Zur Homepage der Dissertation

SPATIAL AND TEMPORAL FACTORS IN VISUAL-AUDITORY INTERACTION

vom Fachbereich 5
Philosophie/Psychologie/Sportwissenschaft
der Universität Oldenburg
zur Erlangung des Grades eines

Doktors der Philosophie

angenommene Dissertation

Heike Heuermann

geboren am 5. Juni 1971 in Twistringen

Erstreferent: Prof. Dr. Hans Colonius
Korreferent: Prof. Dr. Volker Mellert

Tag der Disputation: 20. Dezember 2002

Zusammenfassung

In der vorliegenden Arbeit wurden Latenzen und Trajektorien von Sakkaden auf visuelle und auditorische Ziele in verschiedenen Experimenten untersucht.

In einem auditorischen Lokalisationsexperiment waren die Versuchspersonen aufgefordert, Augenbewegungen auf den wahrgenommenen Herkunftsort des akustischen Reizes auszuführen. Die beobachteten Augenbewegungen hatten häufig einen kurvigen Verlauf, der auf eine oft verzögert einsetzende und häufiger korrigierte Vertikalbewegung des Auges zurückzuführen ist. Bei Augenbewegungen auf visuelle Ziele war dies nicht zu beobachten.

In einem Reaktionszeitexperiment sollten die Versuchspersonen anschließend Augenbewegungen auf einfache visuelle Ziele ausführen. Auf zusätzlich dargebotene akustische Reize sollte hingegen nicht reagiert werden. Sowohl die räumliche Anordnung der Stimuli (horizontaler und vertikaler Abstand) als auch die zeitliche Reizkonfiguration (Intestimulusintervall) wurden variiert. Es zeigte sich, dass Latenzen auf bimodale Stimuli grundsätzlich kürzer waren als bei rein visueller Stimulusdarbietung. Dieser intersensorische Bahnungseffekt war grundsätzlich umso größer, je früher der akustische Reiz relativ zum visuellen Reiz dargeboten wurde und je näher beide Stimuli räumlich zueinander lagen. Der Einfluss des horizontalen Abstands zwischen visuellem und auditorischem Reiz war hierbei unabhängig vom Interstimulusintervall. Im Gegensatz dazu konnte ein Einfluß des vertikalen Abstands nur beobachtet werden, wenn der akustische Reiz vor dem visuellen abgespielt wurde.

Zur Modellierung der Daten wurde das Zwei-Stufen Modell zur visuell-auditorischen Interaktion von Colonius und Arndt (2001) herangezogen. Dieser probabilistische Ansatz beschreibt die getrennte periphäre Verarbeitung visueller und auditorischer Reize als Wettlauf in einer ersten Verarbeitungsstufe. Der Ausgang dieses Wettlaufs entscheidet über das Ausmaß der räumlichen Bahnung in der zweiten, gemeinsamen Verarbeitungsstufe. Es wurde eine Erweiterung des Zwei-Stufen Modells auf zwei räumliche Dimensionen vorgenommen. Die vom Modell geschätzen Werte für die Dauer der periphären und zentralen Verarbeitung stimmen gut mit physiologischen Beobachtungen überein.

Summary

The present work investigates latencies and trajectories of saccades toward visual and auditory targets in various experiments.

In an auditory localization experiment, participants had the task to perform eye movements toward the perceived position of acoustic stimuli. The observed trajectories of the eye movements were often bow-like, which can be explained by the vertical movement, that often starts with a short delay (relative to the horizontal movement) and is corrected for once or twice. Visual target directed eye movements, however, did not show these features.

In a reaction time experiment, the participants were presented with simple visual target stimuli that could be accompanied by accessory auditory stimuli. The task was to perform a saccade to the visual targets while responses to auditory stimuli should be suppressed. The spatial configuration of the stimuli (horizontal and vertical distance) could be varied as well as the temporal arrangement (inter-stimulus interval). It turned out that latencies toward bimodal stimuli were significantly shorter than latencies toward unimodal visual target stimuli. This Intersensory Facilitation Effect was generally the more pronounced, the more the presentation of the auditory accessory stimulus preceded target presentation and the smaller the spatial distance between both stimuli was. The influence of horizontal distance between visual and auditory stimulus turned out to be independent of the interstimulus interval. In contrast, an influence of vertical distance could only be observed if the auditory stimulus was presented before the visual target.

The Two-Stage Model of visual-auditory interaction introduced by Colonius & Arndt (2001) was applied to the data. This probabilistic approach describes at the first stage the peripheral processing of visual and auditory stimuli as a race. The amount of spatial facilitation at the second stage of combined processing is determined by the the outcome of the first stage race. The original Two-Stage Model was extended to two spatial dimensions and fitted to the data of the reaction time experiment. The data fits yielded reasonable estimates for peripheral and central processing times, if compared to data from single cell recordings.

Contents

1 Introduction

Latencies to target stimuli are usually significantly smaller if an additional (non-informative) accessory stimulus is presented in close temporal and/or spatial relationship with the target. Various psychophysical and physiological studies have suggested different explanations for this intersensory facilitation effect (IFE), for example, attentional or warning effects or multisensory information integration.
In humans, interaction between the visual and the auditory system is of special importance, with vision usually dominating perception while audition seems most important for the detection of warning signals. Temporal aspects of various visual-auditory interaction effects, particularly with respect to reaction times, have been investigated quantitatively since the early 60ies. Several models reaching from simple statistical facilitation to attentional effects have been considered to explain the findings (for an early review see Nickerson (1973)). More recently, quantitative analyses of the effect of spatial interstimulus relations on saccadic reaction time (SRT) have been performed (Frens, Van Opstal & Van der Willigen 1995, Harrington & Peck 1998, Hughes, Nelson & Aronchick 1998). A general observation is that the extent of intersensory facilitation increases with spatial proximity. Using pure tones or noise signals, Frens et al. (1995) showed that the perceived stimulus position has a significant influence on visual-auditory interaction. They suggested a linear relation between radial interstimulus distance and the amount of facilitation. Colonius & Arndt (2001) proposed a two-stage model describing both temporal and spatial aspects in visual-auditory interaction. In their study, saccadic reaction times toward visual targets decreased the more the auditory accessory preceded target presentation and the smaller the spatial distance between both stimuli was.

A central role in physiologically based explanations and models for visual-auditory interaction has been assigned to the Deep Layers of the Superior Colliculus (DLSC) (Meredith & Stein 1986). The SC is a brain stem nucleus participating in integrative mechanisms in the visual and visuo-motor system and is of substantial importance for reflexive movements in response to a stimulus. Moreover, it has also been found to be a prominent stage in intersensory integration. Afferents from different modalities converge here, building spatial saliency maps which are in close register with each other. It is worth to be noted at this point that the SC has so far been the only mammalian brain structure showing a *topographically* organized auditory map at all. It remains however unclear how this map is constructed. Unlike the retinotopic maps of the visual and oculomotor system, an internal representation of the auditory environment is based upon the calculation of interaural intensity- and phase-differences and on the analysis of direction-specific spectral cues resulting in a craniocentric reference system.
Interaural time- (or phase-) and intensity difference analysis can be assigned to the EE- ans EI-cells of the Superior Olivary Complex (SO), sending their efferents to the Inferior Colliculus which in turn projects to the SC. Hence, binaural information processing already takes place in subcortical areas, which means that its processing can be assumed to be more “hardwired” and faster. Unfortunately, the details of how auditory elevation judgment is performed and which neural mechanisms exactly are involved are not known yet. It is however clear that the direction-dependent spectral modifications of the signal caused by the listener’s pinna folds (Head Related Transfer Functions, HRTF), represent the substantial cue for localization in the elevation domain. Hence, we deal with a spectral pattern recognition problem of which physiological data indicate that it seems to be performed by a different neural pathway involving thalamic

and cortical areas. This idea is supported by a behavioral study of Hofman, Van Riswick & Van Opstal (1998) showing that participants were able to learn to adequately use a new set of HRTFs (corresponding to a pair of new ears) without losing the capability of correctly localizing with their "genuine ears". Hofmann compared this effect with learning a new language. In another psychophysical study, Frens & Van Opstal (1995) found that auditory saccades are often curved, in contrast to visually evoked eye movements. Auditorily guided trajectories frequently show a strong horizontal trend at first which is supplied by an "elevation correction movement" after a period of about 30 msec. This, too, indicates certain temporal constraints in elevation determination (in contrast to azimuth estimation) and suggests separate mechanisms in the analysis of binaural and monaural location cues. If this holds true, temporal and spatial parameters in visual-auditory interaction should be seen as independent factors, but it can be expected that the amount of specific spatial interaction depend on the SOA actually chosen.

The goal of the present work is to reveal and analyze those aspects of visual and auditory information integration that are involved in processing both azimuth and elevation cues. Therefore, three experiments are executed, providing information over the processing of uni- and bimodal stimuli. In two unimodal auditory experiments, participants are instructed either to perform a target-directed saccade or to give an undirected simple response by turning the eyes to a permanently illuminated point beneath the fixation point (outside the range of any possible target position). In the third experiment, both visual and auditory stimuli are presented, using the focused attention paradigm. Participants here have the task to perform a directed response to the *visual* target as fast and as accurate as possible, while any acoustic signal could be ignored. Both spatial and temporal interstimulus parameters were varied in randomized order during this experiment. Each experiment was performed in a virtual auditory environment and under free field listening conditions.

An extension of the Two-Stage Model by Colonius & Arndt (2001), taking horizontal *and* vertical interstimulus distance as two independent variables of spacial interaction, will be presented. The extended Two Stage Model designates a race of information processing in *three* (instead of two) parallel sensory channels on the first stage: there are one visual and two auditory (azimuth and elevation information) competitors. Like in the original model, integration of spatial information may occur on the second level, in which the amount of integration on the second stage depends on the outcome of the race.

2 Conceptual background

2.1 Models of intersensory interaction: temporal factors

Early findings

Most early descriptions of intersensory effects can be found in the Russian literature, of which works by Urbantschitsch (1888, 1903) shall be taken as representative here. Urbantschitsch investigated several quantitative and qualitative aspects of visual-auditory and auditory-somatosensory interactions. For example, by testing from how far the ticking of a watch could still be perceived by a subject, he found out that auditory threshold was higher when keeping the eyes closed than with open eyes. He furthermore reported that presenting an additional auditory stimulus influenced color perception: high-frequency tones caused colors to be perceived brighter, lower frequency stimuli made them seem darker.
The question of the general *direction* of intersensory effects, i.e. whether accessory stimuli more raised or lowered perception thresholds was discussed quite controversely. For a long time, there was the prevailing assumption that the sensitivity of one sensory organ toward an "adequate" stimulus would be reduced by simultaneous presentation of stimuli in other modalities , in which *"[...] the greater the the stimulus the stronger tends to be the inhibitory power of the corresponding sensation"* (Jacobsen 1911). So, Heymans (1904) found that electrical stimulation of the hand increased auditory detection threshold, with auditory sensitivity becoming worse the more intense the electric shock was. By contrast, Newhall (1923) reported intersensory *facilitation* if auditory click stimuli were presented simultaneously to visual targets (the latter were then judged brighter and more intense). Newhall introduced the idea that attentional effects could be responsible for intersensory effects. It seems however also obvious that the actual choice of target- and accessory stimuli has a crucial impact on the amount and direction of multisensory effects.

Unfortunately, most early studies were too inconsistent in their methodology and their results were thus often not replicable. First systematic quantitative investigations on multimodal integration were performed by Todd (1912). In his complex, far reaching work, he studied manual reaction times presenting light, noise and electrical stimuli either alone, pairwise or all together. Furthermore, the temporal intervals between the stimuli and the order of presentation were varied. In short, his findings can be summarized as follows.

1. There are different reaction times to stimuli from different modalities: reactions toward auditory stimuli are fastest, followed by reaction times to electric shocks, which are in turn shorter than reaction times to visual stimuli.

2. In the case of *simultaneous presentation*, subjects responded faster to a triple of stimuli than to a pair of them. Reactions toward a pair of stimuli are moreover faster than to either of the stimuli. The amount of reaction speed up, if a defined stimulus or pair of stimuli is added, depends on the reaction time to the stimulus (or pair of stimuli, respectively) alone: the shorter it is, the stronger will be the induced reaction speed up.

3. In the case of *successive presentation*, reaction times to multiple stimuli might not be reduced or might even be longer than to unimodal stimulation (as for sound – electric – light presentation in this order with temporal delays, which produced longer reaction times than for light alone). Reactions to multiple stimuli are successively reduced by reducing interstimulus intervals.

Although Todd had already collected all the pieces of information needed to create a stringent theory of multimodal integration, he did not put these findings into a more systematic account.

Energy summation approaches

The innovative work on intersensory facilitation was probably performed by Hershenson (1962) who, inspired by the study of Todd, was able to demonstrate the systematic connection between unimodal and bimodal response times. He measured manual response time (telegraph key pressing) under unimodal visual or auditory and under bimodal stimulation. Participants had the task to react to either stimulus they received first (redundant signals paradigm). Varying the onset asynchrony between visual and auditory stimuli, Hershenson found shortest manual response times for bimodal stimulation if their temporal disalignment corresponded exactly to the difference of the respective unimodal response times (see Figure 1).

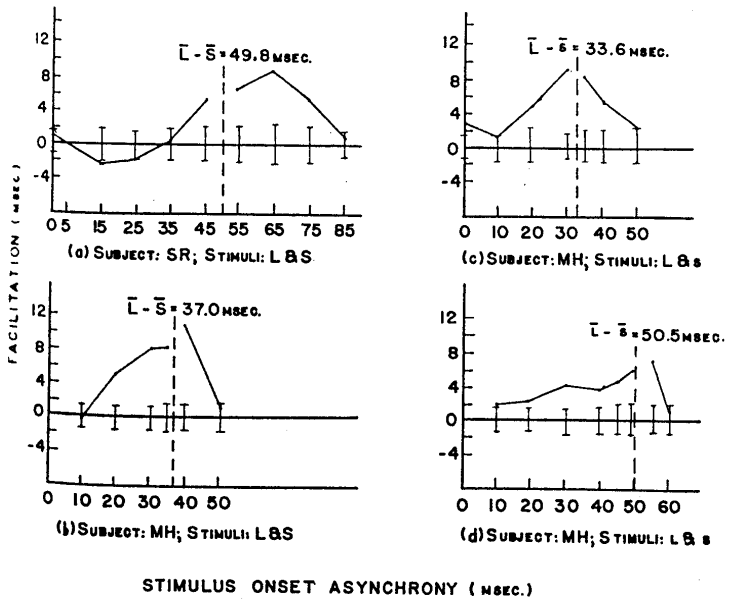


Figure 1: Intersensory facilitation as a function of stimulus onset asynchrony (SOA). Vertical dashed line: difference between mean unimodal visual and mean unimodal auditory reaction time. Facilitation was calculated by substracting bimodal reaction time from unimodal *visual* reaction time if SOA was *greater* than the unimodal reaction time difference and by substracting from auditory reaction time in other cases. Note that facilitation values are maximal for SOAs equal to unimodal reaction time difference. Data from Hershenson (1962)

Figure 2 illustrates why Hershenson claimed there was really a speed up in reaction time and not merely a temporal triggering of bimodal response by the faster auditory response. If there

was no intersensory interaction, the reaction time function should follow the black dotted line which is calculated as follows. The temporal limits for bimodal interaction are given by SOA values between 0 and 40 msec. At SOA=0 msec, bimodal response time equals the measured unimodal auditory reaction time of about 120 msec, i.e. in this case the response is in fact triggered by the acoustic stimulus. For SOA≥40 msec, the auditory stimulus is presented too late to be considered any more, hence the "bimodal" response is a purely visual one with the respective latency of about 160 msec. In between these limits, bimodal response time rises constantly as a function of SOA with a slope of 1.
The red line represents the data found by Hershenson. Obviously, the bimodal response times found are smaller than it could be explained by any "triggering" assumption. The area between the two graphs is the region of facilitation, with the distance between the graphs being a measure for the amount of facilitation at a given SOA. Again, it can be seen that the bimodal effect is strongest with SOA compensating for the proposed neural visual-auditory delay of 40 msec.

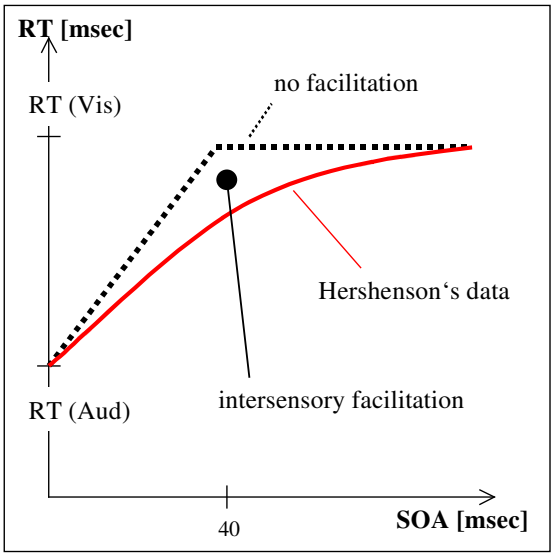
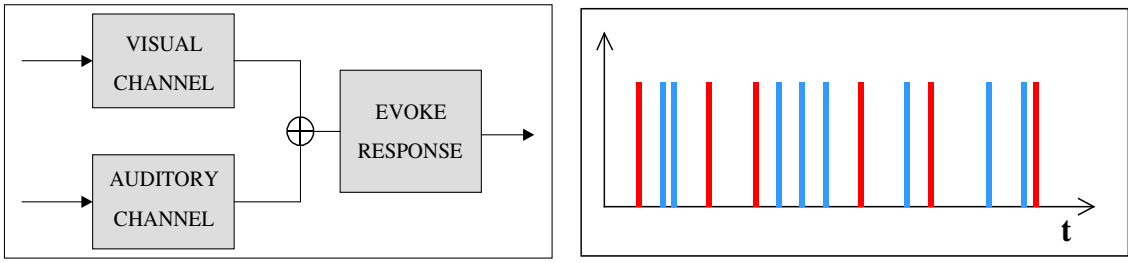


Figure 2: Graphical illustration of the bimodal reaction times measured by Hershenson (solid curve) and the values expected if there was no facilitation (dotted curve). The region between the graphs indicates the amount of facilitation. Compare with Figure 1. After Raab (1962)

Hershenson suggested that neural information from parallel organized sensory channels merges somewhere on a higher processing level and that the resulting summed energy leads to higher sensory arousal and thus to faster responses. Apparently, the "ideal" SOA of 40-50 msec found in the experiments just compensated for the distinct sensory processing times. According to Hershenson, a pure warning effect as unique cause of response speed up should be excluded, as otherwise equivalently high bimodal facilitation effects would have appeared at any temporal offset. Figure 3 shows a sketch of the energy integration model proposed by Hershenson. It was the basis for several other, often more detailed models like the superposition model by Bernstein (1970) or the coactivation models by Miller (1982), Grice, Canham & Boroughs (1984), or Diederich & Colonius (1987). Although the different authors propose different loci of energy integration (so, Bernstein suspects early sensory stages of processing, Miller proposes the decision stage to be the most likely one, and Diederich & Colonius presented data pointing to

effects in the motor component), the general idea is always the same: A response to a stimulus or a set of stimuli is initiated as soon as a certain criterion level is exceeded. Combined energy of multiple activations leads to faster achievement of a set criterion than a single activation can ever (see Figure 3b).



(a) (b)

Figure 3: a: Model of intersensory facilitation due to energy summation across different sensory channels, as suggested by Hershenson. A response is evoked if total neural energy reaches the criterion level. Thus, the summed energies of multiple channels lead to reduced reaction times. b: Another possible depiction of Hershenson's approach as often used in superposition models. Pieces of information (or, in a more physiologically based approach, spikes) from different channels (black and grey) are merged in one channel. If the total number of strokes in the common channel becomes larger than a certain criterion (which is also represented by a number), a response is evoked.

Independent race models

Raab (1962) generally approved the idea of enhanced sensory sensitivity through multimodal information integration, but rather considered it as a possible explanation for perception threshold changes than for response speed up. The latter, he argued, could be completely explained by probabilistic assumptions, if sensory processing times are assumed as (statistically independent) random variables. As the mean of the minimum of two or more random variables is always smaller than or equal to the minimum of the means, the expected reaction time under multimodal stimulation must be smaller than in any of the respective unimodal conditions.
Let unimodal auditory response time be normally distributed with expectation value (i.e. mean reaction time) μ and the unimodal visual response time be normally distributed with expectation value $\mu + d$, in which d denotes a temporal shift in order to account for the longer sensory processing of visual stimuli and/or for stimulus onset asynchrony SOA:

$$RT_A \sim N\ (\mu) \quad \text{and} \quad RT_V \sim N\ (\mu + d).$$

Hence, in the case of physically simultaneous stimulus presentation, d is around 40, if the visual stimulus however precedes the auditory by 50 msec, d is around -10, and so on.
In a redundant signals paradigm, subjects can respond to either stimulus they first perceive, that is, the bimodal reaction time is given by the minimum of the random variables:

$$E(RT_{AV}) = E(min(RT_A, RT_V)).$$

If both distributions do not overlap (for example, in case of large stimulus onset asynchronies), the bimodal reaction time equals the faster of the two unimodal response times. If the distributions however do overlap, bimodal response time is distributed as the minimum function of the two unimodal distributions (see Figure 5). The expectation value of the minimum of multiple distributions is smaller than or equal to the expectation value of either of distribution compared with another, that is

$$\begin{aligned} E(RT_A) &\geq E(min(RT_V, RT_A)) \quad \text{and} \\ E(RT_V) &\geq E(min(RT_V, RT_A)), \end{aligned}$$

in which the the difference between unimodal and bimodal expectation value is the larger, the more the unimodal distributions overlap. From this, it directly follows that the minimum of both expectation values must also be larger or equal to the expectation value of the minimum:

$$min(E(RT_A), E(RT_V)) \geq E(min(RT_A, RT_V)).$$

According to these assumptions, intersensory facilitation is in fact based on multichannel information processing in the brain, but unlike with Hershenson's model, information is not combined. All channels are organized in parallel, clearly separated from each other. Approaches of this kind are called **independent race models**, since the processes in the different sensory channels are regarded as 'competitors' in a race with the first process being finished determining response time.

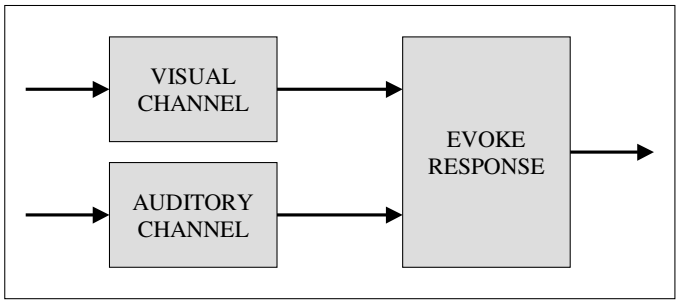


Figure 4: Structural diagram of Raab's Independent Race Model. Like with the energy summation model, sensory processing is assumed to take place in parallel channels, but here the different pieces of information converge independently from each other at a common response evoking stage. The first process arriving starts response activity. Due to the statistical reasons outlined in the text, response activity will be evoked the faster, the more sensory channels are activated.

Figure 5, the plot of a MatLab-based simulation, demonstrates the effect of statistical facilitation graphically. The blue curves represent unimodal response time distributions at different SOA values. The red dotted curve is the minimum density function. As both unimodal distributions more and more overlap (from top to bottom), the minimum distribution shifts toward smaller values. In case of complete overlap, the expectation value of the minimum is visibly smaller than either of the unimodal distributions' means.

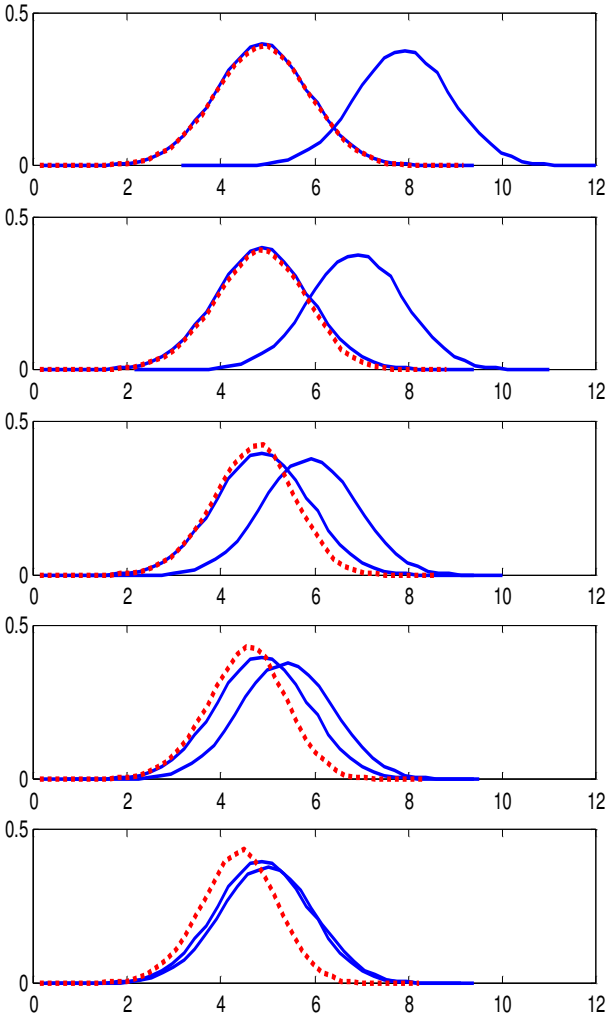


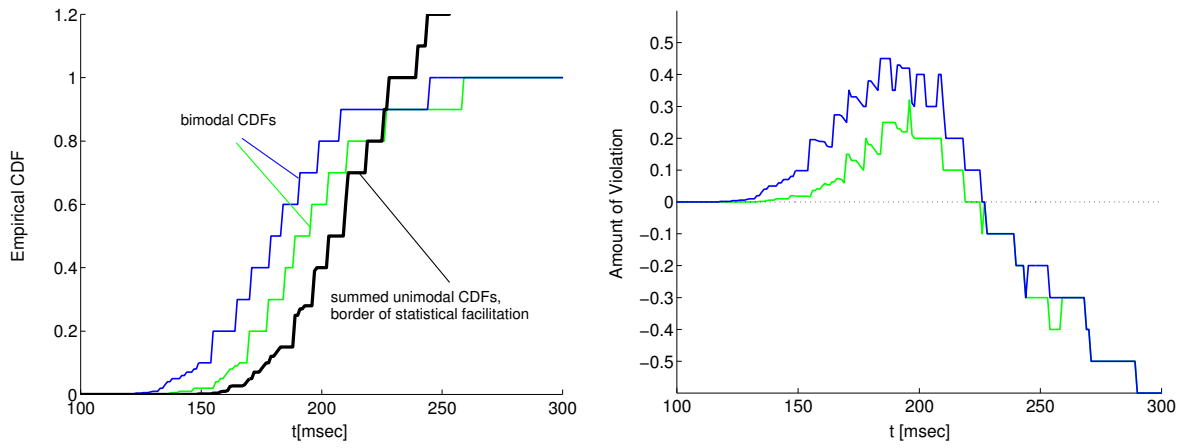
Figure 5: Statistical facilitation due to probability summation. Black lines: distributions of the unimodal response times. Grey dotted line: distribution of the minimum random variable. Obviously, the expectation value for the minimum distribution is less than the expectation value of either unimodal distribution on if there is a large overlap between the unimodal distributions.

Raab's idea is indeed intriguing not only because of its straightforwardness and it is still used in more recent approaches of intersensory modelling – though, it does not explain the whole amount of response enhancement, as could be proved by Miller (1982).
The Miller-inequality Miller determines an upper boundary for statistical facilitation, that is given by the sum of the cumulative distribution functions (CDFs) of the unimodal reaction times:

$$F(RT_{AV}) \leq F(RT_A) + F(RT_V)$$

A comparison with the data of bimodal reaction time experiments however shows that the response speed up is often visibly larger than predicted by an Independent Race approach (see Figure 6). Since statistical reasons as an exclusive assertion have to be rejected, Miller concluded that response facilitation is only completely interpretable with the assumption of intersensory

convergence.



(a) (b)

Figure 6: a: Cumulative Distribution Functions (CDF) of bimodal reaction times (thin lines) and the sum of unimodal cumulative distribution functions (thick line), which is the "border" of statistical facilitation due to the Miller-inequality. b: Amount of violation of Miller's inequality, calculated from the difference of bimodal CDF and summed unimodal CDF ("border function"). It turns out from these plots that pure statistical facilitation has to be ruled out as single explanation for these bimodal reaction times, as the Miller inequality is violated by a large amount. Own data.

Preparation enhancement approaches

A third approach to explain bimodal response speed up, quite contrary to the energy summation assumption, was the arousal hypothesis strongly defended by Nickerson (1973). In his Preparation Enhancement Model, each stimulus processed in its specific sensory channel in order to evoke a specific response. Any other stimuli are processed in parallel and only influence response processing by general activation of some kind of "readiness mechanisms". Information from different sensory channels is not combined.
Nickerson argued that some general (nervous) arousal due to the presentation of an accessory (warning) signal might promote response preparation and so (after computing of the specific response) lead to a faster response *execution* ("response preparation may be proceeded into parallel with the specific response", (Bernstein 1970)). So, he pointed out that there simply was no need for the assumption of multi-channel information convergence but preparation enhancement alone was sufficient to explain all facilitation effects. Nickerson's view is based on the finding that even in a focused attention paradigm, that is if participants are requested to respond to stimuli in one sensory channel only and regard any other stimulus as non-target or distractor, there can still be found a significant response speed up. Intersensory facilitation is also a function of SOA in these cases, but unlike with the Redundant Signals Paradigm, IFE is now

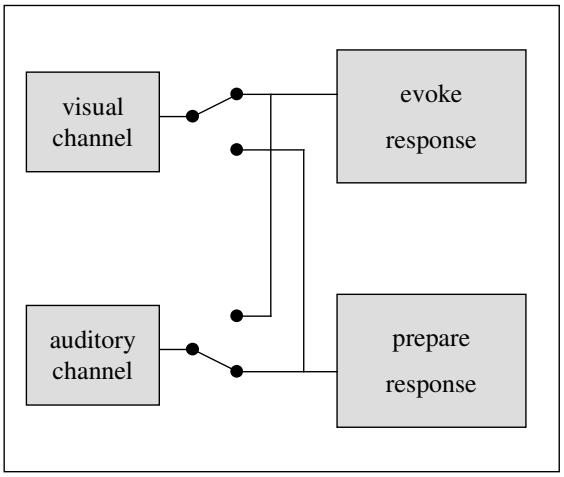
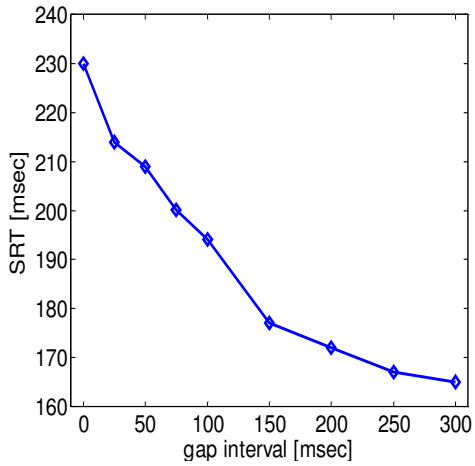


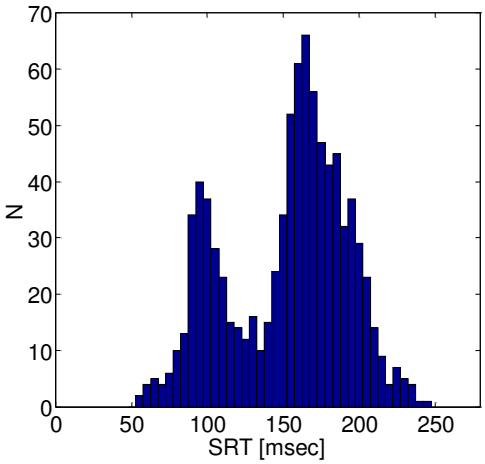
Figure 7: Model of intersensory preparation enhancement as supposed by Nickerson (1973). Only the visual stimulus is able to evoke a response in this case; however, the auditory stimulus can still prepare "response readiness" and thus lead to a significant latency speed up. After Nickerson (1973).

monotonically rising with preceding accessory stimulus.

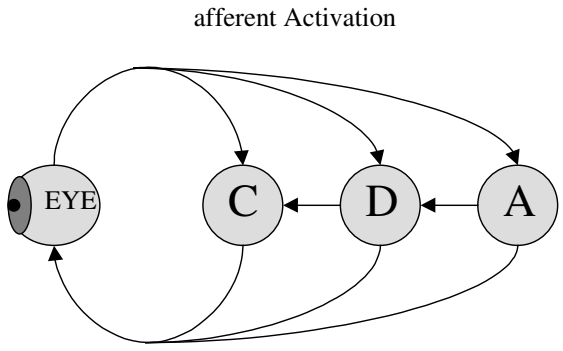
This kind of response acceleration can in fact be found in other, unimodal experiments, too. Nickerson's model shows analogies to ideas of Fischer ((Fischer & Rampsberger 1984, Fischer & Weber 1993) who found reduced saccadic latencies under a quite different kind of paradigm. If in purely visual trials the fixation point disappears before target onset, mean reaction times are reduced dramatically. The amount of this gap-effect is the larger, the longer the temporal delay between fixation offset and target onset is chosen (see Figure 8a).
If not only mean reaction times are regarded but also their distributions (Figure 8b), it becomes clear that the gap-effect can be traced back to the emergence of so-called express-saccades with extremely short latencies (around 100 msec). Fischer regardes them as a separable saccade population that appears due to (pre-motor) fixation release, or in other words by the transition from a fixation (no-answer) mode into a response mode. By this attentional shift, a response can already be prepared in an unspecific way. This idea is up to some degree very alike the preparation enhancement hypothesis by Nickerson, although it is somewhat more complex if regarded in detail.



(a)



(b)



(c)

Figure 8: a: Saccadic latencies as a function of gap duration. Y-Axis: mean reaction times to unimodal visual targets. X-axis: temporal delay between fixation offset and target onset. After Kopecz (1995). b: typical distribution of saccadic latency under the gap paradigm. c: Fischer's Three Loop Model of saccade generation. For an explanation see text.

The Three-Loop Model (1987) by Fischer (Figure 8c) proposes that the preparation of a directed saccade includes at least three operations: attention shift (A), a decision process (D), and calculation of the movement (C). The processes are organized serially, i.e. the execution of one operation cannot be performed before the preceding process is finished. Any Visual input however directly activates all three processes (or process stages), and hence may shorten processing times at all three stages while the general order of processing is kept. In general, it should be possible to transform these assumption to inter-sensory models, too. That means, any input might have an influence on every processing stage, or at least on some of them, but there might

be a combination of different effects in intersensory interaction, too.

Nozawa, Reuter-Lorenz & Hughes (1994) presented a model trying to explain both intersensory and unimodal gap effects in saccadic eye movements. Concluding from the finding that bimodal effects and the gap effect are additive, they proposed parallel organized sensory processing channels terminating in neural summation and subsequent serially connected pre-motor and motor mechanisms, in which gap-effects are caused by pre-motor mechanisms. Figure 9 shall illustrate this graphically.

