University of Twente
Enschede - The Netherlands

The influence of degraded stimuli on audio-visual integration

Alisha Siebold

Supervisor:
Dr. Durk Talsma
Dr. Rob H.J. van der Lubbe

Date: 23-08-2009

Abstract: The integration of input from multiple senses is essential for maintaining an integrated picture of the external environment. The phenomenon of multisensory integration has been subjected to detailed analysis leading to three general principles: the temporal and the spatial rule and the principle of inverse effectiveness (Stein & Meredith, 1993). The objective of the current experiment was to investigate whether degraded auditory and visual stimuli would influence multisensory integration in accordance with the principle of inverse effectiveness at a behavioral level, thus that the best performance is obtained for crossmodal presentation when stimuli are degraded. In a forced-choice categorization task participants were required to identify one of two objects on the basis of auditory, visual or audio-visual features that were partly or fully degraded on some trials. The general results reveal that, with regard to reaction time and accuracy, crossmodal presentation is superior to unimodal presentation and that the best performance was obtained for audio-visual presentation when stimuli are moderately degraded. These data provide behavioral evidence for the principle of inverse effectiveness and suggest that for real-world situations that highly rely on recognition of noisy input from one modality, like air- and car traffic, congruent information presented via additional modalities can enhance recognition.

Key terms: audiovisual processing, inverse effectiveness, degraded information, multisensory integration

Introduction

Like most other species, we perceive and experience events in our environment by processing sensory information that comes in through our sensory channels (Stein & Meredith, 1993). Each of our senses offers unique information that is qualitatively different from that provided by the other senses. For example, like the subjective impressions of pitch and volume have no equivalent in vision, olfaction, degustation and somatosensation, the perception of color can only be discerned through the visual channel. Notwithstanding the substantial inhomogeneity among the sensations that can be apprehended simultaneously, the information from the different senses is usually complementary and we are able to make sense of all the incoming information and maintain an integrated and consistent notion of our external environment. The capacity of using this multisensory information synonymously provides an evolutionary advantage, as the deprivation of one sense allows for compensation by the other senses and decreases sensory uncertainty (Alais & Burr, 2004). Furthermore, multisensory integration can yield information about the external environment by combining sensory input to produce a new representation that is unattainable from unimodal sensory input alone. (O'Hare, 1991).

A number of studies indicate that our capability of merging the inputs across the senses can increase the accuracy of detecting stimuli as well as speeding up reaction times compared to unimodal processing. (Zahn et al., 1978; Miller, 1986; Stein et al., 1989; Perrott et al., 1990; Giray & Ulrich, 1993; Hughes et al., 1994; Frens et al., 1995). An example of multisensory integration of nonlinguistic audio-visual information is the study by Giard and Peronnet conducted in 1999. They investigated the integration of auditory and visual signals in the case of temporal synchrony by means of a forced-choice reaction-time categorization task in which the association of the unrelated crossmodal stimuli had been learned in advance. They combined this behavioral study with the recording of event-related potentials (ERPs) to examine the time course and locations of multisensory integration during object recognition. ERPs were recorded while participants identified one of two objects (object A or object B) on a computer screen by pressing the associated key. The objects were specified on the basis of visual features (a horizontal or a vertical ellipse), or auditory features alone (a 540 Hz or a 560 Hz tone), or the combination of audio-visual features where object A was defined by the 540 Hz tone and the horizontal ellipse and object B was represented by the 560 Hz tone and the vertical ellipse. In accordance with the previously mentioned studies, the behavioral results of the experiment by Giard and Peronnet indicated enhanced responsiveness to crossmodal stimuli compared to unimodal presentation, which provides evidence for audiovisual integration at a behavioral level.

In the light of this fundamental characteristic of multi-sensory processing, knowledge of its underlying mechanisms and neurophysiology are essential for a complete appreciation of our brain function. For example, neuroanatomical studies in primates have identified numerous cortical and subcortical areas involved in the convergence of the different senses. One of the most studied structures of multisensory integration is the superior colliculus, located in the brainstem and involved in navigational processing, orienting behaviors and controlling saccadic eye movements (Wallace et. al, 1996). Neurons that can be excited by input from different senses have been located in deep layers of the superior colliculus in various species, including monkeys (Jay & Sparks, 1984), cats (Gordon, 1973; Meredith &

Stein, 1983; Peck, 1987), and rodents (King & Palmer, 1985; Wallace et al., 1996). By means of a thorough study of the superior colliculus, Stein and Meredith identified three general principles of multisensory integration (Stein & Meredith, 1993). According to the spatial rule (Meredith & Stein 1986; King & Palmer, 1985) and the temporal rule (Meredith et al., 1987; King & Palmer, 1985), multisensory neurons can be excitated beyond a value expected by summation of unimodal stimulation when cues from more than one sensory modality arise from approximately the same location close in time. The principle of inverse effectiveness (Meredith & Stein, 1983) states that this crossmodal integration effect is more articulate or is manifested more frequently when unisensory stimuli are less effective. Likewise, the reverse applies, in that the more effective an unimodal stimulus, the less contribution regarding enhancement of accuracy or responsiveness does a combination of senses provide (Stein et al., 1994). The principle of inverse effectiveness has been consistently confirmed at a behavioral level in both humans and animals as well as at a neurophysiological level (Frassinetti et al., 2005; Bolognini et al., 2005; Perrault et al., 2005; Stanford et al., 2005; Stein et al., 1989, 1996; Welch & Warren, 1986; Meredith & Stein, 1986; Wallace et al., 1996; Wallace et al., 1992). For example, a study by Sumby and Pollack ascertained that our perception of speech for the most part benefits from additional visual features when the auditory signal is flawed by background noise (Sumby & Pollack, 1954). A more recent applied study investigated the effect of bimodal presentation of naturally weak and unclear stimuli (Doll & Hanna, 1989). These researchers examined underwater sonarsystems in which information about approaching water crafts was simultaneously presented visually as well as auditory. With increasing distance between the system and surrounding water crafts, the information in both modalities was gradually impaired by noise. They discovered that crossmodal presentation accounted for a gain of 1.1 dB over unimodal presentation. Thus, while objects were only partially or not at all detectable when presented either visually or auditory alone, they were still recognizable with crossmodal presentation.

