Insperata accident magis saepe quam quae speres.
(Things you do not expect happen more often than things you do expect)   Plautus (ca 200(B.C.)

Project no: 027787

DIRAC

Detection and Identification of Rare Audio-visual Cues

Integrated Project
IST - Priority 2

DELIVERABLE NO: D5.6
Report on electrophysiological correlates of audiovisual temporal-order judgments

Date of deliverable:07-12-31
Actual submission date: 08-02-12

Start date of project: 01.01.2006 Duration: 60 months

Organization name of lead contractor for this deliverable:
Leibniz Institute for Neurobiology

Project co-funded by the European Commission within the Sixth Framework Program (2002-2006)		
Dissemination Level		
PU	Public	
PP	Restricted to other program participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	X
CO	Confidential, only for members of the consortium (including the Commission Services)	

Detection and Identification of Rare Audiovisual Cues

Insperata accident magis saepe quam quae speres.
(Things you do not expect happen more often than things you do expect)   Plautus (ca 200(B.C.)

D5.6 Report On Electrophysiological Correlates Of Audiovisual Temporal-order Judgments

Leibniz Institute for Neurobiology (LIN)

Abstract:
It is currently debated which parameters of auditory and visual stimuli influence the perception of temporal order when such stimuli are presented in close temporal proximity. Previous research has demonstrated that the relative spatial locations and relative intensity relationships have an influence on (1) the perceived temporal order and on (2) the necessary temporal order to achieve perception of simultaneity. Here, using an orthogonal design embedded in an audiovisual temporal-order judgment (TOJ) task, we studied the influence of relative intensities and stimulus duration on temporal-order judgement. We could replicate the known intensity effect and discovered an often overlooked additional confounding stimulus parameter, stimulus duration: Identical increase of duration of auditory and visual stimuli (all other parameters held constant) led to a shift of PSE to smaller values (in the extreme case to negative values). We have also begun to combine a classical temporal-order judgment paradigm with the simultaneous recording of scalp EEG. A preliminary analysis of ERPs revealed audiovisual interaction affects brain activity at both early and late processing stages. The fact that an N2pc component could be demonstrated in our experiment, indicates the possibility for further investigation of how processing of visual stimulus might be modulated in a top-down fashion by processes evoked from stimuli of another sensory modality (e.g. audition).

Table of Content

1. Introduction

In this deliverable we report work done towards the study of EEG correlates of audiovisual temporal-order judgments. As described in the Periodic Progress Report we have shifted the emphasis of this report to a more thorough study of the psychophysics of temporal-order judgment, the necessity of which only became apparent in the course of some pilot experiments. Therefore, the psychophysics, as necessary for the planned research track (cf. Technical Annex and Updated Technical Annex), is now the focus of the present report and the EEG study has a pilot nature.

For the meaningful interpretation of rare events in perception, cognitive systems can take advantage of integrating cues from different sensory modalities. Psychophysical studies in humans have demonstrated various crossmodal interactions like the McGurk effect (McGurk & MacDonald, 1976), auditory driving (Gebhard & Mowbray, 1959; Welch, DuttonHurt & Warren, 1986, Recanzone 2003), and the double-beep illusion (Shams, Kamitani & Shimojo, 2000; Shams, Kamitani & Shimojo, 2002). Many studies further show, that temporal integration plays a crucial role in multisensory processing. In order to perceive a compound, multimodal event, temporal differences in the occurrence of multimodal cues, which are thought to arise mainly from modality-dependent physical transmission and sensory processing times, have to be compensated. Since the second half of the 19th century with Exner (1875), who carried out one of the first bimodal temporal-order judgment (TOJ) tasks in humans (cited in Neumann & Niepel, 2004), there has grown a large body of psychophysical literature dealing with such temporal aspects of multisensory integration. It turns out that in TOJ tasks, the point of subjective equality (PSE), i.e. the relative timing of one stimulus in one sensory modality to a stimulus presented in another modality for being perceived as simultaneous, is not fixed in modality specific way. The actual value of the PSE depends on many factors like stimulus intensities (Neumann & Niepel, 2004), task conditions, and other cognitive factors like attention as revealed by cueing or the prior entry effect (Jaskowski 1996; Neumann and Niepel, 2004; Posner, Snyder and Davidson, 1980; Spence, Shore & Klein, 2001; Titchener, 1908, Neumann, Koch, Niepel and Tappe, 1992; for a review: Neumann and Niepel, 2004; Jaskowski,1996; Johnston and Nishida, 2001). Moreover, exposure to asynchronous events in two different sensory modalities can lead to a recalibration of the PSE. In the processing of audiovisual (AV) stimuli, this recalibration effect is highly pronounced, compared to other modality pairings (c.f. Harrar & Harris, 2008), indicating prominent plastic changes in temporal integration of asynchronous auditory and visual cues (Vroomen, Keetels, de Gelder, & Bertelson, 2004; Fujisaki, Shimojo, Kashino & Nishida, 2004; but see: Hanson, Heron & Whitaker, 2008).
Many studies about crossmodal PSEs claim, however, that there exist general patterns of temporal integration which are specific for two interacting sensory modalities. Though, conflicting results have been found on the relative timing of acoustic and the visual stimuli required for their simultaneous perception. Jakowski et al. (1990; 1996) and Zampini et al. (2003; 2005) found that at the PSE visual stimulation has generally to precede acoustic stimulation, in order to obtain a simultaneous, audiovisual percept. One might conclude from this that physiological and psychological processes leading to detection or perception of a stimulus act more rapidly for a auditory stimulus. On the other hand, Smith (1933),

