Central and peripheral precuing of forced-choice discrimination. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 43A*(4), 859-880. doi:10.1080/14640749108400960

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, 25. 975-979. doi:10.1121/1.1907229.

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics, 16*(2), 409-412. doi:10.3758/BF03203962

Connor, C. E., Egeth, H. E., & Yantis, S. (2004). Visual attention: bottom-up versus top-down. *Current Biology*, *14*(19), 850-852. doi:10.1016/j.cub.2004.09.041

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201-215. doi: 10.1038/nrn755

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers, 34*, 613-617. doi:10.3758/BF03195489

Cusack, R., Carlyon, R. P., & Robertson, I. H. (2000). Neglect between but not within auditory objects. *Journal of Cognitive Neuroscience, 12*(6), 1056-1065. doi:10.1162/089892900563867

Desimone, R. (1999). Visual attention mediated by biased competition in extrastriate visual cortex. In G. W. Humphreys, J. Duncan, & A. Treisman (Eds.), *Attention, space and action: Studies in cognitive neuroscience* (pp. 13-30). New York, NY, US: Oxford University Press.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193-222. doi:10.1146/annurev.ne.18.030195.001205

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70*(1), 80-90. doi:10.1037/h0039515

Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology, 92*(1), 53-78. doi:10.1348/000712601162103

Driver, J., & Spence, C. (1998). Crossmodal attention. *Current Opinion in Neurobiology, 8*(2), 245-253. doi:10.1016/S0959-4388(98)80147-5

Driver, J., & Spence, C. J. (1994). Spatial synergies between auditory and visual attention. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance series. Attention and performance 15: Conscious and nonconscious information processing* (pp. 311-331). Cambridge, MA, US: MIT Press.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113*(4), 501-517. doi:10.1037/0096-3445.113.4.50

Duncan, J., Martens, S., & Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature*, *387*(6635), 808-810. doi:10.1038/42947

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General, 123*(2), 161–177. doi:10.1037/0096-3445.123.2.161

Eimer, M. (1999). Can attention be directed to opposite locations in different modalities? An ERP study. *Clinical Neurophysiology*, *110*(7), 1252-1259. doi:10.1016/S1388-2457(99)00052-8

Eimer, M., & Schröger, E. (1998). ERP effects of intermodal attention and cross-modal links in spatial attention. *Psychophysiology, 35*(3), 313-327. doi:10.1017/S004857729897086X

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. doi:10.3758/BF03193146

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 18*(4), 1030-1044. doi:10.1037/0096-1523.18.4.1030

Gaspelin, N., Ruthruf, E., & Lien, M. C. (2016). The problem of latent attentional capture: Easy visual search conceals capture by task-irrelevant abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(8), 1104-1120. doi:10.1037/xhp0000214

Hancock, P. A., Oron-Gilad, T., & Szalma, J. L. (2007). Elaborations of the multiple-resource theory of attention. In A. F. Kramer, D. A. Wiegmann, & A. Kirlik (Eds.), *Attention: From theory to practice* (pp. 45-56). New York, NY, US: Oxford University Press.

Keil, J., & Senkowski, D. (2018). Neural oscillations orchestrate multisensory processing. *The Neuroscientist*, *24*(6), 609-626. doi:10.1177/1073858418755352

Knudsen, E. I. (2007). Fundamental components of attention. *Annual Review of Neuroscience, 30,* 57-78. doi:10.1146/annurev.neuro.30.051606.094256

Köhlmeier, M. (2011, November 9). Folge 73: Inachos und Eris. In *Mythen – Michael Köhlmeier erzählt Sagen des klassischen Altertums* [Video file]. Retrieved from https://www.br.de/mediathek/podcast/mythen-michael-koehlmeier-erzaehlt-sagen-des-klassischen-altertums/folge-73-inachos-und-eris/41527

Köhlmeier, M. (2013, July 25). Folge 20: Aigisthos. In *Mythen – Michael Köhlmeier erzählt Sagen des klassischen Altertums* [Video file]. Retrieved from https://www.br.de/mediathek/podcast/mythen-michael-koehlmeier-erzaehlt-sagen-des-klassischen-altertums/folge-20-aigisthos/41500

Koppen, C., Levitan, C. A., & Spence, C. (2009). A signal detection study of the Colavita visual dominance effect. *Experimental Brain Research, 196*(3), 353-360. doi:10.1007/s00221-009-1853-y

Kurtz, P., Shapcott, K. A., Kaiser, J., Schmiedt, J. T., & Schmid, M. C. (2017). The influence of endogenous and exogenous spatial attention on decision confidence. *Scientific Reports*, *7*(1), 6431. doi:10.1038/s41598-017-06715-w

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences, 9*(2), 75-82. doi:10.1016/j.tics.2004.12.004

Lix, L. M., Keselman, J. C., & Keselman, H. J. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. *Review of Educational Research, 66*(4), 579-619. doi:10.2307/1170654