Besides a wealth of neurophysiological studies in nonhuman organisms, many behavioral studies on the principle of inverse effectiveness conducted with human participants are confined to human speech perception or applied experiments, as indicated by the two examples cited above (Sumby & Pollack, 1954, Doll & Hanna, 1989). Moreover, for an investigation of the principle of inverse effectiveness, most researchers manipulate the stimulus intensity of sensory input by weakening or attenuating the test stimuli and frequently emphasize only the effect of stimulus intensity in one modality on another instead of a mutual influence of both modalities.

The current study investigates the effect of stimulus intensity as manipulated by degrading stimulus features in both the auditory and visual modality on multisensory integration at a behavioral level. To ensure a broad generalizability, a standardized forced-choice paradigm was employed based on the study by Giard and Peronnet described above (Giard & Peronnet, 1999). The same categorization task as in their study is used as well as the stimuli and object classification into object A and object B. In addition, in order to create various complexity levels of stimulus recognizability, both auditory and visual stimuli were partially or fully degraded by auditory and visual noise, respectively. To overcome the strong contrast between no degradation and complete degradation, 5 different levels of stimulus degradation were created, beginning with an intact stimulus and using ascending intervals of 25% degradation until complete irrecognizability. This scaling ensures a sound differentiation of the different amounts of information that can be extracted from the stimuli in each separate degradation level. Basic indicators for stimulus recognition are accuracy rate, the proportion of objects that are correctly identified, as well as reaction time, the time it takes to execute a response.

As concluded by Giard and Peronnet, a common finding in behavioral studies is that stimuli defined on the basis of crossmodal sensory features are detected or identified more readily and accurately than stimuli specified by unimodal information alone. This effect is generally known as the "redundant-signal effect" (Miller, 1986), as congruent information in the crossmodal condition is provided redundantly by more than one sense. Based on these findings, it is assumed that in the current study reaction times will be faster and accuracy rates higher for stimulus recognition when the object is defined by crossmodal auditory and visual features compared to unimodal auditory or visual features alone.

Second, based on the studies mentioned above that give support to the principle of inverse effectiveness (Frassinetti et al., 2005; Bolognini et al., 2005; Perrault et al., 2005; Stanford et al., 2005; Stein et al., 1989, 1996; Welch & Warren, 1986; Meredith & Stein, 1986; Wallace et al., 1996 ; Wallace et al., 1992), it is expected that stimulus recognition is enhanced for crossmodal stimulus presentation when one of the two or both stimulus modalities are partly degraded. However, considering the idea of a threshold value of activation energy necessary to induce a memory trace, it is assumed that at a given level of degradation, where the accessible information is not sufficient to exceed the threshold value, the benefit of multisensory integration ceases. Thus, it is assumed that for crossmodal stimulus presentation, reaction times and accuracy rates for object recognition will be superior for moderate levels of stimulus degradation compared to more or no degradation and unimodal presentation.

Method Experiment 1

Participants: The sample in the first experiment consisted of 13 participants, who were either volunteers or psychology students at the University of Twente, Enschede. They enrolled through sona-systems, the sampling pool for participants of the University of Twente. Ages ranged from 18 to 26 with a mean age of 20.7 years; 6 (46%) of the participants were male and 7 (54%) were female. All students reported normal or corrected- to- normal stereoscopic vision and normal or corrected- to- normal hearing. Participants were naive to the appearance of the displayed stimuli and the purpose of the experiment. Most of the students received course credit for participation.

Stimuli and apparatus: The stimuli employed in the experiment were auditory signals and visual geometrical objects adopted from the stimuli used by Giard & Peronnet, and Fort et. al as described above (Giard & Peronnet, 1999; Fort et. al., 2002). The auditory signals consisted of a high frequency 540 Hz and a low frequency 560 Hz tone that were presented for 223 ms each, including a 5 ms fade-in and a 5 ms fade-out. Tones were presented via two regular loudspeakers located to the left and right of the monitor. The visual objects consisted of a white vertical and a horizontal ellipse that were centrally presented on a black background screen. Both ellipses were produced by a deformation of a white fixation dot with a diameter of 5 cm. They were identical in size and were formed by a 10% modification of the length of the horizontal and vertical diameters of the circle, respectively.

In both modalities, stimuli were degraded to five different degrees in intervals of 25% that ranged from intact stimuli (0% degradation) to an irrecognizable condition (100%

degradation). Degradation in the auditory modality was achieved by presenting the tones in varying degrees of intensity of background noise, ranging from 0 dB for the intact stimulus features to 96 dB for the 100% degradation condition, with equal intervals of 24 dB for the remaining degradation levels. In the visual modality, degradation was achieved by superimposing a quadrate on the ellipses that contained varying degrees of visual noise in the form of black and white pixels, containing no noise in the intact condition and increasing in equal intervals of 12.5 % of pixels that changed colour from black to white and vice versa. Thus, in the 100% degradation condition, 50% of pixels had changed color.

A standard Pentium IV class computer running E-prime 1.1 experimental software package (Psychology Software Tools, Inc.) was used for stimulus presentation, timing, and acquisition of the necessary response data. Stimuli were presented on a 17 Inch Philips 107-T5 display running at 800 by 600 pixel resolution in 32 bit colour, refreshing at a rate of 60 Hz. The experiment was run in a special secured mode to ensure rigorous response presentation and stimulus registration. Input was given through a standard mouse and keyboard.