Rutschmann and Link (1964), and Neumann et al. (1992) had demonstrated the opposite effect, that acoustic stimulation has to precede visual stimulation at the PSE suggesting that in their TOJ tasks visual perception is faster than auditory perception. To determine PSEs in TOJ-tasks, most of these studies have chosen rather fixed settings of stimulus parameters and task conditions, which differ between studies. This might provide an explanation for the aforementioned conflicting results: In the studies of Jakowski et al. (1990; 1996) and Zampini et al. (2003; 2005), the intensities of the acoustic stimuli were generally higher (> ~70 dB(A)), and the visual stimulus intensities smaller than or similar to the studies of Smith (1933), Rutschmann and Link (1964), and Neumann et al., 1992). Neumann and colleagues (1992) determined audiovisual PSEs for combinations of three different sound intensities and three different light intensities. Their results indicate that the stimulus of higher intensity was perceived earlier than the stimulus of lower intensity, and that the size of this effect is a positive function of the intensity difference between the stimuli.

Despite this intensity effect, Neumann and Niepel (2004) maintained that the general pattern of observation consists of a lead of the visual modality relative to the auditory modality, although with high sound intensities, auditory perception in their experiment finally became faster than visual perception comparable to the results in the experiments of Jakowski et al. (1990; 1996) and Zampini et al. (2003; 2005).

We have argued (Boenke et al. 2007a, b) that an additional factor that might explain the discrepancies of the observed PSE in the different experiments is the duration of the AV stimuli. Notably, those studies concluding that the auditory stimulus has to precede to visual one for perceived simultaneity, had used much shorter stimuli (<10ms) than the studies showing a visual lead by using stimuli of 40 ms in duration. Thus, longer stimulus durations might contribute to perceptual speed in advantage of the visual system.

In the first part of the work presented here we conducted a psychophysical experiment to investigate the combined effects of stimulus duration and intensity on the temporal integration of AV-cues employing an AV temporal order judgment task (TOJ). Our aim was to further explain the observed variability of PSEs obtained in different TOJs across laboratories. We applied AV-stimuli of different durations (9, 40, and 500ms) and used two different light intensities for each duration. From the literature cited above we expected that an increase in light intensity relative to sound intensity and an increase in stimulus duration would decrease an existing relative lead of the auditory modality, or increase an existing lead of the visual modality. Actually, we found a general lead of the auditory relative to the visual modality. However, the increase of light intensity relative to sound intensity and the increase in stimulus duration indeed independently decreased this lead (Boenke, Deliano, & Ohl, 2007a; Boenke, Deliano, & Ohl, 2007b; Boenke, Deliano, & Ohl, in prep.). Possible sources of this effect will be discussed.