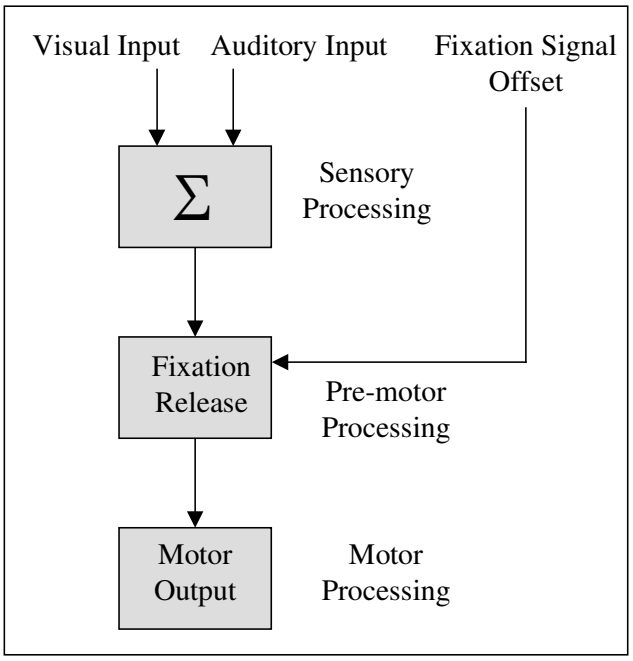


Figure 9: Model of Nozawa et al. attempting to explain both intersensory facilitation and express saccades. Bimodal effects might be explained by summation of sensory energy on the first stage. Warning effects like the gap-effect leading to express saccades find their origin in pre-motor processes on a subsequent stage. Both effects might occur alone or in combination. After Nozawa et al. 1994.

In case of multimodal stimulation, information from different sensory channels is combined at an early common stage like in Hershenson's approach (i.e. there is "real" neural summation and not simple statistical facilitation). Express-saccades emerge due to processes at a pre-motor stage. In analogy to Fischer's Three-Loop Model, this stage is organized *serially* to the stage of (inter-) sensory processing. Fixation point offset leads to activation of saccade preparation (by cutting the gaze to the fixation point) so that a target driven saccade can be performed faster. Hence, fixation signal processing is performed in parallel to stimulus processing (as it is also sensory processing), but fixation release solely occurs if stimulus-driven sensory processing is completed. Thus, the gap becomes effective in a serial manner only.

Nozawa et al. (1994) regarded pre-motor facilitation as merely due to preliminary fixation point offsets. It might however also be possible that the simple presentation of an accessory stimulus

serves as some kind of warning and thus as fixation releasing event. Results of Munoz & Corneil (1995) seem to confirm this assumption, as these authors found less intersensory facilitation if there was an additional temporal gap between fixation and target signal while bimodal effects increased if there was an overlap. With this hypothesis, an additional auditory stimulus might in fact play a double role in visual-auditory interaction experiments. If so, both energy summation and preparation enhancement have to be considered as mechanisms of intersensory interaction and the remaining question would only be in how far these both mechanisms add up or interact in bimodal reaction time experiments.

2.2 Models of intersensory interaction: spatial factors

In most earlier studies on multisensory interaction, main concern was with the influence of the interstimulus interval. Possible effects by spatial arrangement were described only rarely and could hardly be explained by the models existing at those times. However, experiments on choice reaction time (Bernstein, Clark & Edelstein 1967, Bernstein & Edelstein 1971, Simon & Craft 1970) yielded an influence of the spatial stimulus arrangement in so far as accessory stimuli coming from the same side as the target produced facilitation while opposite-side accessories did not. This effect of "stimulus-response compatibility" was taken as a strong hint that total stimulus energy was not an explanation for bisensory facilitation. A shift of attention or response readiness (in the sense of preparation enhancement) toward a certain place, caused by the nontarget, was discussed as a the more plausible explanation. Yet, Nickerson's response-readiness approach (Nickerson 1973) in its primary form does not provide any explicit spatial coding of information either. Moreover, if it were solely warning effects leading to bimodal response time reduction, how can then the influence of an accessory following the target stimulus by 100 msec or more be explained? More recent findings of topographically organized spatial maps in the sensory and motor systems allows a solution in this conflict. If sensory information is not summarized up across all input, but according to the saliency maps' coordinates, the finding of spatial effects in multimodal facilitation can be explained much more easily.
Harrington & Peck (1998) were among the first to demonstrate a systematic relation between spatial interstimulus distance and the amount of intersensory facilitation. As described in the previous section, the extent of violation of the Miller inequality can be taken as measure for sensory information integration. Harrington & Peck investigated saccadic latencies in unimodal and bimodal reaction time experiments using different spatial distances. Their results are displayed in Figure 10. It shows the violation of Miller's inequality by the Cumulative Distribution Function (CDF) difference function introduced in Fig. 6b for each spatial condition in a different plot. It can be seen clearly that the amount of violation (and thus of intersensory facilitation) decreases in a nearly monotonic manner with rising spatial distance.

Comparable results had been presented three years earlier by Frens et al. (1995), who investigated spatial and temporal effects of visual-auditory interaction in saccadic eye movements. Starting from their findings that bimodal saccadic reaction time (SRT) was (a) generally smaller than under unimodal visual condition, (b) decreasing with spatial proximity, and (c) decreasing with increasing temporal gap between (preceding) auditory accessory and visual target presentation, they suggested a physiologically inspired model including mechanisms of general warning as well as summation (or integration) of *spatially encoded* sensory information. In this model based on an approach on saccade generation by Munoz & Wurtz (1993*a*, 1993*b*) (see Figure 11, a prominent role is assigned to the Deep Layers of the Superior Colliculus (DLSC) and the brain stem Omnipause Neurons (OPN). The SC is also a brain stem nucleus participating in integrative mechanisms in the visual and visuo-motor system and is of substantial importance for reflexive movements in response to a stimulus. Afferent nerve fibers from different modalities converge here, building spatial saliency maps which are in close register with each other. Moreover, there are several multimodal cells that respond best if there is input from more than one sensory modality at a time. Cell recordings from the SC have shown that firing rates can be increased massively, if multimodal stimuli were presented that were in close temporal and spatial proximity (for an overview see Meredith & Stein (1986). Any disparity led

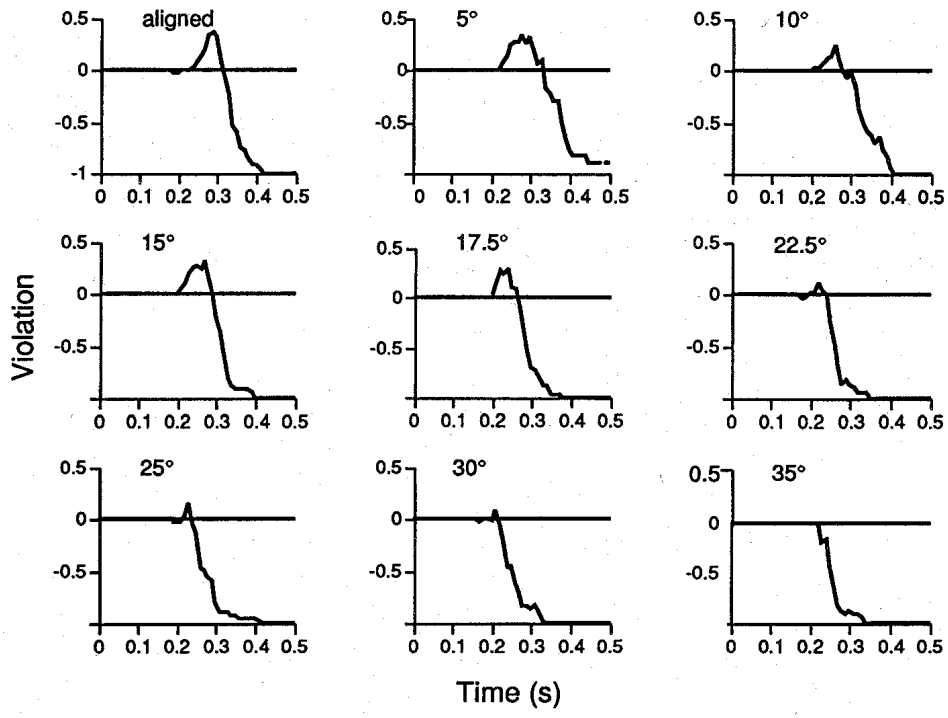


Figure 10: Results from the experiments of Harrington & Peck on spatial influence on intersensory facilitation. The amount of violation of Miller's inequality, which can be regarded as measure for intersensory integration, is calculated as difference between bimodal CDF and summed unimodal CDFs and plotted for different spatial interstimulus distances. It is the larger, the closer both stimuli are presented. From Harrington & Peck (1998).

to a gradual decrease in these cells' activities. Frens et al. propose the following mechanisms and circles for saccade generation in bimodal stimulation. Both visual and auditory information is projected (via different pathways) onto sensory saliency maps in the DLSC. If both stimuli are presented sufficiently close, they might arouse activities in interconnected cell populations (encoding identical positions or regions in space) or in respective multimodal cells. Thus, the joint activity from both stimulations might be much larger than to single stimulation. If a certain threshold is reached, the DLSC cells in turn excite Burst Generator cells that are responsible for saccade generation. Thus, spatially encoded energy summation (i.e. Hershenson's hypothesis in an only slightly modified form) can in fact be put on a physiological basis in saccade generation. A second mechanism for auditory saccade facilitation can be found in inhibitory brain stem circles that are also suspected to play a role in Fischer's Gap-effect. Munoz & Wurtz propose that fixation neurons in the Superior Colliculus detain saccade generation by inhibiting the Burst Generator neurons (via the brain stem Omnipause neurons). Fixation neuron activity may on the other hand be reduced by several factors: fixation offset (not regarded here), general warning effects, or neural activity in the DLSC build-up neurons due to stimulus presentation. Hence, the auditory stimulus may have an impact by two ways. On the one hand, summed activity in the DLSC build-up neurons leads to stronger inhibitory projections to the fixation centers. Moreover, the auditory alone might also decrease fixation neuron activity through unspecific

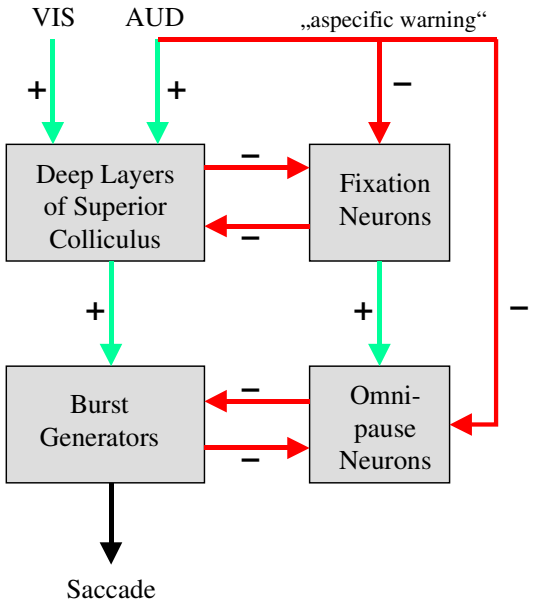


Figure 11: Frens et al.'s model of visual-auditory interaction in saccade programming. For an explanation see text. After Frens, Van Opstal, and Van der Willigen (1995b).

inhibitory input. Any way of fixation neuron inhibition will however lead to faster achievement of Burst Generator threshold activity and thus to generally faster responses. This latter kind of mechanism fits best to the Preparation Enhancement Model proposed by Nickerson (1973).
In short, Frens et al.'s approach considers the auditory accessory stimulus working both as unspecific alerting signal inhibiting fixation-related neurons and thereby indirectly exciting saccade-related centers and as an additional spatial information that might enhance (pre-) motor activity, so that the activation level for a directed saccade is exceeded faster.

Another very useful approach describing temporal and spatial aspects of saccade programming was presented by Findlay & Walker (1999). Their model proposes two parallel streams of information flow and motor command, each running through a hierarchy of levels from high-level cognitive control levels down to brain-stem movement-decision and motor command levels. The crucial concept is the idea of permanent competitive interaction between centers in the so-called WHERE-pathway processing movement commands and fixation-related mechanisms in the WHEN-pathway. The WHEN-mechanisms carry a single-valued (i.e. non spatial) signal and act like a gates; they trigger movement orders in the WHERE-pathway. The WHERE-system contains topographic mappings in form of move- or saliency maps. Both centers are interconnected in inhibitory circuitries. The result is a push-pull situation in which – apart from brief dynamic-equilibrium states – either center completely dominates. These mechanisms confirm that the oculomotor system remains stable (fixating) except for those situations in which an "optimal" saccade is confirmed. A saccade is elicited when activation in the fixate-center decreases below a certain threshold. The metrics of the saccade then depend on the move-centers in WHERE-pathway: the point of maximum activity in the saliency map determines the movement's direction. Several factors may have an influence on the dynamic processes between

fixation and move centers. Central events (like fixation point offset) affect the fixation center only, while peripheral events inhibit fixation activity *and* enhance activity at the corresponding point in the move system's saliency maps. Due to the inhibitory circuitries, activation of the move-centers promotes fixation disengagement even more. Hence, any (accessory) stimulus plays a double role: it adds spatially encoded information to the saliency maps and prepares a response by reducing fixation activity. Although Findlay & Walker originally suggested their model for unimodal visual processing, an extension to multimodal environments has proved to yield reasonable results (Kopecz 1995, Bastian, Riehle, Erlhagen & Schöner 1998, Trappenberg, Dorris, Munoz & Klein 2001).

Taking these assumptions for true, it remains however still unclear how especially the *auditory* spatial information is mapped onto the oculo-motor saliency map in the WHERE-pathway, taking into account the characteristics of the auditory sensory periphery. It is one intention of the present study to investigate these characteristics and integrate them into a model of visual-auditory interaction.

3 Experiments

In the following, three experiments will be described and discussed: a unimodal auditory localization experiment investigating accuracy and latencies of acoustically directed saccades, a unimodal auditory reaction time experiment measuring the time needed for a saccadic response after simple detection of an auditory stimulus (without determining its location), and a bimodal visual-auditory reaction time experiment with visual target stimuli and auditory distractors. All three experiments were performed under two different conditions: in a virtual acoustic environment based on dummy head recordings and in a free field setup using loudspeakers. The experiments demonstrate different characteristics of the visual, the auditory, and the saccadic system and their mutual influence.

3.1 General methods

Participants

In total, two female and five male paid volunteers, aged from 20 to 43 years, took part in the experiments. They were Oldenburg University undergraduate or graduate students and naive as to the purpose of the study (except for HH, who is the author of the study). All participants had normal or corrected-to-normal vision and reported having no hearing problems of any kind. Five participants completed the three experiments using virtual auditory stimuli. Two of these also took part in the respective free field experiments, as did one undergraduate student and the author.

Apparatus

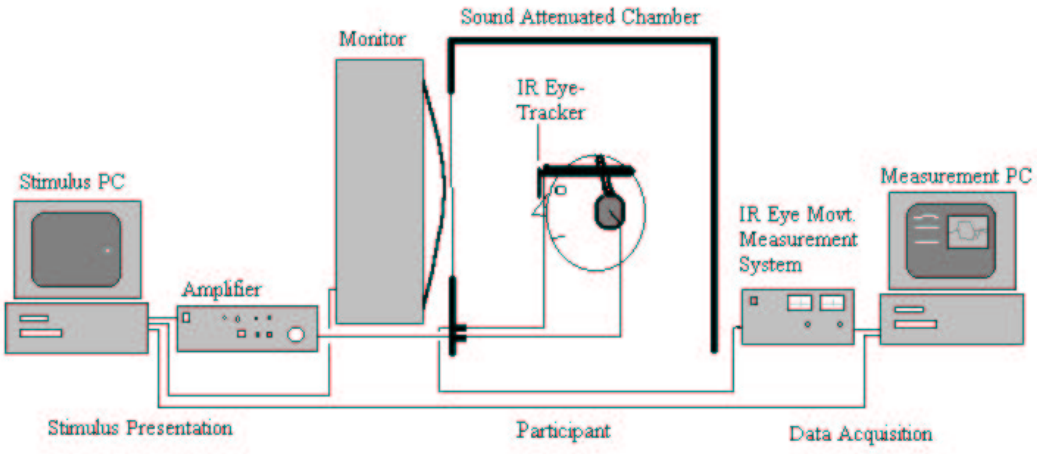


Figure 12: Schematic illustration of the setup used for the experiments using virtual acoustics.

The experiments using a virtual acoustic environment were conducted in a dark and soundproof chamber (1.0 $m \times 1.2\ m \times 1.9\ m$). Visual stimuli were presented on a monitor (NEC XP37, 75 kHz vertical frequency) at a viewing distance of 57 cm. The monitor was placed directly outside the chamber and could be seen through a window which was well sealed against

other sources of light. Sound stimuli were presented via headphones (AKG K500), using a virtual acoustic environment. Presentation of visual and auditory stimuli was controlled by a PC. The temporal arrangement of stimulus presentation and data acquisition were synchronized with the exact presentation time of the visual stimulus determined by the monitor update rate.

In the free field experiments, participants were seated in a dark, sound proof room (3.0 $m \times$ 3 $m \times 2.5$ m). Sound stimuli were white noise signals presented via loudspeakers (Kanton Twin 700) mounted on tripods. Visual stimuli were presented by red light emitting diodes (LED's) fixed at the loudspeakers with a viewing distance of 1.10 m with respect to the participants' heads. One additional red LED served as fixation stimulus and was clamped at a tripod. Visual and auditory stimuli were generated and displayed via TDT System2-components and an additional PM2-module (Tucker-Davis Technologies, Gainesville, Florida) which were controlled by a PC.

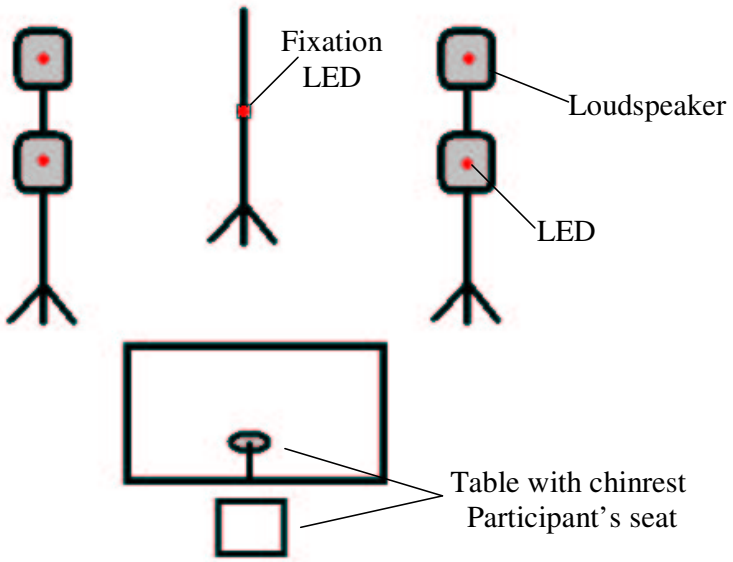


Figure 13: Schematic illustration of the setup used in the free field experiments.

Visual stimuli

White dots of 0.1° in diameter and a luminance of 19 cd/m^2 against dark background (less than 0.01 cd/m^2) served as visual stimuli in the virtual acoustic environment. The red LED's (640 *nm* wavelength) used in the free field setup had an electric power of 120μW and a strong directional characteristic. They were 4 *mm* in diameter, equivalent to 0.2° of viewing angle. All stimuli were presented with a duration of 500 msec at viewing angles of $\pm 25°$ in the horizontal and 0° or 20° in the vertical dimension with respect to the central fixation point.

Auditory stimuli

The sound stimuli in the virtual environment were recorded prior to the experiments by means of a dummy head placed in the sound-proof chamber (the dummy head's pinnae were custom-made at Oldenburg University and formed similar to average human ears). For this purpose, four loudspeakers (Canton Twin 700) were attached at horizontal positions of $\pm 25^\circ$ and vertical eccentricities of 0° or 20° with respect to the dummy head's median plane and ear height. The loudspeakers were removed after the recording. White noise (band-passed within 0.2 - 20 kHz, 3 msec cosine-squared onset-offset ramps) with a duration of 500 msec was used as basic stimulus. It was played once via each loudspeaker, recorded digitally at a sampling frequency of 44100 Hz via microphones (Brüel & Kjær) placed inside the dummy head's ears, and stored in the computer for subsequent presentation. Room characteristics (i.e. reflections by walls or ceiling) were not filtered out, as the experiments were to be conducted with participants taking seat exactly were the dummy head was placed during the recording. During the experiments, the stored noise stimuli were played back by a high-precision sound card (Tahiti, Turtle Beach) with an intensity of 63dB SPL.
In the free field environment, a white noise stimulus (also 0.2 - 20 kHz, 3 msec cosine-squared onset-offset ramps) was generated and put out by the TDT system and presented via the loudspeakers.

Data recording

Data acquisition was identical in both setups. Eye movements were measured with an infrared light reflecting system (IRIS, Scalar Medicals) providing an analog signal of eye position. Eye position data were recorded with LabView on a trial by trial basis, in which the data acquisition PC was triggered by the stimulus PC. Eye positions were furthermore controlled online during the complete session. Due to the calibration procedure and the digitization of the signal, a spatial accuracy of up to 12 min of arc could be achieved. For each trial, 1500 ms were recorded with a sampling frequency of 1 kHz. Horizontal and vertical eye-positions were recorded and stored as separate channels. Hence, the data for one trial consisted of two rows, each of 1500 elements, leading to position-time traces for both azimuth and elevation. Stimulus configuration and calibration data were also stored by a PC. As the IRIS system does not account for head-movements, the participants' heads were stabilized by a chin-rest.

General procedure

Each experimental session started with calibration of the IRIS-system. Participants were instructed to generate an accurate saccade from the central fixation point towards specific peripheral targets (a white dot at one of four positions at $\pm 20^\circ$ horizontal and vertical eccentricity in the virtual acoustic setup and a red LED at $\pm 25^\circ$ horizontal and 0° or 20° vertical eccentricity in free field) and to maintain fixation there as long as the target was visible (about 2 sec). This procedure was repeated between the blocks if needed. All in all, the calibration procedure took about 15 min, during which dark adaptation took place.

In the free field experiments, experimental sessions started immediately after calibration. In the virtual acoustic setup, an additional short training block of about three minutes was inserted between the first calibration and the experimental blocks. In these training blocks, visual and auditory stimuli were presented synchronously and always from of the same direction. Participants should perform an eye movement to the visual stimulus. The results of this pure training block were not used for further analysis and only served for getting the participant acquainted to the use of virtual acoustic stimuli.
The sequences of the experimental trials were quite similar in all three experiments. Each trial started with presentation of the fixation point for a random time interval (800 msec up to 1300 msec). Randomization of the duration of the fixation period was necessary in order to prevent enhanced fixation release (and thus shorter reaction times) simply due to response anticipation by the participant. After the fixation period, the target stimulus appeared. In Experiments 1 and 2, there was only an acoustic target. In Experiment 3, participants should respond to a visual target which could be presented either alone (unimodal case) or accompanied by an acoustic stimulus (bimodal case). In the bimodal experiment, acoustic signals should be ignored and participants were explicitly instructed not to respond to these. In all three experiments trials were separated by time periods of 1500 ms, during which neither a fixation point nor any visual or auditory stimulus was presented.
The participants were instructed tp have a rest between the experimental blocks if they needed. During these periods, they were allowed to remove the eye movement measurement system and move freely, but they had to stay in darkness in order to remain dark adapted.

Response detection

Analysis of the recorded data was performed off-line after the experiment. Self-programmed MatLab-scripts were applied to transform the raw data into calibrated signals and detect saccades on the basis of a preset velocity-criterion ($v \geq 50^{\circ}/sec$). Furthermore, the eye position data of each trial were re-checked visually for proper fixation at the beginning of the trial, for blinks, and for the correct detection of start and endpoint of the detected saccade. Trials with improper fixation or blinks were excluded from further analysis. If necessary, onset and end of the saccade were marked manually. Figure 14 shows an examplary set of eye movements in three successive trials as used in our analysis.

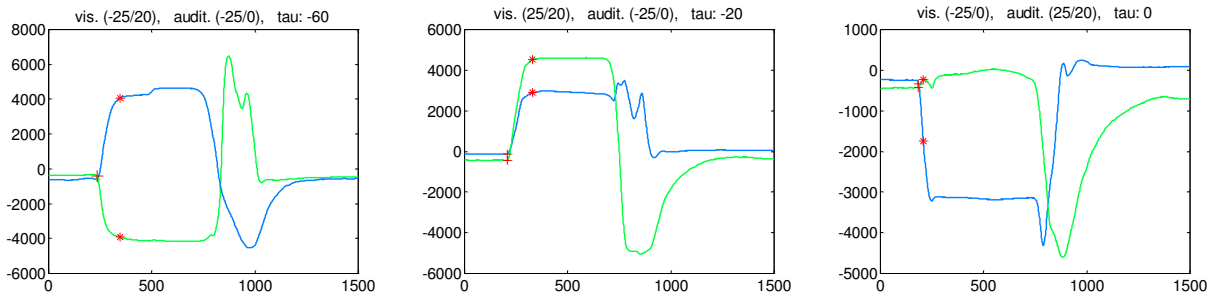


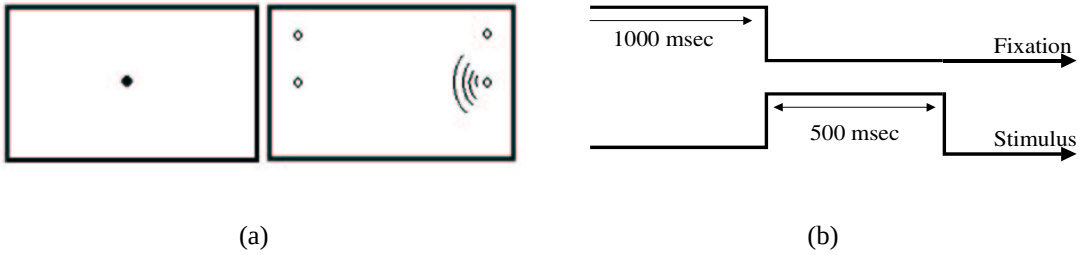
Figure 14: Illustration of eye movement data as displayed graphically for further analysis. Each figure displays 1500 msec of one trial, starting from the moment of target onset. Ordinate: eye position signals in arbitrary units (signal output in mV) which is converted into viewing angle during the analysis. Abscissa: time from target onset in msec. The grey and the black lines show position-time traces of horizontal and vertical eye position across time, respectively. The small marks indicate saccade onsets and endpoints as suggested on the the basis of the velocity criterion. The respective values could be either accepted or manually corrected by the user (as needed in the rightmost trial). At the top of each figure, the positions and SOAs of of the actual trial are displayed.

Reaction times were defined as the time between the onset of the target stimulus and the onset of the saccadic eye movement (in milliseconds). Thus, we use the term “saccadic reaction time” as synonym for saccadic latency. Reaction times of less than 100 msec or more than 500 msec were discarded as anticipatory and misses, respectively. The position of the eye at the beginning of and after the saccade were calculated and stored as horizontal and vertical positions in degree of visual angle relative to the central fixation point.

3.2 Experiment 1: Auditory localization

In Experiment 1, directed saccadic responses toward sound stimuli were investigated. As described above, two different experimental setups were used: a virtual auditory environment based on dummy head recordings and a 'free field' setup using loudspeakers. The main purposes of the first experiment were (1) the evaluation of individual auditory localization performance especially with regard to the use of the virtual environment, (2) a qualitative analysis of auditory guided eye movements in comparison to visually evoked saccades, and (3) an estimation of the processing time needed for a *directed* auditory guided response (not only detecting the presence of a stimulus).

Procedure



(a) (b)

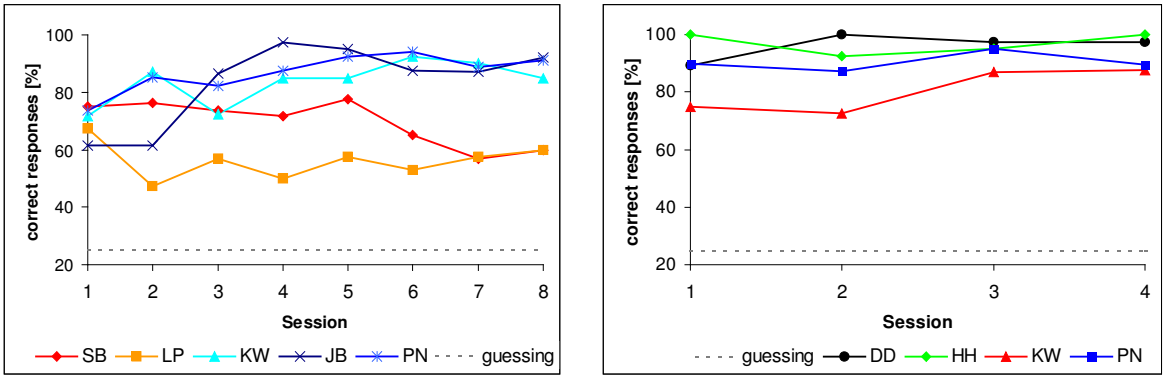
Figure 15: a: Illustration of one unimodal auditory trial in Experiment 1 (auditory localization). Colors are inverted for technical reasons. Stimuli could be presented from any of the four positions indicated here by open circles (not visible in the experiment). b: Chronological order of stimulus presentation.

Auditory targets were presented as described in the general methods section. Each trial began with the presentation of the central visual fixation point, which disappeared after a random period synchronously with the onset of a sound presented from one of four possible positions as indicated in Fig 15. Participants were advised to perform a saccade toward the perceived origin of the sound as accurately and quickly as possible. In each experimental block, either stimulus position was presented ten times in randomized order, making the whole block last about 5 min. Eight auditory localization blocks were performed on eight different days by each participant in the virtual acoustic environment. In the free field setup, each participant took part in four auditory localization blocks.

Results

Localization performance

The first postsaccadic fixation was taken as judgment in the auditory localization task. A saccadic response to an auditory stimulus was rated to be correct if it was within an area of 7° around the mean end position of saccades toward the respective visual target (data collected from Experiment 3).



(a) virtual acoustic environment

(b) loudspeaker environment

Figure 16: Localization performance in a virtual auditory environment(a) and in free field conditions (b) for all participants across sessions.

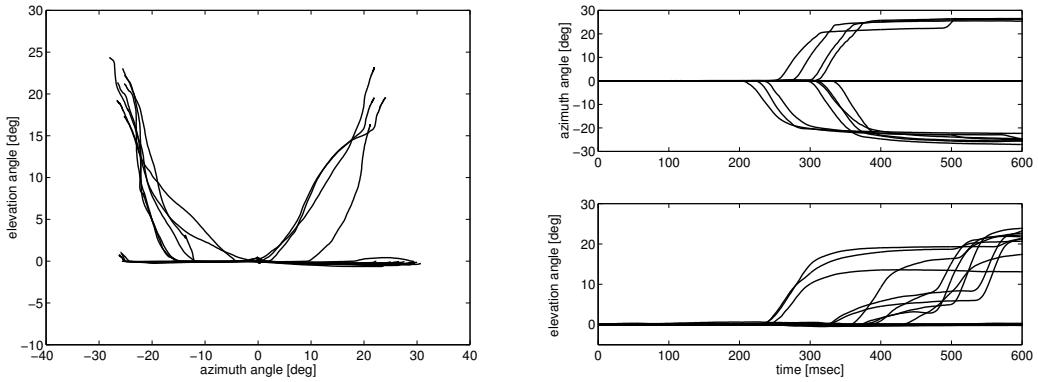
Figure 16 shows the localization performance of each participant in the two auditory environments across the sessions. In the virtual auditory environment, two groups of listeners crystallize. Three of the participants (JB, KW, PN) turn out to be well able to use the virtual acoustic environment, while the two remaining (LP, SB) seem to be only poor localizers. In those participants with high judgment performance an increase of correct answers from session 1 to session 8 can be seen, which might be interpreted as an effect of learning. In contrast, the poor localizer SB becomes less accurate in judgments from in the last sessions. This finding is very surprising, since SB had already taken part in auditory localization experiments (Heuermann & Colonius 1999) and had there demonstrated excellent localization even under dummy-head conditions. A second analysis reveals the reason for this discrepancy. If not primary judgment, but instead the final eye-position is taken to be valid, SB's percentage of correct answers jumps up to over 90%. A closer look at the saccade characteristics of this subject (see below) yields significantly curved trajectories in her auditory guided saccades, in which vertical eye movements often were not elicited until 100 msec after horizontal eye movement onset (i.e. too late to be recognized as primary movement).

The free field data show localization performance values generally comparable to those of the good localizers in the last four sessions in virtual acoustics. However, except for KW, the free field listeners did not seem to need any learning periods, they start with high rates of correct responses (about 90% and more) and remain at that level. Interestingly, participants KW and PN, who took part in both experimental setups and show good localization in the virtual acoustics, are those listeners with least estimation accuracy in the loudspeaker condition. This fits at least to KW's rating that he found loudspeaker sound localization more difficult than virtual sound source localization.

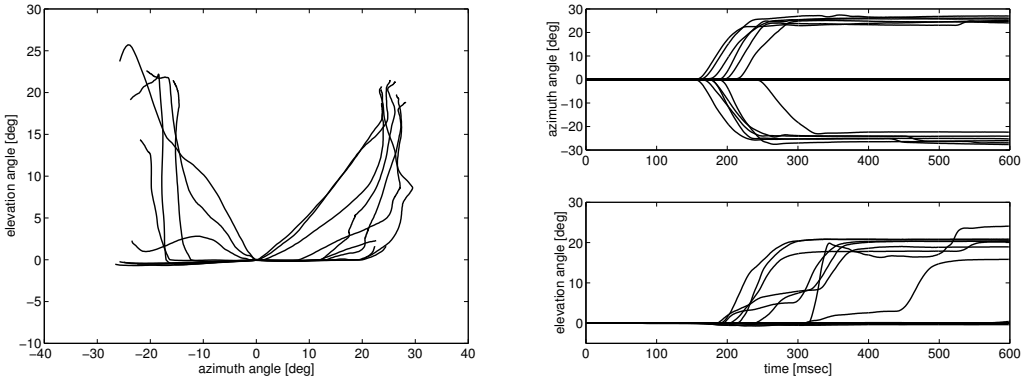
Finally analyzing both the free field and the virtual acoustics data with respect to the type of errors made, it turns out that almost all localization errors are made with respect to elevation estimation. Especially those participants with poor localization performance seem to have large problems in correct vertical judgment.

Saccade and trajectory characteristics

Figures 17 and 18 show some examples of trajectories and position-time traces of auditory guided saccades in the virtual auditory environment and in free field, respectively . A general observation from both experimental setups is that trajectories to positions in the upper hemisphere are quite curved or bow-like.