Procedure: The study has been approved by the institutional ethical committee. Participants were seated in front of a computer screen in a quiet, artificially lit research laboratory room. The viewing distance was approximately 60cm, but it was not explicitly controlled for. Participants received information about the study and instructions regarding the experiment in advance of testing, and were asked to give informed consent. Instructions were given digitally via an introduction screen that required participants in a forced-choice paradigm to associate the high 560 Hz tone with the vertical ellipse (object A) and the low 540 Hz tone with the horizontal ellipse (object B). When a high tone and/or a vertical ellipse were presented, participants were required to press the “A” key (for object A). When a low tone and/or a horizontal ellipse were shown, they were instructed to press the “B” key (for object B). On each separate trial, stimuli were either presented in isolation, thus only one of the two tones or one of the visual objects, or in combination - the audio-visual condition - where a tone was presented simultaneously with a visual feature. On 66 % of trials, stimuli were presented in isolation (33 % for each modality with an even number for both tones and objects) and on 33 % of trials in combination. In the audio-visual condition, stimuli were only presented congruently, thus resembling either object A or object B. On each trial, stimuli could either be fully detectable (0% degradation), completely unrecognizable (100% degradation) or degraded to 25%, 50%, or 75%.

Thus in total, there were three levels of the factor Presentation Mode (auditory versus visual versus audio-visual) and five levels of the factor Stimulus Degradation (0%, 25%, 50%, 75%, and 100%). For the crossmodal presentations, every combination of modality and degradation level was employed. The order of stimulus presentation was randomized across participants. Participants' action and performance were constantly monitored during the whole session via a permanent closed-circuit camera installed behind the participants back.

After completion of the testing session, participants were fully debriefed as to the purpose of the experiment. They were thanked for taking part in the experiment and most of them were granted course credits for participation.

Test phase: The experiment began with the introduction screen described above. By pressing any key, the participants could themselves decide when to initiate the testing session. The test phase consisted of a total of 480 trials equally distributed across 8 blocks. When no response was registered after 1000 ms, participants were informed that no response had been detected and the next trial was initiated automatically. The blocks were separated by a feedback screen which informed participants of their performance, as indicated by mean response times as well as the mean proportion of correct responses made during the completed block as averaged over all trials in the block. This feedback and the accompanied break enabled participants to track their performance, to release their concentration and to direct their attention towards the following block. Participants could themselves determine the length of the break and initiate the next block by pressing any key. This procedure continued until the participants had completed all blocks.

Data Analysis: The data of experiment 1 were merged with E-Prime E-Merge and were subsequently exported using E-DataAid. Data analysis was conducted by means of SPSS 16.0 for Windows. Multivariate repeated measures analyses were run on the data. The factors Stimulus Degradation with five levels (0%, 25%, 50%, 75% and 100%) and Presentation Mode with three levels (auditory signal present versus visual signal present versus audio-visual signals present) served as dependent variables whereas reaction time and accuracy measures each were employed as independent variables. All variables in the experiment were within-subject factors. The data were subsequently analyzed leaving out the 100% degradation condition, as, due to the fact that it did not contain any information, it differed from all other degradation conditions.

Results Experiment 1

A Komolgorov-Smirnov test indicated that the results were distributed normally, confirming the appropriateness of using a parametric multivariate measure for analyzing the data. Two separate multivariate repeated measures analyses (ANOVAs) were run on the data, one with accuracy (in percentage of correct responses) and the other with reaction time (in ms) as dependent variables.

Accuracy measure: The data of experiment 1 were subjected to a repeated measures multivariate ANOVA with level of Stimulus Degradation (0%, 25%, 50%, 75%, and 100%) and Presentation Mode (auditory, visual, and audio-visual) as within-subject variables. The multivariate ANOVA revealed a main effect of Presentation Mode (F (2, 11) = 14.86, p< 0.001) and a significant interaction effect between Presentation Mode and Stimulus Degradation (F (8, 5) = 5.19, p < 0.05). The same analysis with exclusion of the 100% degradation condition revealed only a significant main effect of Presentation Mode with F (2, 11) = 11.68, p< 0.005.

Pairwise comparison of Presentation Mode demonstrated that, in the audio-visual condition, participants obtained higher accuracy rates (M = 0.89) than in both conditions where a stimulus was presented in isolation (M = 0.80 for the auditory and M = 0.86 for the visual condition). However, only the difference between the audio-visual and the auditory condition reached significance (p< 0.05). Neither the difference between the two single presentation conditions (p = 0.47) nor the difference between audio-visual and visual presentation (p = 0.11) were significant.

Pairwise comparison of the interaction between level of Stimulus Degradation and Presentation Mode illustrated that, whereas participants obtained the highest accuracy rates in the audio-visual condition regardless of degradation level, the general superiority of the visual over the auditory condition could not be demonstrated for the 100% degradation level. In this condition, where stimuli were completely irrecognizable, participants obtained the lowest accuracy rates across all presentation modes (M = 0.83). In the audio-visual condition, most stimuli were correctly identified when they were degraded to 25%. For the visual stimulus presentation, the highest accuracy rates were obtained when stimuli were intact (0% degradation), followed by 25% and 75% degradation, as can be seen in Figure 1a.

