Sensory Dominance in an Audiovisual Conflict Situation

Lea K. Hildebrandt

Bachelor thesis

Universiteit Twente

28-08-2009

Supervisors:

Durk Talsma

Rob H.J. van der Lubbe

University of Twente
Enschede - The Netherlands

Table of contents

Abstract

In an audiovisual conflict situation, audition should–according to the modality appropriateness hypothesis–dominate with temporally and vision with spatially distributed input. Out of this proposition the question arises, which modality would dominate if information was presented bimodally distributed across their optimal dimensions. This question was tested by simultaneously exposing participants to a display of dots and a sequence of sounds, of which the number had to be estimated. Additionally, it was tested whether an easily recognizable pattern of dots would have an effect on the estimates. In summary, no expected effects of multimodal interference were found. The results indicate that the current design did not lead to audiovisual integration, which could result from the missing temporal alignment of stimuli.

Sensory Dominance in an Audiovisual Conflict Situation

In everyday live, we often, and mostly unconsciously, rely on more than one sense. An ordinary activity like driving a car requires perceiving the traffic lights and other cars by vision, hearing sirens and hooters but also the own motor sounds and feeling the resistance of the pedals and the steering wheel. Only when all the information perceived via the different sensory modalities is combined appropriately, an adequate reaction–such as pushing the brake pedal–can be carried out. This combination of information is called *multisensory*–or *multimodal–integration.*

According to Ernst and Bülthoff (2004), the purpose of multisensory integration is to reduce the variance of the sensory information and thereby to increase the reliability of the perception. In other words, the perception of information from the environment is facilitated. This effect can be illustrated by considering an obvious example of audiovisual integration–speech. Not only the sounds that one hears but also the movement of the lips, the gesture and facial expression of the speaker are essential to communicate. Speaking on the phone can be more difficult than communicating face to face because of the missing visual cues. The visual cues thus facilitate the processing of speech sounds. Conversely, an auditory stimulus can also facilitate visual perception. Vroomen and de Gelder (2000) discovered the *freezing illusion*: performance at a visual search task is improved by providing a beep simultaneously with the visual target stimulus.

Interestingly, the sensory information has originally very distinct sources–light waves, sound waves, mechanical touch. This raises the question of how this qualitatively different input is integrated in the brain into a coherent reconstruction of the external world–also known as the binding problem (Kalat, 2007).

Early theories of multisensory perception, the race models, stated that information of different modalities would activate different pathways. The fastest of those pathways would then initiate the response (Barutchu, Crewther, & Crewther, 2009).

Even though the mean response to multisensory stimuli would, according to the race models, be faster than that to unisensory stimuli, the individual responses could never be faster than the fastest possible unisensory response. However, Miller (1982) demonstrated that auditory and visual stimuli elicited a faster response when presented together–the redundant signals effect. He suggested the concept of *coactivation:* after initially being processed by distinct systems, the multisensory information merges and therefore a faster response is possible than by unisensory processing alone (Miller, 1982). This idea is neurologically supported by the traditional hierarchical convergence model, which states that, following the unisensory processing in the sensory-specific cortices, multisensory information is integrated in so-called associative cortices (Masterton & Berkley, 1974).

However, recent findings in the neurosciences disprove the pure hierarchical processing: areas that were thought to be modality-specific are also responsible for multisensory integration (Driver & Spence, 2000; Foxe & Schroeder, 2005; Ghanzanfar & Schroeder, 2006; Kayser & Logothetis, 2007; Macaluso, 2006; Shimojo & Shams, 2001). Moreover, implying that multisensory integration occurs in a number of invariant, hard-wired processes, the traditional convergence model is challenged by a more dynamical, functionally oriented alternative (Fort & Giard, 2004; Senkowski, Schneider, Foxe, & Engel, 2008).

Most often, however, multisensory integration is associated with the superior colliculus; a structure located in the brainstem and responsible for reflexive movement (Holmes & Spence, 2005). According to Stein and Meredith (1993) there are three principles for multisensory integration to occur in the superior colliculus: the spatial, the temporal, and the inverse effectiveness rule. The spatial rule states that multisensory stimuli are integrated if they originate at the same spatial location in the environment. Moreover, they should be generated at approximately the same time, as stated by the temporal rule. Finally, the inverse effectiveness rule implies that multisensory integration is more pronounced if one or both stimuli alone are only weakly able to excite a neuron (Holmes & Spence, 2005).