(a) KW, virtual acoustics

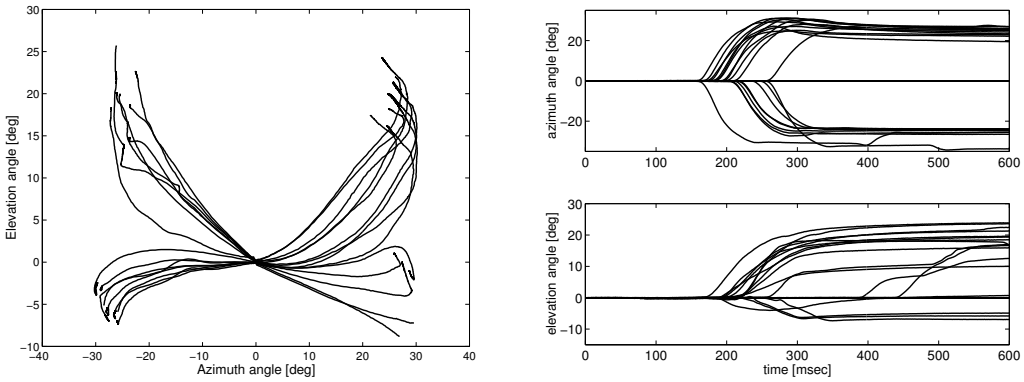


(b) SB, virtual acoustics

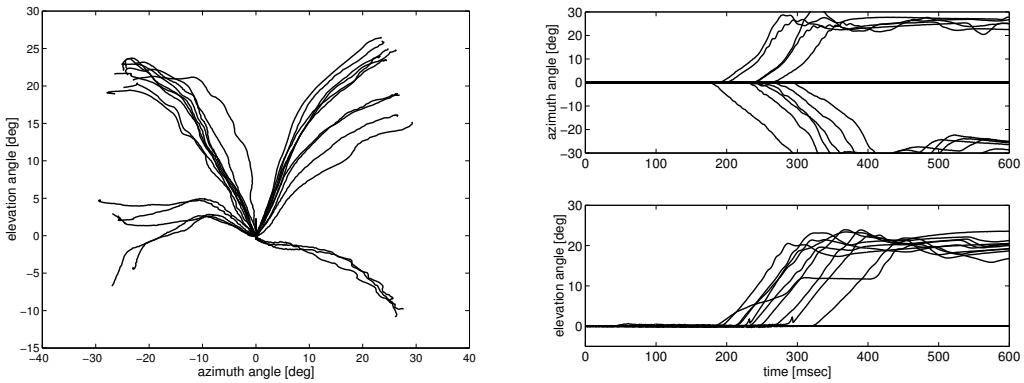
Figure 17: Auditory evoked eye movements in virtual acoustics: trajectories (left panels) and position-time traces (right panels).

In most cases, the movement starts with a definitive horizontal trend which then turns into a more upward directed movement (Fig. 17a, 17b and 18a). There are, however, some participants whose initial eye movements toward elevated stimuli have a pronounced vertical component which ceases afterwards, like DD in Fig. 18b. Considering the respective position-time traces by splitting the movement into horizontal (azimuth angle) and vertical (elevation angle) components reveals that the curve-like nature of auditory saccades can mostly be traced back to the vertical saccade component alone. While horizontal movement seems to be elicited fast

and directly to the position desired, elevation movements often start somewhat later, are less accelerated, and are corrected for once or twice. These corrective eye movements were almost exclusively performed with respect to elevation. A quite interesting observation can be made with the auditory saccades of SB (Figure 17b), whose vertical eye movement often follows the horizontal movement by periods up to 100 msec. Although other participants did not show this pattern that pronounced and so often, the "multi-saccadic responses" can generally be observed.



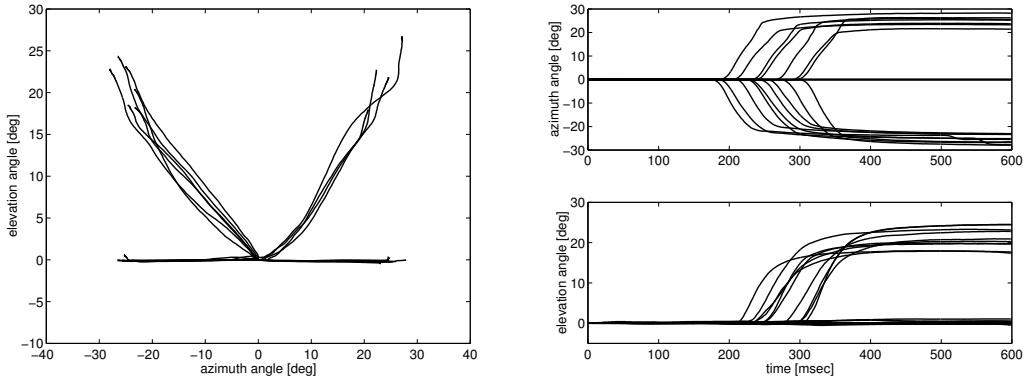
(a) PN, auditory localization, loudspeaker acoustics



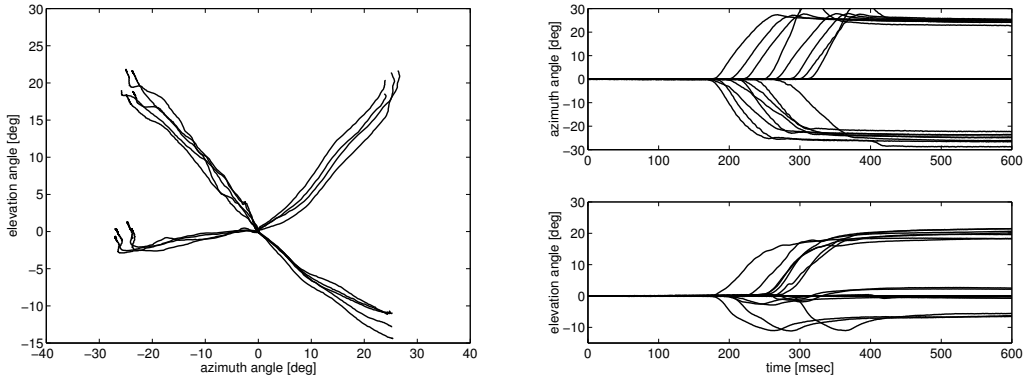
(b) DD, auditory localization, loudspeaker acoustics

Figure 18: Auditory evoked eye movements in free field environment: trajectories (left panels) and position-time traces (right panels).

A comparison with visually guided eye movements (taken from the unimodal visual trials in the third experiment and displayed in Figure 19) reveals that curved eye movements seem to be typical for *auditory* evoked responses only.



(a) KW, visually guided saccades, virtual acoustic setup



(b) PN,visually guided saccades, free field setup

Figure 19: Visually evoked saccades: trajectories (left panels) and position-time traces (right panels).

Saccadic latencies

Table 1 gives an overview of mean SRTs and standard deviations with unimodal auditory stimuli for each stimulus position and participant in the virtual auditory environment. Most auditory evoked saccadic latencies are quite long (when compared to visually evoked saccades, see below and in the literature), and there is a large variance within the response times for each stimulus position. Inter-subject variability with respect to mean SRTs as well as to individual standard deviations is also quite large. Considering the latency distributions and their varying characteristics points up the situation rather well. In some cases, the distributions remind of "typical" saccadic latency distributions for visual stimuli (KW, LP, SB). However, sharply peaked (PN) as well as very broad distributions without a distinctive maximum (JB) can also be observed. As with mean response times, no obvious relation between localization accuracy and the form of the latency distribution can be found.

	Stimulus Position				
Participant	$(-25°/0°)$	$(-25°/20°)$	$(25°/20°)$	$(25°/0°)$	*mean*
JB	479(140)	492(151)	377(103)	387(91)	435
KW	257(94)	248(80)	277(71)	216(67)	237
LP	280(73)	281(63)	296(133)	265(71)	281
PN	179(81)	168(75)	202(87)	195(86)	190
SB	183(56)	185(100)	210(52)	201(50)	195

Table 1: Saccadic latencies (and standard deviations) for auditory localization using a virtual acoustic presentation.

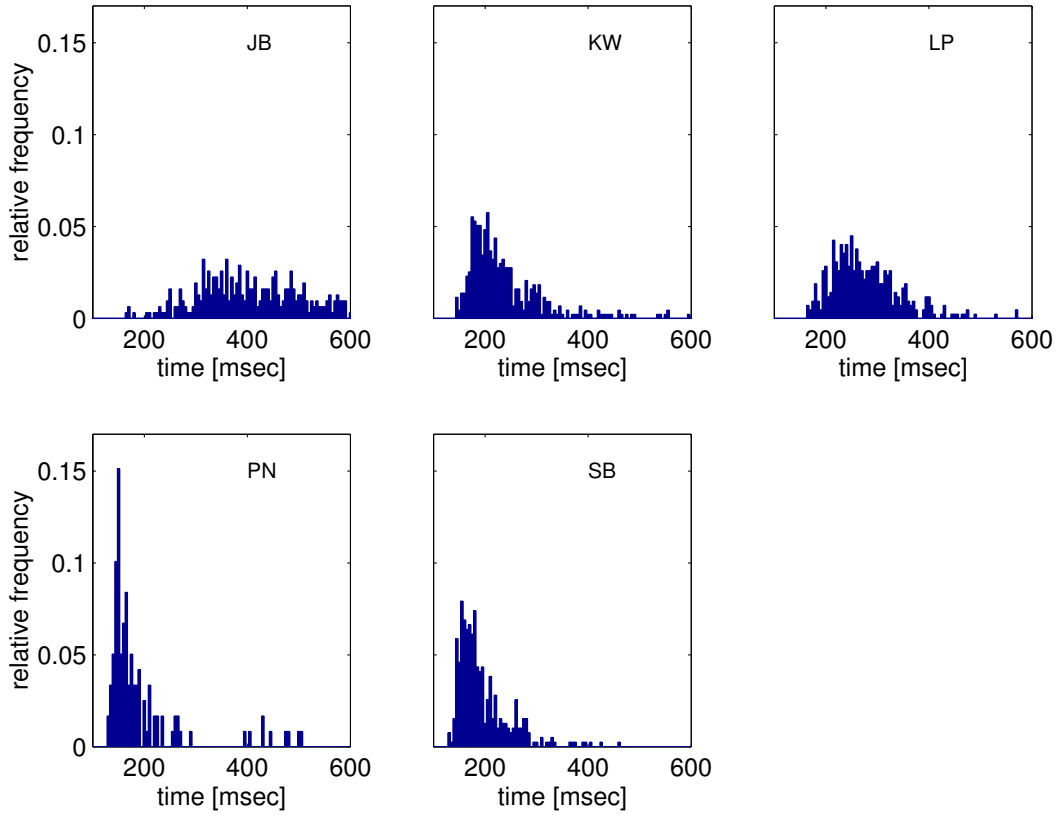


Figure 20: Relative frequency of saccadic reaction times in the auditory localization task. Compare with Table 1.

One-way ANOVAs with post-hoc Scheffé-tests yield a significant influence of laterality, but not of single target positions, on saccadic latency for participants JB ($F(3,348) = 20.4,\ p < 0.001$), KW ($F(3,432) = 6.1,\ p < 0.001$), and SB ($F(3,388) = 3.851,\ p < 0.01$), in which responses to the right are faster than to the left.

In general, the data collected in the loudspeaker environment (Table 2) and Figure 21) yield comparable saccadic reaction times and latency distributions.

	Stimulus Position				
Participant	$(-25°/0°)$	$(-25°/20°)$	$(25°/20°)$	$(25°/0°)$	*mean*
DD	295(57)	266(65)	224(45)	268(78)	254
HH	171(22)	177(18)	176(18)	170(18)	173
KW	242(44)	252(54)	192(30)	201(41)	222
PN	285(108)	314(103)	251(70)	263(90)	278

Table 2: Saccadic latencies (and standard deviations) for free field auditory localization. Compare with Tab. 1. Note that KW and PN participated in both experimental setups.

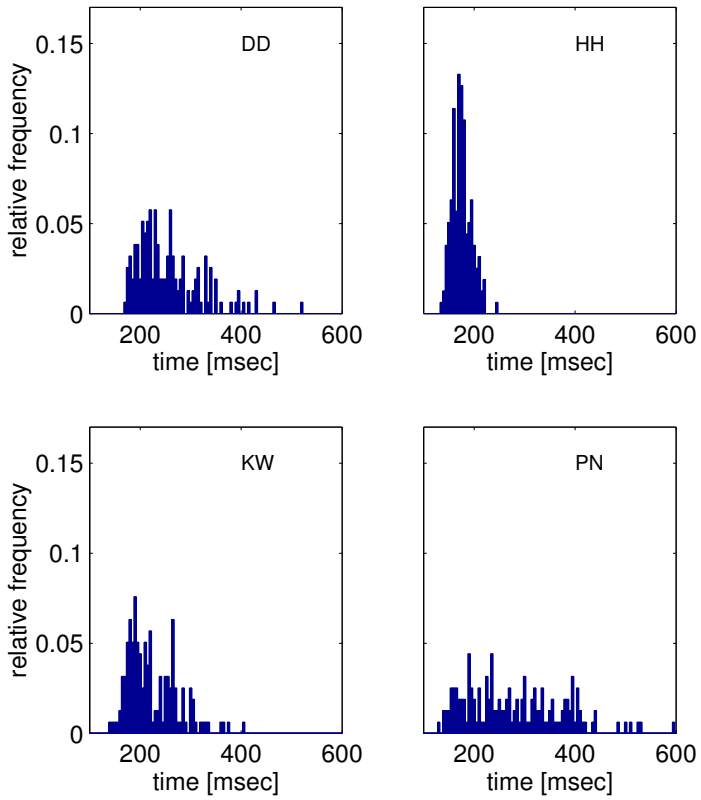


Figure 21: Relative frequency of saccadic reaction times in the auditory localization task. Compare with Table 2.

There is no systematic influence neither of laterality nor of single target positions on saccadic latency, except for KW who again shows an effect of laterality ($F(3,154) = 18.46,\ p < 0.001$) with responses to the right being faster than to the left. As in the virtual acoustic setup, latency distributions vary from sharply peaked to extremely broad (note that the total number of trials was only half as large in the loudspeaker setup, thus the distributions are generally less smooth).

No significant effect of experimental setup (virtual vs. loudspeaker environment) on saccadic latencies an be found (two-tailed t-test, p=0.272). It is, however, somewhat surprising that the latency distributions of PN in the virtual and in the loudspeaker environment differ so distinctively. Although PN's localization performance in both environments is similar, his response times in free field are about 100 msec longer. Moreover, this does not at all fit to his statement of free field localization being much easier than in the virtual acoustic environment.

Discussion of Experiment 1

The results of the auditory localization experiment yield a large bandwidth of strategies and abilities to handle the task, which can be observed with respect to the eye movement trajectories as well as to their latencies. Some participants allow themselves quite a long time to perform an eye movement but are very accurate in their judgments, while others respond very fast but at the same time correctly (if successive movements are also taken into account). There are also some listeners who, despite their comparatively slow reactions, frequently fail to localize correctly. Although most location estimation errors are made with respect to elevation, no significant effects of vertical stimulus eccentricity on saccadic reaction times can be found. There are only effects of laterality in some cases, which are very common. All in all, inter- and intra-subject variabilities of SRTs are quite large, in which SRTs are not related to localization performance. Apparently, auditory guided saccades are not as stereotyped as their visual counterparts: independent of a participant's localization ability, they even differ within one listener depending on what kind of auditory environment was chosen (see PN). A qualitative analysis of trajectory and position time traces reveals that unlike the straight visually evoked saccades, auditory guided eye movements often consist of several successive "miniature-saccades". Auditory eye movements seem to be dividable into a first (quite fast) response, which includes the judged azimuth eccentricity and a rough guess of the elevation component, and one or more "corrective steps" starting some ten milliseconds later with one or more additional vertical movements completing the saccadic response. The initial vertical trend, however, seems to be strategy-dependent. In most participants, there is only a weak elevation component in the first movement, as if elevation estimation is nearly left out at that time. In some other participants the initial movement to elevated stimuli is dominated by a pronounced vertical drift, which would lead to a strong over-estimation of vertical eccentricity if set forth. This might be interpreted as if there was a primary general judgement as to whether the stimulus was elevated or not. Although the pattern of auditory guided eye movements is very individual, it turns out to be quite replicable within each participant and condition. Even if there were significant errors in location estimation (e.g., the free field stimuli from the horizontal plane were frequently judged to come from below, see Fig. 18), the errors themselves were highly reproducible. The curved eye movements were observed in both experimental setups and thus cannot be attributed to possible artifacts of the virtual acoustic environment. Although the poor localizers showed more overall variance in their eye movements, the pattern of curved trajectories could be found with all participants. It therefore seems unlikely that the eye movement patterns observed here coun simply be explained by insufficient localization of the stimuli.

Similar patterns with auditory evoked eye movements have in fact already been observed in earlier studies (Zahn, Abel & Dell'Osso 1978, Zambarbieri, Schmid, Magenes & Prablanc 1982). Jay & Sparks (1995) found multiple saccades in about 20% of human auditory-guided eye movements, but no correcting saccades in visually evoked eye movements. The average intersaccadic interval between initial and corrective movement they found was around 75 msec. These former findings could be replicated well in this study. It could furthermore be proved, to our knowledge for the first time, that virtual acoustic targets evoke similar eye movements.

In order to the interpret the nature of auditory guided saccades, there are two general attempts: Jay and Sparks explained their results with *less pre-motor activity* prior to the execution of acoustically evoked saccades rather than with the complex manner of the transformation of auditory signals into motor coordinates. Frens & Van Opstal (1995) came to the contrary conclusion. These authors investigated the kinematic properties of auditory saccades by decomposing them into two overlapping parts, a (primary) P-movement and a (secondary) S-movement, in which the P-movement has the direction of the initial eye movement and the S-movement is a vertical adjustment. In their data, the vertical component of primary movements correlated significantly with the respective target elevation eccentricity and can thus be interpreted as a first rough estimate, which is postulated to be performed after about 150 msec. The S-movement, which is purely vertical by definition, is then added approximately 30 msec later to compensate for the motor error due to the incomplete first percept of the auditory target position (the whole processing of auditory stimulus position is hypothesized to be complete after about 200 msec). According to Frens & Van Opstal, curved auditory eye movements are to be explained by *characteristics of the auditory sensory rather than the oculomotor system*. The position-time traces of the auditory saccades recorded in the present study confirm their finding that almost only the vertical components are stepped and corrected for while azimuth movement is performed straight and fast.

Subsequent adjustment of "preliminary" saccades which were based on incomplete information processing was suggested by Zambarbieri, Beltrami & Versino (1995) although they hypothesized that this mechanism is a *general feature of all saccades with large eccentricities*, independent of the target's modality. Zambarbieri et al. found a dependency of mean saccadic latencies on the saccades' azimuth eccentricity in both visual and auditory evoked movements. Jay & Sparks (1995) additionally used targets with different elevation positions and thereby made the interesting observation that the longest latencies were responses to auditory stimuli coming from straight above the initial fixation point, *independent of the initial eye position*. Latencies to auditory targets in the upper hemisphere could, on the other hand, massively be reduced if the target also had a horizontal component with respect to the fixation point.
The mean saccadic latencies to auditory targets in this study did not show an effect of vertical eccentricity on saccadic reaction times. However, the elevated targets used here always had a pronounced horizontal eccentricity of 25°. The present data therefore confirms Frens & Van Opstal's hypothesis of ongoing auditory signal processing during auditory guided eye movements, but is also in line with the findings of Zambarbieri et al. and Jay & Sparks. Apparently, the onset of an auditory saccade is triggered by the percept of *any* displacement between eye- and target-position, while the full spatial information (and the resulting motor error to be corrected

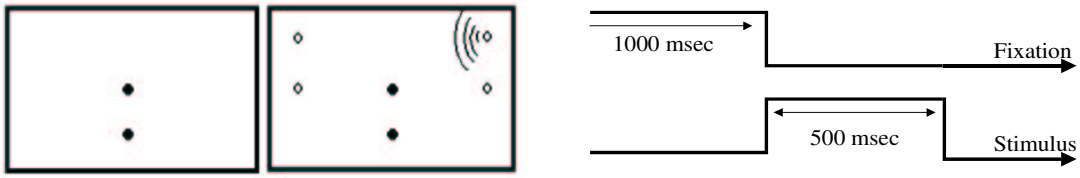
for) is calculated during the eye movement. In most cases, the horizontal movement component is determined quickly and accurately via binaural cues, while the vertical target position has to be computed by a complex analysis of the signal spectrum, which takes more time and may thus be done in successive steps. The continuous “updating” of motor commands according to the actual information might then result in a superposition of movements, or "multiple-look strategy" as described by Hofman & Van Opstal (1998).

Comparing the data of the virtual auditory and the loudspeaker environment, no pronounced differences can be seen for those participants who were able to learn how to cope with the dummy head recorded stimuli. After some training sessions, position judgment performance turned out to be well comparable with free field data. However, some listeners turned out to be unable to use the virtual environment. More training sessions might have enhanced the localization accuracy of these participants. In an elaborate study, during which volunteers permanently wore ear molds across several weeks and repeatedly underwent auditory localization tasks, Hofman et al. (1998) proved that using a new set of Head Related Transfer Functions (equivalent to a set of new outer ears) can be learned within a certain period without losing the ability to localize with the “original” own ears. These findings not only point out the central role of cortical influence in spectral feature analysis, but they also indirectly encourage the use of virtual auditory environments. If a listener is able to adopt to new spectral cues, he should generally be able to adopt to a virtual acoustic environment. On the other hand, no significant changes in correct response rates could be found after eight sessions; the judgments of listener SB even tended to become worse.
For some participants, it might have been helpful to apply individual HRTFs. Earlier studies (e.g. Heuermann & Colonius 1999) have shown that remarkable improvements in localization performance might be achieved by this. However, it has also been shown (e.g. Wightman & Kistler 1989) that some participants are generally poor localizers, including in free field situations. In fact, participant KW showed less localization judgment accuracy in the loudspeaker environment than with the virtual acoustic stimuli, with these results corresponding with his personal rating.
The finding of latencies and trajectories being quite similar in the different experimental setups generally suggest that virtual and “real” auditory stimuli are processed in the same way. However, in some cases we found very broad latency distributions indicating an influence of higher cognitive processes. Since targets were presented from only four different positions, it is conceivable that some listeners turned to an individual strategy to “localize” them, for example by their spectral content. On the other hand, prolonged latencies with flat distributions can be found under both listening conditions (see JB in virtual acoustics and PN in free field).
Hence, the conclusion regarding the usability of virtual acoustic environments in psychophysical experiments might be as follows. In general, the above results support the application of virtual environments. Their setup takes less room and the stimuli are completely controllable. However, preliminary tests of the individual participant’s localization performance are indispensable. Yet, in view of the present data (KW !) and earlier results of Wightman & Kistler (1989), this should be done in *any* environment. More insight in possible differences of virtual and real sound sources in psychophysical investigations might be obtained from the following experiments.

3.3 Experiment 2: Auditory detection

In Experiment 2, the auditory signal was again the target stimulus, however, in contrast to the first experiment, simple responses were chosen (i.e. responses were not target directed, but the same saccadic response had to be given to any auditory target). This experiment was carried out in order to compare the time needed for simple *detection* of an auditory stimulus (i.e. its mere presence independent of position), with the time needed to process spatial information (obtained in Experiment 1). In other words: this experiment yielded the time needed to detect a stimulus and to perform a (possibly pre-programmed) response, but without making a decision about its direction. As responses were performed in the form of saccadic eye movements in both auditory experiments, reaction time variability due to distinct motor times was minimized, making a comparison between detection and localization time possible. The experiment was carried out in both virtual and loudspeaker listening conditions with the same respective participants as in Experiment 1.

Procedure



(a) (b)

Figure 22: a: Illustration of one unimodal auditory trial in Experiment 2 (auditory simple detection). Colors are inverted for technical reasons. Stimuli could be presented from any of the four positions indicated here by open circles (not visible in the experiment). The fixation point and the target point were presented permanently during the trial. b: Temporal order of stimulus presentation.

Two stimuli, the central fixation point and another point 20° below it, were visible throughout each trial. They were either white dots (in virtual acoustics) or red LED's (in the loudspeaker condition). Only auditory targets were presented, from directions identical to those in Experiment 1. The task in this experiment was to fixate the central fixation point until *any* acoustic stimulus was perceived and then look down to the second point as quickly as possible. As in the localization experiment, the different sound stimuli were played in randomized order ten times per session. Eight experimental blocks (5 min each) were performed in the virtual setup, four in the loudspeaker environment, all on different days.

Results

Saccadic latencies

Tables 3 and 4 show mean saccadic latencies and standard deviations for auditory evoked simple reactions in the virtual acoustic and in the loudspeaker condition, respectively.

	Stimulus Position				
Participant	$(-25°/0°)$	$(-25°/20°)$	$(25°/20°)$	$(25°/0°)$	*mean*
JB	142(46)	131(26)	143(31)	138(20)	135
KW	231(87)	233(69)	215(52)	220(71)	225
LP	229(36)	227(41)	221(41)	229(52)	227
PN	157(23)	169(39)	169(39)	171(68)	164
SB	150(28)	143(22)	143(23)	150(56)	146

Table 3: Saccadic latencies (and standard deviations) in auditory detection, virtual acoustic environment.

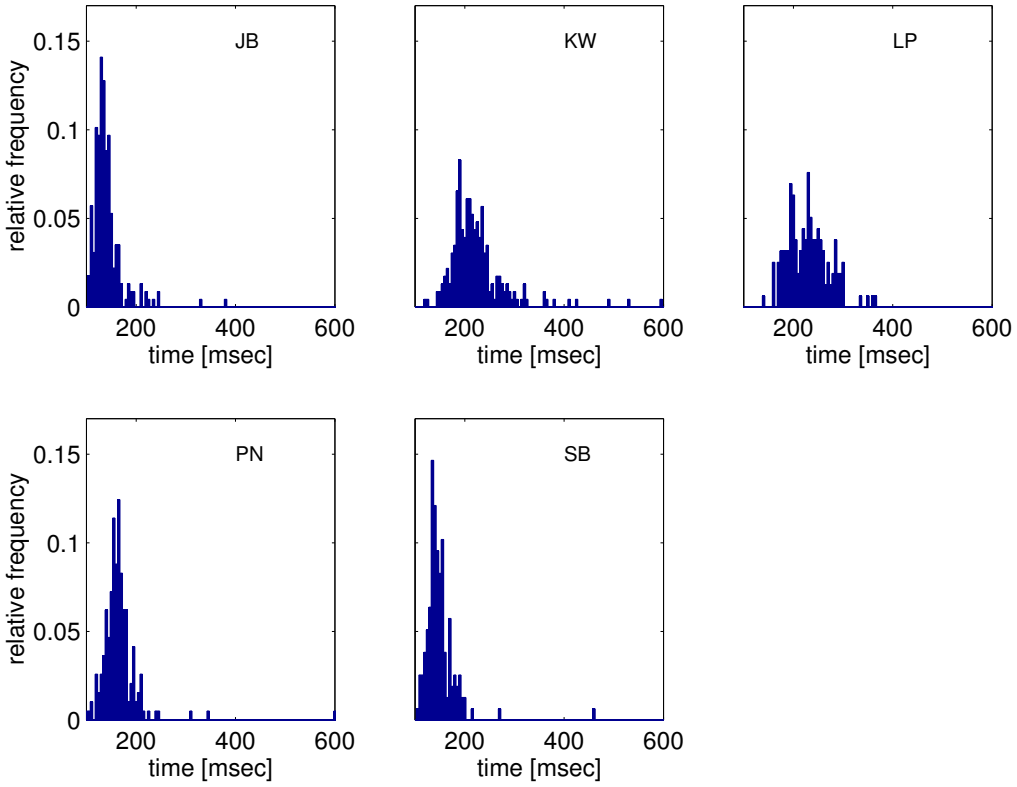


Figure 23: Relative frequency of saccadic reaction times in the auditory simple detection task in virtual acoustics. Compare with Table 3.

Compared with Experiment 1, SRTs are very much shorter for most participants and latency variability with respect to specific target is significantly smaller.

	Stimulus Position				
Participant	$(-25°/0°)$	$(-25°/20°)$	$(25°/20°)$	$(25°/0°)$	*mean*
DD	155(38)	172(62)	172(59)	173(54)	168
HH	138(22)	148(27)	141(27)	142(25)	142
KW	185(29)	215(81)	191(30)	197(57)	198
PN	152(36)	147(20)	149(23)	146(25)	149

Table 4: Saccadic latencies (and standard deviations) in auditory detection, loudspeaker listening condition

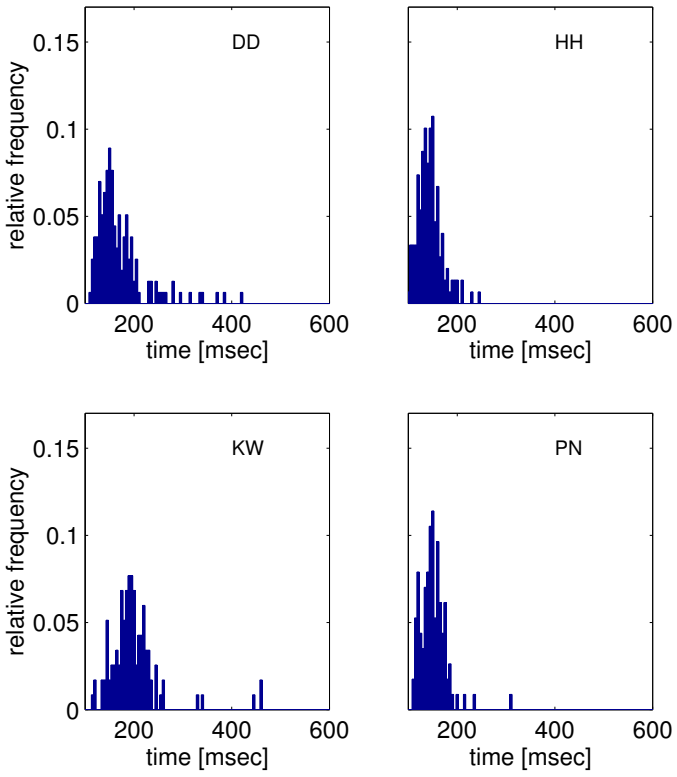


Figure 24: Relative frequency of saccadic reaction times in the auditory simple reaction task in loudspeaker environment. Compare with Table 4.

The same pattern can be derived from the representative latency distribution plots (Figures 23 and 24). For most participants, the saccadic latency distributions for simple auditory detection are roughly comparable to those of the localization task, but they are significantly shifted toward smaller latency values and are somewhat more peak-like. Pronounced changes can, however, be

observed for those listeners who showed only flat latency distributions in the localization task (JB in the virtual environment and PN in the loudspeaker setup). In both cases, their simple responses are performed very fast and with small variance. By contrast, KW's response times in the detection experiment hardly differ from those in the localization experiment, whether in virtual or in free field acoustics. A highly significant effect of task (directed vs. undirected response) was found for all listeners in both experimental setups (all $F's > 13.6$, all $p's < 0.001$), except for KW in the virtual auditory environment ($F(1,625) = 2.0,\ p = 0.153$). In contrast to the auditory localization task, no effect of stimulus position or laterality on simple response times to auditory stimuli could be found for any participant neither in virtual acoustics (all $F's < 1.5$, all $p's > 0.2$) nor in the loudspeaker environment (all $F's < 1.7$, all $p's > 0.2$). Comparing the results of both experimental setups, a slight tendency toward shorter latencies can be observed in the loudspeaker environment, a t-test however did not yield a significant effect ($p = 0.560$).

Discussion of Experiment 2

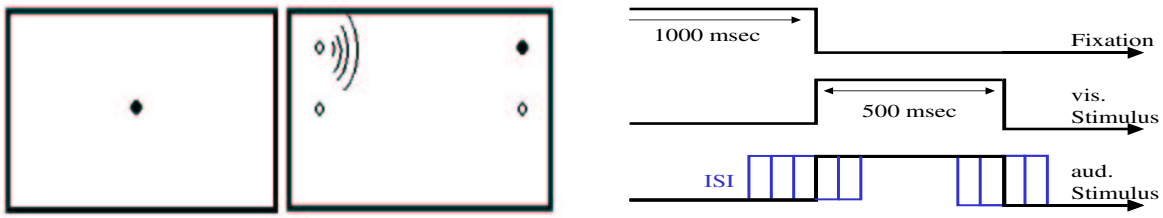
Saccadic latencies in the simple auditory detection task turned out to be reduced when compared to directed auditory responses. This is not surprising, since the detection task did require neither complete analysis of the auditory spatial information (which alone could take up to 200 msec, following Frens & Van Opstal (1995) nor programming of the saccade, as it always had to be directed to the same point, which furthermore was present throughout the trial. As expected, the effect of laterality on saccadic latencies disappeared completely, suggesting motor execution effects being effectively reduced.
However, at least two participants (KW, LP) seemed to hardly be able to ignore spatial information and simply react to the onset of the stimulus, as their reaction times decreased only slightly. These results again show the possible impact on a subject's strategy, which may also reveal itself in a bimodal focused attention task. Participants might be able to follow the instruction and completely ignore the acoustic accessory, or they might not. All in all, it seems that simple detection of an auditory stimulus can take place much faster than localization, which could result in different intersensory effects of general alertness on the one hand and spatially related response facilitation on the other, depending on the respective interstimulus interval.

3.4 Experiment 3: Bimodal reaction time

Experiment 1 revealed the complex way of auditory spatial analysis and its effect on oculomotor processing. The auditory system uses different codes for horizontal and vertical eccentricities which are computed in different pathways consuming different amounts of time. In case of auditory stimulation, the oculomotor map seems to be compiled on the basis of a rough guess which is successively corrected. The mere detection of auditory stimuli takes place much faster, as shown in Experiment 2. The following experiment investigated if and in which way these characteristics of auditory signal processing influence visual-auditory interaction. Given the above assumptions, a significant interrelation between temporal and spatial stimulus parameters could be expected.

Procedure

A visual target stimulus was presented at one of its four possible positions as described in the general methods section. In 80% of the trials, it was accompanied by an auditory stimulus, which could be at the same or at a different position with respect to the visual stimulus. In the remaining 20% of the cases, the visual stimulus was presented alone (unimodal condition). The participant's task was to direct the eyes as fast and as accurately as possible towards the visual stimulus. An auditory stimulus should be ignored in the sense that no reaction should be performed into its direction. Each trial started with the visual fixation stimulus. After a random



(a) (b)

Figure 25: a: Illustration of one bimodal trial (colors are inverted for technical reasons). Visual and auditory stimuli could be presented at either of the four positions indicated by open circles (not visible in the experiment). b: Temporal order of stimulus presentation in bimodal trials.

period of 800 - 1300 msec, the fixation point disappeared, and synchronously the visual target was displayed, possibly accompanied by an auditory stimulus. The interstimulus interval, or stimulus onset asynchrony (SOA), between visual and auditory stimuli varied between -60 msec and +40 msec in steps of 20 msec, in which negative SOAs denote the visual following the auditory stimulus. This results in $4 \times 5 \times 6 = 120$ different combinations of stimuli (4 visual stimulus positions $\times$ (4 auditory stimulus positions + no auditory stimulus) $\times$ 6 SOAs). One bimodal experimental block consisted of the presentation of each of the stimulus combinations in randomized order. All in all, subjects participated in 20-22 bimodal experimental blocks.