To further resolve the opposing results on multisensory integration, it is important to combine psychophysical experiments with neuroimaging and magneto-/electrophysiological approaches. Physiological analysis might thereby reveal common mechanisms underlying different psychophysical effects, or disentangle different mechanisms underlying similar psychophysical observations. Still, carefully conducted, detailed preparatory psychophysical experiments are highly important to find suitable parameters for a pinpointing electrophysical study. Neuroimaging and magneto-/electrophysiological studies (Bushara, Grafman & Hallett, 2001; for a review: Calvert & Thesen, 2004; review: Driver & Noesselt, 2008; Molholm, Ritter, Murray et al., 2002; Murray, Foxe, Higgins et al., 2001) have already shown that crossmodal integration occurs not only in late stages of processing following the integration within single modalities, but can already take place at very early stages and involve primary sensory cortical areas as has been demonstrated in animal models of audiovisual integration by us (Cahill et al. 1996) and others (for review see Kayser and Logothetis 2007). Thus, crossmodal integration is a dynamic process involving a complex network in the brain. Furthermore, it has been shown that processing of multisensory events influences the processing of a subsequent unimodal event (Meylan & Murray, 2007). Together with psychophysical studies, this suggests that multisensory integration depends on both, bottom-up and top-down processes.

In the second part of the work presented here, we will report preliminary data of a combined psychophysical and electrophysiological study. We applied a similar design as in the first part of the work presented here, and employed a spatial AV-TOJ task. Again, participants were asked, irrespective of the sensory modality, to report where, i.e. on which side, they perceived the onset of the first stimulus. Our psychophysical analysis confirmed Zampini's (2003) observation that in such a spatial TOJ, the bimodal temporal resolution measured by the just noticeable difference (JND) is smaller (higher JNDs) than the unimodal temporal resolution, especially than the visual.
Using an ERP approach, we analyzed as a first step general differences between unimodal and bimodal processing. In many EEG approaches for revealing audiovisual interactions the term AV-(A+V) is used and a non-zero result is interpreted as AV interaction. However, this analysis method has often been criticized, because in case of a third, unknown but unspecific factor "C" in all three conditions, this factor would be overestimated by subtracting it one times more than it is exists in the AV condition (Calvert & Thesen, 2004; Gondan et al., 2005; Teder-Sälejärvi, McDonald, di Russo, & Hillyard, 2002). Similar to the approach by Gondan and colleagues (2005) we avoided this problem by using the term (AV+VA)-(AA+VV).

1.1 Experiment 1 – Role of Stimulus Duration in Audiovisual Temporal Order Judgments

1.1.1 Methods

1.1.1.1 Participants

Twelve healthy participants (6 male) aged between 19 and 32 years (mean 23.6 ± 4.6 SED) took part in the experiment. All participants were right-handed and reported normal or corrected-to-normal vision. They were instructed about the task and gave their written consent. The study was approved by the ethical committee of the Otto-von-Guericke University in Magdeburg conforming to the 1964 Declaration of Helsinki.

1.1.1.2 Apparatus and Stimuli

The experiment was carried out in a dark and sound attenuated room with an ambient noise level of ~29 dB(A). To avoid unintended reflection of light, the setup was covered with a black velvet cloth. A green light-emitting diode (LED, 0.5 cd/m2) was centred in front of the participants at eye-level and a distance of about 165cm, and was used for fixation. Left and right to the centre of the green fixation LED two boxes were placed symmetrically slightly below eye-level. Each of these two boxes contained a sound source (two identical speakers with white coverage) and a light source (white LED). The white LED was placed on top of the speaker. In front of the speaker was an aperture (diameter 4cm). Thus, the sound and an indirect light signal could be presented at the same location in space. The distance between the centre of the aperture of each box and the centre of the fixation LED was approximately 38cm, yielding a visual angle of ~12.5° between the fixation and the side of AV stimulus presentation (Figure 1). The acoustic stimuli consisted of bursts of white noise with 1.5 ms onset and offset ramps. The visual stimuli consisted of flashes emitted by the white LEDs. Responses could be given via a specific designed hand-held response box. Stimulus presentation and recording was controlled by a program written with the psychtoolbox (PT-2) in the Matlab R14 environment, which generated the signals using a National Instruments card (PCI-6071E) with a presentation accuracy better than 1ms and stored the recorded responses to disk on an IBM 486-compatible microcomputer.