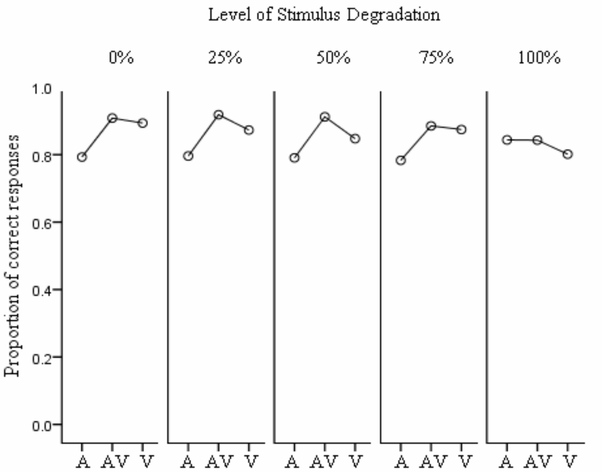

Figure 1a: The effect of level of Stimulus Degradation and Presentation Mode on accuracy rates. Most objects were correctly identified in the audio-visual condition, in particular when stimuli are moderately degraded.

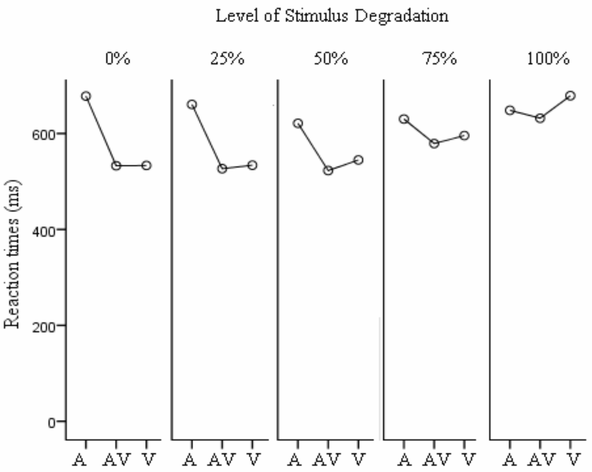

Figure 1b: The effect of Presentation Mode and level of Stimulus Degradation on reaction times. Reaction times were fastest in the audio-visual condition. When stimuli are highly degraded, they are not as readily detected as when degraded to a lesser degree

Reaction time measure: The same repeated measures ANOVA with level of Stimulus Degradation and Presentation Mode as independent within-subject factors that was run on the accuracy data was also conducted for the reaction time measures. It revealed a main effect for Presentation Mode (F (2, 11) = 23.51, p< 0.001). However, the interaction between Presentation Mode and level of Stimulus Degradation did not reach significance with F (8, 5) = 3.37, p< 0.098. The same repeated measures analysis without the 100% degradation condition revealed a main effect of Presentation Mode (F (2, 11) = 24.52, p< 0.001) as well as a significant interaction between Presentation Mode and Stimulus Degradation (F (6, 7) = 4.42, p< 0.05).

Pairwise comparison of Presentation Mode illustrated that participants responded fastest when stimuli where presented in both modalities simultaneously (M = 558.72 ms) compared to unimodal presentation (M = 647.75 ms for auditory and M = 577.43 ms for visual presentation). Only the auditory condition differed significantly from the other two conditions, as indicated by a comparison of mean differences. For both the audio-visual and the visual condition, reaction times were fastest and almost equally high for the stimuli that are intact (0% degradation) or masked for either 25% or 50% (M = 581.47, M = 573.79, and M = 563.07, respectively), followed by the 75% degradation condition (M = 601.71) and the condition where stimuli are completely masked (M = 653.11). These findings are illustrated in Figure 1b.

Discussion Experiment 1

As expected on the basis of general findings from behavioral and neurophysiological studies of multisensory integration (Miller, 1986; Giray & Ulrich, 1993; Hughes et al., 1994), the present results indicate that subjects detected and identified objects more rapidly as well as more accurately when both visual and auditory features were presented simultaneously compared to unimodal presentation. Furthermore, for both the reaction time and the accuracy data, the worst performance was obtained in the auditory unimodal condition. This might be due to a discrepancy in the difficulty of relying on visual and auditory features for response selection. Following the testing session, participants reported identifying objects on the basis of the unimodal auditory stimuli to be more demanding than specifying objects on the basis of visual features.

For the second hypothesis, it was expected that, in accordance with the principle of inverse effectiveness, in the audio-visual condition subjects' performance is superior to all other conditions when stimuli are moderately degraded. Indeed, participants recognized most objects correctly in the crossmodal condition when the stimuli were moderately degraded (25%). For the reaction time measure, the highest responsiveness was likewise obtained in the crossmodal condition when stimuli were degraded for 50%. This trend could be confirmed by a second ANOVA in which the 100% degradation condition was excluded. Thus, when the information that is accessible from the stimulus of one modality is incomplete, the contribution of a simultaneously presented stimulus from another modality enhances accuracy and speed of object recognition beyond a value that is obtained when both stimuli are intact.

Experiment 2

The results of experiment 1 raise some questions regarding the procedure and some general issues about the method employed. Reports by subjects as well as the difference in performance between visual and auditory features suggest that the auditory features were harder to discriminate than the visual features. Furthermore, many subjects reported the use of the “A” and “B” keys for response registration as requiring an uncomfortable hand position. Lastly, participants had to learn the association between the visual and auditory features and the classification into objects A and B with the the associated key press during the actual testing session, leading to a reduced performance in the first trials. These findings prompted a second experiment, in which those three aspects were addressed and refined.

Method Experiment 2

Participants: For the second experiment, a new sample was drawn which consisted of 16 participants, who were either volunteers or psychology students at the University of Twente, Enschede. They enrolled through sona-systems, the sampling pool for participants of the University of Twente. Ages ranged from 18 to 57 with a mean age of 25; 5 (31%) of the participants were male and 11 (69%) were female. All students reported normal or corrected-to- normal stereoscopic vision and normal hearing abilities. As in experiment 1, participants were naive to the appearance of the displayed stimuli and the purpose of the experiment. Most of the students received course credit for participation. Due to technical difficulties during the data acquisition phase that call into question the reliability of the results, four participants were excluded from any analysis.