These principles apply similarly for multisensory integration in the cortex (Stein & Wallace, 1996). However, it seems that the superior colliculus processes mostly spatial aspects of a stimulus (“where”), while the cortex is responsible for higher-level perception (“what”) (Fort & Giard, 2004; Stein & Meredith, 1993).

Concerning the behavioral consequences, multisensory integration can lead to a superadditive perception, which means that “the multisensory whole is greater than the sum of its unisensory parts” (Holmes & Spence, 2005, p. 763). This is in line with the inverse effectiveness rule and the concept of coactivation (or rather the redundant signals effect).

To determine the relative influence of each modality on audiovisual perception, research was conducted with conflict situations where the stimuli presented in both modalities were not congruent. Known as the *ventriloquist effect* is the phenomenon that, while watching television, speech sounds seem to originate from the lips of the actors instead of the loudspeakers of the TV set (Alais & Burr, 2004). This effect is explained by *visual capture*, which means that vision dominates over other modalities during multisensory perception (Welch & Warren, 1980). However, illusions such as the sound-induced flash illusion (Shams, Kamitani, & Shimojo, 2002) refute the dominance of vision. For this effect to appear, participants are exposed shortly to one circle and two rapidly succeeding beeps. The result of this design is that two circles are perceived. This illusion demonstrates that audition can dominate vision. In addition, examples have been proposed where no dominance of one modality was found but rather interference between both modalities. During their experiment, McGurk and MacDonald (1976) presented the participants simultaneously a face with moving lips and a speech sound. The lip movement matched the repetitive articulation of the syllable ‘ga’ whereas the audio track consisted of a repetitive pronunciation of the syllable ‘ba’. Surprisingly, the great majority of the participants reported to have perceived the syllable ‘da’. The auditory and the visual information thus merged to produce a new perception that had not been provided to either modality (McGurk & MacDonald, 1976). To summarize, there is no single dominating modality. Rather, the dominance seems to depend on the task requirements and stimulus properties, such as the spatial and temporal dimensions of the stimuli.

The *modality appropriateness hypothesis* states, simply put, that mainly the information from the modality that is most precise is used (Andersen, Tiippana, & Sams, 2005; Lederman & Klatzky, 2004; Shams, Kamitani, & Shimojo, 2004; Welch & Warren, 1980; Witten & Knudsen, 2005). Vision has a high spatial resolution because the activation of the retinal receptor cells reflects the spatial location of the stimulus in the environment. Audition, on the other hand, is more reliable regarding temporal processing (Shams et al., 2004; Witten & Knudsen, 2005). This hypothesis seems to be valid only as long as no noise degrades the stimuli (Burr & Alais, 2006). Otherwise, the *information reliability hypothesis*, which states that the modality that receives more reliable information will dominate, can be used to explain sensory dominance (Andersen et al., 2005).

Shams et al (2002) rejected the modality appropriateness hypothesis on the basis that the two visual flashes accompanied by one sound did not lead to a fusion illusion of the flashes. However, Andersen, Tiippana and Sams (2004) did find this effect. The hypothesis is further supported by the ventriloquism effect, where vision captures audition in a spatial task. The so-called temporal ventriloquism effect demonstrates that the opposite prediction implied by the modality appropriateness hypothesis: audition dominates vision in a temporal task (Morein-Zamir, Soto-Faraco, & Kingstone, 2003). Morein-Zamir and colleagues (2003) tested whether two sounds accompanying two flashes of light would influence the temporal order judgment of the lights. Indeed, depending on the occurrence in time of the sounds relative to the lights, the lights were perceived as either ‘pulled together’ or ‘pulled apart’. The modality appropriate hypothesis seems thus to provide a valid framework for the study of (non-degraded) multisensory integration. Yet, in the mentioned experiments, usually stimuli were distributed either spatially or temporally. According to the modality appropriateness hypothesis, one modality is therefore always clearly expected to dominate.

The purpose of this thesis is to investigate what happens if the visual information is distributed spatially and the auditory information temporally. In this case, both modalities should dominate, leading to a conflict. Consequently, the outcome of this conflict could be either a mutual interference between both modalities or the domination of one of the modalities.

To investigate whether this interference occurs and to identify the nature of the interaction, a numerosity judgment task is used. Philippi, van Erp and Werkhoven (2008) have used a similar task to specify the effects of temporally presented audiovisual stimuli. However, their main goal was to show that underestimates occurring in unimodal situations would decrease in multimodal conditions. For this end, they used congruent auditory and visual stimuli. In contrast, the purpose of the current experiment is to compare the influence of incongruent stimuli. Nevertheless, the research of Philippi et al (2008) provides specifications of the number of stimuli required; in their experiment, between 2 and 10 sounds were used and the interval between stimuli was 20 to 320 ms.