1.1.1.3 Design

We applied a TOJ-task using the method of constant stimuli. In the literature there generally exist two different types of TOJ tasks. Participants can either be asked to report the modality they perceived first, when light and sound are presented at a single or two separate spatial locations (modality AV-TOJ; see Zampini et al., 2003 for comments on spatial confounds in earlier work when using two spatially different sources of stimulation), or, alternatively, can be asked to report the location of first stimulus on- or offset when light and sound are delivered from two spatially different sources (spatial AV-TOJ). Here we decided for the latter task and asked the participants to report on which side (left or right) the first stimulus occurred avoiding a bias towards a modality (see Spence et al., 2001; Zampini et al., 2003). In each trial, one acoustic (A) and one visual (V) stimulus was presented (AV-stimulus pair). The first stimulus was presented either on the left or the right side, and the second stimulus on the other side, respectively (Figure 2). If the first stimulus was acoustic the second stimulus was visual, and vice versa. The applied AV stimuli had the following stimulus onset asynchronies (SOAs): -240ms, -140ms, -100ms, -60ms, -20ms, 20ms, 60ms, 100ms, 140ms, and 240ms(Figure 2). Negative values indicated that the acoustic stimulus onset occurred first. There were two orthogonal factors: duration (9ms, 40ms, and 500ms) and intensity (high intensity versus low intensity of the visual signal) (Figure 3). The first two durations were chosen in accordance with former studies (see above), and the longest duration of 500 ms was added as a control for temporal summation effects. The intensity of the noise-bursts was kept constant at 49.8 dB(A). For the light-flashes an intensity of 0.64 cd/m2 was chosen for the high intensity (VH) condition, and 0.14 cd/m2 for the low intensity (VL) condition. Stimulus intensities had been previously determined in a pilot-study to obtain a good level of comfort for the participants with respect to the 500 ms light-flash of high intensity, and the 500 ms noise-burst. Stimuli were all supra-threshold.
Each possible stimulus configuration was repeated 20 times, with configurations presented in pseudo randomized fashion, in such a way that no more than 3 identical configurations occurred in succession. The experiment was performed on one day, and divided into 4 experimental blocks yielding a total of 2400 trials (5 SOAs [±240ms, 140ms, 100ms, 60ms, 20ms] x 2 stimulus first [A/V] x 2 side [left/right] x 3 durations [9ms, 40ms, 500ms] x 2 intensity [A-VL/A-VH] x 20 repetitions).

1.1.1.4 Procedure

The green fixation LED was constantly illuminated throughout the experiment and participants were instructed to maintain their view on the LED during the stimulus presentation. In each trial, participants were asked report on which side (left or right) they perceived the first out of two stimulus onsets by pressing the button of a handheld response device on the corresponding side with their left or right thumb. Participants were asked for accuracy but not to wait too long for a response. Further, they were instructed to make their best guess in cases of uncertainty about the correct sequence. The first trial started 4-5 seconds after a warning signal. The following trials were self paced. After a response given by pressing one of the defined response buttons, an intertrial-interval was initiated with a duration between 1.5 s and 2 s chosen randomly from a uniform distribution.
Each experiment was preceded by an approximately 10 minute long training block, which also served as adaptation phase for the participants. In this training block the participants were asked to indicate their response verbally and feedback was given by the instructor. To avoid any immeasurable technical intensity bias for the shortest duration between the left and the right AV-source, the boxes containing loudspeakers and LEDs were swapped in half of the participants.

1.1.2 Results

There was no response bias towards the left or the right side, as no significant differences could be found between the proportions of correct responses on the left and the right side.
For every participant and for each of the 6 conditions (2 light intensities x 3 durations), psychometric functions were calculated for the "imaginary" response "vision first". Thereby, two separate psychometric functions were calculated for stimulus presentation on the left and the right side, respectively. As evidence for similar response probabilities on the left and the right side, a rank correlation between corresponding psychometric functions for the left and the right side was calculated at a significance level of $p<0.01$. Two participants failed to obtain significant rank correlation indicating a perceptual bias towards one side. Because in these participants psychometric functions were also flat which further indicates an overall bad task performance, they were excluded from further analysis.
To increase the number of observations and to estimate the PSE independently from the side of stimulus presentation, the data of the left- and the right-side psychometric functions were finally collapsed in the remaining participants. For determining the PSE, the resulting single psychometric function (with in total of 40 observations for each data point) was estimated using a Bayesian inference procedure after Kuss et al. (2005). An example of the psychometric functions for the different experimental conditions obtained this way is shown in Figure 5 for one participant. For all participants and experimental conditions the PSEs determined from these psychometric functions are displayed in Figure 4. The mean and the standard error of the PSEs across subjects for each experimental condition are given in Table 1. For the determined PSEs, a two-way between participants ANOVA with the main factors "Duration" (9ms, 40ms, and 500ms) and "Intensity" [high (VH) versus low (VL)] was calculated. Both main effects could be statistically verified: the low intensity condition (VL) and the high intensity condition (VH) differed by 36ms whereas the mean PSE across all durations in the VL condition (mean: 75ms) was significantly more shifted in favor of the auditory modality (visual stimulation has to occur earlier in order to be perceived as simultaneous) than the mean PSE across all durations in the VH condition (mean: 39 ms)