Stimuli and apparatus: The auditory stimuli in experiment 2 were identical to those employed in experiment 1. Due to the fact that after completion of experiment 1 participants evaluated the auditory stimuli to be harder to discriminate than the visual stimuli, it was decided to adapt the level of difficulty of discrimination of the visual stimuli to that of the auditory stimuli. The reason for adapting the difficulty of the visual stimuli to that of the auditory instead of the other way around is justified by the overall high accuracy rates obtained in experiment 1. Degradation levels and the technique of degrading stimuli resembled those of experiment 1.

The hard- and software employed for data presentation and response acquisition remained equally unchanged for experiment 2.

Procedure: The same research lab as in experiment 1 was used for testing.

Participants received a short instruction via the computer screen and gave informed consent. For the second experiment, only minor changes were introduced to the forced-choice paradigm and procedure of experiment 1. One of these modifications concerned the implementation of a familiarization phase proceeding the actual testing session. This was to enable the participants to get used to the stimuli and to learn the audio-visual association between the tones and the objects as well as the related key press in advance of testing. Furthermore, the familiarization phase prevented the introduction of a learning effect during the testing session that would have been characterized by a gradual decline in reaction time over trials. Due to the fact that many participants that had completed experiment 1 reported the use of the "A" and "B" keys for entering responses as requiring to engage them in an uncomfortable hand position, in the second experiment the "A" and "B" keys were replaced by the "Z" and "M" keys, respectively. This enabled participants to hold their right and left hand on an equal height. Thus, taken together, three changes were made to experiment 1: the difficulty of discriminating the visual objects was matched to that of the auditory signals; a familiarization phase was introduced; and, for reasons of convenience, the response keys were changed from "A" and "B" to "Z" and "M" , respectively.

After completion of the testing session, participants were fully debriefed as to the purpose of the experiment. They were thanked for taking part in the experiment and most of them were granted course credits for participation.

Familiarization phase: The familiarization phase began with an introduction screen that informed the participants of the task. The same stimuli were employed as in experiment 1, except for the modification of the visual stimuli that were matched to the difficulty of the auditory stimuli. The practice phase consisted of 2 blocks, containing 60 trials each. The order of trials was distributed randomly across the blocks and the blocks were randomized across participants. After completion of the familiarization phase, the test session was initiated by the researcher.

Test phase: The stimuli and task were the same as in the familiarization phase. The test phase consisted of a total of 480 trials equally distributed across 6 blocks. Again, the single blocks were separated by a feedback screen that contained information about the individual participant's performance of the completed block, as indicated by mean reaction time and percentage of correct responses given. The procedure continued until responses to all 480 trials were registered.

Data Analysis: The data of the remaining 11 participants were merged with E-Prime E-Merge. With E-DataAid the data were exported and subsequently analyzed by use of SPSS 16.0 for Windows. As in experiment 1, two separate repeated measures analyses with the factors Stimulus Degradation with five levels (0%, 25%, 50%, 75% and 100%) and Presentation Mode with three levels (auditory signal present versus visual signal present versus audio-visual signals present) were conducted, one on the reaction time measures and the other on the accuracy rates. As before, the data were subsequently analyzed leaving out the 100% degradation condition.

Results Experiment 2

The data obtained from the familiarization phase were not subjected to any analysis because the training session solely served the subjects for familiarizing with the stimuli and learning the association between the features representing the two different objects A and B.

As in the first experiment, the results were distributed normally which legitimated the use of a parametric multivariate measure for analyzing the data. Again, two separate multivariate repeated measures analyses were run on the data with reaction time (in ms) and accuracy (in percentage of correct responses) as independent within-subject variables.

Accuracy measure: The same 5 x 3 (0%, 25%, 50%, 75%, and 100% degradation; auditory, visual, and audio-visual presentation) repeated measures ANOVA that has been conducted on the data of experiment 1 was also run on the accuracy data of experiment 2. In general, the mean percentages of correct responses given were slightly lower than in the first experiment, especially for the visual stimuli, which is probably due to the adjustment of the discrimination level of the visual stimuli to that of the more demanding auditory signals.

As in the analysis of the accuracy data for experiment 1, the multivariate ANOVA revealed a main effect of Presentation Mode (F (2, 9) = 24.45, p< 0.001). However, in contrast to experiment 1 the interaction between Presentation Mode and Stimulus Degradation did not reach significance (F (8, 3) = 6.77, p = 0.072), yet a trend in the expected direction was present. The same ANOVA was conducted leaving out the 100% degradation condition. It revealed only a significant main effect of Presentation Mode with F

(2, 9) = 13.21, p< 0.005).

Pairwise comparison of Presentation Mode resembled the findings of experiment 1, by demonstrating that participants gave most responses correctly during audio-visual presentation (M = 0.87), as compared to the visual and the auditory condition with 79.7% and 75.1% of accurate responses, respectively. Mean differences of the presentation mode conditions revealed that only the difference between the auditory and the audio-visual stimulus presentation reached statistical significance (p<0.05). A graph of these findings can be seen in Figure 2a.

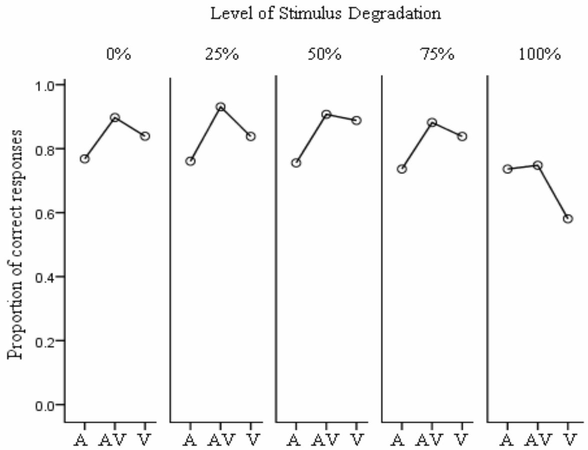

Figure 2a: The effect of Presentation Mode and level of Stimulus Degradation on accuracy data. Subjects gave most responses correctly with crossmodal presentation when stimuli were completely irrecognizable, accuracy rates dropped to just above chance level in the visual condition.