Results

Saccade and trajectory characteristics

Only primary saccades directed toward the respective visual target were included in further data analysis. A bimodal response was rated as correct if the first postsaccadic fixation was within an area of 5° around the mean end position in the respective unimodal trials. Some examples of saccades toward visual targets under bimodal stimulation are plotted in Figure 26. Comparing

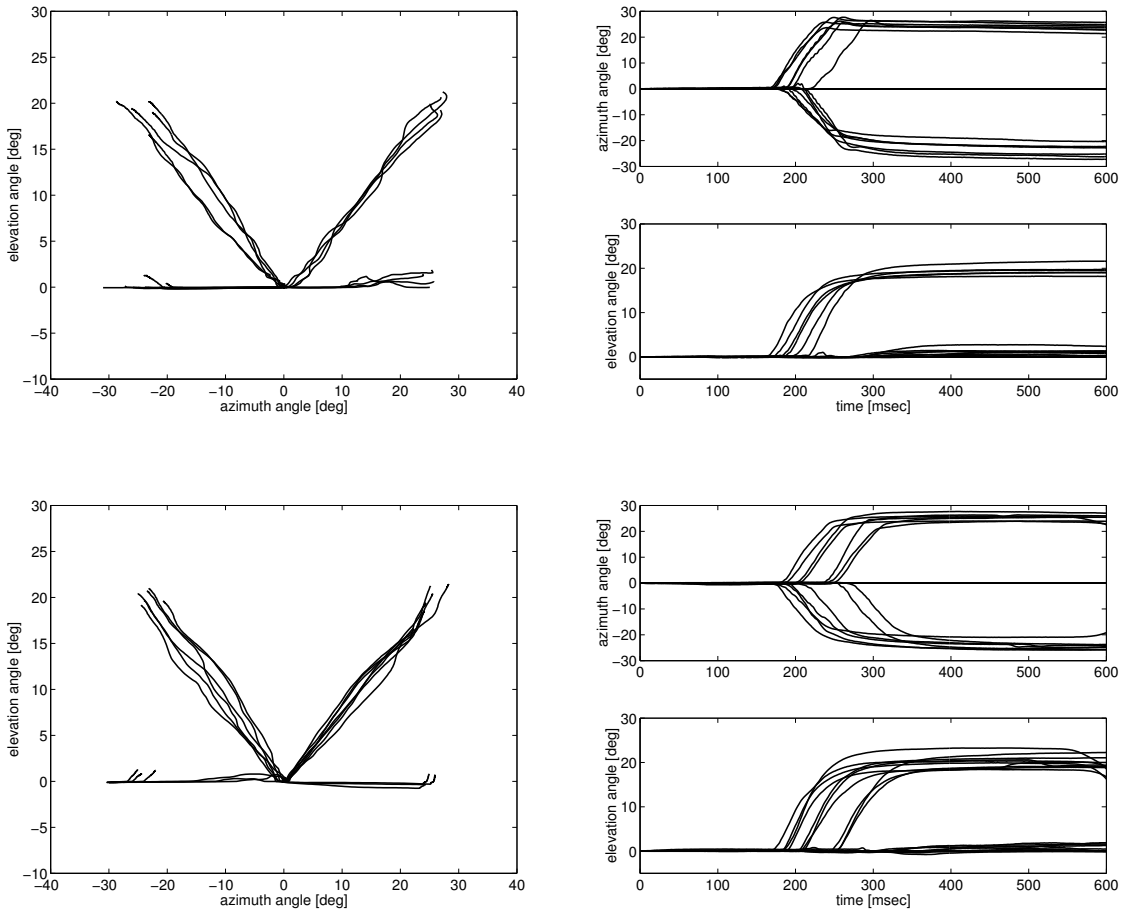
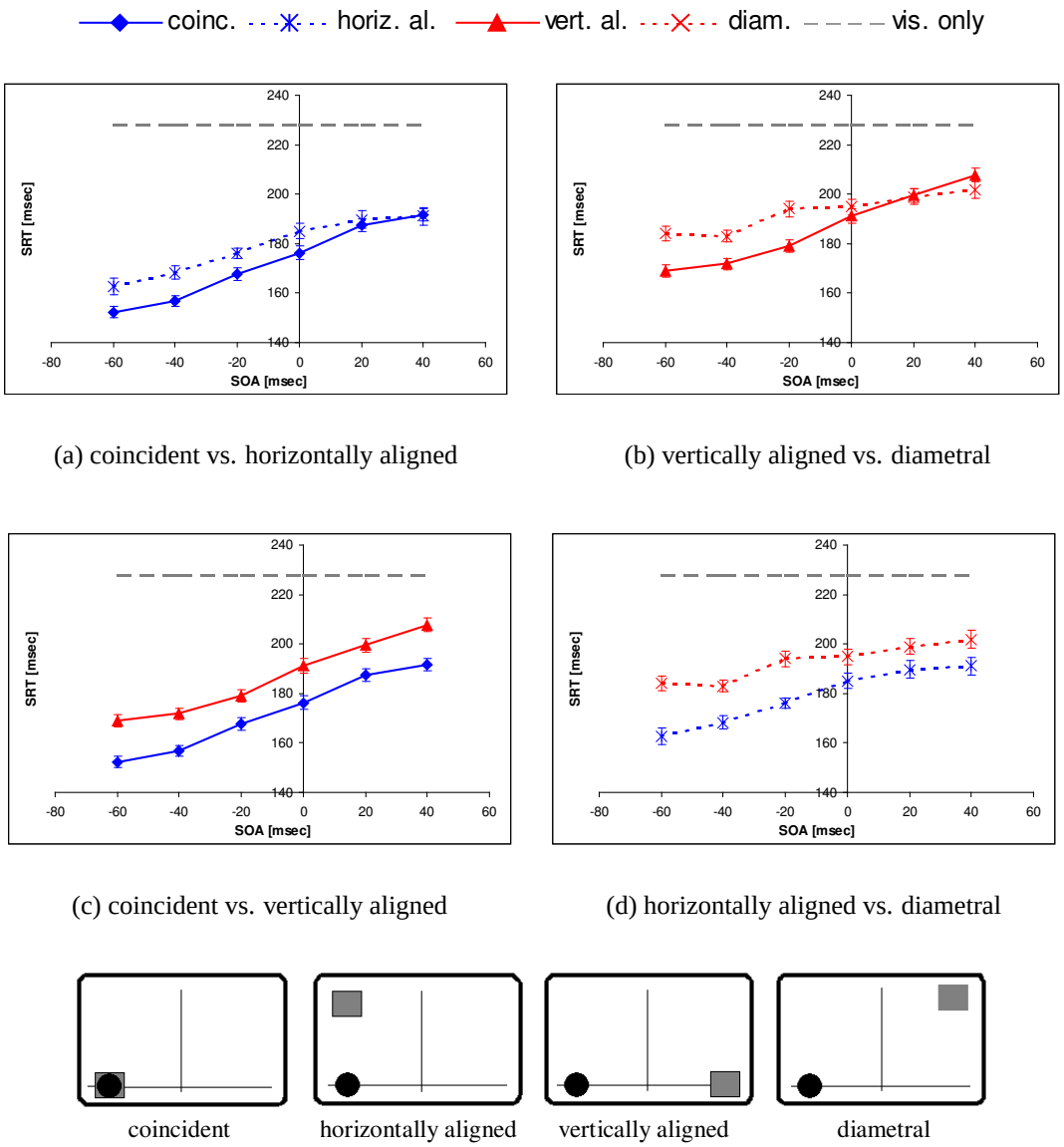


Figure 26: Examples of visually guided saccades in bimodal condition.

the characteristics of unimodal and bimodal visually guided saccades (see also Figure 19) shows that neither trajectories nor position-time traces differ distinctively. Bimodal visually guided saccades are highly reproducible, direct and rarely corrected for, just like their unimodal counterparts. If compared with the acoustically evoked eye movements (see Figures 17 and 18), it becomes clear that visually guided saccades, also under bimodal stimulation, hardly ever have the curved form of auditory guided eye movements. Apparently, the saccades in the bimodal task are purely visually driven. The participants in this experiment were able to follow the task and suppress any response toward the auditory stimulus.

Saccadic latencies

Figure 27 compares mean saccadic latencies for participant JB (good auditory localizer in virtual acoustic setup) across SOA for different stimulus configurations.



(a) coincident vs. horizontally aligned

(b) vertically aligned vs. diametral

(c) coincident vs. vertically aligned

(d) horizontally aligned vs. diametral

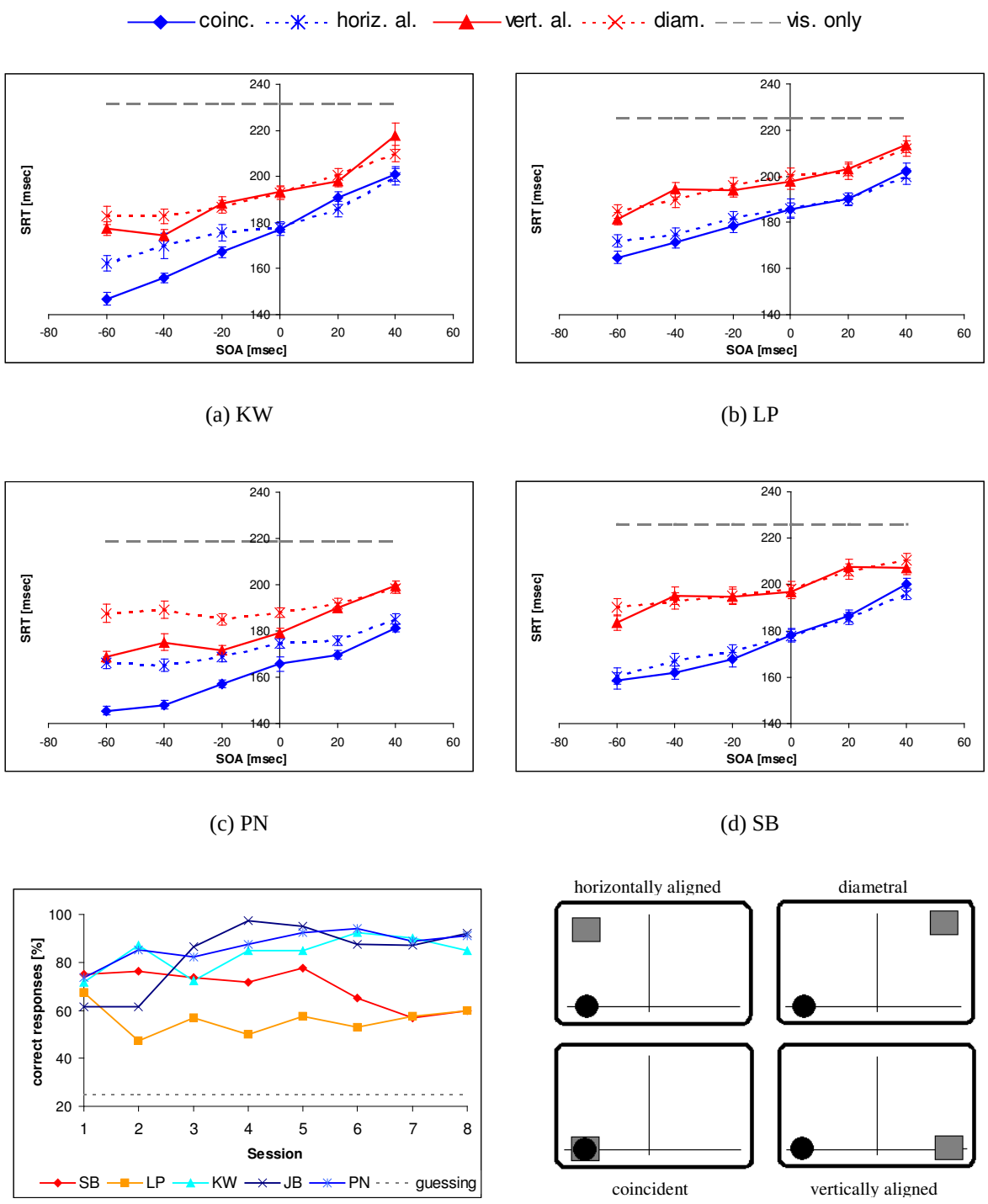
Figure 27: Saccadic latencies under different spatial (graphs) and temporal (abscissa) conditions for participant JB, as described in the text. a and b: Effects of vertical interstimulus distance. Vertical distance has an effect on saccadic latencies only for negative SOAs. c and d: Effects of horizontal interstimulus distanced. These remain constant across SOAs in the range considered here. Possible stimulus configurations for one target position are displayed at the bottom. All four target positions were used.

Six different stimulus onset asynchronies and five spatial configurations were possible: “visual only” (unimodal case), “coincident” (visual and auditory stimuli presented at the same location, spatial distance 0°), “horizontally aligned” (both stimuli having the same azimuth, but distinct

elevation components, distance = 20°), “vertically aligned” (the opposite combination, 50°), and "diametral" (stimuli both horizontally and vertically disaligned, distance 54°), also see Fig. 27d. The data were pooled across target position.

Three main observations can be made with respect to the bimodal saccadic reaction times. First, saccadic latencies were always shorter with bimodal stimulus presentation than with unimodal visual stimulation, even if the accessory stimulus came from positions more than 50° away from the target and/or was presented 40 msec after the visual target. There was no inhibiting influence of the auditory stimulus. Secondly, the strength of the intersensory facilitation effect (IFE, defined as the difference between unimodal and bimodal latencies) decreased monotonically with SOA. This holds for all spatial stimulus configurations used here. Thirdly, IFE also decreased with increasing spatial distance, although this effect was somewhat more complex. An influence of spatial distance could be found within both the horizontal and the vertical dimension. However, effects of vertical stimulus eccentricity were more pronounced the more the auditory stimulus precedes target onset (Figure 27, panels a and b). Saccadic reaction times to coincident stimuli were faster than to stimuli which differ in their vertical components – but only if the auditory accessory stimulus preceded the visual target. By contrast, effects of horizontal distance seemed to be independent of SOA: the difference between saccadic latencies for coincident stimuli and vertically aligned stimuli did not vary significantly across SOA (panel c), neither did the SRT-difference between horizontally aligned and diametrically presented stimuli (panel d). Hence, if the auditory stimulus succeeded the visual target onset only azimuthal distance seemed to play a role, while for negative SOAs both horizontal and vertical distance components are taken into account.

Considered across all participants in the virtual auditory environment, saccadic latencies are always significantly smaller in the bimodal than in the unimodal case (all $F's > 136$, all $p < 0.001$) and the intersensory facilitation effect decreases monotonically with SOA, independent of spatial configuration. Two-way ANOVAs of the bimodal trials revealed highly significant main effects of both SOA and spatial stimulus configuration (all $F's > 41, p < 0.001$ for any participant) and an interaction between both for participants JB ($F(15,1683) = 2.3$, $p < 0.030$), KW ($F(15,1914) = 2.4$, $p < 0.020$), PN ($F(15,2001) = 4.2$, $p < 0.001$), and SB ($F(15,1764) = 2.0$, $p < 0.014$). For the data of JB, KW, and PN, subsequent Student-Newman-Keuls tests ($\alpha = 0.05$) revealed an effect of vertical eccentricity for negative SOAs, while for positive SOAs the conditions “coincident” and “horizontally aligned” form a homogeneous subgroup as do “vertically aligned” and “diametral”. Figure 28 shows mean saccadic latencies for each participant separately across all temporal and spatial conditions. For JB and KW, the patterns of the latency-curves follow that of Fig 27, while for two other participants (LP, SB), only an effect of horizontal interstimulus distance is found to be significant. However, if these data are compared to those of Experiment 1 (panel e), it turns out that LP and SB are just those participants who were poor auditory localizers. Obviously these two could not use the auditory spatial information to its full extent (i.e. they simply did not notify the elevation component of the sound) and thus did not produce a significant effect of vertical interstimulus distance.



(a) KW

(b) LP

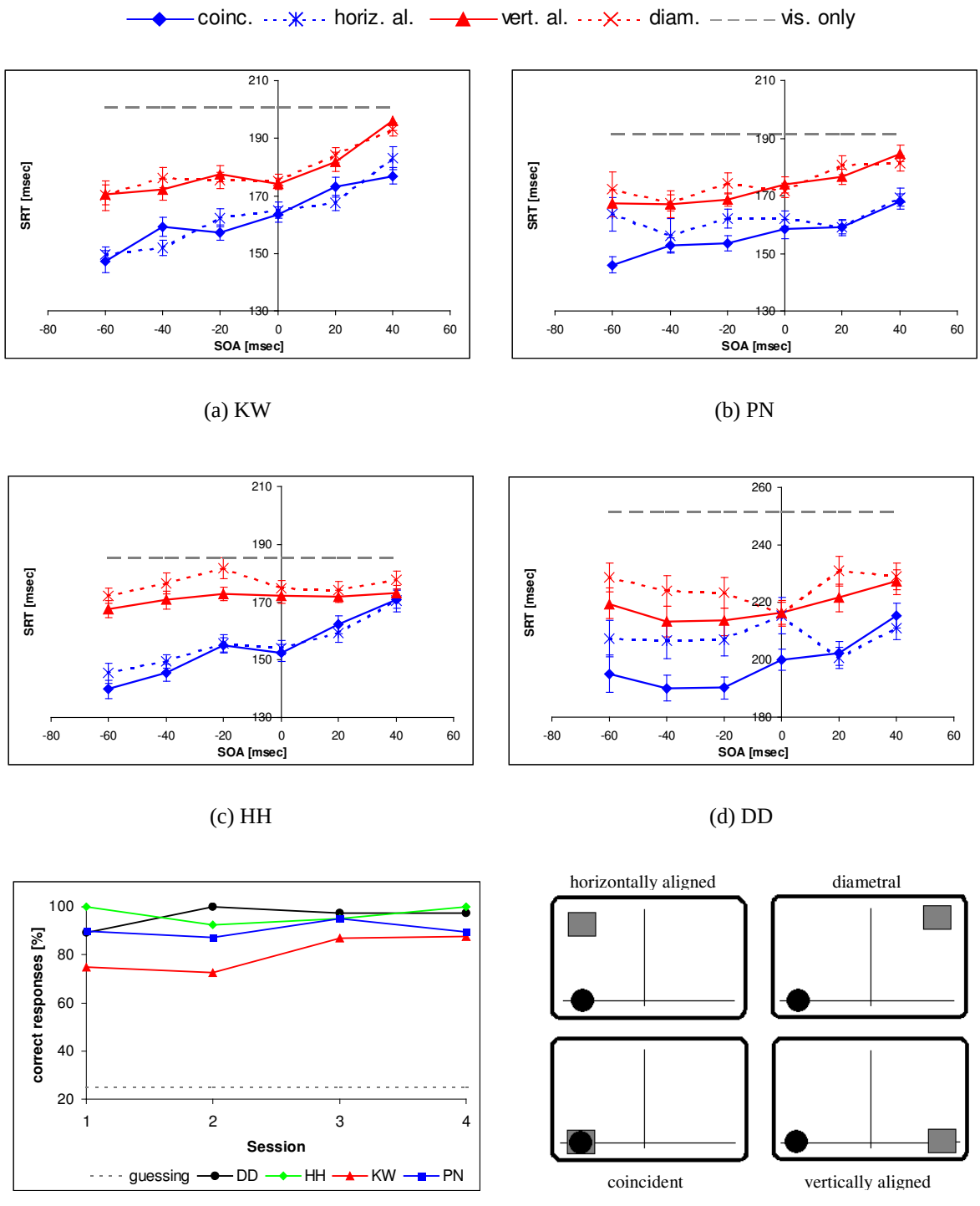
(c) PN

(d) SB

(e) auditory localization performance

Figure 28: Saccadic latencies under different spatial (graphs) and temporal (abscissa) conditions for the other four participants. Compare with auditory localization-performance (panel e) and note that these participants with poor localization performance (LP, SB) show no vertical SOA-effect.

The free field data showed generally comparable effects, as can be seen in Figure 29.



(a) KW

(b) PN

(c) HH

(d) DD

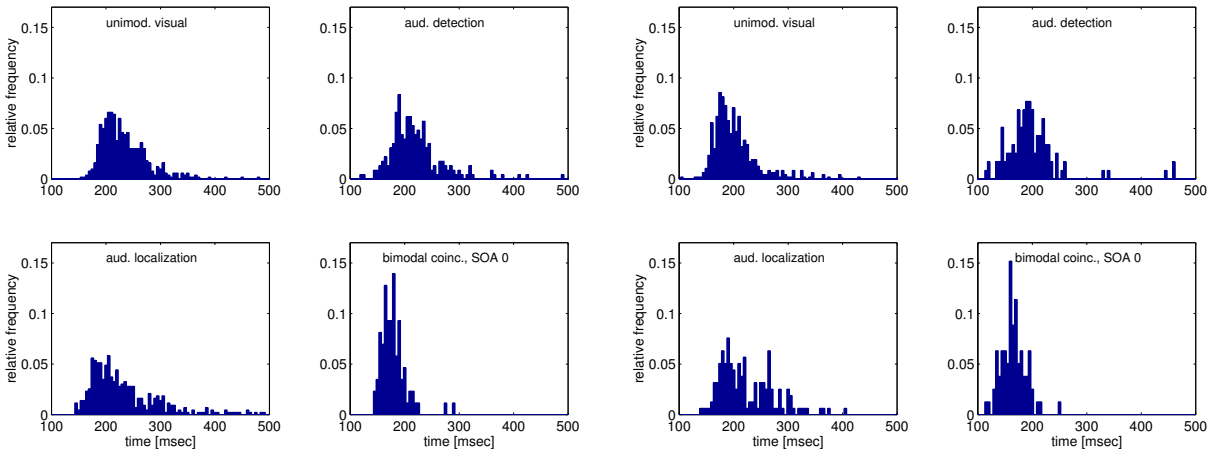
(e) auditory localization performance

Figure 29: Saccadic latencies under different spatial (graphs) and temporal (x-axis) conditions for four participants in the free field experiment. Note the different scalings of the ordinates in panel a-d in comparison with Figure 28.

Again, reaction times were significantly shorter in the bimodal than in the unimodal case (all $F's > 217$, all $p < 0.001$) and the IFE decreased monotonically with SOA, independent of spatial configuration. Two-way ANOVAs of data of the bimodal trials revealed main effects of both SOA and spatial stimulus configuration (all $F's > 3.2$, all $p's < 0.01$ regarding the factor SOA and all $F's > 17$, all $p's < 0.001$ regarding the factor spatial configuration, for each participant), but a significant interaction only for participant HH ($F(15,1613) = 3.2$, $p < 0.001$). A significant effect of vertical eccentricity on saccadic latency was found at negative SOAs in three out of four participants (DD: $F(1,973) = 17.7$, $p < 0.001$, HH: $F(1,815) = 3.15$, $p < 0.001$, and PN: $F(1,768) = 8.49$, $p = 0.004$), but there were no significant vertical distance effects for $SOAs \geq 0$). Again, there is a relation between auditory localization performance and vertical distance effects in bimodal interaction: participant KW, who showed no vertical distance effect, also revealed low auditory localization performance concerning elevation estimation.

Latency distributions

A comparison of the trajectories of the saccades under various stimulus conditions in Experiments 1 and 3 revealed that the responses in the bimodal trials were decidedly visually driven. Nevertheless, the auditory accessory stimulus had a significant influence on mean saccadic latencies, as shown in Figures 26, 28, and 29. An analysis not only of means, but also of the latency distributions for the different conditions might allow further insight into the mechanisms of these bimodal facilitation effects. Figure 30 shows distributions of saccadic latencies of directed and undirected auditorily guided responses, of unimodal, and of bimodal visual responses (coincident stimuli with SOA = 0 msec) for participant KW in the virtual acoustic and in the free field setup, respectively. Further comparative latency distribution plots can be found in the Appendix.



(a) KW, virtual acoustic environment

(b) KW, loudspeaker environment

Figure 30: Latency distributions for saccadic responses of participant KW in the auditory localization, auditory detection, unimodal, and bimodal visual reaction time experiments.

Results are again quite consistent in both setups. The respective distributions are generally very similar, although the free field latencies tend to be somewhat shorter. The differences between auditory and bimodal visually evoked reaction times are explicit. Bimodal latencies furthermore show significantly less variance than unimodal visual response times and thus resemble the auditory detection time distributions for some participants (such as JB or HH, see appendix) – however, they are not identical. The bimodal latency distributions displayed here are more peak-like than those for undirected auditory responses of the same participant and their maximum can be found at smaller SRT-values. Indeed, this is not the case for all bimodal distributions, as can be observed in Figure 31. Bimodal latency distributions become flatter with growing interstimulus distances and SOAs. They furthermore shift toward larger latency values. Hence, under 'less ideal' conditions, auditory simple responses are performed faster than a bimodal response. An influence of the different positions of the accessory stimulus on the *form* of the latency distributions does not become apparent at any SOA (except for an obvious general shift to slower latencies). Figure 31 demonstrates that bimodal responses are not simply temporally triggered by the perception of an auditory accessory stimulus, just as their trajectories are not directed by it.

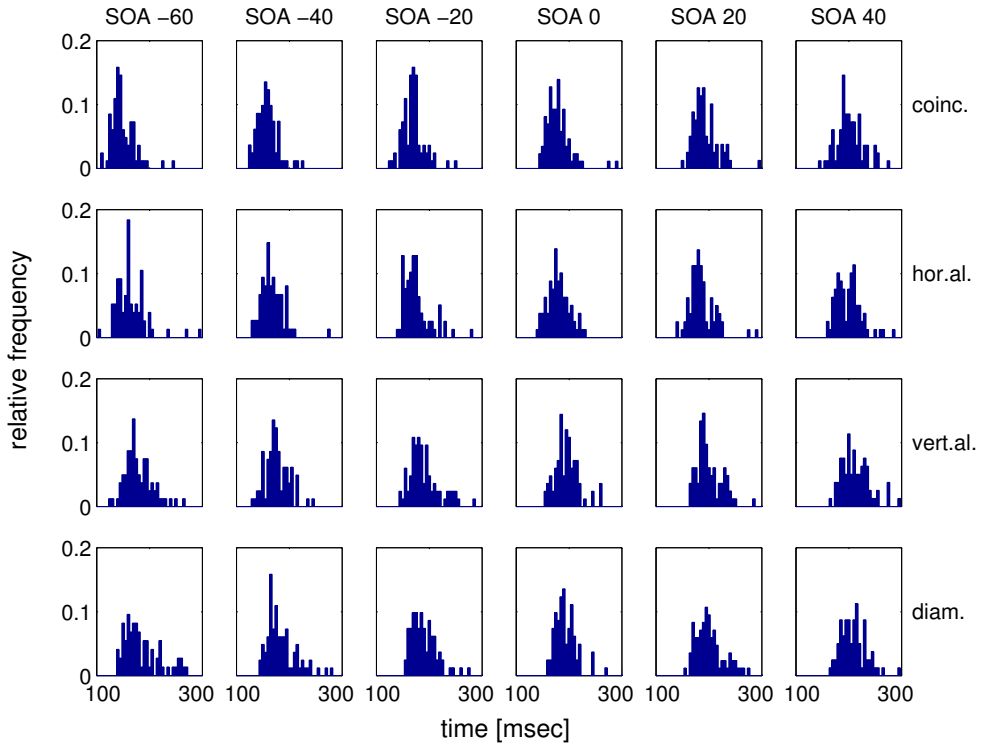


Figure 31: Latency distributions of saccadic responses by participant KW in the bimodal reaction time experiment (virtual acoustic setup). The various temporal conditions are displayed in different columns, spatial conditions in rows. Further bimodal latency distributions can be found in the Appendix.

A very useful application of saccadic latency distributions is to determine the magnitude of intersensory integration effects by analyzing them in terms of Miller's Race Model Inequality. In section 2.1, it has been outlined that the upper limit for bimodal response acceleration in a parallel system, i.e. simply by statistical reasons, is represented by the sum of the cumulative distribution functions (CDF) of the unimodal latencies, that is

$$P(RT_{AV} < t) \leq P(RT_A < t) + P(RT_V < t).$$

Any violation of this inequality rules out strict separate-activation approaches assuming no bimodal information integration. In equivalence to the above formula, the magnitude of violation is given by

$$Violation = CDF(\text{bimodal SRT}) - [\, CDF(\text{visual SRT}) + CDF(\text{auditory SRT}) \,].$$

In the following analysis, we will use the saccadic latency distributions for *directed* auditory responses, as the responses in the bimodal experiment were also directed ones.

The concept of statistical facilitation in Separate Activation or Independent Race Models had originally been developed for explaining response acceleration in so-called *divided attention tasks*, where participants were allowed to react toward whichever stimulus they first perceived. The possibility of an application to data collected in a *focussed attention paradigm* (like in the present study) is not so straightforward and the topic has been discussed controversely. The problem is as follows: in a divided attention task, both stimuli separately activate a response evoking mechanism and the finishing of the first process alone determines the response time. There is no need for any information to be exchanged between the different sensory channels. If an onset asynchrony is introduced to be (say, the auditory signal starts 40 msec prior to the visual stimulus), the bimodal reaction times simply follow the unimodal distribution of the first stimulus, see Figure 5. But what happens if the participant is not allowed to respond to the first stimulus? Is the activation induced by the arrival of this accessory stimulus maintained and added to the target stimulus activation? Or is it suppressed, since it is non-valid? The latter is evidently not the case, as clear latency reductions can be observed in these situations. It becomes clear that the Miller Inequality cannot be used in a focussed attention paradigm without making further assumptions about the internal processes, except for one situation.
If the SOA is chosen to be zero, it can be observed that bimodal latencies in a divided attention task are shorter or equal to bimodal latencies in the corresponding focussed attention task. This makes sense not only due to purely statistical assumptions (see above), but also when considering that sensory processing of auditory stimuli takes place faster than visual processing. Hence, if a violation of Miller's Inequality (Miller 1982)can be found in a focussed attention paradigm with SOA=0msec, even the more violation can be expected in a divided attention task in the same situation. In other words, a violation can be regarded as clear proof for intersensory integration in such a case. Figure 32 shows the results of an analysis of the present data at SOA=0msec with respect to the Miller Inequality.

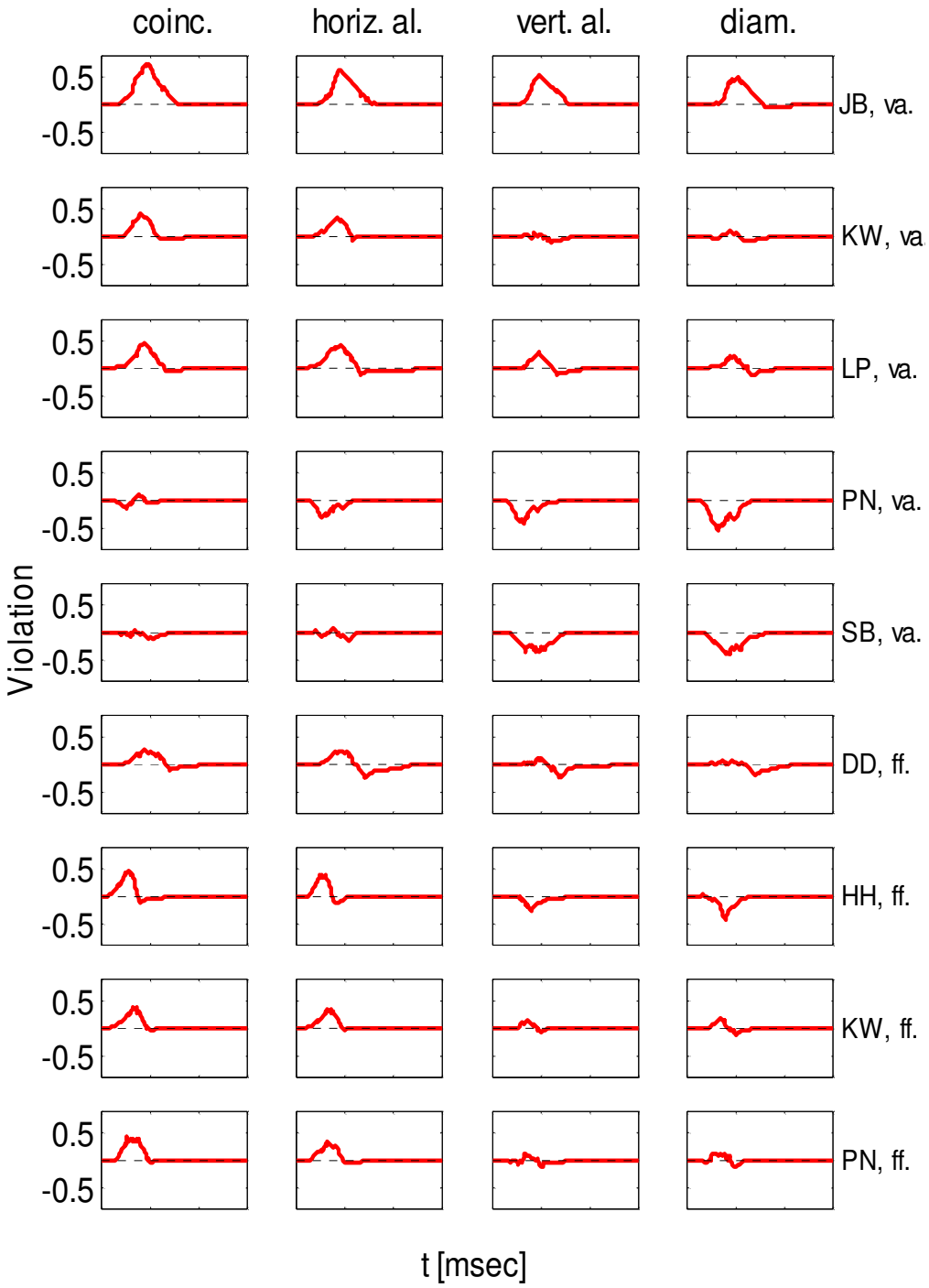


Figure 32: Violation of Miller Inequality at SOA=0 msec for all spatial conditions (columns) and participants (rows) in both experimental setups (va: virtual acoustic, ff: free field). Spatial distance increases from left to right.

A significant violation of the Race Model Inequality can be found for almost all participants in the virtual acoustics as well as in the loudspeaker setup. The magnitude of the violation depends on the spatial interstimulus distance. In most cases, there is only a violation if the accessory is presented ipsilaterally to the target. Hence, at least for these cases it can be concluded that sensory information is merged in the nervous system and thereby leads to reduced processing times at later processing stages.

Discussion of Experiment 3

The data obtained in this study display the strong effect of an additional auditory stimulus in an experimental task that requires participants to concentrate on and exclusively react to a visual target. Surprisingly, the accessory signal never seemed to work as a real distractor at all: bimodal saccadic latencies were always shorter than unimodal ones, even if visual and auditory stimuli were presented contralaterally and more than 50° apart. The strength of intersensory facilitation simply fades with rising interstimulus distance. This finding might in part be due to the fact that there were no catch trials (i.e. trials in which only the auditory distractor was presented). The fact that the auditory stimulus was always accompanied by a target made it a relevant alerting cue. Introducing catch trials should therefore significantly reduce the amount of response acceleration and also reveal inhibition.

Nevertheless, there are several reasons why the finding of latency reduction under bimodal stimulation can hardly be attributed to the possibility that participants had reacted toward the acoustic stimulus. Firstly, the position of the acoustic stimulus did not enable participants to predict target position. Secondly, comparing trajectories, it turns out that visually guided saccades under unimodal and bimodal conditions do not differ significantly. They are performed straight toward the target and are highly reproducible. Curved trajectories and a generally larger scatter in the movements seem, for their part, to be characteristics of auditory guided saccades making these clearly separable from visually directed responses. The same conclusion can be made in the temporal domain on the basis of latency distributions. Although the distributions of bimodal visually guided saccades are often narrower than those of unimodal visual responses, bimodal responses cannot be simply triggered. In some cases, bimodal responses are even faster than those in the simple auditory detection task and their distributions are even narrower. Finally, applying the data to the Miller-Inequality proves that the response acceleration found here cannot simply be explained by statistical considerations, but that intersensory information integration has to be assumed.

At first glance, the finding of rising IFE with an increasing temporal gap between auditory and visual stimulus presentations seems contradictory to the idea of spatial facilitation through neural summation. An interstimulus interval of about +40 msec, in which the auditory stimulus *succeeds* visual stimulus presentation, should just make up for the "neural delay" between visual and auditory transmission and thus maximize spatial facilitation. However, physiological data of Superior Colliculus single cell recordings show time windows up to 1500 msec for visual-auditory neurons (with an optimum interval of about 100 msec), within which neural response enhancement by bimodal inputs is possible (Meredith, Nemitz & Stein 1987, Stein & Meredith 1993). Incoming sensory signals sharing the same receptive field might thus have quite a

long time interval and still enhance the cell's response. An interpretation of spatial integration might also be made under a larger-scale view, looking at SC saliency maps. Incoming auditory information leads to arousal of buildup neurons, which then fades during a certain decay time. If sensory input from a nearby target stimulus arrives within that time, motor programming is enhanced by the already existing arousal and the threshold for response initiation is reached faster. If auditory and visual stimuli come from different directions they excite different areas of the saliency map and thus do not interact. In that case, however, a preceding auditory stimulus might still inhibit fixation neurons which in turn leads to a general enhancement of buildup neuron activity.
These hypotheses closely follow the assumptions of Frens et al. (1995) and are in line with the Findlay-Walker (1999) framework of separate WHEN and WHERE programming of saccades described in section 2.2. It can be assumed that the later the accessory stimulus is presented with respect to the target, the stronger spatially unspecific effects in the WHEN-circuitries dominate over effects of neural summation within the WHERE-pathway.