Lu, Z.-L., & Dosher, B. A. (2000). Spatial attention: Different mechanisms for central and peripheral temporal precues? *Journal of Experimental Psychology: Human Perception and Performance, 26*(5), 1534-1548. doi:10.1037/0096-1523.26.5.1534

Marinato, G., & Baldauf, D. (2019). Object-based attention in complex, naturalistic auditory streams. *Scientific Reports*, *9*(1), 2854. doi:10.1038/s41598-019-39166-6

Marois, R., & Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in Cognitive Sciences, 9*(6), 296-305. doi:10.1016/j.tics.2005.04.010

McDonald, J. J., Teder-Sälejärvi, W. A., Russo, F. D., & Hillyard, S. A. (2003). Neural substrates of perceptual enhancement by cross-modal spatial attention. *Journal of Cognitive Neuroscience*, *15*(1), 10-19. doi:10.1162/089892903321107783

Müller, H. J., & Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention: Time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance, 15*(2), 315-330. doi:10.1037/0096-1523.15.2.315

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442. doi:10.1163/156856897X 00366

Posner, M. I. (1978). Chronometric explorations of mind. Hillsdale, NJ: Erlbaum.

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-processing account of its origins and significance. *Psychological Review, 83*(2), 157-171. doi:10.1037/0033-295X.83.2.157

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109*(2), 160-174. doi:10.1037/0096-3445.109.2.160

Prinzmetal, W., McCool, C., & Park, S. (2005). Attention: Reaction Time and Accuracy Reveal Different Mechanisms. *Journal of Experimental Psychology: General, 134*(1), 73-92. doi:10.1037/0096-3445.134.1.73

Prinzmetal, W., Park, S., & Garrett, R. (2005). Involuntary attention and identification accuracy. *Perception & Psychophysics, 67*(8), 1344-1353. doi:10.3758/BF03193639

Prinzmetal, W., Zvinyatskovskiy, A., Gutierrez, P., & Dilem, L. (2009). Voluntary and involuntary attention have different consequences: The effect of perceptual difficulty. *The Quarterly Journal of Experimental Psychology, 62*(2), 352-369. doi:10.1080/17470210801954892

Roque, N. A., Wright, T. J., & Boot, W. R. (2016). Do different attention capture paradigms measure different types of capture? *Attention, Perception, & Psychophysics, 78*(7), 2014-2030. doi:10.3758/s13414-016-1117-4

Sarter, N. B. (2007). Multiple-resource theory as a basis for multimodal interface design: Success stories, qualifications, and research needs. In A. F. Kramer, D. A. Wiegmann, & A. Kirlik (Eds.), *Attention: From theory to practice* (pp. 187-195). Oxford, NY, US: Oxford University Press.

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 6*(4), 147-151. doi:10.1027/1614-2241/a000016

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition, 80*(1-2), 1-46. doi:10.1016/S0010-0277(00)00152-9

Serences, J. T., Schwarzbach, J., Courtney, S. M., Golay, X., & Yantis, S. (2004). Control of object-based attention in human cortex. *Cerebral Cortex, 14*(12), 1346-1357. doi:10.1093/cercor/bhh095

Serences, J. T., Shomstein, S., Leber, A. B., Golay, X., Egeth, H. E., & Yantis, S. (2005). Coordination of voluntary and stimulus-driven attentional control in human cortex. *Psychological Science, 16*(2), 114-122. doi: 10.1111/j.0956-7976.2005.00791.x

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences, 12*(5), 182–186. doi:10.1016/j.tics.2008.02.003

Shomstein, S., & Yantis, S. (2006). Parietal cortex mediates voluntary control of spatial and nonspatial auditory attention. *The Journal of Neuroscience, 26*(2), 435-439. doi:10.1523/JNEUROSCI.4408-05.2006

Spence, C. (2010). Crossmodal attention. *Scholarpedia, 5*(5). 6309. doi:10.4249/scholarpedia.6309

Spence, C., & Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 22*(4), 1005-1030. doi:10.1037/0096-1523.22.4.1005

Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, *59*(1), 1-22. doi: 10.3758/BF03206843

Spence, C., & Parise, C. (2010). Prior-entry: A review. *Consciousness and Cognition*, *19*(1), 364-379. doi:10.1016/j.concog.2009.12.001

Spence, C., & Read, L. (2003). Speech shadowing while driving: On the difficulty of splitting attention between eye and ear. *Psychological Science, 14*(3), 251-256. doi: 10.1111/1467-9280.02439

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149. doi:10.3758/BF03207704

Teder-Sälejärvi, W. A., Münte, T. F., Sperlich, F. J., & Hillyard, S. A. (1999). Intra-modal and cross-modal spatial attention to auditory and visual stimuli. An event-related brain potential study. *Cognitive Brain Research, 8*(3), 327-343. doi:10.1016/S0926-6410(99)00037-3