[Intensity: $F(1,9)=95.4$, $p<.001$]. For the duration conditions the effect was smaller but statistically reliable [$F(1,9)=4.5$, $p<.05$]. The longer the duration the more the PSE was shifted in favor of the visual modality (visual stimulation has to occur less early in order to be perceived as simultaneous). In the VL- and the VH condition, PSEs were shifted 12-14ms for the step from 9ms to 40ms duration, and 3-4ms for the step from 40ms to 500ms duration (see Figure 5 and Table 1). No interaction was found [$F(1,9)=.07$, $p=0.94$] suggesting independence of the intensity and the duration effect.

1.1.3 Discussion

By applying an orthogonal design in an AV-TOJ task we addressed the question, whether the duration or the relative intensity of applied stimuli in an audiovisual temporal order judgment task (AV-TOJ) has any influence on the perception of their sequence. Our results show that intensity and duration play a role in the temporal perception of AV-stimuli pairs. We found a strong effect for relative intensity, which is in accordance with results reported before (Neumann and Niepel, 2004). However, we measured a positive PSE of in the range of 32-84ms indicating a temporal lead of the auditory modality (see Table 1). Looking closer to the intensities of the Neumann et al. study (1992) their medium intensities is in the same range as our low intensity condition (0.1 cd/m2 and 48 dB (A) versus 0.14 cd/m2 and 49.8 dB (A)). However, whereas Neumann et al. (1992) found a negative PSE of ~ -15ms (estimated from their Figure 3) we found in our study the PSE for stimuli with 40ms duration to be at 72ms. This might be in part explained by the fact that visual stimuli presented in the periphery like in our study are processed slower compared to foveal presentation, as it was the case in the study of Neumann et al. (1992). Also the size of our visual stimuli was smaller, which could also contribute to the observed differences. Furthermore, it is well known that there exist high interindividual differences in TOJ-tasks. As in our study (n=10) and in the study of Neumann et al. (1992) the number of participants (n=6) was small, this could also explain the differences. This suggests that besides intensities and durations other factors, neither controlled in our, nor in Neumann's study, play a role in temporal AV integration.
Most importantly, we could demonstrate for the first time that different durations of AV-stimuli have an influence on the perception of their temporal order. Although the observed duration effect seems to be less pronounced than the intensity effect it is statistically reliable for the set of parameters chosen. The question remains, why the PSE shifts in favor of the visual modality with longer stimulus durations, and not in favor of the auditory modality? To answer this question we took closer look on the level of single participants and their individual perception. As shown in Figure 6, in most of the participants, the visual stimulus had to be presented long before the auditory stimulus in order to be perceived as simultaneous. This lead of the auditory modality generally decreased when the duration of AV stimulus pairs was elongated. However, in one single participant (denoted G) the opposite was true: the acoustic stimuli had to precede the visual, in order to evoke the impression of simultaneity. Furthermore, with longer AV-stimulus the necessary lead of the visual stimulus decreased. This raises the interesting question, whether instead of a modality-specific universal pattern, the observed changes in the PSE rather reflect the compensation of delays towards a PSE at 0ms, irrespective of the particular sensory modality.
To shed light on this question we designed a follow-up experiment which will be carried out in the near future. The rationale behind this experiment is to employ the same design as in the current experiment, but to reduce the auditory intensity significantly. The aim is to reverse the

observed pattern: "audition leads vision", i.e. to obtain a significant number of cases, in which the auditory stimulus has to precede the visual stimulus for perceived simultaneity. Then we could investigate the direction of PSE shifts with a change in duration and intensity, and determine whether this shift is modality specific or acts for the compensation of the existing delay.