However, the complex pattern of different spatial effects at various SOAs (which are no plain distance effects) need some more explanation. It turns out that if the auditory accessory stimulus precedes the visual target by a sufficient temporal interval, vertical as well as horizontal interstimulus distance has to be taken into account. In contrast, if the auditory stimulus follows the visual signal, only horizontal distance is of importance, any vertical eccentricity seems not to be perceived.
Horizontal and vertical auditory stimulus position seem to be features which might be used in combination (in the case of negative SOAs) or alone (the horizontal distance in the case of positive SOAs). In Experiments 1 and 2, it turned out that the detection of an auditory stimulus seems to take place faster than localization, and that in localization the evaluation of horizontal position is often faster than estimating vertical position. In auditory guided eye movements this results into stepped position-time traces or bow-like trajectories due to subsequent information update. In the bimodal Experiment 3 such an auditory update cannot be performed once the visually guided saccade has been calculated. Hence, what is crucial in this situation is the amount of auditory information (or which feature) arrives before the visual sensory processing is completed.

Frens & Van Opstal (1995) could convincingly show that it is the *perceived* and not the physical interstimulus distance which determines the magnitude of intersensory facilitation. They used pure tones as acoustic accessory stimuli at distinct positions, including elevation. However, the elevation eccentricity of a narrow-band sound signal is usually not perceived correctly, it is rather determined by its center-frequency irrespective of the real physical position (Blauert 1983). This is exactly what Frens et al. found in their auditory localization task. In a bimodal focused attention task performed subsequently, it turned out that the amount of facilitation only depended on the horizontal distance between visual and auditory stimuli and that any vertical component, although physically valid, was not used. In some respects, we have a comparable situation in our study, with the difference that the (complete) perception of the acoustic signal is restricted not by spectral stimulus properties but by temporal limitations in the analysis, namely the task to perform a response as soon as the visual target is perceived.

The results of the present study give interesting insight into the performance and characteristics of the auditory system in (involuntary) localization of sound sources and projection of information onto the oculomotor map for the purpose of interaction with the processing of a visual target stimulus. These findings should be taken into consideration in the modelling of such processes.

4 Bottom-up and top-down processes in visual-auditory interaction

In the three experiments described above it turned out that the perceived location of an auditory stimulus changes during the process of localization. While the pure presence of an acoustic stimulus is detected very fast (see 3.3), its perceived location seems to have a specific time course. Azimuthal position of auditory signals is evaluated on the basis of interaural time and intensity differences in the upper Olivary Complex early in auditory stimulus processing. As a consequence, information about horizontal position is available early on. In contrast to this, elevation has to be calculated from spectral cues induced by the listener's pinna folds. An analysis of these auditory spectral patterns in primates is probably carried out by cortical regions in a serial manner (Recanzone, Guard, Phan & Su 2000). The fact that the spectral content of the auditory signal is not known to the listener in advance, makes the computation of elevation from the modified stimulus spectrum a time consuming process. Apart from that, the correct calculation of a signal's spectrum is always a time consuming process due to physical reasons (signal processing theory). The more exact the spectral information of a signal is wanted, the longer is the time interval needed for its analysis . Hence, a sufficiently long period of an auditory stimulus has to be analyzed before obtaining a reliable estimate of its elevation. This phenomenon can directly be observed in behavioral experiments, like in Experiment 1 of the present work. In an auditory localization task by Frens & Van Opstal (1995) it was shown that very short auditory stimuli with durations up to 10 ms are localized incorrectly. The elevation component of these stimuli is underestimated, whereas the azimuthal position is still perceived correctly.

Experiment 3 (3.4) investigated whether and how the time course of auditory processing affects intersensory facilitation. Two points have to be made with respect to the differences between a localization task with auditory stimuli and a focussed attention task investigating intersensory facilitation in visually evoked responses. First, in the latter task attention is withdrawn from the auditory accessory stimulus. Although Alho (1992) has shown that physical features of auditory stimuli are extensively processed even in the absence of attention, differences in processing might occur. Second, intersensory facilitation can also be elicited by the pure presence of an accessory stimulus, i.e. at least part of the IFE effect may be based on detection of the auditory stimulus and thus independent of localization. Arndt and Colonius (submitted) have shown that only a part of the components of IFE depends on spatial distance between stimuli, whereas other components are independent of localization. On the other hand, Experiment 3 revealed that a remarkable amount of the IFE depends on the *perceived* distance between visual target and accessory auditory stimulus. Frens et al. (1995) explicitly introduced two facilitating mechanisms in their model on visual auditory interaction, one spatially specific and one unspecific. Although these components of IFE are superposed, we predict the IFE to depend on the state of auditory processing that is reached in the moment when the response to an imperative stimulus is elicited or executed.

Using the focused attention paradigm, it is possible to manipulate the state of processing by presenting an accessory auditory stimulus with different temporal delays (SOAs) with respect to the visual target stimulus. In Experiment 3 (3.4, SOAs were chosen such that for certain SOAs the localization of the auditory stimulus could be assumed to be incomplete when the (visually guided) response was elicited. We now particularly aim to test the hypothesis, whether in case of a lack of auditory accessory information top-down or bottom-up processes determine the perceived spatial distance between visual target and auditory accessory. Three hypotheses are to be tested:

1. Top-down processes provide a specific default value, i.e. in case of incomplete signal processing the auditory stimulus is always perceived at a certain position. This position might be "naturally given" (e.g. at ear height) or may be influenced by long term or short term experience and by expectation.

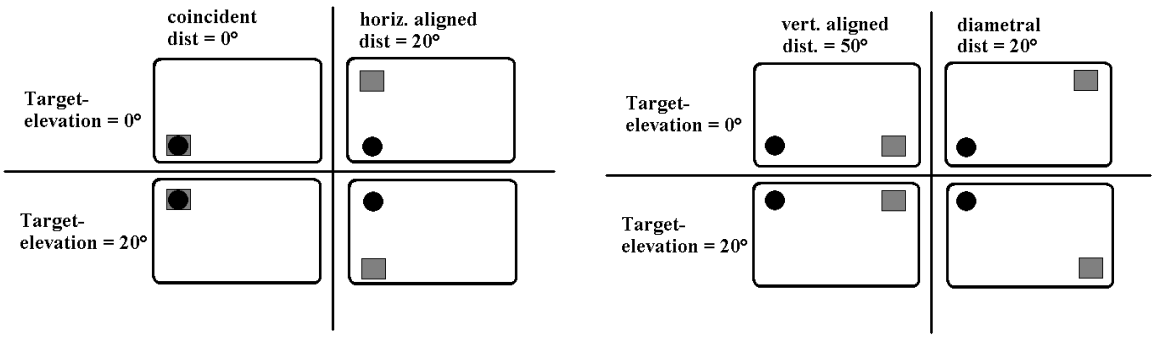
2. Bottom-up processes provide a preliminary estimate of auditory stimulus elevation. As suggested by Hofman & Van Opstal (1998) a "first estimate value", based on rough short term spectral analysis may be calculated and improved subsequently.

3. Neither top-down nor bottom-up processes provide any clue. In this case, the elevation of the auditory stimulus is not fixed at all, which would lead to a blurred representation along the vertical axis at a specific azimuth.

The three hypotheses lead to different expectations about the IFE data in the case of incomplete processing of the auditory stimulus. No assumption about auditory elevation (Hypothesis 3), i.e. a blurred, vertically elongated representation, yields approximately equal IFE values for all target elevations and vertical distance conditions. If the listener turns to a default value (Hypothesis 1), the exact magnitude of IFE depends on which default value is actually chosen. In any case however, under incomplete processing of auditory location, IFE should be rather determined by target position, as the listener's percept of auditory elevation is "fixed" to default and variation in perceived distance should merely depend on variation in target elevation. In contrast, under the assumption that the listener roughly estimates the elevation of the auditory stimulus (Hypothesis 2), IFE should be dependent on (physical) stimulus distance, as some information (albeit unreliable) about auditory vertical position is available even at this early stage of processing.

4.1 Possible elevation assumptions in virtual acoustics

In order to test these hypotheses, the data of Experiment 3 is analyzed in a slightly different manner than there. Since the regarded IFE is a relative function between unimodal and bimodal visually evoked saccadic latencies, it is possible to directly compare the data of all participants in one experimental setup. Data were therefore pooled across subjects (see Figure 34a). As we further want to concentrate on effects of vertical distance, results of ipsi- and contralateral stimulus presentation are observed separately, that is, IFE values are compared for those stimulus pairs with identical horizontal component (coincident vs. horizontally aligned stimuli on the one hand and vertically aligned vs. diametrally positioned stimuli on the other). Moreover, we regard responses toward targets in the horizontal plane separately from those to elevated targets, as illustrated in Figure 33. These results are displayed in Figure 34b and c.



(a) ipsilateral stimulus presentation

(b) contralateral stimulus presentation

Figure 33: Graphical illustration of data analysis. Black circles indicate target position, grey rectangles stand for the accessory's location. Stimulus combinations displayed in panel a correspond to Figure 34b, those in panel c to the data of Figure 34b.

Figure 34a shows IFE as a function of SOA and spatial distance pooled across all target positions, comparable to Figures 28 and 29 in Experiment 3, but here the data was pooled across all participants. In analogy to those figures, it can be seen that vertical interstimulus distance is effective for $SOAs < 0$ only. Furthermore IFE decreases with increasing SOA, which is due to the superposition with an spatially unspecific effect already found in other studies, probably a warning effect. In the following, it will be our attempt to exclude these unspecific SOA effects by investigating *relative* changes between the different IFE-function only.
We therefore calculate *IFE differences*, which indicate the changing effect of one specified factor (interstimulus distance, target elevation, or auditory elevation) across SOAs. Comparing the time courses of IFE difference functions might reveal interactions among stimulus parameters. Any other factors (especially general warning effects) will, on the other hand, be discarded through calculating the differences, given independency of the parameters.

We will use the following denotation of IFE at different spatial positions

IFE_{VA}	IFE for spatially coincident stimuli with elevation 0°
IFE^{VA}	IFE for spatially coincident stimuli with elevation 20° both
IFE_V^A	IFE for horizontally aligned stimuli with target elevation 0° and auditory accessory elevation 20°
IFE_A^V	IFE for horizontally aligned stimuli with target elevation 20° and auditory accessory elevation 0°

Interhemispheric effects will not be investigated further, i.e. only ipsilateral stimulus combinations as plotted in Figure 34b are taken into account in the following analysis.

Figure 34b (ipsilateral stimulus combinations, compare with Figure 33a) reveals that there are in fact effects of target position on the IFE that seen to intermingle with the influence of vertical spatial distance. For $SOAs < 0$, distance effects clearly dominate IFE, as IFE_{VA} and IFE^{VA} do not differ significantly from each other, neither do IFE_V^A and IFE_A^V. In the case of positive SOA, the situation is just the other way. IFE values for responses to elevated visual targets are now larger than those for saccades within the horizontal plane., i.e. $IFE^{VA} = IFE_A^V$ and $IFE_{VA} = IFE_V^A$. Vertical interstimulus distance does not play a role any more. Figure 34c shows the same situation for contralateral stimulus combinations (see Figure 33b). The results are in general comparable, although IFE values are smaller and standard errors are larger. Nevertheless, we will concentrate on the data from ipsilateral stimulus presentation in the following analyses of relative IFE effects.

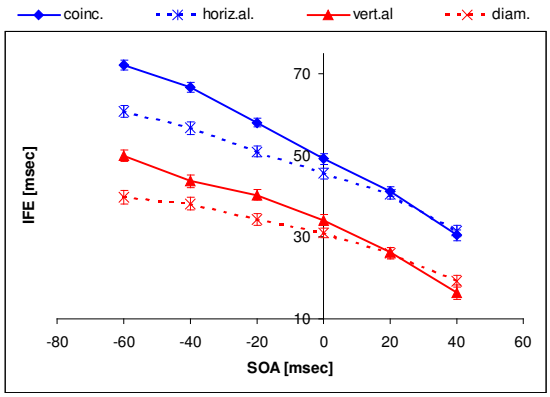
The hypothesis that, in the case of incomplete stimulus processing, no assumptions about the elevation of the auditory stimulus are made at all (Hypothesis 3) can be ruled out directly on the basis of the data. A comparison of the amount of IFE for positive SOAs shows differences between target elevations of 0° and 20° (Figure 34b), which should not occur according to Hypothesis 3. Moreover, the data support the "default value" over the "first estimate" hypothesis. According to the "first estimate" assumption (Hypothesis 2) the amount of IFE should be determined by the position of the auditory stimulus with respect to the visual not only for negative but also for positive SOAs. The latter is not the case. Rather, the IFE seems to depend on the position of the target per sé.

Spatial effects for different target positions

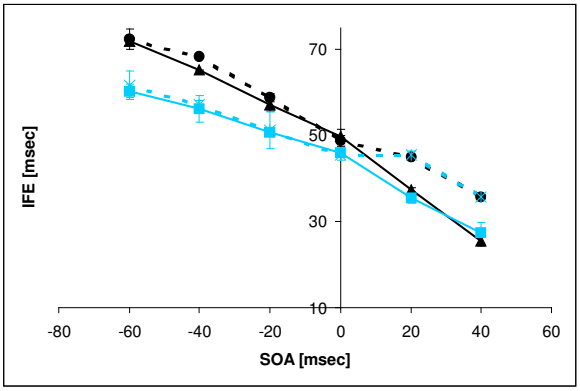
The influence of (changing) perceived interstimulus distance on IFE at given target positions across SOA can be observed by calculating the difference between IFE for coincident and IFE for horizontally aligned stimuli at defined target elevation:

$$\begin{aligned} \Delta_V &= IFE_{VA} - IFE_V^A \\ \Delta^V &= IFE^{VA} - IFE_A^V \end{aligned} \tag{1}$$

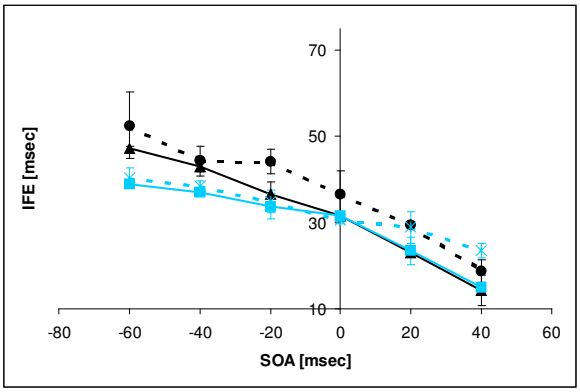
Like the IFE, the different Δs are functions of SOA. Positive values indicate larger spatial facilitation for coincident stimulus presentation compared to vertically separated stimuli, while values around zero suggest that physical distance between target and accessory does not affect reaction times. The two difference functions are displayed graphically in Figure 35a. The graphs show that the influence of interstimulus distance on IFE decreases with increasing SOA.



(a) pooled across hemispheres



(b) ipsilateral stimuli only



(c) contralateral stimuli only

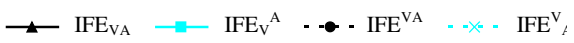
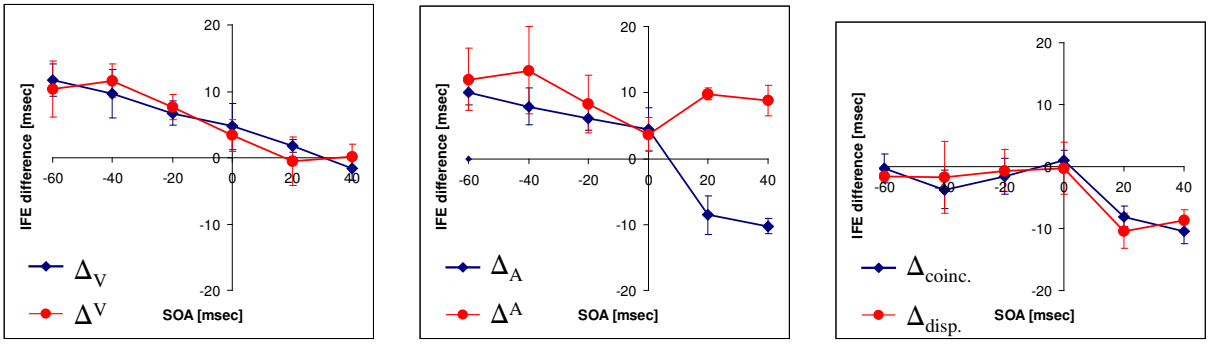


Figure 34: IFE (and standard errors) in bimodal saccadic latencies across SOA, from Experiment 3, virtual acoustic environment. Data was pooled across al participants. a: across all target positions, equivalent to the presentation of Figure 28 b: for ipsilateral stimulus presentation only, c: for contralateral stimulus presentation only (not analyzed further).

The benefit of spatial coincidence is only given for preceding accessories and vanishes at $SOA > 0$. This effect can be found for target elevation 0° and 20°. Moreover, the curves for 0° and 20° target elevation are nearly identical, which means that the changes of disparity effects across SOA are independent of absolute target position. Our results indicate that either the vertical distance is simply ignored (since not recognizable) at positive SOAs, or that the location of the auditory appears to the listener to have approximately the same distance to the target in both configurations (coincident or vertically disparate), irrespective of physical distances.



(a) Effects of different target elevations

(b) Effects of different auditory elevations

(c) Effects of absolute stimulus configuration

Figure 35: IFE differences in bimodal saccadic latencies across SOA, ipsilateral stimulus presentation in virtual acoustic environment. Data was pooled across al participants.

These findings are generally compatible with all three hypotheses we are testing. For a blurred representation without localized elevation (Hypothesis 3) one would predict exactly this outcome. However, this hypothesis was already ruled out on the basis of Figure 34b. The data also imply that the perceived elevation is the same for accessory auditory stimuli presented from 0° and from 20° elevation at a given target position. Note that this does not automatically mean that the auditory accessory is localized just midway between both possible positions. Theoretically, it might be perceived at any elevation. If, for example, a listener judges a sound coming from the horizontal plane in a certain situation, $IFE_{VA} - IFE_V^A = 0$ as well as $IFE^{VA} - IFE_A^V = 0$, as either the perceived vertically disparate stimuli are then perceived as being presented coincident (in the first case of target elevation being zero) or a coincident stimulus pair is judged as vertically disparate (like in the second case). Similar assumptions can be made for a perceived elevation of 20° or any other value.

While such ideas are plausible for a default value assumption (Hypothesis 1) they seem somewhat contradictory to the theory of a first estimate (Hypothesis 2). One might however argue that there is less confidence in a judgement based on a first rough estimate, and that there is a larger variance in the elevation estimates in these cases. A decrease in vertical distance effects at positive SOAs could then be explained by the fact that elevations of 0° and 20° are simply not as distinguishable as with a preceding auditory stimulus.

Spatial effects for different auditory elevations

To get further insight in the perceived vertical position of the auditory stimulus we now compare IFE for coincident and vertically disparate stimuli with fixed auditory positions:

$$\begin{aligned} \Delta_A &= IFE_{VA} - IFE_A^V \\ \Delta^A &= IFE^{VA} - IFE_V^A \end{aligned} \tag{2}$$

Like above, positive Δ-values indicate higher IFE for coincident that for vertically disparate stimulus presentation. The respective curves are plotted in Fig. 35b. Unlike with *target* elevation, spatial facilitation differs remarkably across SOA for Δ-functions corresponding to 0° auditory elevation. With 20° auditory elevation, IFE differences are nearly the same for negative and positive SOAs. All difference values are clearly above zero, which indicates that IFE is larger for spatially coincident stimuli than for vertically separated stimuli *or* that IFE is larger for visual targets with 20° elevation than for visual targets with 0° elevation, or both.
In contrast to this, IFE differences between coincident and vertically separated stimuli strongly decline at positive SOAs when the auditory accessory is presented from the horizontal plane. This means that the typically found benefit of coincident stimuli compared to spatially separate stimuli reverses for positive SOAs. This reversal could be interpreted as an advantage in IFE for visual targets with 20° elevation compared to visual targets with 0° elevation. At negative SOAs, i.e., when the auditory stimulus is processed completely before the response is elicited, this effect may be dominated by the well known benefit for coincident stimuli.

The results are compatible with two different interpretations for positive SOAs: Either the auditory is perceived closer to the upper target than to the lower target, regardless of its physical position. Alternatively, the position of the auditory is perceived midway between the possible target locations, but for saccades to targets with 20° elevation IFE is larger than for saccades within the horizontal plane.
Hypothesis 3 is clearly ruled out by the IFE difference functions observed here. The idea of a blurred representation implies that IFE is the same for all target positions at positive SOAs, an assumption which is severely violated here in two points. First, the values for both curves should be identical, which they obviously are not. Second, the curves should approach zero at positive SOAs, what they do not. As discussed above, Hypotheses 1 and 2 are in line with the results under the assumption that either the auditory signal is perceived closer to the upper target or that visual targets with 20° elevation benefit more from the auditory accessory than do those in the horizontal plane.

Effects of absolute stimulus configuration

To investigate, whether the higher benefit for targets at 20° is restricted to positive SOAs or equivalently can be observed for negative SOAs, we calculate the IFE differences between target elevations separately for coincident and disparate stimuli. The respective difference functions are

$$\begin{aligned} \Delta_{\text{coinc.}} &= IFE_{VA} - IFE^{VA} \\ \Delta_{\text{disp.}} &= IFE_{V}^{A} - IFE_{A}^{V} \end{aligned} \tag{3}$$

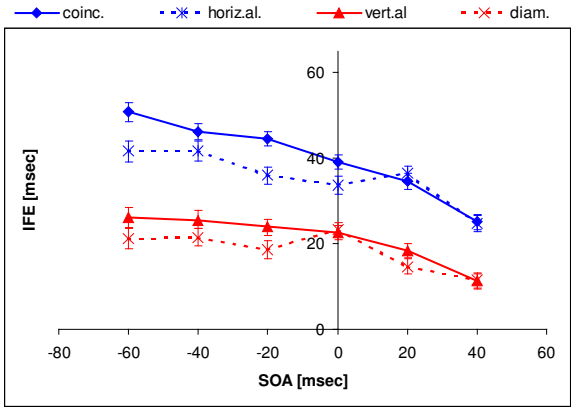
Nonzero Δ-values indicate that, at a given spatial distance, a particular stimulus configuration (i.e. target's and accessory's position and not only their relation) leads to a stronger IFE than another. Fig. 35c shows that difference functions have values around zero for negative SOAs, whereas they decrease at positive SOAs. Hence, target or accessory position do not influence spatial IFE, as long as the the visual target precedes the auditory accessory. Here the strength of IFE depends on physical distance between visual target and auditory accessory. In case of positive SOAs however, the IFE difference function changes remarkably to negative values. In this situation, intersensory effects are larger for elevated targets, independent of physical spatial distance between target and accessory. This finding further supports the assumption that the listeners here turned to a default elevation value when they were not able to judge the elevation of the auditory stimulus by spectral cue analysis. Again, the auditory signal is either perceived nearer to the 20° target or visual targets with 20° elevation benefit more from the auditory accessory than do those in the horizontal plane. The fact that both IFE difference functions have about the same values at positive SOAs further support the notion that the default, resp. first estimate is the same for accessory stimuli presented from 0° and from 20° elevation.

4.2 Possible elevation assumptions in free field

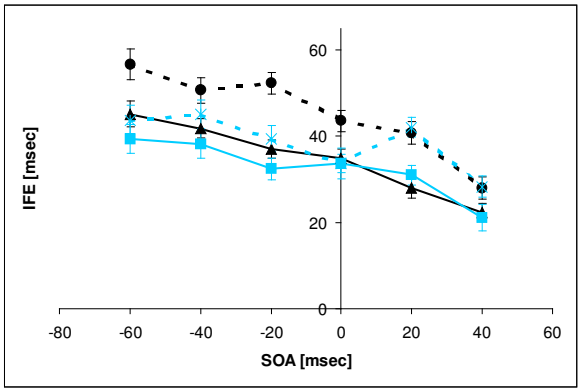
So far, no pronounced differences between using a virtual or a real auditory environment have shown up in the analyses of our Experiments. Localization performance was comparable at least after the participants in the virtual environment took some training, the auditory detection task yielded no differences, and the spatial effects in the bimodal reaction time experiment were also comparable. However, there have also been some hints that using virtual and loudspeaker acoustics are not exactly the same. So, KW was a good localizer in the virtual setup, but surprisingly he was not so fair in the loudspeaker experiment. In fact, both listeners who were used to the virtual environment had a decreased performance in the free field setup. Moreover, the second participant who attended in both environments, PN, showed significantly longer latencies in the free field localization task. Both findings might indicate that the participants in the virtual setup had learned and got used to some strategy how to localize in an unusual environment and that this new strategy was not the same as "natural" sound localization because it included certain assumptions regarding the stimulus environment (which positions are possible, how often does a stimulus come from a certain place, how exactly do they sound, and so on). The surprising finding of an elevated "default-elevation" somewhere between 10° and 20° in the bimodal experiments under virtual acoustic listening conditions could be explained by these assumptions. A comparative analysis of vertical distance effects in the free field might yield some more insight.

Figure 36 shows the IFE data collected in the loudspeaker setup. Overall, the results are well comparable with those in the virtual environment. Again, an effect of vertical spatial disparity is only present for $SOA < 0$. Curves for stimulus positions with varying horizontal components only are nearly parallel while those those for stimuli with identical horizontal, but different vertical component merge at positive SOA (Figure 36a). The curves are somewhat flatter than those from the virtual environment and IFE values are generally smaller. This might be interpreted as a ceiling effect, since unimodal visually guided saccades are mostly significantly smaller in the free field experiment (see Experiment 3).
If the IFE functions are split up with regard to target positions (Figure 36b and c), it turns however out that there is a difference between the data collected in virtual acoustics and in free field. For positive SOA, the results are like in Figure 36. The IFE seems to be completely determined by target position, no influence of vertical distance becomes visible. At negative SOAs, it can again be found that $IFE_{VA} > IFE_V^A$ and $IFE^{VA} > IFE_A^V$ (in analogy to Figure 36a), but this effect is superposed by the target position effect which had disappeared for negative $SOAs < 0$ in the virtual environment. Spatial spatial effects are much stronger for elevated than for non-elevated targets and there is a general shift toward higher IFE values in case of elevated targets across all SOAs: the curves for IFE_{VA} and IFE^{VA} are nearly parallel here. Similar observations can be made with vertically disparate stimulus presentation, although the effects are too various and differences are too weak to allow a reliable analysis.

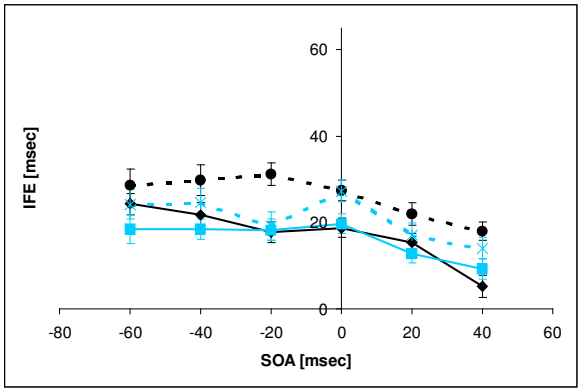
Hypothesis 3 can again be ruled out on the basis of this data, as IFE values still differ significantly at positive SOA when there is no obvious influence of vertical distance any more. However, the influence of target position, which had strongly supported the "default assumption" or top-down hypothesis in the virtual acoustic setup, becomes visible throughout the whole free field experiment at any SOA, what makes this argument less powerful. The data provide a strong hint for a facilitating effect of the accessory that is independent of spatial interstimulus distance, but dependent on target position, and that is therefore no general unspecific warning effect as described above.



(a) pooled across hemispheres



(b) ipsilateral stimuli only



(c) contralateral stimuli only

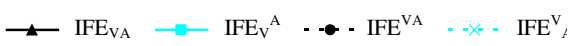
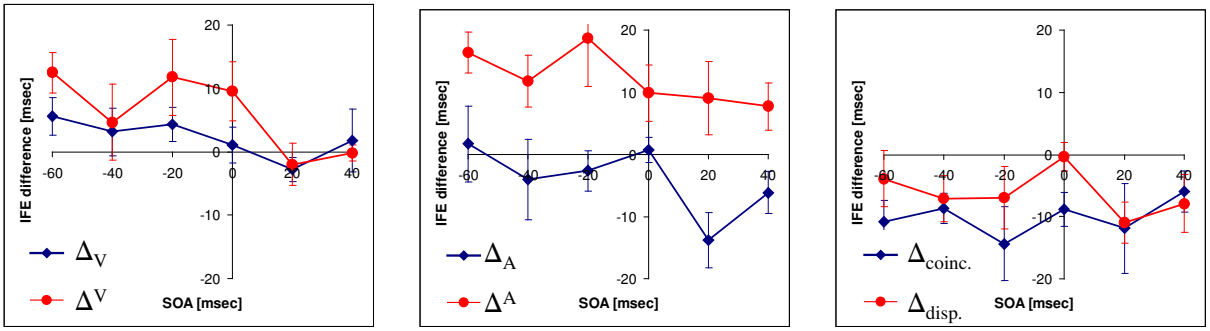


Figure 36: IFE (and standard errors) in bimodal saccadic latencies across SOA, from Experiment 3, virtual acoustic environment. Data was pooled across al participants. a: across all target positions, equivalent to the presentation of Figure 28 b: for ipsilateral stimulus presentation only, c: for contralateral stimulus presentation only, data not used in further analyses.



(a) Effects of different target elevations

(b) Effects of different auditory elevations

(c) Effects of absolute stimulus configuration

Figure 37: IFE differences in bimodal saccadic latencies across SOA, ipsilateral stimulus presentation in virtual acoustic environment. Data was pooled across all participants.

Regarding IFE difference functions (Figure 37), it can first of all be seen that standard errors are much larger in free field than in the virtual acoustic environment. Regardless of this, the difference curves are generally not that smooth as in Figure 35. Both findings might partly be due to the smaller number of participants (four instead of five), but are rather to be interpreted as a sign of less stereotyped responses of the participants in the loudspeaker setup.

Spatial effects for different target positions

Figure 37a shows that, like in the virtual acoustic environment, there is no distance effect for neither target position if the auditory stimulus is presented after the visual. For a preceding accessory, there is a small vertical distance effect for targets from the horizontal plane, but a much larger one for elevated target positions. Contrary to the virtual environment, the curves differ remarkably here. However, the general trend of decreasing IFE differences with increasing SOA is still slightly visible for both target positions in the present figure.
Regarding the hypotheses about bottom-up or top-down processes in auditory localization, the data for positive SOA is very similar to that collected in the virtual environment and can thus be interpreted in the same manner. An explanation of the results for $SOA \leq 0$ is however somewhat more difficult. Disparity effects across SOA are not independent of target elevation any more, but there is an interaction or a superposition between both influence factors. The fact that $IFE^{VA} > IFE_{VA}$ values, which seems to be the main reason for the different Δ-functions (compare with Figure 36b) might be due to larger IFEs either in case of elevated targets or with elevated auditory accessories, or both. From these difference functions it is however not possible to distinguish between the two possible factors.

Spatial effects for different auditory elevations

Although it does not seem so at a fist glance at Figure 37b, a closer look reveals that IFE-difference functions with varying *auditory* elevation yields generally comparable curves in both experimental environments. However, the Δ_A-curve representing combinations with auditory

stimuli coming from the horizontal plane is shifted downward, especially for negative SOAs. By comparing this plot with Figure 36b, it turns out that this finding can be explained by smaller values for IFE_{VA}. Again, target position is revealed as a factor of significant impact in the free field experiments. Different facilitation effects for responses to elevated vs. non-elevated targets can however not explain the pronounced decrease of Δ_A for positive SOAs found in both experimental environments.

Effects of absolute stimulus configuration

Figure 37c probably reveals the most remarkable difference between the two experimental setups. Unlike with the virtual auditory environment, it here turns out that the higher IFE benefit for responses to targets at 20° elevation is not restricted to positive SOAs, but can be observed with relative constancy across all SOAs. Moreover, this effect is nearly equally strong for coincident and disparate stimulus combinations, i.e. although target position dependent, it seems to be rather independent of vertical interstimulus distance.

4.3 Discussion: bottom-up or top-down ?

The results show that an auditory stimulus facilitates the response towards the target stimulus even if it is presented as long as 40 ms after the visual target. However, the intersensory facilitation effect is determined by different spatial stimulus features depending on whether the accessory auditory is presented before or after target presentation. If the accessory precedes the target or is presented simultaneously the IFE is influenced by the spatial distance between visual and auditory stimulus. This result is in line with the results of Frens et al. (1995) and Colonius & Arndt (2001). In contrast to this, the physical vertical distance does not contribute to IFE if the accessory is presented later than the target (Figures 34a and 36). This pattern of IFE changes in dependence on spatial distance above is superposed by a warning effect which evokes a continuous decrease of IFE with increasing SOA. Moreover, the IFE is strongly influenced by target position, especially under free field listening conditions. In our experiments, those targets with 20° elevation showed higher IFEs than those with 0° elevation, in which these effects could be observed across all SOAs in the the loudspeaker setup, but solely for positive SOAs in the virtual environment.

In order to further investigate the influence of certain spatial factors in visual-auditory interaction (target position, perceived auditory position, absolute stimulus configuration), relative IFE changes were calculated in which general warning effects disappear. Three hypotheses about the perceived auditory vertical position in case of insufficient analysis of sensory information were tested by this method. Based on our results, the assumption that no or a blurred representation of auditory elevation is established can be discarded. The higher IFE for the 20° target elevation compared to 0° elevation rules out this hypothesis.

Our IFE data are in principle compatible with both the assumption of a bottom-up process providing a first rough estimate of auditory elevation and the hypothesis of a top-down process yielding a default value which is independent of the physical position of the stimulus. The assumption of a bottom-up process turned however out to be rather implausible. Our analyses provide evidence that, if the auditory stimulus is presented after the visual target, it is perceived with a specific elevation unequal to zero. A spectral analysis of the auditory stimuli used within the first 10, 20, and 50 msec lead to the conclusion that this finding could not be led back to a stimulus induced artefact. Although more noisy due to the shorter time windows, the spectral contents of these early portions of the stimuli did not differ from the spectral content of the complete stimuli. We therefore assume that a first estimate does not play a role in the intersensory facilitation effects we investigated in this experiment. Rather, a top-down process providing a default elevation in case of incomplete stimulus processing is fully compatible with our data. A default should always lead to the same perceived elevation regardless of physical stimulus position and that is what we found in our experiments.

However, the higher IFE for targets at 20° elevation compared to those at 0° require an explanation. As stated above, two interpretations are possible. Either the default elevation is located nearer to the upper target, i.e. between 10° and 20°, or targets at 20° elevation benefit more than those at 0° from the accessory auditory stimulus, or both. A stronger IFE for targets with 20° elevation is conceivable, given the fact that an oblique saccade is necessary to direct gaze to those target locations. While horizontal eye movements (target elevation 0°)

are controlled by only one pair of muscles, all three pairs of eye muscles are active in oblique saccades. As discussed e.g. by Diederich & Colonius (1987), intersensory facilitation may influence motor processes. This might evoke stronger effects when more muscles are active. But why does this play a role only in positive SOAs if a virtual display is used while it seems to be predominating in the loudspeaker setup?

A possible explanation for the findings in virtual acoustics could be that the intersensory effect on motor functions occurs early after the accessory signal is detected and decays over time. Then, highest facilitation would have to be expected if the accessory is detected just before the response is elicited. This interpretation requires further investigation, particularly with regard to the different effects found in the loudspeaker setup.