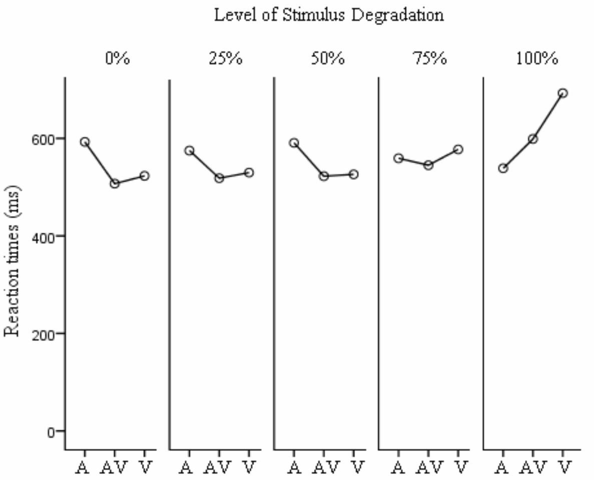

Figure 2b: The effect of level of Stimulus Degradation and Presentation Mode on reaction times. Contrary to expectations, responses were fastest in the crossmodal condition when stimuli were intact.

Reaction time measure: A repeated measures ANOVA was run on the reaction time data with level of Stimulus Degradation and Presentation Mode as independent within-subject factors. Mean reaction times were generally lower than in experiment 1. The fact that, compared to experiment 1, participants were quicker to make a response while at the same time responses were on average slightly less accurate, suggests a trade-off between accuracy and reaction time. Participants seem to adopt different strategies of responding, either laying more emphasis on speeded responses, where the likelihood for making a wrong response is high, or on responding accurately, which seems to require more time.

In conformity with the findings of experiment 1, the multivariate ANOVA displayed a main effect of Presentation Mode (F (2, 9) = 7.13, p< 0.05). In contrast to experiment 1, this effect is significant only at the 0.05 level. Unlike experiment 1, a significant interaction between Presentation Mode and level of Stimulus Degradation was demonstrated (F (8, 3) = 17.87, p< 0.05). As before, participants gave the slowest responses when stimuli were irrecognizable (M = 610.13 ms). However, contrary to expectations, fastest responses were obtained in the audio-visual presentation condition when stimuli were intact and not with moderate degradation. The same ANOVA with exclusion of the 100% degradation level likewise revealed a main effect of Presentation Mode (F (2, 9) = 25.12, p< 0.001), as well as a significant interaction effect (F (6, 5) = 5.15, p< 0.05).

Pairwise comparison of the Presentation Mode conditions yielded a superior responsiveness for the audio-visual condition (M = 538.38 ms) as compared to the two unimodal presentation conditions with a mean reaction time of 571.31 ms for the auditory and 569.93 ms for the visual condition. These results are in accordance with the data of experiment 1. However, in contrast to the earlier findings, where only the auditory condition differed significantly from the other two conditions, mean differences indicated that in experiment 2 the crossmodally presented stimuli are significantly different from the two unimodal presentation conditions.

Regarding the interaction between level of Stimulus Degradation and Presentation Mode, pairwise comparisons revealed that, contrary to expectations, participants gave the fastest responses in the crossmodal condition when objects are intact (507.17 ms). When objects are completely irrecognizable as indicated by a degradation of 100%, responsiveness was highest in the visual condition (692.88 ms). A graph of these findings is depicted in Figure 2b.

Discussion Experiment 2

With regard to the first hypothesis, which stated that objects defined on the basis of crossmodal cues are recognized faster and more accurately compared to unimodal presentation, the results of experiment 2 resemble those of experiment 1. Simultaneous presentation of auditory and visual cues enhanced participants' performance over unimodal auditory or visual presentation. As before, the slowest and most inaccurate responses were given in the auditory condition. However, the performance difference in the two unimodal conditions was minimal compared to experiment 1, indicating that the adjustment in complexity of the visual features to that of the auditory features has been successful. An additional analysis without the 100% degradation condition did confirm the previous results, lending strong support to the assumptions made in hypothesis 1.

Contrary to the expectations made in hypothesis 2 based on the evidence for the principle of inverse effectiveness, there was no significant superiority of accuracy for audio-visual presentation for moderately degraded stimulus features over the unimodal presentation conditions and the remaining degradation conditions. Likewise, the subsequent analysis of the data exclusive of the 100% degradation condition did not provide support for the second hypothesis. For the accuracy data, the results obtained were in the expected direction, not significant though. However, the effect was close to a significance level of 0.05, and might have reached significance if the data of four subjects had not had to be excluded from any analyses. Thus, with the inclusion of additional participants, the effect would have been stronger, lending support to the second hypothesis. An additional explanation for the sparse significance might lie in the fact that the complexity of the visual features was increased for experiment 2, enhancing the difficulty of object discrimination.

As opposed to the accuracy data, the interaction between Presentation Mode and level of Stimulus Degradation did reach significance for the reaction time measure, not in the expected direction though. In contrast to the principle of inverse effectiveness, subjects' responsiveness was highest in the crossmodal presentation condition when objects were intact. One potential explanation for this finding might be that reaction times were influenced by altering the response keys for experiment 2 from “A” and “B” to “Z” and “M” , which allowed for more comfort in hand position in experiment 2 as reported by participants. This is in accordance with a general decrease in reaction times for experiment 2 and might as well have led to a different pattern of results. A second feasible assertion for the results of the reaction time data obtained contrary to the expectations from hypothesis 2 and the findings of experiment 1 is the introduction of a familiarization phase in experiment 2. As participants were allowed to get accustomed to the features associated with object A and B, they did not have to learn the associations during the testing session as in experiment 1, leading to a heightened overall responsiveness and might as well have led to a response pattern contrary to expectations. As with the adjustment of response keys, the introduction of a training session hampers a direct comparison of the results from experiment 1 and 2.