Owing of the obligatory incongruent nature of the current experiment, it was chosen to use a design consisting of a primary auditory stimulus and a secondary visual stimulus. The former was designed as a sequence of which the number of beeps had to be guessed by the participants, whereas the latter consisted of spatially distributed dots. In order to direct attention also to the visual modality and therefore to avoid suppressing the perception of the visual stimuli, catch trials were used in which the participants were required to report the number of dots instead of the number of beeps. The analysis of these trials should also reveal whether possibly found effects hold for the reversed direction, thus whether vision can be modified by audition. Taken together, the analyses of both the experimental and the catch trials could allow a better understanding of the dominance in the depicted conflict situation.

On the basis of the spatial and temporal principles of Stein and Meredith (1993), the stimuli are presented in close spatial proximity and during a short time window that is given by the duration of the sound-sequences. It is not possible to exactly align both stimuli due to the different qualities of the visual and auditory stimuli; however, multisensory integration does not necessarily depend on a simultaneous onset or the exact same location, but rather on close spatial and temporal proximity (Stein & Meredith, 1993).

According to the modality appropriateness hypothesis, it is expected that the stimuli presented visually and auditory will interfere. More precisely, in the proposed experiments the estimate of the number of beeps per sequence is expected to be modulated towards the number of displayed dots. The opposite should hold when the participants are required to report the number of dots presented to the visual modality.

An additional effect is expected based on the form of the spatial distribution of the visual display. It is expected that dots ordered in a dice-like fashion will be easier to perceive than dots distributed randomly over the display, and thus have a stronger influence. Based on the facilitatory effect of multimodal stimuli, it is expected that either display of dots will lead to facilitation and therefore a more accurate response than in the absence of dots. This effect is assumed to be more pronounced for the large numbers of sounds presented, as the difficulty of estimating this number increases. Similar to the effects found by Burr and Alais (2006) with degraded stimuli, this would lead to an decreased reliance on the auditory domain and thus a greater influence of the visually presented information.

Finally, it is expected that the displayed number of dots will interact with the pattern just described. It is assumed that especially for large numbers of dots, a dice-like pattern will be beneficial, due to the recognizability.

Experiment 1

Methods

Participants. Sixteen students of the University of Twente participated in this experiment, whereof eight were female. The mean age of the participants was 20 with an age range of 18 to 26. Most of the students were recruited via the participants' pool of the University of Twente and thus received credits for their participation, others participated voluntarily. All had normal or corrected-to-normal vision and hearing and were naïve to the purpose of the experiment. Before the experiment began, all participants gave written informed consent to take part in the study. Data of thirteen of the participants were valid and could be used for further analysis. The experiment was approved by the ethics committee of the faculty of behavioral sciences of the University of Twente.

Stimuli and apparatus. The visual stimuli were composed of red and white dots of a diameter of 15 pixels. In the experimental trials, three different conditions were used: dice-pattern, random-pattern and absent-pattern. In the dice-pattern condition, the dots were ordered in a dice-like fashion. In the random-pattern condition, the dots were ordered randomly on the display according to an algorithm. Between five and ten dots were displayed per trial in both conditions. In the third condition, absent-pattern, no visual stimuli were displayed. All the conditions were randomly (without replacement) but equally frequent displayed. Furthermore, approximately one third of the trials were catch trials, which were also presented at random without replacement. In these trials, the dots presented were red. The auditory stimuli consisted of sequences of beeps at a frequency of 1000Hz. The sequences were composed of between five and ten beeps, each separated by 150 ms, measured onset to onset. This interval was chosen so that the participants were unable to count the number but instead had to make a guess.

The experiments were conducted using Pentium IV computers in the laboratories of the faculty of behavioral sciences of the University of Twente. To present the stimuli, the software E-prime 1.1 (Psychology Software Tools) was used. All visual stimuli were displayed on a 17 inch Philips 107-T5 monitor with a refresh rate of 60Hz. The resolution was set to 800x600 and the color depth to 32 bit. Because it is necessary that the stimuli come from approximately the same spatial location (spatial rule) for audiovisual integration to occur, it was chosen not to use headphones to present the auditory stimuli but the speakers integrated in the base of the monitor. The exact number of beeps and dots per trial were presented at random with the premise that all combinations are presented equally frequent.