Here, a reason might be found in the usage of the new targets. The presentation of *red* target stimuli instead of white ones could be problematical, since it is has been reported that the different photo-receptors are not distributed equally across the retina. Generally, cone density is highest in the foveal pit and falls rapidly outside the fovea to a fairly even density in the peripheral retina (Curcio et al., 1987). However, S-cones (blue light sensitive) have a different distribution with lowest density in the fovea and a maximum density on the foveal slope (Ahnelt et al. 1987). So far, analogous data has not been presented for human retina M-cones (green light sensitive) or L-cones (red light sensitive), since actually no method is known for distinguishing them morphologically. In the monkey retina, Marc and Sperling (1977) could however show that L-cones occur at about 33% of the cones throughout the whole retina. These findings suggest that using red stimuli instead of white ones should not be of major impact in the present experiments.

A rather significant factor seems the larger diameter of the LED's, which probably led to the massively reduced unimodal SRT's that in turn yielded weaker IFE's. It is known from physiological data (Stein & Meredith 1993) as well as from psychophysical experiments (Frens & Van Opstal 1995) that response enhancement by intersensory integration is the more pronounced, the weaker the target stimulus is. It might therefore be that with the weaker target signals in the virtual acoustics setup, spatial effects were more pronounced than general motor enhancement and therefore dominated the findings. In the free field setup however, the higher target intensity might have led to a decreased spatial influence (in fact, IFE was generally smaller in the loudspeaker setup and therewith the spatial effects) so that then motor components in the IFE turned out to be relatively more effective. Concluding, the differences in the result might simply be interpreted as a strong hint for the mutual influence of various separate mechanisms in visual-auditory interaction.
Unfortunately, the strong influence of target-elevation covers all other effects in the free field experiment, so that no more analyses with regard to vertical distance effects are possible with this data set. Some more investigations in the loudspeaker setup, using weaker targets, might be useful.

In a situation in which spatial factors are the more effective, the assumption of an auditory default elevation between 10° and 20°can however not be ruled out, even if it seems somewhat implausible. Other investigators have found default values near to zero elevation ((Hofman

& Van Opstal 1998, Frens & Van Opstal 1995) with auditory stimuli arranged symmetrically around the horizontal meridian. Given our experimental setup with auditory stimuli at 0° and 20° elevation, it is plausible to assume that a default is set midway between the two possible stimulus positions, i.e. at 10° elevation. Probably, the default is not predetermined, but depends on the current (experimental) situation and the stimulus positions which are to be expected in this environment. Thus, expectations of the subject and knowledge about the environment play a crucial role in the estimation of the elevation of auditory stimuli. However, which factor might shift the default to a more elevated position regrettably remains unclear and also deserves further study.

5 Modelling visual-auditory interaction in two-dimensional space

Traditionally, investigations about multisensory interaction and response facilitation were carried out measuring simple, i.e. undirected (mostly manual) responses. Studies investigating response speed up in target directed (eye-) movements were, on the other hand, for a long time confined to intra-modal effects. Konrad, Rea, Olin & Colliver (1989) were among the first to report about saccadic response facilitation with an auditory distractor that did *"not appear to be due to alerting"*, but *"probably due to an auditory input to the superior colliculus which decreased threshold for initiating a saccadic eye movement.*" More detailed quantitative analyses on spatial and/or temporal factors in visual-auditory interaction and attempts to find explanatory mechanisms followed soon. The common finding was that the strength of intersensory facilitation increases with spatial proximity between target and accessory and if the auditory accessory precedes target presentation within a certain time interval (see earlier chapters of this study for an overview). Current models attempting to describe and explain these phenomena usually designate several stages of unimodal and bimodal processing from the peripheral information transfer up to response execution and therefore provide various places and mechanisms of interaction.

In classical psychophysical modelling, processing times are often represented by random variables. Earlier models of visual-auditory interaction like the initial Independent Race Model of Raab (1962) provide random variables that stand for complete response times (to either unimodal stimulus), while recent approaches are more refined and distinguish between sensory processing, decision processes, motor programming processes, and so on. The different stages of processing can be organized in parallel or serially. There may be serially combined stages with sub-processes organized in a parallel manner and vice versa, with each of the sub-processes being represented by its own random variable. Today, most authors in this field of research agree to the hypothesis that early sensory processing is performed in parallel and that sensory information is then integrated somehow in order to be further processed by more central and by motor-related systems. One possible mechanism of intersensory interaction might be found at that stage of early sensory information combination. Motor programming, on the other hand, is assumed to take place at other stages that are separate from and organized in series with sensory processing. The accessory stimulus might however also have an (unspecific) influence on these processes that can be inhibitory as well as excitatory. Spatially unspecific facilitating mechanisms like general arousal (as proposed by Bertelson & Tisseyre (1969)) or preparation enhancement (Nickerson 1973) can be assumed to work within these circuitries. The model of Nozawa et al. (1994) which was described in Section 2.1, see Figure 9, page 12) is a typical representative of these approaches.
Stochastic models are well suited to simulate response behavior of organisms with various levels of resolution – from the activity of single cells via defined subsystems (like the Superior Colliculus) up to patterns of complete responses. In a living system, whether a response is evoked and how strong this response is does not only depend on the stimulating input, but also on internal random variation. Hence, the most natural way of dealing with this 'internal noise' is in fact an approach based on probabilistic assumptions. Unfortunately, most stochastic

approaches in response time modelling are confined to the temporal domain, i.e. they try to explain effects of the *temporal* stimulus constellation only. So far, effects of stimulus *position* have only rarely been simulated quantitatively. Recently, the Multi-Channel Diffusion Model, a dynamic probabilistic approach developed and introduced by Diederich (1995), was expanded to model responses in a choice reaction experiment using multiple visual stimuli and auditory accessories. Target choice as well as the latencies of the saccadic responses could be simulated well in this study (Troidl 2002). The Two-Stage Model of visual and auditory interaction in saccades by Colonius & Arndt (2001) demonstrates another way of integrating spatial and temporal factors in bimodal response times easily in a stochastic approach. It will be described further in the following section.

Another family of models is that of the dynamic system approaches, which are mainly applied to properties of the saccadic system and are often closely related to the more physiologically motivated hypotheses on saccade generation. They designate one or several set(s) of differential equations representing the states of sub-processes within the saccadic system that permanently interact with each other. Like in the Findlay-Walker framework (see Section 2.2), they make use of saliency maps that spatially encode neural activity. A dynamic function represents the system's internal state at a specific coordinate of the saliency map, which has to reach a defined threshold value if a saccade is to be initiated. The system's dynamics is influenced by internal dynamics (time constant and other pre-set parameters), interactions with the environment (lateral inhibition), and by exogenous (sensory stimulation) and endogenous (p.e. intention) input. Simulations are performed by multi-layered neural networks.
The Neural Field Models (Kopecz 1995, Bastian et al. 1998) are typical representatives of this class of models in this context. Based on early work by Amari (1977) who initially tried to simulate unimodal saccadic latency distributions, the Neural Field approach has been expanded to simulating especially SC oculomotor behavior for different types of saccade-related neurons like fixation, buildup, and burst neurons. Neural field dynamics include the spatial relevance of multiple stimuli as well as their temporal relationship. In a recent paper, Trappenberg et al. (2001) demonstrate the wide variety of their Neural Field Model's properties and abilities in various paradigms including gap effect, the use of (visual) distractors, or contingencies. Although these models have so far mainly been used for simulating unimodal visual effects only, but there is no obvious reason why its principles might not be transferred to visual-auditory interaction, too. The idea of applying dynamic fields generally seems a promising approach in modelling bimodal saccadic latency. However, although differential equations are a powerful tool for describing systems' behavior in nature, they lack one crucial feature: random. Dynamic systems irresistibly follow the rules of the equations through which they are defined. These are generally motivated by the natural laws of thermodynamics (especially in learning systems), nevertheless this approach is a thoroughly deterministic one. Thence, it does neither simulate nor is it able to explain things like spontaneous activity of neurons or their general variance regarding the responses to a defined input. Indeed, Ratcliff, Van Zandt & McCoon (1999) have suggested the introduction of an additional (normally distributed) noise variable simulating internal fluctuations. So far, applications have however been rare and limited to a small number of simulated phenomena.

Therefore, the integration of temporal and two-dimensional spatial stimulus properties as investigated in Experiment 3 shall here be attempted using a probabilistic approach. We will apply the Two-Stage Model of (Colonius & Arndt 2001), which has shown up to well account for spatial and temporal factors and their mutual influence in visual-auditory interaction. So far, the Two-Stage model has been tested with stimuli from positions within the horizontal plane only. In the following, we will present an extension into two-dimensional space by two ways in order to comprises the different processes in auditory localization outlined above. In a first step, we will introduce an additional variable representing the sensory processing time for vertical auditory position which is suspected to differ from auditory horizontal information processing, or the time needed for simple auditory detection. Further, we will test the assumption of different distance effects within the horizontal and the vertical plane by comparing two approaches, the first assuming one common distance parameter and the second assuming individual spatial parameters for each dimension.

5.1 The Two Stage Model of Colonius & Arndt

The Two-Stage Model is a stochastic approach to describe the temporal properties of peripheral and central processes from modality-specific stimulus reception to response performance in human saccadic responses. The basic assumptions are very simple. Initial processing is triggered individually in each sensory channel by the arrival of a specific stimulus at the sensory organ. The complete peripheral processing of sensory information is assumed to be performed independently in separate channels. A second processing stage, which is arranged in series with the peripheral first stage, comprises motor computation and execution [1]. Particularly, integration of sensory information from the various modalities takes place at this common sensory-motor stage. All processes and sub-processes are represented by random variable, which will be specified more precisely below. Figure 38 shows a sketch of the Two-Stage Model's basic architecture, further details and more specified assumptions in order to quantitatively test the model will be successively described in the following.

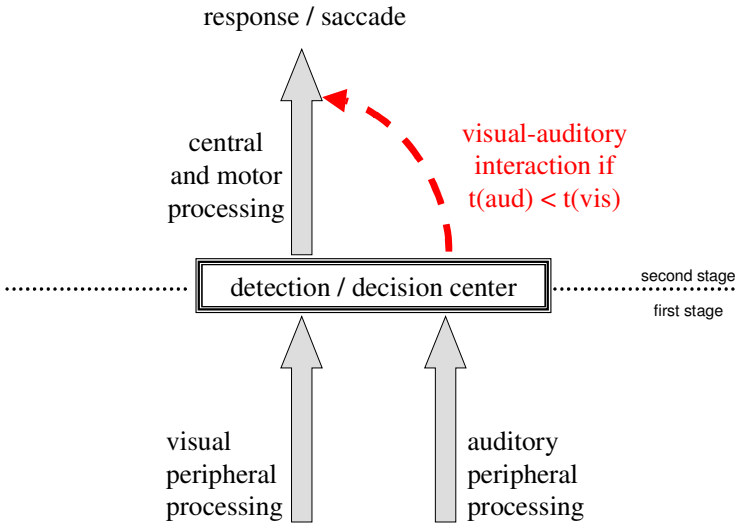


Figure 38: Illustration of the Two-Stage Model as suggested by Colonius & Arndt (2001). Peripheral sensory processing (bottom) is assumed to be performed in parallel channels, which denote the first stage of the model. It is finished with the arrival of either stimulus (MIN-version) or the the imperative stimulus (V-version) at a decision level. The second stage processes are then initiated. They involve motor computation and execution. Note that no sensory information integration can occur at the first stage. Bimodal interaction can affect the second-stage processing, if the accessory stimulus' information arrives the detection/decision stage prior to the target information. Thus, temporal and spatial affects are clearly separated by the two stages. See text for further description.

[1] This implies the assumption of constant motor execution times which, although not justified by properties of the oculomotor system, shall here be taken for true for reasons of simplicity.

As the first stage is organized in parallel, there is no mutual influence between visual and auditory stimulus processing at that level. The sensory channels however finally terminate into a common central stage where the various pieces of information may interact. Whether they do and with which strength depends on the outcome of the first stage. An initial independent race as described by Raab (1962) is assumed to take place in the peripheral channels. However, in the present approach, the race between the modalities is only one sub-process involved in the pure *detection* of the signals. In case of visually guided responses, further processes will be affected by interaction between visual and auditory stimulus information *if* the auditory stimulus has reached the common stage faster than the imperative visual stimulus. If the visual stimulus reaches the detection stage earlier, it will be further processed as if there were no additional stimuli. Hence, only the temporal alignment between the stimuli is of importance at the first stage, but not their spatial relationship.
The second stage of more central processing, in which the sensory information from the peripheral processes might be integrated, starts as soon as the first stage is terminated. The original model of Colonius & Arndt distinguishes between the V-version and the MIN-version, which stand for two different termination rules. In the V-version, the first stage is finished by the arrival of the imperative visual stimulus, while in the MIN-version it is the first of both stimuli that initiates the second stage. The first idea follows the assumption that the target stimulus must be detected prior to further oculomotor activity in order to elicit an appropriate response. The second approach allows *temporal* triggering of the next higher stage already before the target coordinates are known. This would result into (spatially unspecific) enhanced response preparation, maybe comparable to mechanisms reckoned to be responsible for the gap-effect. In this study, only the V-version will be used for further development.

As mentioned above, processes do not interact within the first stage, so that only temporal features have to be considered here. These do however play a crucial role. If there is a temporal gap in stimulus presentation, the probability that the stimulus being presented first also wins the first-stage race rises significantly. The duration of second-stage processing is influenced by intersensory interaction through neural summation *only if* the accessory stimulus is the winner of the afferent processing race. The amount of the interaction effect then depends on the spatial distance between target and accessory. It is assumed to be maximal if both stimuli are coincident (distance zero) and to decrease with decreasing proximity. In case of large distances, there might even result a negative value, i.e. bimodal response inhibition.
It is a crucial feature of the Two-Stage Model that the amount of intersensory effects induced by second-stage information integration solely depends on the spatial configuration of the stimuli. Therefore, temporal and spatial effects are clearly separated by the two processing stages. It should however be kept in mind that temporal factors have an *indirect* effect on the amount of second-stage interaction, as the order of stimulus arrival at the detection stage determines whether there will be any interaction at all. Since we deal with a probabilistic approach, it follows that the *mean* bimodal reaction times (and thus the mean IFE) depend on spatial as well as on temporal factors. This point will be illustrated in more detail in the following section.

So far, the Two-Stage Model has been applied to one-dimensional spatial problems only. If we extend our region of interest into two-dimensional space, we have to consider the different coding of horizontal and vertical spatial information within the auditory system, as discussed

above. We therefore introduce another sensory processing channel (or, first-stage race competitor) being responsible for *vertical* auditory position calculation. Since visual spatial information is encoded retinotopically, no further specifications have to be made with regard to visual processing. All in all, we now have three peripheral processes: one calculation visual information, one for auditory azimuth and the third for auditory elevation information. Central and motor processing is assumed to take place as before, but now being dependent on the outcome of a three-competitor race.

In the style of Colonius & Arndt, the basic assumptions of the extended Two-Stage Model used here can then be summarized in four statements:

1. Parallel first stage processing: initial sensory processes are modality-specific and take place in separate parallel channels. The processes in the different channels are statistically independent. This assumption is rather restrictive, especially with regard to the separate processing of auditory azimuth and elevation processing. However, separate channels for different modalities are quite plausible, given the completely different ways of signal analysis and decoding in the visual and the auditory system. On the other hand, there is physiological evidence that neural processing of spectral features and binaural cue analysis is performed in different pathways from the Nucleus Cochlearis on, which is the first in afferent auditory processing after the Cochlea itself. Although it cannot be taken for granted that there is not information exchange at all until auditory information is merged with visual, the assumption of separate pathways shall serve here as a first simplifying approach.

2. Termination of the first stage: peripheral processing is finished as soon as the visual target information reaches the common detection center. This is equivalent to the V-version of the Colonius-Arndt approach. The common detection center, from which further central processes like motor computation and execution start, might be interpreted as analogue to the Deep Layers of the Superior Colliculus (DLSC). The crucial role of the SC, especially the Deep Layers, in (a) programming saccadic eye movements and (b) intersensory integration suggests this approach, which is commonly shared in recent models (Van Opstal & Van Gisbergen 1989, Frens & Van Opstal 1995, Trappenberg et al. 2001). Although the importance of other systems like the Frontal Eye Fields (FEF) in saccade initiation shall not be underemphasized, we will mainly come back to SC-related models and assumptions while interpreting the following data fits.

3. Separation of spatial and temporal factors: the spatial configuration of the stimuli has no influence on the first-stage processing, but may affect bimodal processing time in the second motor programming stage. Conversely, the temporal order of stimulus presentation does not directly affect second-stage processing. Visual-auditory interaction however underlies the necessary condition that at least part of the accessory auditory (i.e. horizontal or vertical) information has reached the detection stage prior to the visual. The sufficient condition is that there is an appropriate spatial and temporal relationship for information integration. More precisely: if an auditory stimulus is presented too early, it will probably not be relevant for subsequent visual processing. If it is presented from spatial positions too far away from the visual target, it might be simply ineffective rather than inhibitory

or even facilitating[2].

4. Second-stage intersensory interaction: if one of the auditory competitors is the winner of the first-stage race, second-stage processing may be affected by bimodal information integration. Intersensory effects may yield a speed-up in processing time (facilitation), but they might also prolong it (inhibition), depending on the spatial configuration.

[2]Note that these latter depictions shall be regarded as purely hypothetical examples in order to make the situation somewhat clearer. They are neither explicit experimental observations nor expectations.

5.2 Formal description of the extended Two-Stage Model

Distribution-free assumptions

In the following, a more detailed description of the extended Two-Stage Model shall be presented. We will start with some more general statements following out of the above assumptions. In order to test the model quantitatively, we will thereafter specify these statements by assuming concrete random variable distributions which we will fit to the data of Experiment 3. Let the various peripheral and central processing times be represented by non-negative random variables with finite means and variances and the following denotation:

W_1	first stage, or sensory, processing time
W_2	second stage, or central, processing time
V	visual sensory processing time
A	auditory azimuth sensory processing time
E	auditory elevation sensory processing time
τ	stimulus onset asynchrony (SOA) in msec
RT_V	observed saccadic response time when only the visual stimulus is presented
$RT_{VA,\tau}$	observed saccadic response time to visual stimulus with auditory accessory presented at SOA τ

The observed saccadic response time is then then assumed to be the sum of the first-stage sensory processing time and the central, sensory-motor processing time

$$RT_{VA,\tau} = W_1 + W_2. \tag{4}$$

Since we follow the V-version of the Two-Stage model, the first-stage processing time W_1 is simply the visual peripheral processing time V. The calculation of the second-stage processing time is somewhat more complex, since unlike with the peripheral processing times, W_1 and W_2 are only *conditionally* independent. As outlined above, W_2 depends on the outcome of the first-stage race, that is we have to regard the different orders of arrival of V, A, and E.

In general, two different situations might be distinguished: the case of no interaction, when the visual stimulus is processed faster than either of the auditory components, and the case of given interaction, if one of the auditory competitors wins the race. Let I describe the general event of intersensory interaction, i.e.

$$\begin{aligned} I &= \{(A+\tau) \cup (E+\tau) \geq V\} \\ I^c &= \{V > (A+\tau) \cup (E+\tau)\}. \end{aligned} \tag{5}$$

The (cumulative) bimodal reaction time distribution is a mixture of the two distributions for either of the complementary events

$$\begin{aligned} P(RT_{VA,\tau} \leq t) &= P(I) \cdot P(RT_{VA,\tau} \leq t|I) \ + \ P(I^c) \cdot P(RT_{VA,\tau} \leq t|I^c) \\ &= P(I) \cdot P(RT_{VA,\tau} \leq t|I) \ + \ [1-P(I)] \cdot P(RT_{VA,\tau} \leq t|I^c) \\ &= P(RT_{VA,\tau} \leq t|I^c) \\ &\quad - \underbrace{P(I) \cdot \left[P(RT_{VA,\tau} \leq t|I^c) - P(RT_{VA,\tau} \leq t|I) \right]}_{\text{IFE}}. \end{aligned} \tag{6}$$

In order to make more detailed predictions, it is necessary to further specify the possible amount of spatial interaction which, in turn, depends on the exact order of arrival of the three first-stage competitors. As it was pointed out above, intersensory interaction is a function of the perceived interstimulus distance and hence of the perceived position of the auditory stimulus. Auditory accessory information can feed into the common processing stage until the visual spatial information has reached it. Thereby, the four following incidents are possible:

I_{AE}	incident of visual-auditory interaction in both dimensions, occurring if $(A+\tau < E+\tau < V)$ or $(E+\tau < A+\tau < V)$
I_A	incident of visual-auditory interaction only in the **a**zimuth dimension, occurring if $(A+\tau < V < E+\tau)$
I_E	incident of visual-auditory interaction only in the **e**levation dimension, occurring if $(E+\tau < V < A+\tau)$
I^c	incident of no interaction, i.e. normal central processing of the visual stimulus if $(V < A+\tau < E+\tau)$ or $(V < E+\tau < A+\tau)$

As the peripheral processing times are statistically independent, the joined specific incidents are equal to the general incident of interaction, $I = I_{AE} \cup I_A \cup I_E$, and therefore

$$P(I) = P(I_{AE}) \ + \ P(I_A) \ + \ P(I_E). \tag{7}$$

The distribution-free predictions for bimodal expectation values can be calculated in a similar way as the probabilities in Equation 6

$$\begin{aligned} E(RT_{VA,\tau}) &= E(W_1) + E(W_2) \\ &= E(W_1) \ + \ P(I) \cdot E(W_2|I) \ + \ P(I^c) \cdot E(W_2|I^c) \\ &= E(W_1) \ + \ P(I) \cdot E(W_2|I) \ + \ [1-P(I)] \cdot E(W_2|I^c) \\ &= \underbrace{E(W_1) \ + \ E(W_2|I^c)}_{\text{unimodal SRT}} \ - \ P(I) \cdot \underbrace{[E(W_2|I^c) - E(W_2|I)]}_{\equiv \Delta}, \end{aligned} \tag{8}$$

in which the spatial interaction parameter Δ can now be split up into three components in analogy to I:

$$IFE \ = \ P(I) \cdot \Delta \ = \ P(I_{AE}) \cdot \Delta_{AE} + P(I_A) \cdot \Delta_A + P(I_E) \cdot \Delta_E \tag{9}$$

Specification of sub-process distributions

In order to test the model and its predictions by fitting it to the data of Experiment 3, we will now have to make further assumptions about the processing time distributions within the various sub-processes. Here we again follow Colonius & Arndt by assuming exponentially distributed durations of the peripheral processes, that is

$$\begin{aligned} V &\sim \lambda_V \exp(-\lambda_V \cdot t), \quad \mathrm{E(V)} = 1/\lambda_V \\ A &\sim \lambda_A \exp(-\lambda_A \cdot t), \quad \mathrm{E(A)} = 1/\lambda_A \\ E &\sim \lambda_E \exp(-\lambda_E \cdot t), \quad \mathrm{E(E)} = 1/\lambda_E . \end{aligned} \tag{10}$$

In the V-Version, the first stage processing time is simply the expectation value of the visual processing time:

$$E(W_1) = E(V) = 1/\lambda_V . \tag{11}$$

After termination of the first stage, central processing at the second stage is initiated. The second stage is also described by a random variable. In the case of a simple unimodal visual target, its duration is represented by a normally distributed random variable

$$W_2(\text{unimodal}) \sim N\ (\mu). \tag{12}$$

In the case of bimodal stimulation however, the integration of spatial sensory input might result into different central processing times. The magnitude of second stage facilitation is determined by the spatial relationship of the stimuli. In our model, we assume a linear relationship between the *perceived* stimulus positions and the amount of spatial intersensory facilitation. Temporal stimulus parameters do not directly affect the second stage processing. However, the amount of possible spatial influence is dependent on the temporal interstimulus relationship, insofar as the probability of any kind of spatial interaction depends on the arrival times of the visual and auditory information.

The probability of each of the incidents simply is the probability that the respective auditory information (horizontal and/or vertical position) is processed faster than the visual stimulus. It depends on the stimulus onset asynchrony (SOA), denoted by the parameter τ, in which the cases of positive and negative SOA have to be regarded separately. The probability of any order of arrival, or incident of spatial interaction, can be calculated using the respective Fubini-integral (see Appendix). Hence, the probability of visual-auditory interaction in both dimensions yields out of

$$\begin{aligned} \pi_{AE}(\tau) &:= P(I_{AE}) = P(A+\tau < E+\tau < V) + P(E+\tau < A+\tau < V) \\ &= \begin{cases} 1 - \frac{\lambda_V\, e^{\lambda_E \tau}}{\lambda_A+\lambda_V} - \frac{\lambda_V\, e^{\lambda_A \tau}}{\lambda_E+\lambda_V} + \frac{\lambda_V\, e^{(\lambda_A+\lambda_E)\tau}}{\lambda_A+\lambda_E+\lambda_V} & \text{for} \quad \tau < 0 \\ \frac{\lambda_A\, \lambda_E\, e^{-\lambda_V \tau}}{\lambda_A+\lambda_E+\lambda_V} \left(\frac{1}{\lambda_A+\lambda_V} + \frac{1}{\lambda_E+\lambda_V}\right) & \text{for} \quad \tau \geq 0 \end{cases} \end{aligned} \tag{13}$$

The probability of visual-auditory interaction only within the azimuth dimension is

$$\begin{aligned} \pi_A(\tau) &:= P(I_A) = P(A+\tau < V < E+\tau) \\ &= \begin{cases} \frac{\lambda_V\, e^{\lambda_E \tau}}{\lambda_E+\lambda_V} - \frac{\lambda_V\, e^{(\lambda_A+\lambda_E)\tau}}{\lambda_A+\lambda_E+\lambda_V} & \text{for} \quad \tau < 0 \\ \frac{\lambda_A\, \lambda_V\, e^{-\lambda_V \tau}}{(\lambda_E+\lambda_V)(\lambda_A+\lambda_E+\lambda_V)} & \text{for} \quad \tau \geq 0 \end{cases} \end{aligned} \tag{14}$$

Analogously, the probability for vertical spatial interaction results into

$$\begin{aligned} \pi_E(\tau) &:= P(I_E) = P(E+\tau < V < A+\tau) \\ &= \begin{cases} \frac{\lambda_V\ e^{\lambda_A \tau}}{\lambda_A+\lambda_V} - \frac{\lambda_V\ e^{(\lambda_A+\lambda_E)\tau}}{\lambda_A+\lambda_E+\lambda_V} & \text{for} \quad \tau < 0 \\ \frac{\lambda_E\ \lambda_V\ e^{-\lambda_V \tau}}{(\lambda_A+\lambda_V)(\lambda_A+\lambda_E+\lambda_V)} & \text{for} \quad \tau \geq 0 \end{cases} \end{aligned} \tag{15}$$

Central processing duration is then calculated as follows.
In the case that neither of the acoustic "competitors" reaches the second stage before the visual, no spatial interaction takes place, that is

$$(W_2|I^c) = W_2(\text{unimodal}). \tag{16}$$

Central processing time of the visual stimulus is then equal to the second stage duration in the unimodal case, which is assumed to be normally distributed with expectation value μ. If intersensory interaction does however occur, central processing time is changed by a quantity represented by Δ (see Equation 8),

$$\Delta = E(W_2|I^c) - E(W_2|I) \tag{17}$$

in which the incident I denominates the set of all possible incidents of interaction. The spatial interaction parameter is assumed to be linearly dependent on the *perceived* distance between the stimuli, that is

$$\begin{aligned} \Delta_{AE} &= (\Delta|I_{AE}) &:= b - a \cdot \sqrt{\gamma_A^2 + \gamma_E^2} \\ \Delta_A &= (\Delta|I_A) &:= b - a \cdot \gamma_A \\ \Delta_E &= (\Delta|I_E) &:= b - a \cdot \gamma_E \end{aligned} \tag{18}$$

in which γ_A and γ_E are the spatial distances in azimuth and elevation. In analogy to the above assumptions, each of the three Δ-functions occurs at its SOA-given probability π_A, π_E, or π_{AE} respectively and combines to the overall spatial interaction effect:

$$\pi(\tau) \cdot \Delta = \pi_{AE}(\tau) \cdot \Delta_{AE} \ + \ \pi_A(\tau) \cdot \Delta_A \ + \ \pi_E(\tau) \cdot \Delta_E \tag{19}$$

with $\pi = \pi_A + \pi_E + \pi_{AE}$. The expected saccadic reaction time can be deduced easily by inserting the respective formulas and expectation values. Recapitulating Equation 8 again leads to

$$\begin{aligned} E(RT_{VA,\tau}) &= E(V) + E(W_2|I^C) - P(I) \cdot \Delta \\ &= E(RT_V) - \underbrace{(\pi_{AE} \cdot \Delta_{AE} + \pi_A \cdot \Delta_A + \pi_E \cdot \Delta_E)}_{\text{IFE}}. \end{aligned} \tag{20}$$

This formula can easily be applied to experimental data in order to test the model.

5.3 Data fits to the extended Two-Stage Model

Altogether, six parameters were fitted to 26 data points for each participant:

λ_V	intensity parameter for visual peripheral processing time with $E(V) = 1/\lambda_V$
λ_A	intensity parameter for auditory azimuth peripheral processing time with $E(A) = 1/\lambda_A$
λ_E	intensity parameter for auditory elevation peripheral processing time with $E(E) = 1/\lambda_E$
μ	central visual processing time
a	slope-parameter in the Δ-functions
b	intercept-parameter in the Δ-functions.

The intercept parameter b can be regarded as the maximum IFE which occurs if the stimuli are presented spatially coincident. The slope parameter a might then be interpreted as aspect ration encoding an individual's internal elongation or compression of space.

The MatLab-implemented Levenberg-Marquardt-algorithm, a least-mean square method for medium-scale optimization, was used to fit the model-function to our data. All parameters were estimated for each participant separately using mean saccadic latencies of the various spatial and temporal conditions. Unimodal *auditory* latencies from Experiments 1 and 2 were used as further side conditions. In a simplifying manner, we therefore assumed that central processing time (estimated by μ) does not differ significantly under processing of visually and auditory guided responses. An undirected response toward an acoustic target was estimated by $1/\lambda_A + \mu$, auditory localization by $1/\lambda_E + \mu$. Since these assumptions are very simplifying and as we are mainly interested in modelling bimodal visually guided responses, the latter estimates were weighted down with a factor of $1/5$ in order to reduce their influence on the whole data fit.

Despite the relatively high number of parameters chosen, the data fits turned out to be extremely stable to variation of initial parameter values. Although various combinations of initial values were tried out with every individual data set, the algorithm almost always returned to the same solutions of (subject-specific) parameter estimates.

First data fits to the extended Two-Stage Model: results and discussion

The curve to be fitted is given by the function

$$RT_{VA,\tau} = \frac{1}{\lambda_V} + \mu - \underbrace{\pi_{AE} \cdot (b - a\sqrt{\gamma_A^2 + \gamma_E^2})}_{\text{IFE in both dim.}} - \underbrace{\pi_A \cdot (b - a \cdot \gamma_A)}_{\text{IFE in azim. dim.}} - \underbrace{\pi_E \cdot (b - a \cdot \gamma_E)}_{\text{IFE in elev. dim.}},$$

in which π_{AE}, π_A, and π_E are the interaction probabilities calculated in Equations 13, 14, and 15 and γ_A and γ_E denote physical distances in azimuth and elevation dimension, respectively. Parameter estimates and residual sum-of-squares are presented in Tables 5 (data from virtual acoustic setup) and 6 (free field data). In order to allow an easier interpretation, the intensity

parameters of the peripheral processes are replaced by their inverse values, which directly represent the estimated mean processing times.

First Fit to virtual acoustics data

Partic.	$V[ms]$	$A[ms]$	$E[ms]$	$t_D[ms]$	$\mu[ms]$	a	b	$\sum_{Res^2}$
JB	140	45	340	40	94	0.40	76	792
KW	77	70	87	39	156	0.62	85	336
LP	89	87	148	55	139	0.54	69	203
PN	114	60	65	31	105	0.65	72	454
SB	112	30	101	23	111	0.60	70	467

Table 5: First fit of the extended Two Stage Model to data of Experiment 3 in the virtual acoustic setup.

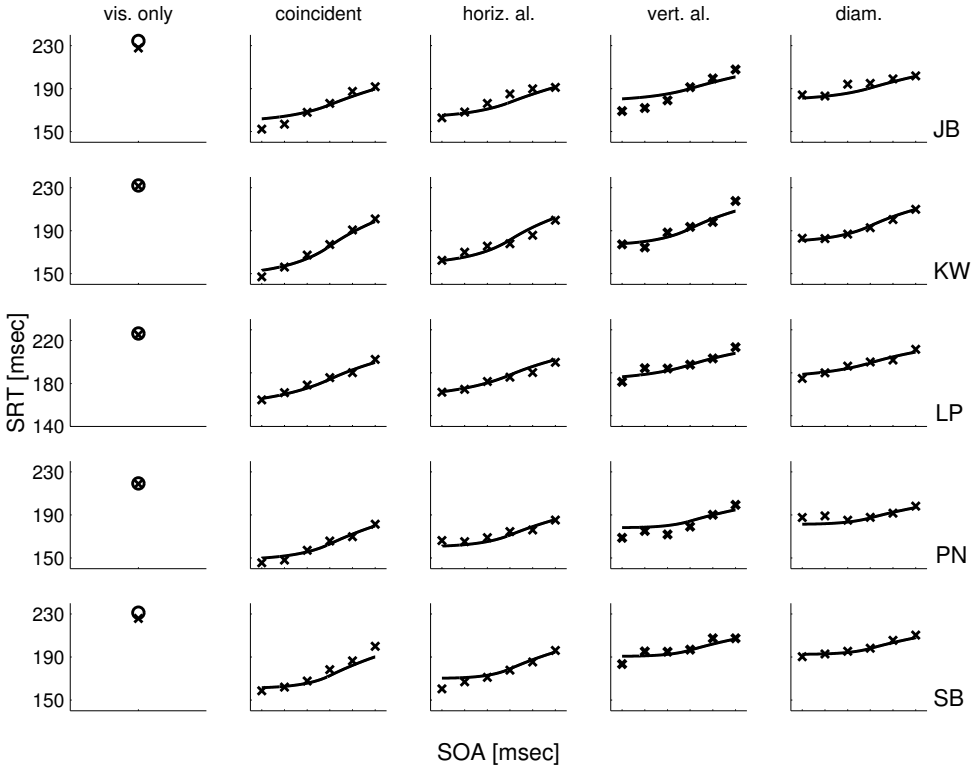


Figure 39: Plots to the first fit of the extended Two Stage Model to the virtual acoustics data (Table 5). Results for different participants are displayed in rows, different spatial interstimulus relations in columns.

First Fit to loudspeaker setup data

Partic.	$V[ms]$	$A[ms]$	$E[ms]$	$t_D[ms]$	$\mu[ms]$	a	b	$\sum_{Res^2}$
DD	85	3	76	3	162	0.48	52	794
HH	80	38	84	26	104	0.59	42	371
KW	79	75	112	45	123	0.57	61	395
PN	129	87	213	62	62	0.53	47	283

Table 6: First fit of the extended Two Stage Model to the free field data.