1.2 Experiment 2 – Preliminary Electrophysiological Data

1.2.1 Methods

1.2.1.1 Participants

Five healthy participants (1 female) aged between 23 and 29 years (mean 25.8 ± 2.7 SED) took part in the experiment. All participants were right-handed and reported normal or corrected-to-normal vision. Participants were instructed about the task and gave their written consent. The ethical committee of the Otto-von-Guericke University in Magdeburg had given its permission for the study, and conformed to the 1964 Declaration of Helsinki.

1.2.1.2 Apparatus, Stimuli, Design and Procedure

The apparatus, stimuli, design and procedure were the same as in the psychophysical experiment reported above with minor changes as follows: To obtain a reasonable signal to noise ratio for an ERP analysis, the number of collected trials were increased to 60 per condition. Therefore, to keep the overall time of the experiments within limits, only one duration (9ms) and a fixed intensity (audio: 60dB(A); video: 0.64 cd/m2) was selected. Also, only 3 stimulus onset asynchronies (SOAs) were presented (20ms, 55ms, 90ms). Additionally to the crossmodal conditions, two unimodal conditions, one with visual and one with acoustic stimulation alone, were added. This resulted in a total number of 24 experimental conditions. The experiment was carried out in one day, and was divided into 6 Blocks with 10 repetitions per experimental condition (lasting about 12 minutes each giving a total of 1440 trials per participant (3 SOAs [±90ms, 55ms, 20ms] x 4 conditions [AV, VA, VV, AA] x 2 side [left/right] x 60 repetitions).

1.2.1.3 Electrophysiological Recording

EEG was recorded from 61 Ag+/AgCl--electrodes according to the international 10/10 system using high input impedance amplifier (10MΩ, BrainAmp, Brain Products GmbH, Munich, Germany) and electrodes mounted on an electrode cap (M 11, FMS, Munich, Germany). Eye blinks were monitored using two additional electrodes placed below and next to the right eye. An additional electrode was placed on the nose tip and served for possible re-referencing. Electrodes AFz and FCz were used as ground and physical reference, respectively. Electrode impedances were reduced below 6 kΩ before data acquisition. Data were recorded with 0.1 Hz high-pass and 100 Hz low-pass filter. Data were digitized at 1000 Hz.

1.2.1.4 Data Analysis

The EEG-signal was filtered using a Butterworth zero-phase filter with a low cut-off frequency of 0.3 Hz, 24 dB/oct and a high cut-off frequency of 30 Hz, 24 dB/oct. Epochs of 1024 ms length were generated including a baseline of 200 ms before stimulus onset. Using a

semi-automatic artifact rejection procedure, we excluded epochs with absolute voltage difference exceeding 15 µV between two neighboring sampling points and with amplitude outside +70 or -70 µV. Epochs with eye blinks were discarded. The rejected epochs were evenly distributed across the 24 experimental conditions. For a preliminary analysis, the 24 experimental conditions were combined such that they formed two main groups: bimodal (AV and VA) and unimodal (AA and VV). The data were re-referenced to the nose electrode and corrected to the 200 ms baseline preceding stimulus onset. In the next step, difference waves were calculated by subtracting the unimodal from the bimodal condition in each participant.
A grand average of the unimodal, and the bimodal condition, and of the difference between these conditions was calculated across all 5 participants. As a first orientation, differences between the bimodal and the unimodal condition were identified by inspecting the topography and the time courses of the voltage maps of the grand average of the difference waves. In a next step, by using the voltage maps of the unimodal and bimodal condition, electrodes corresponding to local maxima or local minima within the identified region of maximum differences were identified. By this method we determined the following deflections of interest (in brackets are the time point of the local extreme reported and the chosen time window within which the mean activity was calculated): Vision: N1 (peak: 202ms, time window: 185-215ms) with minimum at electrode P7, N2 (peak: 260ms, time window: 240-280ms) with minimum at PO7. Audio: N1 (peak: 96ms, time window: 90-110ms) with minimum at FCz, P2 (peak: 229ms, time window: 210-250ms). Finally, a late difference between the bimodal and the unimodal condition could be identified at parietal electrode-sites reaching its maximum at Pz at 543ms (time window: 400-700ms). If not mentioned otherwise the statistics is derived from a paired t-test between the unimodal and the bimodal condition.