Task and procedure. The participants were greeted and seated in one laboratory at a distance of approximately 70 cm to the computer monitor. After completing a different experiment, a short verbal introduction to the experiment was given. Furthermore, an instruction was displayed on the monitor. The participants were free to choose when to start the experiment by pushing the space bar on the keyboard. During the experiment, all the participants were exposed to all conditions: all possible combinations of beeps and dots, all three patterns of dots, and the catch trials. During the regular experimental trials, the participants were asked to report the number of beeps in the sequence by pressing the adequate key on the keyboard (e.g. '7' for seven, '0' for ten beeps). Since only the accuracy of the response was of interest, no time limit was given. The following trial started two seconds after the participants made a response. During the catch trials, which were marked by red dots, the participants had to report the number of dots displayed instead of the number of beeps heard. The participants completed six blocks of 108 trials, which results from the combination of six possible sound conditions (number of beeps), six possible dot conditions (number of dots), and three pattern conditions. Between the blocks, the participants were allowed to pause. By pressing the space bar they could chose independently when to continue with the following block. The number of catch trials slightly varied per participant. Overall, the duration of the experiment was approximately 45 minutes.

Data analysis. The data recorded were the key pressed by the participant (5, 6, 7, 8, 9, or 0). First of all, it needs to be mentioned that the data were analyzed separately for the six numbers of beeps. The reason for this handling is that responses are not comparable between the different numbers of beeps because of a range effect. Due to the constriction of the possible responses to a range of six natural numbers, it is conceivable that these responses are biased towards the mean of the scale.

The necessary data were extracted using the applications E-DataAid and E-Merge of E-prime (Psychology Software Tools) and converted into Microsoft Excel pivot tables. After the tables were imported to SPSS Statistics (version 16) and for normal distribution was tested, the data were analyzed by performing six univariate repeated measures ANOVAs with number of dots (5,6,7,8,9,10) and pattern (yes, no, absent) as factors. In the cases where sphericity could not be assumed, the significance probability was corrected using the Greenhouse-Geisser estimate. Subsequently, the catch trials were analyzed using a univariate 2 x 6 (pattern x sounds) repeated measures ANOVA. Effects were further analyzed using Bonferroni pairwise comparisons.

Results and discussion

The results of the six repeated measures analyses of variance are displayed in table 1. In this table, the test statistics and the significance probabilities are shown. These data are presented for the main effects of pattern and number of dots, and the interaction effects. As can be seen, none of the effects reaches significance when compared to a Bonferroni-corrected significance level of $\alpha = 0.0083$.

Table 1.

Summary of Analyses of Variance for Effects of Visual Pattern and Stimuli on Auditory Perception

	Sounds per sequence					
Source	5	6	7	8	9	10
Pattern						
F(2,24)	4.544	0.325	1.474	2.061	1.738	2.477
p	0.048[a]	0.727	0.253	0.154	0.218[a]	0.109
Dots						
F(5,60)	0.747	1.036	0.577	0.842	0.556	0.582
p	0.502[a]	0.381[a]	0.616[a]	0.490[a]	0.661[a]	0.637[a]
Pattern x dots						
F(10,120)	1.316	2.092	0.757	0.610	1.603	0.753
p	0.286[a]	0.131[a]	0.536[a]	0.654[a]	0.198[a]	0.580[a]

[a]Greenhouse-Geisser correction for sphericity is applied.

The analyses of the catch trials using repeated measures ANOVAs revealed in one of the six analyses a main effect of pattern (see table 2). The subsequent pairwise comparison confirmed this effect and indicated that the dice-pattern condition leads to a higher response than the random pattern. These results are in line with the expectations, showing that a dice-like pattern leads to a more accurate response for the highest numbers of dots. These results confirm that a pattern might facilitate visual perception but the auditory stimuli did not have an influence on the perception of the visual stimuli.

Table 2.

Summary of Analyses of Variance for Effects of Visual Pattern and Auditory Stimuli on Visual Perception

Source	Dots per trial					
	5	6	7	8	9	10
Pattern						
F(1,12)	0.022	6.169	6.310	3.275	0.294	28.055[*]
p	0.886[a]	0.030[a]	0.029[a]	0.108[a]	0.599[a]	0.000[a]
Sounds						
F(5,60)	1.836	0.820	1.699	1.401	0.548	0.384
p	0.194[a]	0.490[a]	0.184[a]	0.270[a]	0.632[a]	0.660[a]
Pattern x sounds						
F(10,120)	0.357	0.870	3.931	0.480	0.424	0.789
p	0.746[a]	0.432[a]	0.018[a]	0.670[a]	0.685[a]	0.511[a]

[a]Greenhouse-Geisser correction for sphericity is applied.
[*]$p < 0.0083$

It seems thus that no cross-modal influence occurs. These results contradict our expectations based on the modality appropriateness hypothesis. Accordingly, it was expected that spatial information presented visually and temporal information presented auditory will interfere, thus leading to a modification of the response. Specifically, since the response was required to be given on the basis of the auditory information, it was expected that the visual information would influence this response.