Treisman, A. (1986). Features and objects in visual processing. *Scientific American, 255*(5), 114B-125. doi:10.1038/scientificamerican1186-114B

Vogel, E. K., & Luck, S. J. (2002). Delayed working memory consolidation during the attentional blink. *Psychonomic Bulletin & Review, 9*(4), 739-743. doi:10.3758/BF03196329

Wahn, B., & König, P. (2017). Is attentional resource allocation across sensory modalities task-dependent?. *Advances in Cognitive Psychology, 13*(1), 83-96. doi:10.5709/acp-0209-2

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors, 50*(3), 449-454. doi:10.1518/001872008X288394

Appendix A

Abstract

Previous studies investigated the influence of spatially localized, brief sounds on simultaneous visual-spatial attention and could demonstrate the existence of cross-modal effects. However, little is known about the effect of continuous, lateralized auditory stimulation. So far, cross-modal effects have been demonstrated mainly in studies with cueing paradigms, less often in dual-task situations. In addition, we wanted to investigate whether a semantically different complex content of auditory stimulation influences visuospatial attention.

We therefore performed an experiment (N = 19) in which the participants should react quickly to peripheral visual stimuli. In parallel, a continuous auditory stream of semantically simple (LSC condition) or complex (HSC condition) stimuli was presented bilaterally via loudspeakers, while the participants should only actively listen to one relevant side, blockwise. First, auditory stimuli were sequences of random tonepips, in which participants should count silent omissions in between. Second, a spoken story was presented, followed by a comprehension questionnaire after each block.

We observed significantly shorter response times in spatially congruent trials, compared to spatially incongruent trials, but no change in participants' sensitivity towards the detection of visual stimuli. We could also not determine any influence of semantically different complex auditory content. Our results indicate that attention towards a continuous auditory signal affects speed of visual processing in a dual-task situation. However, the observed effect is not as substantial as reported in previous studies. For future research, we recommend applying a more difficult and top-down attention demanding task in the visual modality as well.

Keywords: Multisensory attention, cross-modal attention, audiospatial, visuospatial, divided attention, spatial attention, dual-task situation

Zusammenfassung

Frühere Studien untersuchten den Einfluss von lokalisierten, kurzen Tönen auf die gleichzeitige, visuell-räumliche Aufmerksamkeit. Dabei konnte die Existenz crossmodaler Effekte nachgewiesen werden. Es ist jedoch wenig über die Wirkung einer kontinuierlichen, auditiven Stimulation aus der Peripherie bekannt. Bisher konnten kreuzmodale Effekte vor allem in Studien mit Hinweisreiz-Paradigmen nachgewiesen werden, seltener in dual-task Situationen. Zusätzlich wollten wir untersuchen, ob ein semantisch verschieden komplexer Inhalt der auditiven Stimulation einen Effekt auf die visuell räumliche Aufmerksamkeit hat.

Wir haben daher ein Experiment ($N = 19$) realisiert, in welchem die TeilnehmerInnen schnell auf periphere, visuelle Zielreize reagieren sollten. Gleichzeitig wurde eine kontinuierliche Abfolge von Reizen mit semantisch einfachem (LSC Bedingung) oder komplexem (HSC Bedingung) Inhalt bilateral über Lautsprecher präsentiert, wobei nur auf eine zielrelevante Seite pro Block geachtet werden sollte. Auditive Reize bestanden in der ersten Bedingung aus zufälligen Tonsequenzen, bei denen die TeilnehmerInnen Tonauslassungen zählen sollten. In der zweiten Bedingung wurde eine gesprochene Geschichte präsentiert, gefolgt von einigen Fragen zum inhaltlichen Verständnis.

Wir beobachteten signifikant geringere Antwortzeiten in räumlich kongruenten Versuchsdurchgängen verglichen mit räumlich inkongruenten Durchgängen, jedoch keine Veränderungen in der Entdeckungsleistung gegenüber visuellen Reizen. Wir konnten keinen Einfluss der unterschiedlichen semantischen Komplexität der auditiven Stimulation nachweisen. Unsere Ergebnisse deuten jedoch darauf hin, dass die endogen orientierte Aufmerksamkeit für ein kontinuierliches auditorisches Signal die visuelle Verarbeitungsgeschwindigkeit in einem dual-task beeinträchtigt. Dieser Effekt ist jedoch nicht so substantiell wie die zuvor berichtete Wirkung in vorangehenden Studien. Wir empfehlen für zukünftige Forschung, auch in der visuellen Modalität eine schwierigere top-down Aufgabe durchzuführen.

Schlüsselwörter: multisensorische Aufmerksamkeit, kreuzmodale Aufmerksamkeit, audiospatial, visuospatial, geteilte Aufmerksamkeit, räumliche Aufmerksamkeit, dual-task Sitation