General Discussion

The objective of the current study was to investigate whether degraded auditory and visual stimuli would influence multisensory integration in accordance with the principle of inverse effectiveness at a behavioral level. In a forced-choice categorization task participants were required to identify one of two objects on the basis of auditory, visual or audio-visual features that were partly or fully degraded on some trials.

Based on general findings of multisensory integration (Miller, 1986; Giray & Ulrich, 1993; Hughes et al., 1994), it was expected that crossmodal audio-visual presentation is superior to unimodal auditory and visual presentation, as indicated by increased accuracy and speeded response times. This hypothesis was confirmed for both measures in either experiment.

According to the principle of inverse effectiveness (Meredith & Stein, 1993), the second hypothesis stated that participants obtained the best performance regarding reaction time and accuracy rates for crossmodal audio-visual presentation when stimuli are moderately degraded. The findings of experiment 1 confirmed the hypothesis for both measures. The accuracy data of experiment 2 were in the expected direction but did not reach significance whereas the reaction time data did not validate the hypothesis. Thus, it seems that under some conditions, participants' recognition of a degraded stimulus presented in one modality can be increased by simultaneous presentation of a related stimulus feature presented in another modality, even when this feature is likewise degraded.

To explain the insignificant results of the accuracy data for experiment 2, it is assumed that the adjustment in complexity of the visual features makes response decisions more difficult. Furthermore the exclusion of four participants due to difficulties during response acquisition might have lowered statistical power. However, these explanations cannot account for the reaction time data of experiment 2 and the unexpected findings for hypothesis 2 in both experiments. An interpretation could be given by some general problems in experimental and stimulus design in the current study. In both experiments it was not controlled for hands and fingers that subjects used to enter responses which might have differentially influenced responsiveness but not accuracy. The introduction of a familiarization phase in experiment 2 led to a diminished learning effect during the testing session that should have been more articulate in experiment 1. To permit a valid comparison between the experiments it should be considered to re-analyze the data leaving out the first block of experiment 1 in which subjects learned the association between the object features that was achieved through the training session in experiment 2. Furthermore, a separate comparison of the first half with the second half of all trials should be taken into consideration for each experiment to account for a general effect of training and a potential effect of fatigue during the second half of trials.

The finding that, for both experiments, the best performances were obtained for moderately degraded objects regardless of modality of presentation, casts doubt about the validity of results confirming the principle of inverse effectiveness of hypothesis 2. Crossmodal presentation of object features is expected to provide a behavioral advantage in recognition over unimodal presentation for moderately degraded objects. However, the fact that subjects performance was generally better for moderately degraded objects regardless of unimodal or crossmodal presentation, questions the advantage of crossmodal presentation. On the other hand, this outcome together with the findings from the condition in which stimuli are completely degraded, that subjects nevertheless obtained accuracy rates as high as 70%, indicate problems inherent to the experimental design. Two possible explanations would be that a particular sequence was adhesive in the way the randomization of trials was computed or that participants employed an effective guessing strategy. A more feasible explanation is that even with 100% stimulus degradation, subjects were still able to detect movements of the deformation of the dot to either of the two ellipses which would call into question the procedure employed for degrading stimuli and the mask itself. It could also be possible that the way the degradation was achieved is not in conformity with or different from the way the brain naturally refines and processes degraded information in our day-to-day life. Therefore, additional research is needed to validate the hypotheses with different and more complex objects, with more than two associations between stimulus features, and with a different means of degrading stimuli. Instead of superimposing black and white pixels to achieve a uniform degradation, objects could be masked by degrading various parts of the objects as done in studies by Biederman (Biederman, 1987).

Furthermore, it would be interesting to investigate the influence of attention on multisensory integration with degraded objects, as it is indicated that attention as well as audio-visual integration provide a perceptual enhancement by raising sensitivity to certain external events, suggesting common mechanisms underlying both crossmodal and attentional processes (Driver & Spence, 1998; Macaluso et al., 2000; Weissman et al., 2004; Talsma et al., 2006). The effect of attention could directly be tested by instructing subjects to attend to just auditory, just visual, or both features presented simultaneously. When subjects are not explicitly given instructions on which features to focus their attention, as in the current study, it seems that it is the uncertainty of individual modalities, as evoked by various levels of degradation that determines to what extent information from each modality is considered when identifying an object, which is similar to the principle of inverse effectiveness.

Another influence of the relative contribution of visual and auditory features in multisensory integration is the extent to which subjects are considered to be visually or auditory dominant, having a tendency to generally base their decisions on visual over auditory features and vice versa. In the current study, it was not controlled for this possible effect. In a follow-up study subjects could in a pretest be classified as either visually or auditory dominant and the groups subsequently compared.

Finally, an explanation for the principle of inverse effectiveness might be that most stimuli that we are confronted with in our external environment are not pure and contain a particular amount of noise. For example, without any technical modification we usually perceive harmonics, not pure tones. Thus, it seems that our brain is used to some sort of noise embedded in the input that we experience, leading to a general superiority of noisy input compared to pure information. It would be interesting to investigate whether musicians, who are schooled in the perception of pure tones, are superior in the detection of undegraded crossmodally presented stimuli over degraded ones, showing a lessened or no effect of the inverse effectiveness rule. To what extent these questions can be answered and the hypotheses be validated will be the challenge of future research.