One possible reason for not obtaining the expected results could be that the tone sequences presented were too short. When asked at the end of the experiments, some of the participants indicated that they had been able to count the beeps and therefore did not have to make a guess. However, it was the intention of the experiment that the participants should estimate the number of beeps. By actively counting the beeps it might be possible that most of the attentional capacity was allocated to the beeps and thus the visual stimuli were consciously suppressed. To exclude the possibility of counting the beeps, a second experiment was conducted. In this experiment, the number of beeps was increased, ranging from seven to twelve beeps per sequence.

Experiment 2

Methods

Participants. Fifteen subjects participated in the second experiment. Of those, eleven were female and again most of them were students of the University of Twente that received credits for participation. The mean age was 25 with an age range of 18 to 57. The participants all had normal or corrected to normal vision and hearing and gave informed consent. Due to technical problems, data of only nine participants could be used for analysis.

Stimuli and apparatus. The same apparatus was used as in the first experiment. The stimuli did not differ from those of the first experiment, except that the number of beeps and dots presented was increased to the range of seven to twelve. As in experiment 1, the visual stimuli were randomly presented in a dice-like pattern, without a pattern or the visual display was absent. Furthermore, approximately one third of the trials were catch trials where the dots were colored red and the participants were asked to report the number of dots.

Task and procedure. Again, the subjects first participated in a different experiment. Subsequently, a short verbal introduction was given which was accompanied by a written instruction on the monitor. The participants could choose when to start the first trial by pressing the space bar. The task of the participants was the same as in experiment 1. In the experimental trials, where the dots were colored white, the participants had to report an estimate of the number of beeps in a sequence. When red dots were presented, they were required to report the number of dots. In both cases the answer recorded was the key pressed on the keyboard, whereby '0' ten, '1' eleven and '2' twelve represented. Again, the participants had to complete six blocks of 108 experimental trials each, which took approximately 45 minutes.

Data analysis. The responses of the experimental trials were again analyzed separately for the different numbers of beeps displayed in a sequence. The same software and procedure was used as in experiment 1. Thus, after having tested for normal distribution, six 3 x 6 (pattern x number of dots) univariate repeated measures ANOVAs were performed with SPSS. The catch trials were analyzed using a 2 x 6 (pattern x number of sounds) repeated measures ANOVA and pairwise comparisons according to Bonferroni.

Results and discussion

The results of the six separate ANOVAs are summarized in table 3. Again, the test statistics and the significance probabilities for the main effects of pattern and dots and the interaction effects between pattern and dots are displayed. The columns show the six separate analyses.

Table 3.

Summary of Analyses of Variance for Effects of Visual Pattern and Stimuli on Auditory Perception

	Sounds per sequence					
Source	7	8	9	10	11	12
Pattern						
$F(2,10)$	0.203	4.284	5.986	0.088	0.097	2.430
p	0.819[a]	0.035[a]	0.036[a]	0.916[a]	0.908[a]	0.162[a]
Dots						
$F(5,25)$	0.100	1.168	1.369	1.140	1.545	2.320
p	0.959[a]	0.345[a]	0.299[a]	0.353[a]	0.232[a]	0.145[a]

Source	Sounds per sequence					
	7	8	9	10	11	12
Pattern x dots						
F(10,50)	1.220	1.667	0.457	1.192	0.602	0.369
p	0.327[a]	0.194[a]	0.689[a]	0.333[a]	0.612[a]	0.768[a]

[a]Greenhouse-Geisser correction for sphericity is applied.

Considering the Bonferroni correction for the six analyses, none of these results reaches significance. These results indicate that, as shown in experiment 1, a visual display of dots did not modify the estimate of the sounds. Further, the pattern of the visual display did not have an influence either. In conclusion, the increased number of stimuli did not lead to the expected effects.

The analyses of the catch trials revealed two main effects of pattern: in trials with 8 dots and in trials with 12 dots (see table 4). In contrast to experiment 1, the pattern facilitates the perception of the dots not only for the highest numbers, but rather for the extremes. The trials where the dots were ordered randomly showed a great range effect because the responses deviated towards the mean of the range. This effect was reduced by displaying the dots in a dice-like fashion. In other words: the variance of the trials in the dice-pattern condition was smaller than that in the random-pattern condition. This effect is only visible for the most extreme numbers of dots, as the variance in the center of the range deviates towards both ends and the variance thus cancels itself out when averaged. This effect was furthermore not found for the trials with 7 dots. A possible reason is the small number of dots, which could also be easily recognizable when displayed randomly distributed.