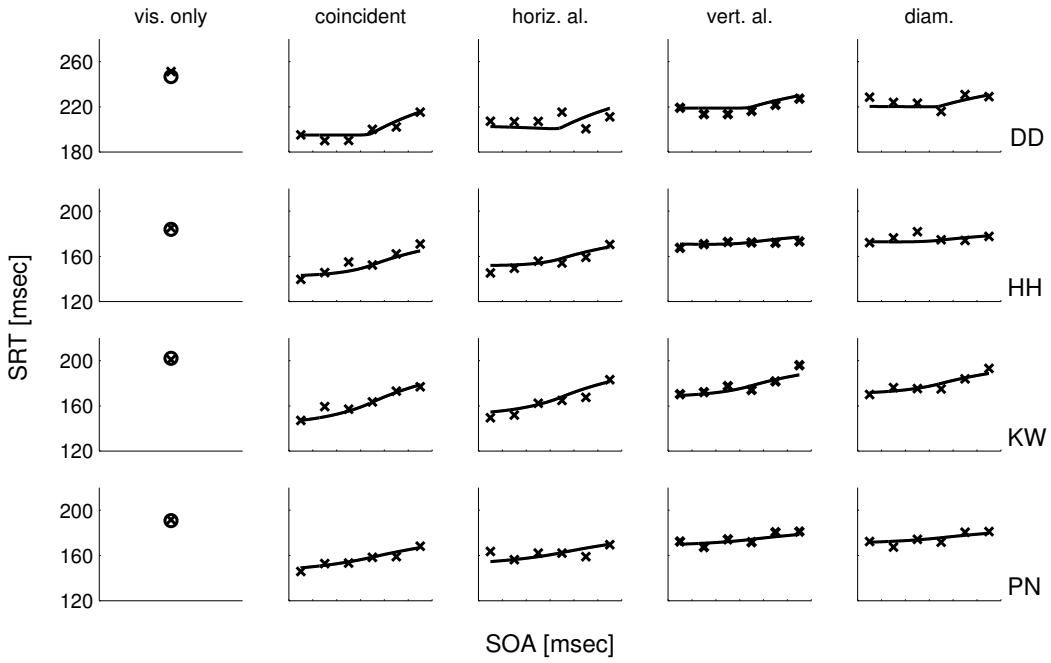


Figure 40: Plots to the first fit of the extended Two Stage Model to the free field data (Table 6). Results for different participants are displayed in rows, different spatial interstimulus relations in columns. Note the different absolute values of the ordinate if compared to the virtual acoustics data.

At this point, it has to be made clear, that the term “auditory peripheral processing time” *does not mean* detection of each cue, but the duration until the respective piece of information has completely been transformed to the oculomotor map in order to be combined with the visual target information. The *detection* of the auditory stimulus happens as soon as the first piece of auditory information has reached the central stage, thus the detection time t_d is determined by $t_d = \min(A,E) = \frac{1}{\lambda_A+\lambda_E}$, if we assume exponentially distributed peripheral processing times. The values of t_d, calculated by the above way, are also displayed in the Tables. The respective plots of the data fits (Figure 39 and 40 respectively) can be found directly below the tables.

It can be seen by the plots that the optimized curves fit the data points quite fair in most cases. Significance and plausibility of the estimates can be verified by comparing the values to behavioral and neurophysiological data.
Estimates for visual peripheral processing ($1/\lambda_V$) lie between 77 and 140 msec and are therefore in part larger than to be expected on the basis of physiologically derived data found in the literature. Recordings of DLSC-layers in the cat yielded visual latencies between 55 msec to 120 msec (Meredith et al. 1987). The same can be said with respect to the estimated auditory detection time t_D. According to Meredith et al., first auditory activation of cat-DLSC cells can be observed after about 10 msec to 30 msec. The estimates calculated here often clearly exceed this expected range, varying between 3 and 62 msec, which additionally means quite remarkable deviations. On the other hand, the latter finding reflects the large inter-subject variability in latencies of directed auditory responses (Experiment 1, see Tables 1 and 2 on page 28 and 29.
Estimated values for auditory elevation processing are always larger than those for azimuth processing. Moreover, auditory azimuth processing is usually estimated to take less time than visual sensory processing, which in turn is often faster than auditory elevation processing. These relations are in accordance with our expectations based on the bimodal reaction time data. Comparing the fits of the two experimental setups it turns out that estimates for visual sensory and central processing times are in the same range in both conditions while estimated auditory processing times differ more. However, there is a generally larger variance in the free field auditory estimates, which makes further interpretations difficult. The spatial IFE parameters *a* and *b* are very close for all participants in both setups, in which the values are generally smaller in the free field data fittings. This simply reflects the finding of smaller facilitation effects in the latter condition. The estimated slope of the Δ-function, represented by *a*, was well above zero in all cases, reflecting decreasing spatial facilitation with increasing distance. The parameter b, denoting maximum spatial facilitation in case of coincident stimulus presentation, is found to take values between 42 and 61 with the loudspeaker setup data and between 69 and 85 for the virtual acoustic situation. These values correspond exactly to the maximum strength of IFE found in the experiments (see Figures 34 and 36). Usually, those participants with smaller IFEs also yielded smaller b-estimates. The slope-parameter a, which can also be interpreted as some kind of internal aspect ration, shows however less systematics. In general, is is somewhat smaller in the loudspeaker setup data, but there is no obvious relation between a-values and spatial characteristic an individual's IFE.

This latter ascertainment is still somewhat dissatisfactory. Although the model yielded good fits and generally plausible estimates, the relation between IFE and spatial distance lacks insofar as those participants with pronounced spatial influence on IFE (like JB in virtual acoustics or DD in the loudspeaker setup) did not always yield larger *a*-values than those with only weak spatial influence (like SB in virtual acoustics and KW in the loudspeaker setup). One reason for this might be found in the choice of only one parameter accounting for spatial distance. So far, we assumed a dependency of the IFE on the *radial* (perceived) distance between visual and auditory stimuli. This approach fits to retinotopically encoded space, which can be assumed in case of visual stimuli, but might not be grated for auditory space. If we take the hypothesis for true that there are different neural pathways for azimuth and elevation processing, it can be supposed that there are different calibrations for the respective eccentricities, or in other words: horizontal and vertical distances might be perceived in a distorted manner if compared

to physical distances and thus lead to distance effects other than modelled so far. Even given the case of equal distance perception in both dimensions, there might be individually pronounced reliability in auditory spatial processing, which would again yield different distance effects for different dimensions and/or participants. Therefore, the assumption of separate spatial parameters for azimuth and elevation distance seem reasonable.

The easiest way include the assumption of different horizontal and vertical distance effects in the present model is to assume different slope parameters for the two dimensions. This approach will be presented in the following section.

Fits with two separate spatial parameters: results and discussion

In the initial model, the function denoting spatial facilitation, Δ, was assumed to be linearly dependent on the physical distance between the stimuli, i.e. $\Delta = b - a \cdot dist$. In the following, we will assume linear dependencies *within each dimension* with possibly different slope parameters. Similar to the parameter a, which is multiplied with the physical horizontal distance we introduce a parameter e which is accordingly multiplied with the vertical distance. Hence, a and e may be interpreted as internal calibration or aspect ratio of space, as already discussed above. For example, a defined vertical distance might be perceived larger than a physically equal horizontal distance, or vertical distance should simply affect spatial facilitation more pronounced than the respective horizontal distance. Hence, the slope estimate for vertical dimension, e, should be larger than the slope estimate in the horizontal dimension. The following function is now to be fitted to the data.

$$RT_{VA,\tau} = \frac{1}{\lambda_V} + \mu - \underbrace{\pi_{AE} \cdot \left(b - \sqrt{(a \cdot \gamma_A)^2 + (e \cdot \gamma_E)^2} \right)}_{\text{IFE in both dim.}} - \underbrace{\pi_A \cdot (b - a \cdot \gamma_A)}_{\text{IFE in azim. dim.}} - \underbrace{\pi_E \cdot (b - e \cdot \gamma_E)}_{\text{IFE in elev. dim.}}$$

The results are presented in Tables 7 and 8 with the respective plots, Figures 41 and 42 displayed below them.

Again, the optimization procedure turned out to be outmost stable against variations of initial parameter choice. The quality of the new fits is similarly convincing and in some cases, significant reductions of residual sum-of-squares could even be achieved. Comparing Tables 7 and 8 with Tables 5 and 6, it can be seen that the introduction of a second slope parameter has no major influence on the estimated peripheral or central processing times. The parameters a (azimuth slope) and b (intercept parameter, maximum facilitation) remain nearly unchanged, too, although a here only encodes the strength of spatial factors in the azimuth dimension.
Regarding the novel parameter e, there are pronounced inter-individual differences. So, SB and LP (virtual acoustics) reveal quite small values of e, which might be interpreted as poor resolution of vertical eccentricity. In fact, these both participants showed only low localization performance, too. Hence, even if auditory vertical information is processed fast enough to contribute to visual-auditory interaction effects (as indicated by SB's estimates for $1/\lambda_V$ and $1/\lambda_E$), there will be no significant effects, since vertical disparity is hardly perceived. Higher estimates for the vertical slope parameter can however be found especially for JB (virtual acoustics), PN (both setups), DD and KW (loudspeaker setup). This might explain to some degree the quite contradicting findings of JB being a good localizer with pronounced vertical distance effects on the one hand, however with extremely large $1/\lambda_E$ estimates on the other. The longer auditory processing might simply be compensated by the excellent spatial resolution or by higher reliability of the judgment which leads to larger spatial effects in case of interaction. Generally, e estimates are clearly larger in the free field setup. This seems plausible, as auditory stimuli might better be located in the free field than in a virtual environment. In general, participants with larger e-values have shown to be good localizers too. This finding supports the interpretation of a and e specifying auditory space encoding rather than being solely bimodal interaction parameters.

Different spatial parameters, virtual acoustic setup

Partic.	$V[ms]$	$A[ms]$	$E[ms]$	$t_D[ms]$	$\mu[ms]$	a	e	b	$\sum_{Res^2}$
JB	138	43	342	38	96	0.40	1.43	79	619
KW	77	70	87	39	156	0.62	0.64	85	336
LP	88	88	145	55	139	0.54	0.27	68	182
PN	104	50	82	31	114	0.62	1.22	75	286
SB	114	32	90	24	118	0.58	0.09	68	284

Table 7: Second fit of the extended Two Stage Model to the data from Experiment 3 in the virtual acoustic setup. Different spatial parameters are used to account for possibly different strengths of spatial effects within the horizontal and the vertical plane.

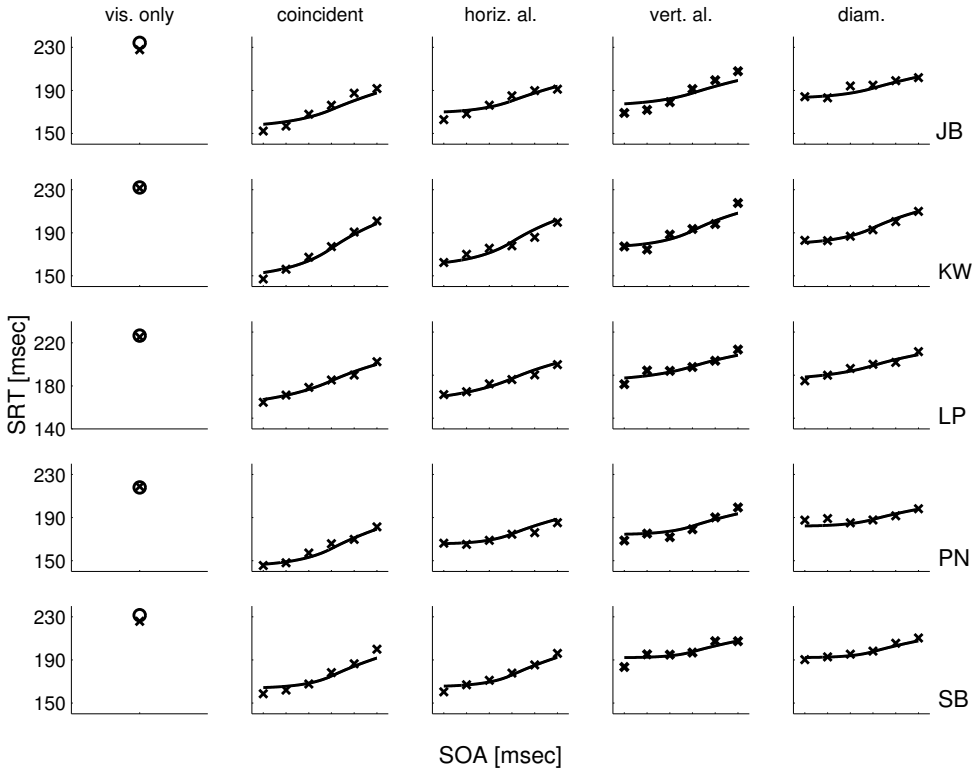


Figure 41: Plots to the second fit of the extended Two Stage Model to the virtual acoustics data (Table 7). Results for different participants are displayed in rows, different spatial interstimulus relations in columns.

Different spatial parameters, loudspeaker setup

Partic.	$V[ms]$	$A[ms]$	$E[ms]$	$t_D[ms]$	$\mu[ms]$	a	e	b	$\sum_{Res^2}$
DD	85	4	97	4	162	0.48	1.19	55	590
HH	85	43	79	28	99	0.59	0.25	41	324
KW	77	74	97	42	124	0.54	1.54	56	288
PN	111	70	199	52	78	0.51	1.98	46	288

Table 8: Second fit of the extended Two Stage Model to the free field data. Different spatial parameters are used to account for possibly different strengths of spatial effects within the horizontal and the vertical plane. Note the different absolute values of the ordinate if compared to the virtual acoustics data.

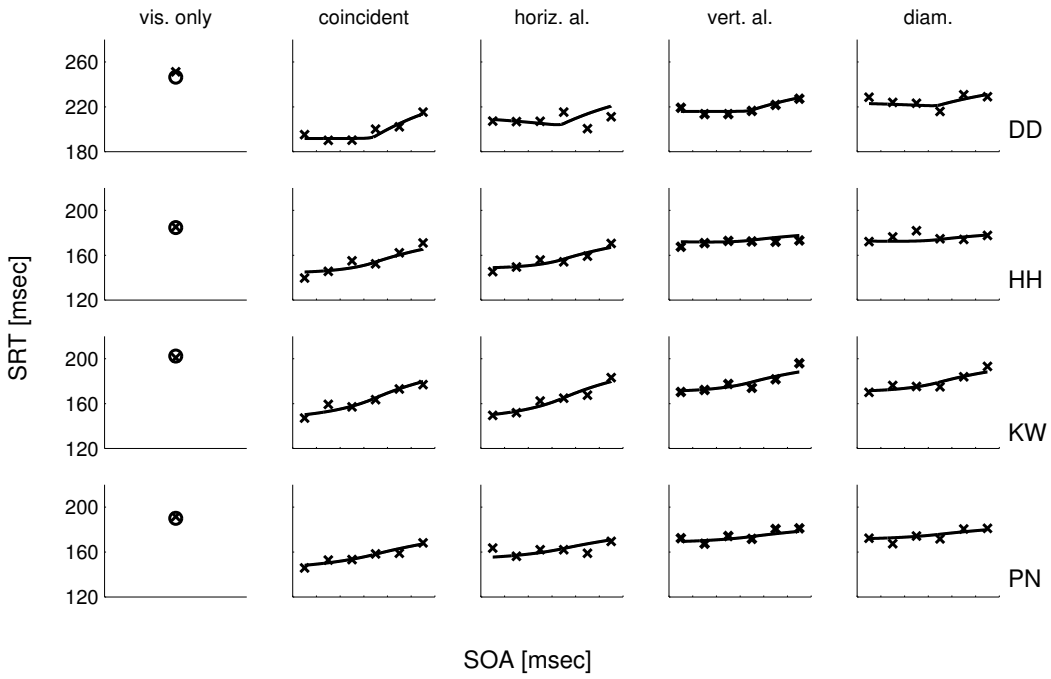


Figure 42: Plots to the second fit of the extended Two Stage Model to the free field data (Table 8). Results for different participants are displayed in rows, different spatial interstimulus relations in columns.

5.4 General discussion of the extended Two-Stage Model

Both versions of the extended Two-Stage Model of Colonius & Arndt yield generally encouraging results. The simple extension of introducing separate processing channels for auditory processing in the horizontal and the vertical dimension made it possible to fairly account for the two-dimensional intersensory facilitation effects we found. In that, the parameter estimation procedure turned out to be very stable across all participants in both experimental setups. The estimated parameters were comparable with physiological data at least up to some degree. Estimated auditory processing times were however often larger than expected by looking at data from cell recordings. On the other hand, individual parameter estimates were mostly in line with the respective behavioral data like localization performance and auditory response time distributions. This latter finding strongly supports further use and development of the Two-Stage Model.

In that connection, one might ask in how far the introduction of an additional spatial parameter *e* makes sense. On the one hand, the simple assumption of a linear relationship between spatial distance and the amount of IFE at a given SOA already yielded convincing estimates being were able to account for a big part of the spatial effects we had found. On the other hand, assuming separate aspect ratios for horizontal and vertical distance could in fact significantly better the fits for some participants, especially those with strong spatial effects. Particularly the clearly observable interaction between the peripheral processing time estimate $1/\lambda_E$ and the spatial parameter *e* has to be observed carefully. So far, it is not clear whether these two parameters are really representatives of two aspects in auditory processing – speed and accuracy – as it might seem at the first glance, or whether they only interact in these data fits with one parameter simply compensating for the insufficient estimate of the other. Large-scale simulations might give an answer to this question. The primary goal of the present study was the general investigation of vertical distance effects compared to effects within the horizontal plane. The next vital step should now be the collection of more data with various spatial combinations in order to further test the present model and its parameters in more detail.

Moreover, general (i.e. spatially unspecific) effects of the auditory accessory are to be taken into account. Since the main intention of the present modelling approach was the inclusion of two-dimensional spatial factors, purely temporal aspects have somewhat been put to the background. The model in fact stresses the crucial role of the temporal order of presentation, but this is only done in connection with the consequences in the spatial domain. This should however not be done in the long run. The MIN-version of the Colonius-Arndt Model may be considered in this connection.

6 Summary and conclusion

Goal of the present study was the investigation of visual-auditory interaction in two-dimensional space and its effect on visually guided saccadic responses. Thereby, the characteristics of the auditory system, precisely the fact that horizontal and vertical spatial information is processed in different manners and probably by separate sensory channels, should be taken into account. Moreover, the usability of virtual acoustic environments in a non-localization task was examined. So far, the quality of virtual environments has mainly been defined by listeners' judgment accuracies in auditory localization experiments. What however might happen in a task that keeps attention away from the auditory stimulus was unclear. Nevertheless this is an important question, regarding the various fields of applications of a virtual acoustic environment in psychophysical experiments.
The Two-Stage Model of Colonius & Arndt (2001) was applied and extended, in order to account for the different processes in auditory spatial analysis and the consequentially different effects of horizontal and vertical interstimulus distance. Two extensions of the original Colonius-Arndt approach were tested. The first modification accounts for different afferent processing times of horizontal and vertical location information in the auditory system. In the second version, the possibility of different influences of a defined horizontal or vertical interstimulus distance on the magnitude of intersensory facilitation was considered.

Characteristics of visually and auditory guided eye movements were verified by analyzing trajectories, position-time traces and latency distributions of target directed saccades. It thereby turned out that auditory target directed eye movements differed significantly from their visually evoked counterparts in all three domains. Auditory trajectories were often curved or bow-like. An observation of the respective position-time traces revealed that this can mainly be attributed to the vertical component of the movement, which often starts later, is sometimes less accelerated and is corrected for once or twice in many cases. Auditory evoked responses were generally not that stereotyped as visually guided saccades, which concerns trajectories as well as latencies. The latter were often broadly distributed in the auditory localization experiment, but not in an auditory detection task, where participants should perform an always identic, non target-directed response as soon as any auditory stimulus was perceived. All these results could be found in both listening conditions and were not related to individual auditory localization performance.

Since target positions were the same with visual and auditory stimuli, the saccade characteristics described above can be attributed to properties of the auditory sensory system rather than to oculomotor features. In the literature, it has been discussed whether the bow-like nature observed in some saccades was connected to stimulus position or rather modality specific. The present data strongly support the latter assumption. The step-like nature of especially the vertical component might then be explained by a *"multiple-look strategy"* (Hofman & Van Opstal 1998) due to subsequent refinement of auditory spectral features analysis, which is the basis for vertical eccentricity estimation. By contrast, the horizontal position of an auditory target is calculated on the basis of binaural cues in early stages of sensory processing. It can therefore be assumed to be completed practically instantaneously. These assumptions are reflected pretty good by one participant's statement of how he localized sound in the present study (loudspeaker setup): *"Left-right decisions are very easy, I simply know the position, I don't know why. Up-*

down distinction is harder, I always have to think a little bit of it, but it gets better from session to session.[...] The upper loudspeakers sound somewhat different than the lower ones, don't they?".
In general, the above findings are in line with earlier studies, e.g. by Frens & Van Opstal (1995) or Zambarbieri et al. (1982). However, the present work reveals for the first time that similar effects can be observed using a virtual acoustic environment. On the other hand, the analyses go further than the "classical" localization experiments simply comparing judgment acuity in free field and virtual listening conditions. Since not only the final location estimations are compared, but complete eye movements, some more inferences can be made with regard to the process of auditory localization. The similarities of auditory guided eye movements in both experimental setups thereby vindicate the further use of virtual acoustic displays.

The consequences of the specific time course of auditory spatial processing, which has become obvious in analyzing acoustically guided eye movements, were investigated in a bimodal focused attention task with visual targets and auditory distractors. First it was verified that participants really followed the instruction of suppressing reactions toward auditory stimuli. Trajectories of bimodal visually evoked saccades were similar to their unimodal visual counterparts. Their latency distributions were however different insofar as they were narrower and shifted toward smaller latency values. Although the bimodal latency distributions sometimes resembled those of the auditory simple detection task, it can be assumed that bimodal saccades are not simply triggered by the perception of an auditory stimulus. Applying the Miller-inequality proved that statistical reasons have to be ruled out as solely explanation for the speed-up of bimodal saccades. In other words, participants were faster in the bimodal task than they could have been if they had simply responded to the first unimodally perceived stimulus. It has therefore to be assumed that (1) participants in fact responded to the visual target in the bimodal task and that (2) nevertheless, the auditory accessory had a significant influence on saccadic latencies, but not on saccade metrics.
Although participants obviously reacted toward the visual targets only, saccadic latencies were significantly shorter in the bimodal trials, in which this intersensory facilitation effect (IFE) was the more pronounced, the earlier the accessory was presented with respect to the target. Spatial effects turned out to be more complex. Generally, saccadic latencies were the smaller the closer both stimuli were located. However, unlike with horizontal distance, vertical displacement was affective for negative interstimulus intervals only. With positive SOA, bimodal saccadic reaction times did not differ for stimulus configurations with different *vertical* distances only.

The results might be interpreted in terms of the framework of Findlay & Walker (1999), who proposed parallel WHEN- and WHERE-pathways in visual processing and saccade programming. Similar to the approach of Frens et al. (1995), it might be suggested that the auditory signal has a double-function here. On the one hand, it acts as unspecific warning-signal by inhibiting SC fixation neurons and brainstem omnipause-neurons and therefore allows general preparation enhancement. On the other hand, bimodal information integration within the sensory and oculomotor saliency maps in the Deep Layers of the Superior Colliculus (DLSC) can lead to enhanced activation of the affected neurons and therefore to faster responses.
This scheme can in fact explain both findings, the spatial effects through neural summation and the rising strength of intersensory facilitation with a preceding accessory. At a first glance, max-

imum effects should be expected if the auditory stimulus follows the visual stimulus by about 40 msec, since this time lag would just compensate for the differen afferent processing times of the two sensory systems. Assuming additional "alerting" effects as outlined above, makes one aspect of visual-auditory interaction quite clear.

The different spatial effects within the horizontal and the vertical plane however require some more explanation. The characteristic bow-like form of many auditory guided saccades were explained by assuming different pathways of auditory horizontal and vertical spatial information processing. It is proximate at this point to assume the same mechanisms acting in visual-auditory information integration, too. If auditory horizontal information is available on the stage of the Superior Colliculus early on, it can be assumed to be effective even at positive SOA. Elevation information however, which depends on the more time-consuming analysis of spectral patterns, is not available so fast and thus only becomes effective if the auditory stimulus occurs earlier with respect to target presentation. Therefore, the results of the unimodal and the bimodal experiments in this study are in good accordance.

The results of the bimodal experiments in virtual and in the loudspeaker acoustics were comparable regarding the influence of the factors SOA, horizontal distance and vertical distance. However, if effects of vertical distance on the IFE are observed in combination with possible target position effects, significant differences emerge between the results collected in the two setups. It turned out as a general finding, that for positive SOA the influence of vertical distance ceases. Aside from horizontal distance effects, the amount of IFE is then completely determined by the vertical position of the target, in which elevated target positions yield larger IFE values. At negative SOA however, this *target-position* effect is dominated by *vertical distance* effects. IFEs for spatially coincident stimulus combinations are then equal, irrespective of target position. Yet, this is only true for the data collected in the virtual acoustic setup.In the loudspeaker environment, the vertical distance effect at negative SOA is still well observable, but it is clearly superimposed by the influence of target position across all SOA.
The findings might be explained by the fact that saccadic responses were generally faster in the loudspeaker setup. As a consequence, the strength of the IFE was generally smaller, which could be explained as a ceiling effect. Probably, the use of LEDs in the loudspeaker setup instead of small white dots presented on a monitor like in the virtual acoustic environment, is a major explanation for this finding. From the literature (Stein & Meredith 1993, Frens et al. 1995), it is known that the amount of spatial facilitation decreases with rising stimulus intensity. One might follow that in the loudspeaker setup using LEDs, the amount of spatial facilitation was severely reduced. More general motor effects might however still have had a strong influence which revealed itself especially in those movements that took most oculomotor effort: those to elevated targets. These reflections might indicate a possible way to distinguish between intersensory facilitation due to sensory integration on the one hand and preparation enhancement on the other.

The extended version of the Two-Stage Model of Colonius & Arndt, which was applied to the reaction time data of the bimodal experiment, is a stochastic approach for modelling processes in peripheral sensory and more central oculomotor programming stages. All processes and sub-processes are represented by random variables. The present model assumes parallel

peripheral processing in separate channels for visual, auditory azimuth and auditory elevation information at the first stage, which is finished as soon as the visual stimulus has reached a common detection and/or decision center. The second stage collects information from the different sensory channels until the first stage processes are terminated, and then prepares the oculomotor response. Second-stage processing times can be reduced significantly by intersensory integration, if at least some of the auditory information has reached the central stage before the visual. Hence, we assume an independent race as proposed by Raab (1962) at the first stage, in which the outcome determines second stage processing. The magnitude of the second-stage IFE is assumed to depend on the *perceived* spatial interstimulus distance. Depending on how much auditory information is available, either horizontal, vertical, or both information is taken into account. Two different approaches were tested in this context. Fist, the spatial was assumed to depend linearly on *radial* distance. In a second attempt, separate spatial "strength" factors were introduce for either dimension.

At a more physiological level, first-stage processes might correspond to afferent processes within the sensory periphery, therefore the assumption of separate channels seems justified. Since recent behavioral and physiological data (Heffner 1997, Schupp, Mrsic-Flogel & King 2001) strongly suggest separate processing of binaural (horizontal) and spectral (vertical) auditory cues, too, the assumption of three divided channels in two-dimensional spatial interaction seems reasonable. The common stage might be interpreted as the Superior Colliculus, which has turned out to play a major role in multimodal integration as well as in saccade programming.

The data fits of both approaches yielded reasonable values for visual processing times, while auditory processing times were often over-estimated, if compared to data from single cell recordings of the cat SC. The order of magnitude for the three estimates was however as expected. Furthermore, the estimates corresponded well to the individual behavioral data. The introduction of a second spatial distance parameter, which allowed to account for different strengths of spatial effects within the horizontal and the vertical plane, yielded significantly better data fits in some cases, in which the estimated values again corresponded well to behavioral data. However, unlike with other model parameters, the additional spatial parameter showed less systematics and is therefore only hardly interpretable. More data have therefore to be collected and applied to the model, in which especially vertical spatial factors should be investigated quantitatively, in order to test the merits of either of the two approaches.

References

Ahnelt, P. K., Kolb, H. & Pflug, R. (1987), 'Identification of a subtype of cone photoreceptor, likely to be blue sensitive, in the human retina', *J Comp Neurol* **255**(1), 18–34.

Alho, K. (1992), 'Selective attention in auditory processing as reflected by event-related brain potentials', *Psychophysiology* **29**(3), 247–263.

Amari, S. (1977), 'Dynamics of pattern formation in lateral-inhibition type neural fields', *Biol Cybern* **3**(27), 77–87.

Bastian, A., Riehle, A., Erlhagen, W. & Schöner, G. (1998), 'Prior information preshapes the population representation of movement direction in motor cortex', *NeuroReport* **9**, 315–319.

Bernstein, I. (1970), 'Can we see and hear at the same time? some recent studies of intersensory facilitation of reaction time.', *Acta Psychol* **33**, 21–35.

Bernstein, I. & Edelstein, B. (1971), 'Effects of some variations in auditory input upon visual choice reaction time', *J Exp Psychol* **87**(2), 241–247.

Bernstein, I. H., Clark, M. & Edelstein, B. (1967), 'Effects of auditory signals on visual reaction time', *J Exp Psychol* **80**(3), 567–569.

Bertelson, P. & Tisseyre, E. (1969), 'The time course of preparation: Confirmatory results with visual and auditory warning signals', *Acta Psychol* **30**, 145–154.

Blauert, J. (1983), *Spatial Hearing: The psychophysics of human sound localization.*, MIT Press, Cambridge.

Colonius, H. & Arndt, P. (2001), 'A two-stage model for visual-auditory interaction in saccadic latencies', *Percept Psychophys* **63**(1), 126–147.

Curcio, C. A., Sloan, K. R. J., O., P., Hendrickson, A. & Kalina, R. E. (1987), 'Distribution of cones in human and monkey retina: individual variability and radial asymmetry', *Science* **236**(4801), 579–582.

Diederich, A. (1995), 'Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models', *J Math Psychol* **52**, 197–215.

Diederich, A. & Colonius, H. (1987), 'Intersensory facilitation in the motor component?', *Psychol Res* **49**, 23–29.

Findlay, J. & Walker, R. (1999), 'A model of saccade generation based on parallel processing and competitive inhibition', *Behavioral and Brain Sciences* **22**, 661–721.

Fischer, B. (1987), 'The preparation of visually guided saccades', *Rev Physiol Biochem Pharmacol* **106**, 1–35.

Fischer, B. & Rampsberger, E. (1984), 'Human express saccades: extremely short reaction times of goal directed eye movements', *Exp Br Res* **55**, 232–242.

Fischer, B. & Weber, H. (1993), 'Express saccades and visual attention', *Behav Brain Sci* **16**, 553–610.

Frens, M. A. & Van Opstal, A. J. (1995), 'A quantitative study of auditory-evoked saccadic eye movement in two dimensions', *Exp Br Res* **107**, 103–117.

Frens, M. A., Van Opstal, A. & Van der Willigen, R. (1995), 'Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements', *Percept Psychophys* **57**(6), 802–816.

Grice, G. R., Canham, L. & Boroughs, J. M. (1984), 'Combination rule for redundant information in reaction time tasks with divided attention', *Percept Psychophys* **35**(5), 451–463.

Harrington, L. & Peck, C. (1998), 'Spatial disparity affects visual-auditory interactions in human sensorymotor processing', *Exp Br Res* **122**, 247–252.

Heffner, H. E. (1997), 'The role of macaque auditory cortex in sound localization', *Acta Otolaryngol Suppl* **532**, 22–27.

Hershenson, M. (1962), 'Reaction time as a measure of intersensory facilitation', *J Exp Psychol* **63**(3), 289–293.

Heuermann, H. & Colonius, H. (1999), Localization experiments with saccadic responses in virtual auditory environments, *in* T. Dau, V. Hohmann & B. Kollmeier, eds, 'Psychophysics, physiology and models of hearing', World Scientific Publishing, Singapore, pp. 89–92.

Heymans, G. (1904), 'Untersuchungen über psychische Hemmung: V. Die Verdrängung von Schallempfindungen durch elektrische Hautempfindungen', *Zsch f Psychol* **34**, 15–28.

Hofman, P. M. & Van Opstal, A. J. (1998), 'Spectro-temporal factors in two-dimensional human sound localization', *J Acoust Soc Am* **103**(5), 2634–48.

Hofman, P. M., Van Riswick, J. G. & Van Opstal, A. J. (1998), 'Relearning sound localization with new ears', *Nature Neurosc* **1**, 417–421.

Hughes, H., Nelson, M. & Aronchick, D. (1998), 'Spatial characteristic of visual-auditory summation in human saccades', *Vision Res* **38**, 3955–3963.

Jacobsen, E. (1911), 'Experiments on the inhibition of sensations', *Psychol Rev* **18**, 24–53.

Jay, M. F. & Sparks, D. L. (1995), *Localization of auditory and visual targets for the initiation of saccadic eye movements*, The Cognitive Neurosciences, Gazzaniga, M.S. (Ed.), Massachusetts: MIT Press, pp. 351–374.

Konrad, H. R., Rea, C., Olin, B. & Colliver, J. (1989), 'Simultaneous auditory stimuli shorten saccade latencies', *Laryngoscope* **99**(12), 1230–1232.

Kopecz, K. (1995), 'Saccadic reaction times in gap/overlap paradigms: A model based on integration of intentional and visual information on neural, dynamic fields', *Vision Res* **35**, 2911–2925.

Marc, R. E. & Sperling, H. G. (1977), 'Chromatic organization of primate cones', *Science* **196**(4288), 454–456.

Meredith, M. A., Nemitz, J. & Stein, B. (1987), 'Determinants of multisensory integration in superior colliculus neurons I: temporal factors', *J Neurosc* **7**(10), 3215–3229.

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

Miller, J. (1982), 'Divided attention: Evidence for coactivation with redundant signals', *Cognitive Psychology* **114**, 247–279.

Munoz, D. & Corneil, B. (1995), 'Evidence for interactions between target selection and visual fixation for saccade generation in humans', *Exp Br Res* **103**(1), 168–173.

Munoz, D. & Wurtz, R. (1993*a*), 'Fixation cells in monkey superior colliculus I: Characteristics of cell discharge', *J Neurophysiol* **70**(2), 559–575.

Munoz, D. & Wurtz, R. (1993*b*), 'Fixation cells in monkey superior colliculus II. reversible activation and deactivation', *J Neurophysiol* **70**(2), 576–589.

Newhall, S. (1923), 'Effects of attention on the intensity of cutaneous pressure and on visual brightness', *Arch of Psychol* **61**.

Nickerson, N. H. (1973), 'Intersensory facilitation of reaction time: Energy summation or preparation enhancement', *Psychol Rev* **80**(6), 489–509.

Nozawa, G., Reuter-Lorenz, P. & Hughes, H. (1994), 'Parallel and serial processes in the human oculomotor system: bimodal integration and express saccades', *Biol Cybern* **72**, 19–34.

Raab, D. H. (1962), 'Statistical facilitation of simple reaction times', *Transac NY Ac Sc* **24**, 574–590.

Ratcliff, R., Van Zandt, T. & McCoon, G. (1999), 'Connectionist and diffusion models of reaction time', *Psychol Rev* **106**, 261–300.

Recanzone, G., Guard, D. C., Phan, M. L. & Su, T. K. (2000), 'Correlation between the activity of single auditory cortical neurons and sound-localization behaviour in the macaque monkey', *J Neurophysiol* **83**, 2723–2739.

Schupp, J. W. H., Mrsic-Flogel, T. & King, A. (2001), 'Linear processing of spatial cues in primary auditory cortex', *Nature* **414**, 200–204.

Simon, J. R. & Craft, J. L. (1970), 'Effects of an irrelevant auditory stimulus on visual choice reaction time', *J Exp Psychol* **86**(2), 272–274.

Stein, B. A. & Meredith, M. A. (1993), *The Merging of the Senses*, A Bradford Book, MIT Press, Cambridge.