Conclusion

Besides a replication of general findings on multisensory integration, the main finding of the current study is that under specific conditons, crossmodal presentation of audio-visual features enhances performance of accuracy and reaction time for moderately degraded objects. Thus, despite some inconsistencies in the results that should be addressed in future studies, the findings of the current study indicate that the principle of inverse effectiveness, which is commonly validated by means of weak or attenuated stimuli, can also be confirmed with degraded stimuli. Additionally, an enhancement of recognition of degraded stimulus features presented in one modality is not only achieved by features presented via another modality, but even when these features are likewise degraded. Besides adding to the literature of multisensory integration, these findings have implications for our day-to-day life, in particular for situations that highly rely on noisy audio-visual information like transport and car- and air traffic.

References

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

Biederman, I. (1987). Recognition-by-Components: A Theory of Human Image Understanding. *Psychological Review, 94, 115-147.*

Bolognini, N., Frassinetti, F., Serino, A., & Ladavas, E. (2005). Acoustical vision of below threshold stimuli: interaction among spatially converging audiovisual inputs. *Experimental Brain Research*, 160, 273 – 282.

Buzsaki, G. (1989). Two-stage model of memory trace formation: a role for "noisy" brain states. *Neuroscience,* 31, 551 - 570.

Calvert, G.A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex,* 11, 1110 - 1123.

Doll, T.J. & Hanna, T.E. (1989). Enhanced detection with bimodal sonar displays. *Human Factors*, 31, 539-550.

Driver, J. & Spence, C. (1998) Crossmodal attention. *Current Opinion in Neurobiology*, 8, 245-253.

Fort, A., Delpuech, C., Pernier, J., & Giard, M.H. (2002). Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cerebral Cortex,* 12, 1031 - 1039.

Frassinetti, F., Bolognini, N., Bottari, D., Bonora, A., & Ladavas, E. (2005). Audiovisual integration in patients with visual deficit. *Journal of Cognitive Neuroscience*, 17, 1442 - 1452.

Frens, M.A., Van, O.A., & Vander, W.R. (1995). Spatial and temporal factors determine auditory–visual interactions in human saccadic eye movements. *Perceptual Psychophysics,* 57, 802 - 816.

Giard, M.H. & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, 11, 473 - 490.

Giray, M., & Ulrich, R. (1993). Motor coactivation revealed by response force in divided and focused attention. *Journal of Experimental Psychology: Human Perception and Performance,* 19, 1278 - 1291.

Gordon, B. G. (1973). Receptive fields in the deep layers of the cat superior colliculus. *Journal of Neurophysiology,* 36, 157 - 178.

Hughes, H.C., Reuter, L.P., Nozawa, G., & Fendrich, R. (1994). Visual–auditory interactions in sensorimotor processing: saccades versus manual responses. *Journal of Experimental Psychology : Human Perception and Performance,* 20, 131 - 153.

Jay, M. F., & Sparks, D. L. (1984). Auditory receptive fields in primate superior colliculus shift with changes in eye position. *Nature,* 309, 345 - 347.

King, A. J., & Palmer, A. R. (1985). Integration of visual and auditory information in bimodal neurons in the guinea-pig superior colliculus. *Experimental Brain Research,* 60, 492 - 500.

Macaluso, E., Frith, C.D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science*, 289, 1206 – 1208.

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

Meredith, M.A., & Stein, B.E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. Journal of *Neurophysiology*, 56, 640-662.

Miller, J.O. (1986). Time course of coactivation in bimodal divided attention. *Perceptual Psychophysics,* 40, 331 - 343.

O'Hare, J.J. (1991). Perceptual integration. *Journal of the Washington Academy of Sciences,* 81, 44 - 59.

Peck, C. K. (1987). Auditory interactions in cat’s superior colliculus: Their role in the control of gaze. *Brain Research,* 420, 162 - 166.

Perrault, T.J., Vaughan,J.W., Stein, B.E.,& Wallace, M.T. (2005) Superior colliculus neurons use distinct operational modes in the integration of multisensory stimuli. *Journal of Neurophysiology*, 93, 2575 – 2586.

Perrott, D.R., Saberi, K., Brown, K., & Strybel, T.Z. (1990). Auditory psychomotor coordination and visual search performance. *Perceptual Psychophysics,* 48, 214 - 226.

Stanford, T.R., Quessy, S., & Stein, B.E. (2005). Evaluating the operations underlying multisensory integration in the cat superior colliculus. *The Journal of Neuroscience*, 25, 6499-6508.

Stein, B.E., Meredith, M.A., Huneycutt, W.S., & McDade, L. (1989). Behavioral indices of multisensory integration: orientation to visual cues is affected by auditory stimuli. *Journal of Cognitive Neuroscience,* 1, 12 - 24.

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

Sumby, W.H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212 – 215.

Talsma, D., Doty, T.J., Strowd, R., & Woldorff, M.G. (2006). Attentional capacity for processing concurrent stimuli is larger across sensory modalities than within a modality. *Psychophysiology*, 43, 541 - 549.

Wallace, M.T., Meredith, M.A., & Stein, B.E. (1992). Integration of multiple sensory modalities in cat cortex. *Experimental Brain Research*, 91, 484 – 488.

Wallace, M.T., Wilkinson, L.K., & Stein, B.E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of Neurophysiology,* 76, 1246 - 1266.

Weissman, D.H., Warner, L.M., & Woldorff, M.G. (2004). The neural mechanisms for minimizing cross-modal distraction. *The Journal of Neuroscience*, 24, 10941-10949.

Welch, R.B. & Warren, D.H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological bulletin*, 3, 638-667.

Welch, R.B., & Warren, D.H. (1986). *Intersensory interactions Handbook of perception and human performance*. New York, Wiley-Interscience, 1 - 36.

Zahn, J.R., Abel, L.A., & Dell'Osso, L.F. (1978). Audio-ocular response characteristics. *Sensory Processes,* 2, 32 - 37.