Table 4.

Summary of Analyses of Variance for Effects of Visual Pattern and Auditory Stimuli on Visual Perception

Source	Dots per trial					
	7	8	9	10	11	12
Pattern						
F(1,6)	7.288	17.913	0.995	0.002	0.081	45.076
p	0.027	0.004	0.348	0.962	0.785	0.001
Sounds						
F(5,30)	0.702	0.736	0.695	0.574	0.420	5.146
p	0.522	0.515	0.543	0.608	0.719	0.012
Pattern x sounds						
F(10,80)	0.702	0.874	0.667	0.630	0.825	4.990
p	0.522	0.465	0.548	0.586	0.486	0.013

[a]Greenhouse-Geisser correction for sphericity is applied.
$^{*}p < 0.05$

Furthermore, the effect of the beeps and the interaction effect almost reach significance for the 12-dots trials. A closer visual analysis of the responses made by the participants revealed that this result could be due to individual outliers in the data of three of the nine participants.

Generally, these analyses show similar results as experiment 1: The visual display of dots does not affect the perception of sounds (experimental trials) and vice versa (catch trials).

Furthermore, in both cases no interaction effect with pattern was found. However, there seems to be an effect of pattern in the catch trials, which confirms that dots ordered in patterns are more easily recognizable.

General Discussion

According to the modality appropriateness hypothesis, vision dominates over audition when multisensory information is presented spatially and audition over vision in temporal tasks. Both parts of this hypothesis were confirmed by a number of well-known effects, such as the sound-induced illusory flash (2002), the ventriloquism effect, and the temporal ventriloquism effect (Vroomen & de Gelder, 2000). However, to my knowledge, no study yet has investigated the effects of information simultaneously presented in both respective dominant modes. It was hypothesized that information presented spatially distributed to the visual modality and temporally distributed to the auditory modality would lead to a conflict in multisensory integration. The objective of this study was to test this hypothesis and detect whether one modality would dominate. With the present experimental design, it was expected that vision would capture audition and, thus, the number estimates of auditory stimuli should be influenced by the visual signals.

However, the results suggest that this is not the case. The participants were well able to report the number of beeps in a sequence. No influence of the number of dots displayed was observable in the conducted experiments.

Furthermore, it was expected that the pattern in which the dots were displayed would have an effect on the response. In the dice-pattern condition, the dots were displayed in a dice-like fashion, which was expected to facilitate the perception of the stimuli. In the random-pattern condition, the dots were randomly ordered on the display. Although it was expected that in the random-pattern condition the influence of the dots would be less distinctive than in the dice-pattern condition, there should have been still more modification by the visual domain than in the absent-pattern condition. This expectation was based on the fact that multimodal presentation facilitates accuracy relative to unimodal presentation. This effect could not be found in either experiment.

Finally, an interaction effect between dots and pattern was hypothesized. Specifically, it was expected that the dice-pattern condition would have a stronger effect for high numbers of dots. This hypothesis was also not confirmed in the experiments.

A possible interpretation of these results, which would be in line with the hypotheses, is that the auditory modality strongly dominates in the case of conflict. For this finding to be confirmed, the responses in the catch trials–where the modality to be reported was vision–should be strongly modified by the number of beeps per sequence. However, the analysis of the catch trials did not support this finding. In contrast, the participants were also well able to report the number of dots displayed. The auditory modality did not have an influence on the response, and therefore general auditory dominance can be excluded. The analysis of the catch trials confirmed further that stimuli ordered in a pattern lead to a more accurate perception of the visual display. A dice-like pattern seems to decrease the range effect through decreasing the variance of responses. Thus, it should be valid to conclude that a pattern leads to a more accurate perception of the number of dots.

These findings, however, lead to the conclusion that the participants were able to process the information presented to both modalities separately. The catch trials indicated that, especially when ordered, the number of dots were easily recognizable. The same finding for the recognition of the auditory stimuli resulted from the analysis of the experimental trials. However, there was no influence of either modality on the other. Thus, no interaction occurred between the modalities; this gives rise to the assumption that no multimodal integration occurred. Moreover, the responses to the trials with a visual display did not differ from those to the trials without any dots, which indicates that both stimuli where processed unisensory.