Todd, J. W. (1912), Reaction to multiple stimuli, *in* R. S. Woodworth, ed., 'Archives of Psychology: No. 25. Columbia Contributions to Philosophy and Psychology, Vol. XXI', Science Press, New York.

Trappenberg, T. P., Dorris, M. C., Munoz, D. P. & Klein, R. M. (2001), 'A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus', *J Cogn Neurosci* **13**(2), 256–7.

Troidl, K. (2002), The influence of an auditory accessory stimulus on target choice and reaction time with two visual stimuli, Doctoral dissertation thesis, Oldenburg University.

Urbantschitsch, V. (1888), 'Ueber den Einfluss einer Sinneserregung auf die uebrigen Sinnesempfindungen', *Arch ges Psychol* **42**, 154–182.

Urbantschitsch, V. (1903), 'Ueber die Beinflussung subjektiver Gesichtsempfindungen', *Arch ges Psychol* **94**, 347–448.

Van Opstal, A. & Van Gisbergen, J. A. M. (1989), 'A nonlinear model for collicular spatial interactions underlying the metrical properties of electrically elicited saccades', *Biol Cybern* **60**(3), 171–83.

Wightman, F. & Kistler, D. (1989), 'Headphone simulation of free-field listening. II: Psychophysical validation', *J Acoust Soc Am* **85**(2), 868–878.

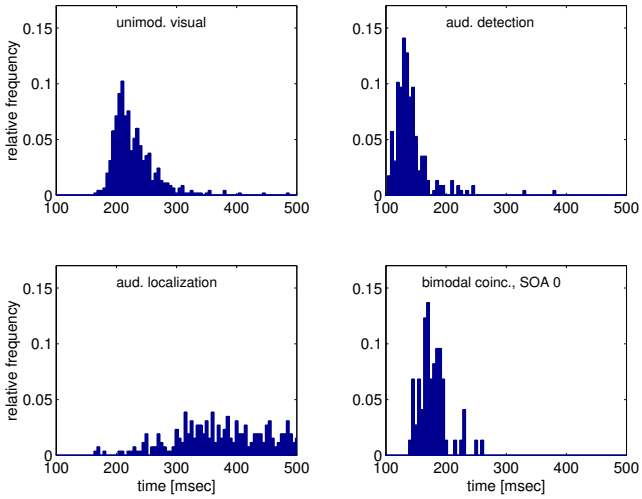
Zahn, J., Abel, L. & Dell'Osso, L. (1978), 'Audio-ocular response characteristics', *Sensory Processes* **2**, 32–27.

Zambarbieri, D., Beltrami, G. & Versino, M. (1995), 'Saccade latency toward auditory targets depends on the relative position of the sound source with respect to the eyes', *Vision Res* **35**(23/24), 3305–3312.

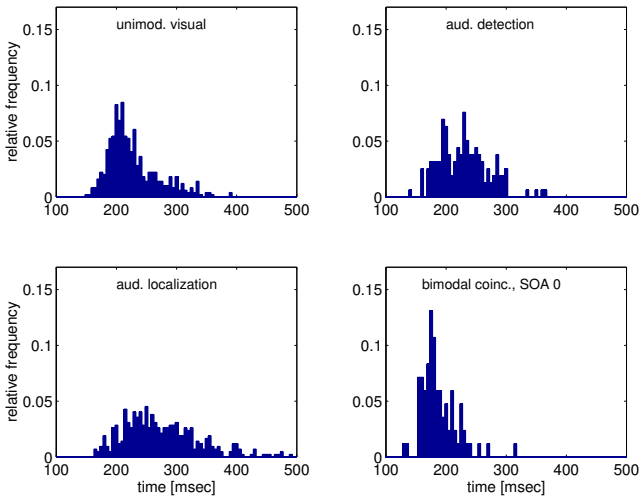
Zambarbieri, D., Schmid, R., Magenes, G. & Prablanc, C. (1982), 'Saccadic responses evoked by presentation of visual and auditory targets', *Exp Br Res* **47**, 417–427.

A Comparison of latency distributions in the various tasks

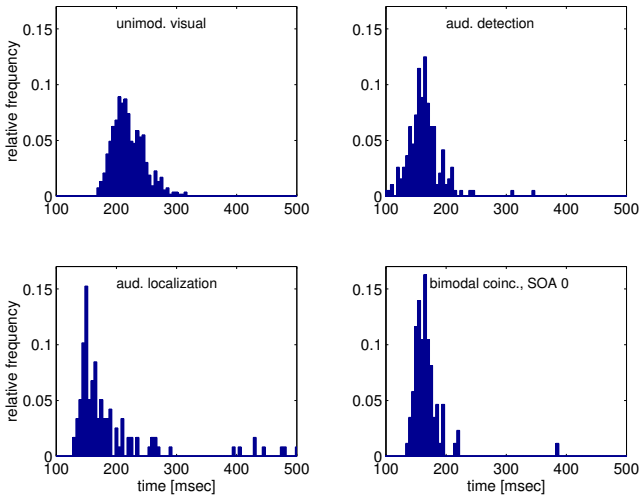
Virtual acoustic setup



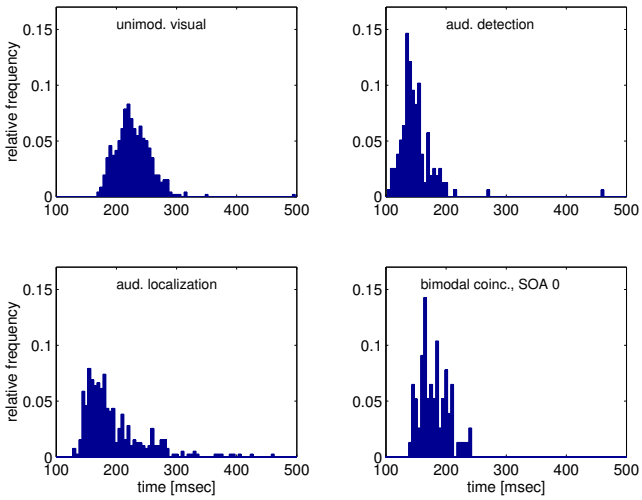
participant JB, virtual acoustic setup



participant LP, virtual acoustic setup

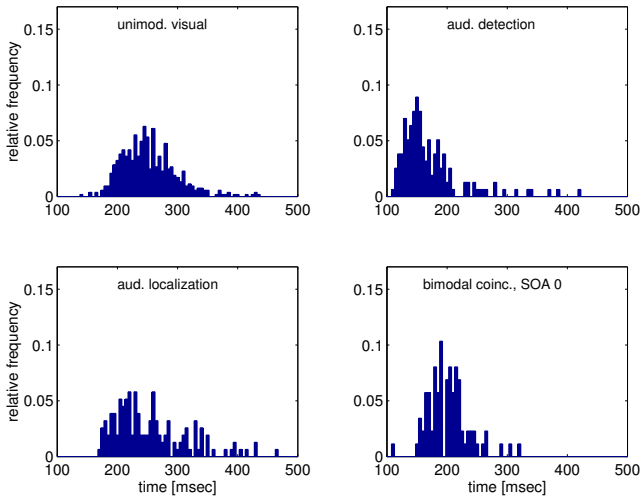


participant PN, virtual acoustic setup

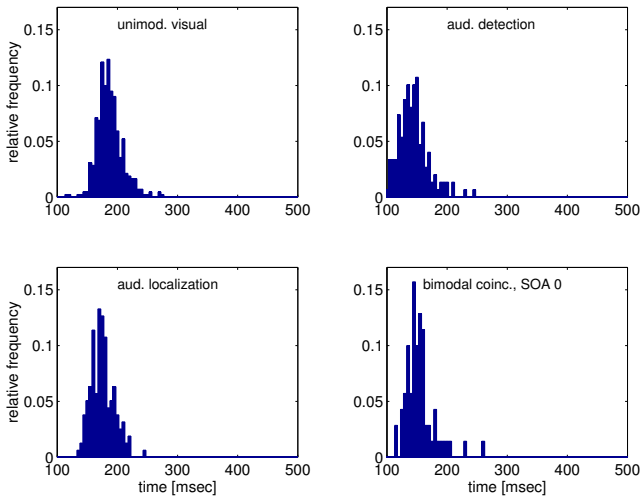


participant SB, virtual acoustic setup

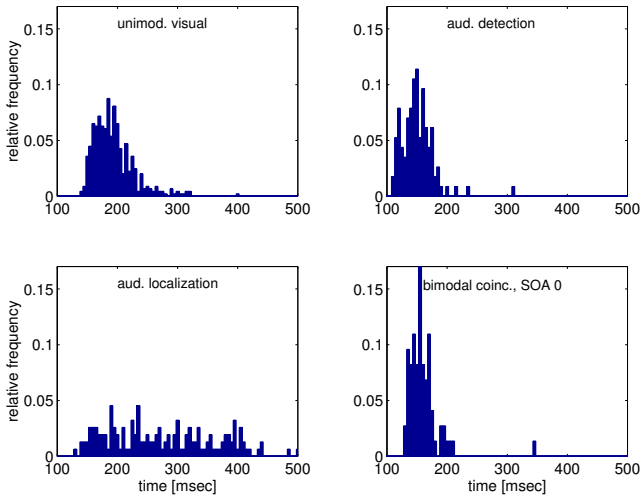
Loudspeaker setup



participant DD, loudspeaker setup



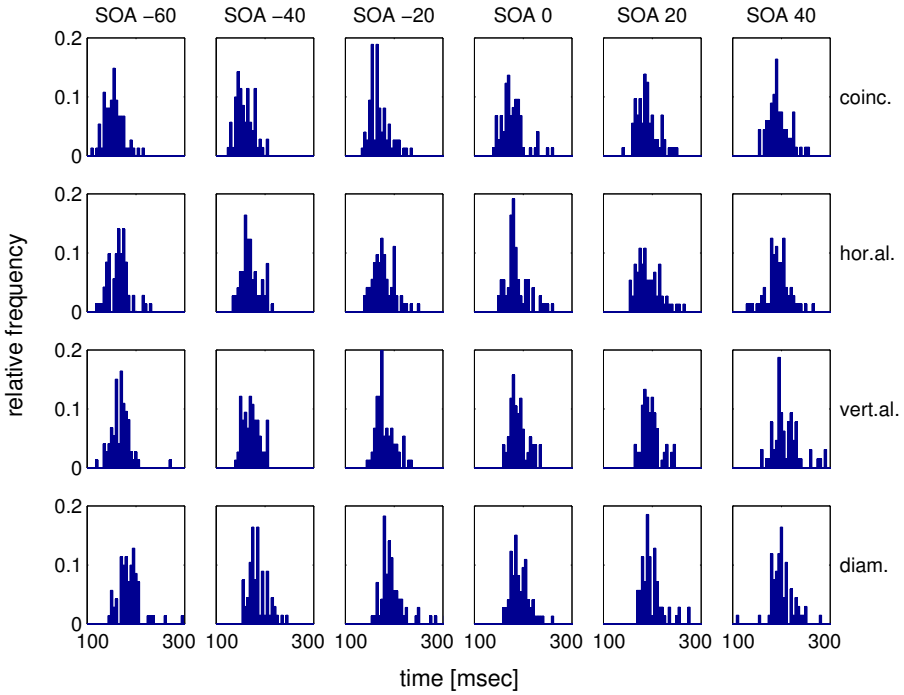
participant HH, loudspeaker setup



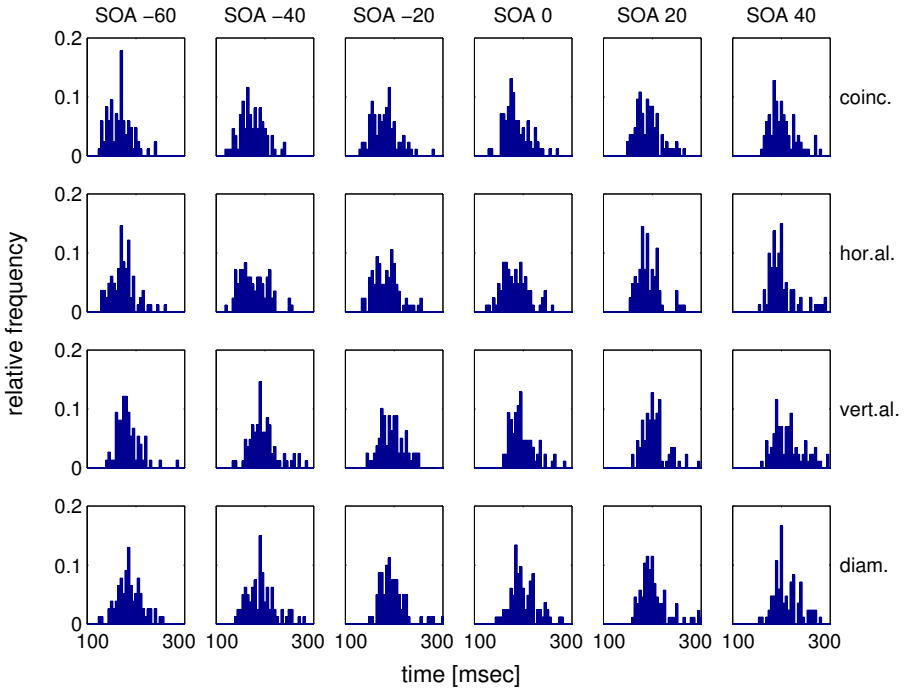
participant PN, loudspeaker setup

B Latency distributions in the bimodal reaction time task

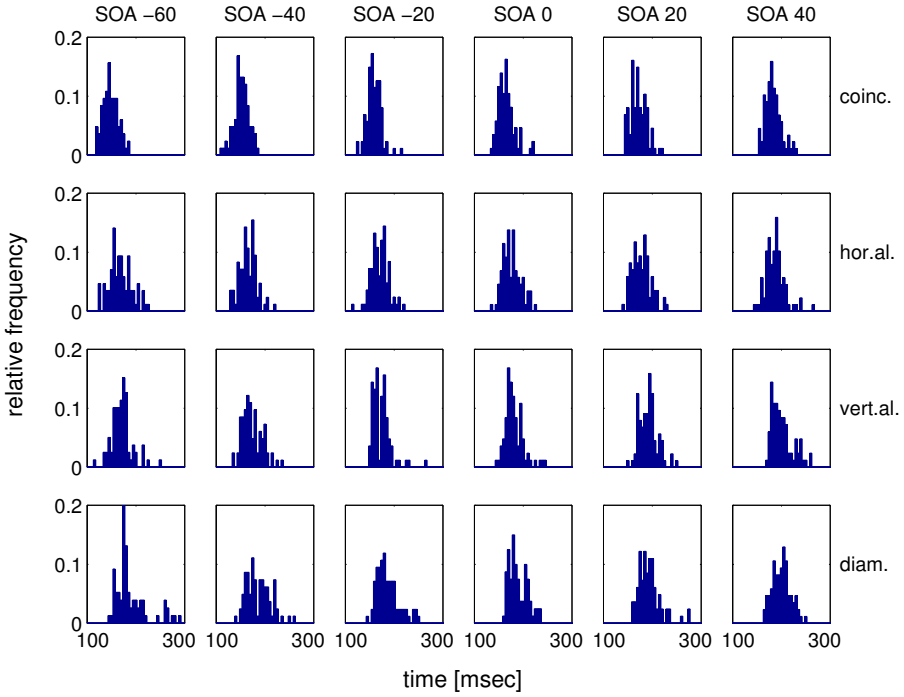
Virtual acoustic setup



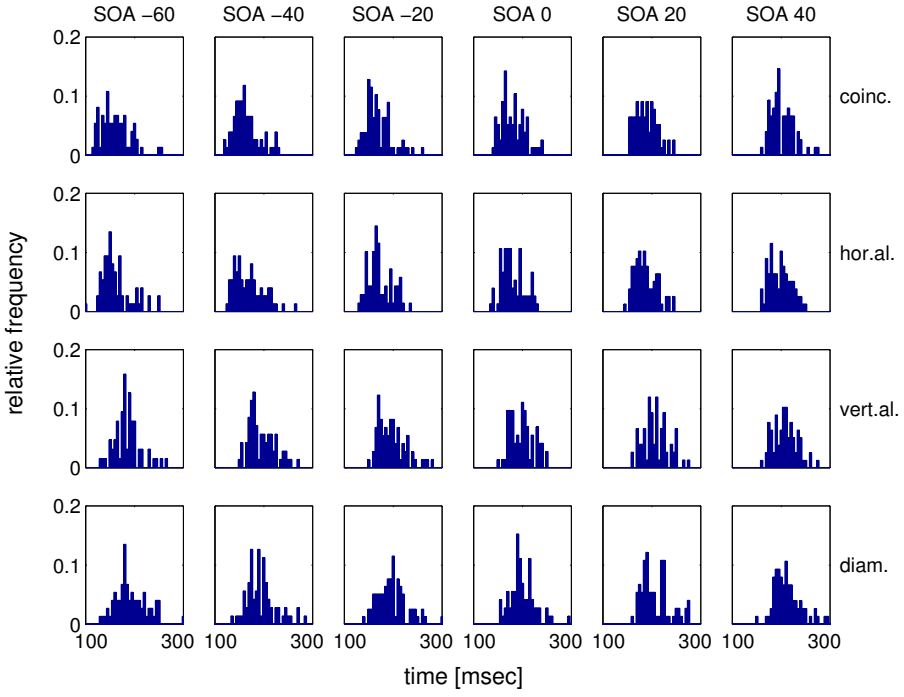
participant JB, virtual acoustic setup



participant LP, virtual acoustic setup

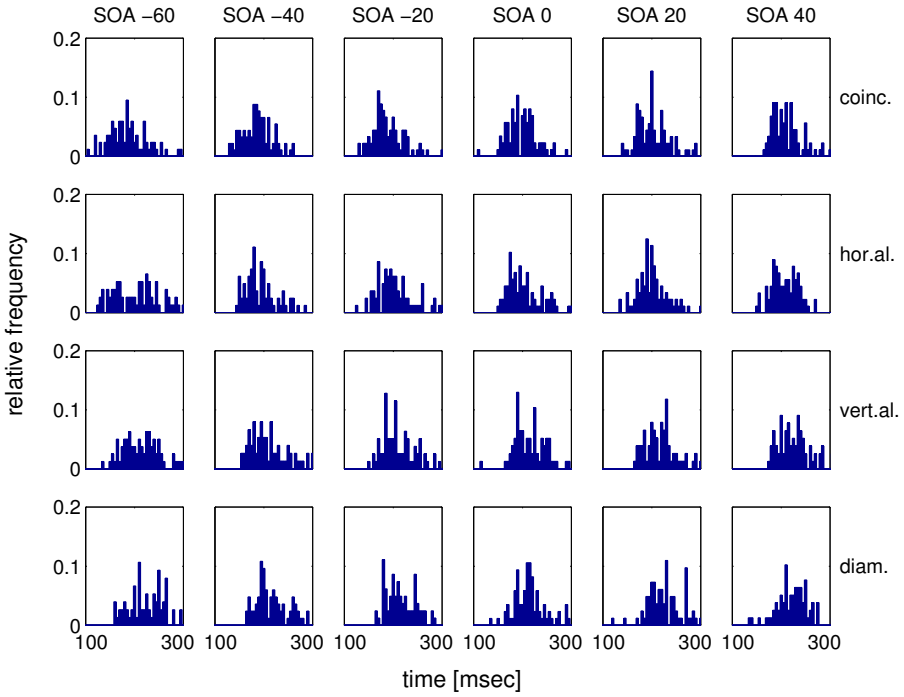


participant PN, virtual acoustic setup

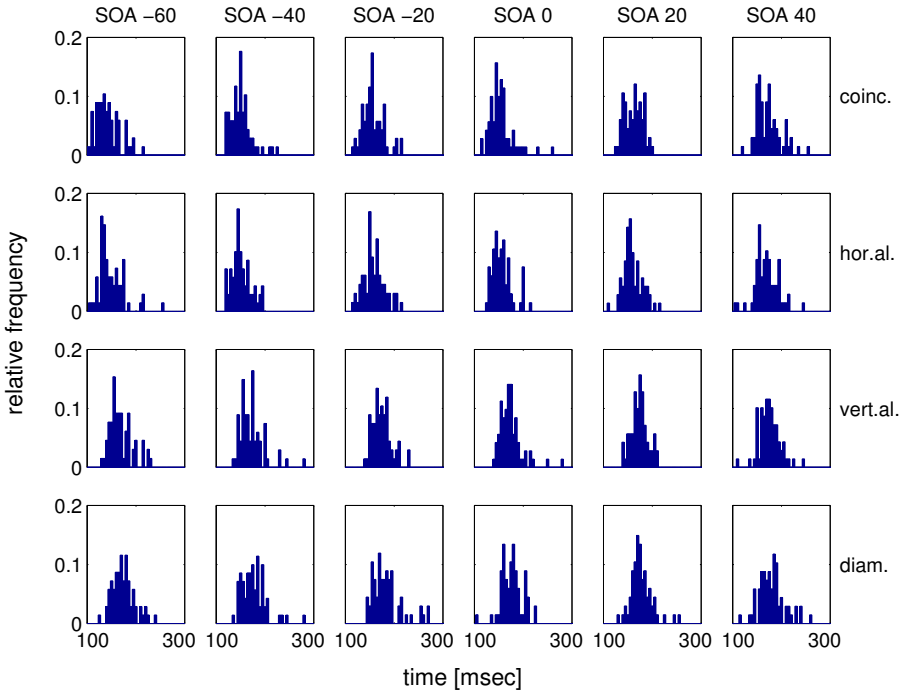


participant SB, virtual acoustic setup

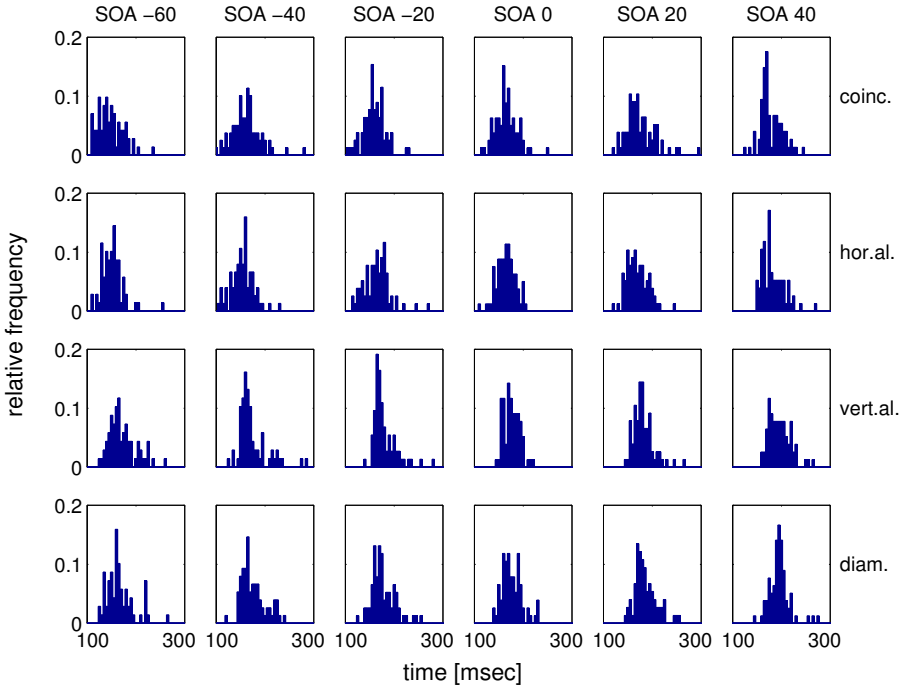
Loudspeaker setup



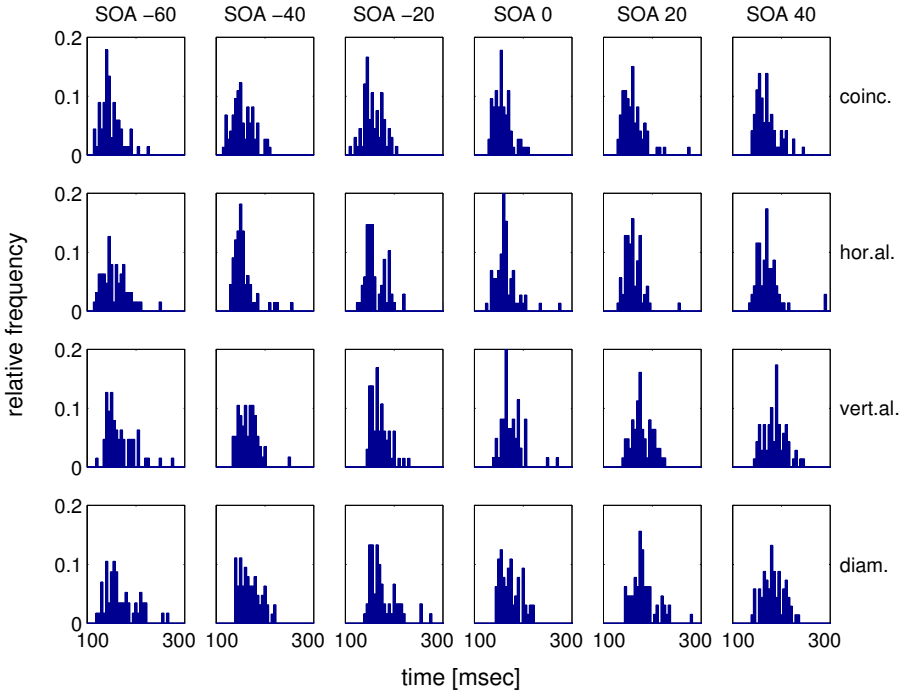
participant DD, loudspeaker setup



participant HH, loudspeaker setup



participant KW, loudspeaker setup



participant PN, loudspeaker setup

C Calculation of interaction probabilities

Positive SOA

In case of positive SOA (auditory stimulus following visual stimulus presentation), i.e. $\tau > 0$, the Fubini-integrals can be calculated by

$$\begin{aligned} P(A+\tau<E+\tau<V) &= \int_{\tau}^{\infty}\int_{0}^{v-\tau}\int_{0}^{e} \lambda_A\, e^{-\lambda_A a}\, \lambda_E\, e^{-\lambda_E e}\, \lambda_V\, e^{-\lambda_V v}\, da\, de\, dv \\ &= \frac{e^{-\lambda_V \tau}\lambda_A\lambda_E}{(\lambda_E+\lambda_V)\,(\lambda_A+\lambda_E+\lambda_V)} \end{aligned}$$

$$\begin{aligned} P(E+\tau<A+\tau<V) &= \int_{\tau}^{\infty}\int_{0}^{v-\tau}\int_{0}^{a} \lambda_E\, e^{-\lambda_E e}\, \lambda_A\, e^{-\lambda_A a}\, \lambda_V\, e^{-\lambda_V v}\, de\, da\, dv \\ &= \frac{e^{-\lambda_V \tau}\lambda_A\lambda_E}{(\lambda_A+\lambda_V)\,(\lambda_A+\lambda_E+\lambda_V)} \end{aligned}$$

$$\begin{aligned} P(A+\tau<V<E+\tau) &= \int_{0}^{\infty}\int_{\tau}^{e+\tau}\int_{0}^{v-\tau} \lambda_A\, e^{-\lambda_A a}\, \lambda_V\, e^{-\lambda_V v}\, \lambda_E\, e^{-\lambda_E e}\, da\, dv\, de \\ &= \frac{e^{-\lambda_V \tau}\lambda_A\lambda_V}{(\lambda_E+\lambda_V)\,(\lambda_A+\lambda_E+\lambda_V)} \end{aligned}$$

$$\begin{aligned} P(E+\tau<V<A+\tau) &= \int_{0}^{\infty}\int_{\tau}^{a+\tau}\int_{0}^{v-\tau} \lambda_E\, e^{-\lambda_E e}\, \lambda_V\, e^{-\lambda_V v}\, \lambda_A\, e^{-\lambda_A a}\, de\, dv\, da \\ &= \frac{e^{-\lambda_V \tau}\lambda_E\lambda_V}{(\lambda_A+\lambda_V)\,(\lambda_A+\lambda_E+\lambda_V)} \end{aligned}$$

Negative SOA

In case of negative SOA, i.e. $\tau < 0$, the easiest way to derive the respective integrals is substituting τ by $\zeta = -\tau$. That is, we again use an exponent that is > 0, which simplifies our calculation a lot.

$$\begin{aligned} P(A-\zeta < E-\zeta < V) &= \int_0^\infty \int_0^{v+\zeta} \int_0^e \lambda_A\, e^{-\lambda_A a}\, \lambda_E\, e^{-\lambda_E e}\, \lambda_V\, e^{-\lambda_V v}\, da\, de\, dv \\ &= \frac{\lambda_A}{\lambda_A+\lambda_E} - \frac{e^{-\lambda_E \zeta}\, \lambda_V}{\lambda_E+\lambda_V} + \frac{e^{-(\lambda_A+\lambda_E)\zeta}\, \lambda_V\, \lambda_E}{(\lambda_A+\lambda_E+\lambda_V)\,(\lambda_A+\lambda_E)} \end{aligned}$$

$$\begin{aligned} P(E-\zeta < A-\zeta < V) &= \int_0^\infty \int_0^{v+\zeta} \int_0^a \lambda_E\, e^{-\lambda_E e}\, \lambda_A\, e^{-\lambda_A a}\, \lambda_V\, e^{-\lambda_V v}\, de\, da\, dv \\ &= \frac{\lambda_E}{\lambda_A+\lambda_E} - \frac{e^{-\lambda_A \zeta}\, \lambda_V}{\lambda_A+\lambda_V} + \frac{e^{-(\lambda_A+\lambda_E)\zeta}\, \lambda_V\, \lambda_A}{(\lambda_A+\lambda_E+\lambda_V)\,(\lambda_A+\lambda_E)} \end{aligned}$$

$$\begin{aligned} P(A-\zeta < V < E-\zeta) &= \int_\zeta^\infty \int_0^{e-\zeta} \int_0^{v+\zeta} \lambda_A\, e^{-\lambda_A a}\, \lambda_V\, e^{-\lambda_V v}\, \lambda_E\, e^{-\lambda_E e}\, da\, dv\, de \\ &= \frac{e^{-\lambda_E \zeta}\, \lambda_V}{\lambda_E+\lambda_V} - \frac{e^{-(\lambda_A+\lambda_E)\zeta}\, \lambda_V}{\lambda_A+\lambda_E+\lambda_V} \end{aligned}$$

$$\begin{aligned} P(E-\zeta < V < A-\zeta) &= \int_\zeta^\infty \int_0^{a-\zeta} \int_0^{v+\zeta} \lambda_E\, e^{-\lambda_E e}\, \lambda_V\, e^{-\lambda_V v}\, \lambda_A\, e^{-\lambda_A a}\, de\, dv\, da \\ &= \frac{e^{-\lambda_A \zeta}\, \lambda_V}{\lambda_A+\lambda_V} - \frac{e^{-(\lambda_A+\lambda_E)\zeta}\, \lambda_V}{\lambda_A+\lambda_E+\lambda_V} \end{aligned}$$

Finally replacing ζ by $-\tau$ results into the probability values for preceding auditory stimuli.

EIDESSTATTLICHE VERSICHERUNG

Ich versichere hiermit, daß ich meine Dissertation *Temporal and spatial factors in visual-auditory interaction* ohne unerlaubte Quellen angefertigt und mich keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe. Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Oldenburg, den 19. Februar 2002

Lebenslauf

Heike Heuermann
geb. am 05. 06. 1971 in Twistringen

Schule

07/1977–07/1981	Grundschule in Twistringen
08/1981–05/1990	Gymnasium Unserer Lieben Frau in Vechta bis zum Abitur

Studium

10/1990–09/1997	Studium der Physik an der Carl von Ossietzky Universität Oldenburg
05/1997–07/1998	Studentische Hilfskraft am Institut für Kognitionsforschung des FB 5 der Universität Oldenburg
	Anfertigen der Diplomarbeit mit dem Titel *'Vergleichende Untersuchung der Lokalisierbarkeit akustischer Stimuli unter Freifeld- und virtuellen Bedingungen mittels gerichteter Augen- und Handbewegungen'* Erstgutachter: Prof. Dr. Dr. B. Kollmeier Zweitgutachter: Prof. Dr. V. Mellert
09/1998	Abschluß der Diplomprüfung
10/1998–02/2003	Promotionsstipendiatin des interdisziplinären Graduiertenkollegs Psychoakustik an der Universität Oldenburg
	zeitgleich Studium der Psychologie
10/2002–12/2002	Anfertigen der Hausarbeit mit dem Titel *'Dynamische Modelle menschlicher Entwicklung'* als zusätzlicher Nachweis einschlägiger wissenschaftlicher Kenntnisse im Fach Psychologie Erstgutachterin: Prof. Dr. G. Szagun Zweitgutachter: Prof. Dr. K.-P. Walcher
02/2003	Einreichen der Dissertationsschrift *'Spatial and temporal factors in visual-auditory interaction'* Betreuer: Prof. Dr. H. Colonius Korreferent: Prof. Dr. V. Mellert
12/2003	Abschluß des Promotionsvorhabens mit der Disputation

Oldenburg, den 14. Februar 2003

Danksagung

Die vorliegende Arbeit wurde ermöglicht durch ein DFG-Stipendium im Rahmen des Graduiertenkollegs ‘Psychoakustik’.

Mein Dank gilt all denen, die zum Gelingen meiner Promotion beigetragen haben. Dies gilt vor allem für

- Hans Colonius, meinen Betreuer, für die Motivation zu meiner wissenschaftlichen Arbeit und für die langjährige Förderung derselben
- Petra Arndt und Gisela Szagun für die unzähligen (nicht immer ganz ernsten) Gespräche und Diskussionen und das daraus resultierende anregende Arbeitsklima
- Karin Troidl, Holle Kirchner, Annika Åkerfeldt, Jale Özyurt, Claudia Steinbrink, Sandra Tabeling und allen weiteren KollegInnen und MitstreiterInnen, von denen immer mindestens eine(r) ein offenes Ohr für meine Sorgen hatte
- Roland Rutschmann für seine TEX-Tipps und so manchen lockeren Kommentar
- Adele Diederich, ohne die meine Disputation in 2002 nicht mehr stattgefunden hätte, für ihre hilfreichen Anmerkungen zu verschiedenen Modellierungsansätzen
- Arno Schilling und Wolfgang Gerken für die (programmier-) technische Unterstützung
- Annika Åkerfeldt und Monika Niemann, die meine Dissertation im Rekordtempo korrekturgelesen haben
- meine Versuchspersonen, ohne deren Geduld und Ausdauer ich meine Experimente nicht hätte durchführen können
- Klaus-Peter Walcher für seine Unterstützung im Rahmen des Promotionsverfahrens

Als wichtigen Sponsor möchte ich auf keinen Fall die OE Controlling der Deutschen Genossenschafts-Hypothekenbank unerwähnt lassen:

- Herrn Hajo Eggers, der es mir nicht nur über Monate hinweg erlaubt hat, einen voll ausgestatteten Arbeitsplatz in der DG Hyp zu nutzen, sondern dessen Laptop auch anlässlich meiner Diputation hervorragende Dienste geleistet hat
- Chanh Kunz, der mir darüber hinaus den einen oder anderen Extra-Wunsch erfüllt hat
- Henning Klages, Joachim Kropp und viele weitere “Kollegen", die mir immer das Gefühl gegeben haben, ein willkommener Gast zu sein

Vor allem aber danke ich meinem Mann Hans-Jürgen. Einerseits für die Leitung der Messungen bei der Versuchsperson HH und für die kritische Diskussion meiner Arbeiten. Vor allem aber dafür, dass er meine Anfälle von Arbeitswut ebenso geduldig ertragen hat wie er immer wieder (mit Erfolg) versucht hat, mich aus den unvermeidlichen “Frustphasen” herauszuholen. Ohne Dich hätte ich vieles nicht geschafft.