There are a number of possible explanations why no expected effects were found. The most evident reason would be that there are simply no effects and the hypotheses are refuted. This would lead to the conclusion that no conflict between the modalities arises. One justification for this interpretation could be that stimuli distributed across the two dimensions are possibly processed separately: temporal information in the superior colliculus and spatial information in the cortex. However, considering the above mentioned missing multisensory integration of the auditory and visual stimuli, it seems premature to reject the main hypothesis–that a conflict occurs–on the basis of this interpretation. Rather, it might be beneficial to take a look at some other factors that might account for the results.

Therefore, the experimental design needs to be considered. The purpose of the experiment was to present the stimuli differently: one spatially and one temporally distributed. Although a spatial alignment is not a necessary condition for multisensory integration to occur (Noesselt, Fendrich, Bonath, Tyll, & Heinze, 2005), temporal alignment generally is (Stein & Meredith, 1993). Although both stimuli were presented during a certain time window that would allow multisensory integration, it was in this experimental design not possible to align the visual stimulus temporally with the auditory stimulus. Without this alignment, it could be possible that the different stimuli were not perceived as belonging together, and thus no audiovisual integration could occur.

Another problem might have been the lack of a time limit to respond. The research question emanates from the assumption that the *estimation* of numbers would be multimodally influenced. With the unlimited response time, the participants were able to elaborate consciously afterwards. This idea is supported by the fact that some participants reported to have been able to instantly count the beeps or to retrieve the sequence from the phonological loop. This could have led to a late allocation of attentional capacity to the auditory modality, suppressing the visual information.

A second problem related to attention might be that, by instruction, the attention of the participants was directed towards one modality, usually audition. Again, this was inevitable due to the experimental design. Using the catch trials, it was hoped to increase uncertainty as to which stimulus needs to be reported and thus to avoid early filtering. Still, the criterion of which number to report was an outstanding feature of the dots: the color. This could have led to an early suppression of the superfluous modality. The fact that the participants were well able to report the number of stimuli in the required modality suggests that the participants did focus their attention on one modality. This is, according to the *directed attention hypothesis,* as stated by Warren (1979) and referred by Andersen et al. (2004), another indication that the stimuli in both modalities were not perceived as one object.

Furthermore, the low number of stimuli could have had an effect on the results. Although the results of the two experiments conducted did not differ considerably and the number of beeps and dots presented per trial was similar to the ones used in another experiment (Philippi et al., 2008); there were still participants in experiment 2 who reported to have counted the beeps. To account for the differences between participants in counting ability, a larger number of stimuli could lead to the intended guessing by all participants.

In conclusion, I would consider a number of modifications for a subsequent experiment. To begin with, I would use a task where both stimuli are distributed across both dimensions. This could be achieved by using a number of loudspeakers located around a monitor and several locations on this monitor. Temporally, both stimuli should be displayed in a sequence. Therefore, the mode of presentation is the same for both modalities, which might increase the possibility of multisensory integration. Another advantage of this design is that the participants would not need to direct their attention by instruction to one modality, but rather attend to both. The task of the participants could be to detect changes in either the modality or the dimension (spatial or temporal). Such a task would further allow for the use of only two keys on the keyboard, which could be located more convenient. This aspect supports the use of a time limit and the measurement of reaction times. As mentioned above, pressure of time might inhibit conscious elaboration, which is further supported by the fact that, in the proposed forced-choice task, no counting is necessary. In my opinion, a measure of reaction times would augment the results of such an experiment, because multisensory facilitation can also be visible in shorter reaction times, as indicated by coactivation (Miller, 1982). Practically, stimuli would always be presented to both modalities simultaneously and changes would occur in either one modality or both. These changes could occur across either dimension–spatially or temporally. In particular, the trials where the stimuli changed in both modalities would be of interest. Compared to the unimodal change trials, the congruent multimodal trials (that is, stimuli change on the same dimension) should be processed faster and more accurate as suggested by coactivation. However, the incongruent multimodal stimuli, especially the ones where both stimuli change in the dominant dimension, could reveal which modality dominates in the case of this conflict. It would be beneficial to look at the trials where the participants did not give a correct response, but rather pressed only one key. Although this experimental design does not necessarily lead to multisensory integration–the changes are not always congruent–it could reveal nonetheless important insights over the proposed conflict situation, which might apply as well for multisensory integration.

Conclusions

The objective of this thesis was to determine the effect of simultaneously displayed visual and auditory stimuli distributed in a spatial and temporal manner, respectively. Since stimuli were presented to both modalities distributed across their dominant dimension–as specified by the modality appropriateness hypothesis–it was expected that this experimental design would lead to a conflict. The outcome of this conflict would shed light on the underlying processes of audiovisual integration; specifically which modality would dominate in this case and what the relative influence of either modality would be. Particularly, it was expected that visually presented dots would interfere with auditory sequences of sounds and further that a pattern of the dots would enhance this effect. However, no effects were found.

According to the results of the experiments, there seems to be no interference between the two modalities. This interpretation was supported by the analyses of the catch trials. Nevertheless, these findings do not necessarily reject the hypothesis, which tests whether the described conflict occurs. Rather, it seems that the chosen experimental design did not lead to the expected multisensory integration. This interpretation would confirm the importance of spatial and temporally alignment for multisensory integration to occur. Further research with an adjusted design should be conducted in order to test the hypothesis.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology , 14*, 257-262.

Andersen, T. S., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research , 21*, 301-308.

Andersen, T. S., Tiippana, K., & Sams, M. (2005). Maximum likelihood integration of rapid flashes and beeps. *Neuroscience Letters , 380*, 155-160.

Barutchu, A., Crewther, D. P., & Crewther, S. P. (2009). The race that precedes coactivation: Development of multisensory facilitation in children. *Developmental Science , 12*, 464-473.

Burr, D., & Alais, D. (2006). Combining visual and auditory information. *Progress in Brain Research , 155*, 243-258.

Driver, J., & Spence, C. (2000). Multisensory perception: Beyond modularity and convergence. *Current Biology , 10*, R731-R735.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences , 8*, 162-169.

Fort, A., & Giard, M.-H. (2004). Multiple electrophysiological mechanisms of audiovisual integration in human perception. In G. E. Calvert, C. E. Spence, & B. E. Stein, *The handbook of multisensory processes* (S. 503-514). Cambridge, MA: MIT Press.

Foxe, J. J., & Schroeder, C. E. (2005). The case for feedforward multisensory convergence during early cortical processing. *NeuroReport , 16*, 419-423.

Ghanzanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Science , 10*, 278-285.

Holmes, N. P., & Spence, C. (2005). Multisensory integration: Space, time and superadditivity. *Current Biology , 15*, 762-764.

Kalat, J. W. (2007). *Biological psychology* (9th ed.). Belmont, CA: Thomson Wadsworth.

Kayser, C., & Logothetis, N. K. (2007). Do early sensory cortices integrate cross-modal information? *Brain Structure and Function , 212*, 121-132.

Lederman, S. J., & Klatzky, R. L. (2004). Multisensory texture perception. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (S. 107-122). Cambridge, MA: MIT Press.

Macaluso, E. (2006). Multisensory processing in sensory-specific cortical areas. ‘*Neruoscientist , 12*, 327-338.

Masterton, R. B., & Berkley, M. A. (1974). Brain-function: Changing ideas on role of sensory, motor, and association cortex in behavior. *Annual Review of Psychology , 25*, 277-312.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature , 264*, 746-748.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology , 14*, 247-279.

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research , 17*, 154-163.

Noesselt, T., Fendrich, R., Bonath, B., Tyll, S., & Heinze, H.-J. (2005). Closer in time when farther in space--Spatial factors in audiovisual temporal integration. *Cognitive Brain Research , 25*, 443-458.

Philippi, T. G., van Erp, J. B., & Werkhoven, P. J. (2008). Multisensory temporal numerosity judgment. *Brain Research , 1242*, 116-125.

Senkowski, D., Schneider, T. R., Foxe, J. J., & Engel, A. K. (2008). Crossmodal binding through neural coherence: Implications for multisensory processing. *Trends in Neuroscience , 31*, 401-409.

Shams, L., Kamitani, Y., & Shimojo, S. (2004). Modulation of visual perception by sound. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 27-34). Cambridge, MA: MIT Press.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research , 14,* 147-152.

Shimojo, S., & Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interaction. *Current Opinion in Neurobiology , 11,* 505-509.

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

Stein, B. E., & Wallace, M. T. (1996). Comparisons of cross-modality integration in midbrain and cortex. *Progress in Brain Research , 112,* 289-299.

Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance , 26,* 1583-1590.

Warren, D. H. (1979). Spatial localization under conflict conditions: Is there a single explanation? *Perception , 8,* 323-337.

Welch, R. B., & Warren, D. H. (1980). Immediate response to intersensory discrepancy. *Psychological Bulletin , 88,* 638-667.

Witten, I. B., & Knudsen, E. I. (2005). Why seeing is believing: Merging auditory and visual worlds. *Neuron , 48,* 489-496.