1.2.2 Results

1.2.2.1 Behavioral Data

To compare our results with Zampini et al. (2003), we also converted the proportions of "right first" responses at intermediate SOA to equivalent z-scores and fitted an affine function to the data (for a grand average of the behavioural results see Figure 13). We compared the condition visual same (Vs), visual different (Vd), auditory same (As), and auditory different (Ad). Unlike Zampini and colleagues (2003), we found no differences between the points of subjective equalites (PSEs) by applying a paired t-test.
Analyzing the just noticeable differences (JND) by calculating the slopes of the linear regression lines we found that the experimental condition visual same (Vs) could be better distinguished in their temporal order than all other conditions (Vs versus Vd, $t(1,4)=3.86$, $p<0.05$, Vs versus Ad, $t(1,4)=3.5$, $p<0.05$, and Vs versus As, $t(1,4)=5.53$, $p<0.01$; see Figure 13). This result is in line with the results from Zampini and colleagues.

1.2.2.2 Electrophysiological Data

Visual components

N1

The audiovisual interaction found on this component was marginally significant. The N1 amplitude was larger in the unimodal condition (-3.32 ± 1.66 µV) compared to the bimodal condition (2.39 ± 1.87 µV, t(1,4)=-2.74, $p = 0.05$, see Figure 7).

N2

We could demonstrate a strong audiovisual interaction on the visual N2 component. The amplitude was larger in the unimodal condition (-2.85 ± 3.19 µV) compared to the bimodal condition (-0.64 ± 2.65 µV, t(1,4)=-5.24, $p < 0.01$, see Figure 8).

Auditory components

N1

With the currently small number of participants we were not able to validate statistically the audiovisual interaction on the auditory N1 component. However, in opposition to the visual N1 component the amplitude in the bimodal condition (-2.78 ± 2.91 µV) was more negative than in the unimodal condition (-1.82 ± 1.86 µV, t(1,4)=-1.95, $p = 0.12$, see Figure 9).

P1

The auditory P1 amplitude was significantly larger in the bimodal condition (4.1 ± 1.25 µV) compared to the unimodal condition (2.14 ± 0.88 µV, t(1,4)=-3.51, $p < 0.05$, see Figure 10).

Late modulation

P3

We could statistically validate an audiovisual interaction on the late P3 component. The maximally activation was at a parietal site (electrode Pz). The amplitude was larger in the unimodal condition (7.89 ± 2.37 µV) compared to the bimodal condition (6.43 ± 2 µV, t(1,4)=3.28, $p < 0.05$, see Figure 11).

1.2.3 Discussion

Comparing our behavioural results to the study of Zampini et al. (2003) we found a similar result with respect to the JNDs, i.e. that the unimodal visual condition (Vs) tends to have a temporal higher resolution than all other conditions when embedded in a spatial-TOJ paradigm. However, whereas in the Zampini et al. (2003) study most of the participants showed the pattern that the visual stimulus had to precede the acoustic stimulus for perceived simultaneity, we could not confirm this observation. This might be due to our small group size, or other effects discussed in the experiment of part 1.

In a preliminary electrophysiological analysis of our spatial audiovisual (AV) TOJ task we could reveal AV interactions at various stages of the processing stream in the auditory as well as the visual modality, but also at later stages. To our knowledge, similar electrophysiological correlates in an AV-TOJ task have not been shown before. Though, Teder-Sälejärvi et al. (2005) show effects of auditory cueing in a visual TOJ task, investigating the prior entry effect and Khallafalla et al. (1999) used fMRI as method of choice, and Meylan and Murray (2007) compared the effect of single flashes on a subsequent multisensory event. Because our analysis is preliminary, it is too early to draw strong conclusions about the modulation of AV processing we have found. Still, our finding of a modulation on early components already indicates different discrimination processing for unimodal and bimodal stimulation. The modulation on the visual N2 further might be explained by a N2pc indicating spatial attention or rapid shifts of spatial attention in a visual task (Woodman & Luck, 1999). The N2pc component further seems to be a promising tool for investigating attentional factors within spatial AV-TOJ tasks. Finally, this experiment must be seen as a first step in developing an EEG design to uncover the neuronal mechanism underlying the effects of stimulus intensity and duration in a spatial AV-TOJ task described in the first part of this work (Boenke et al., 2007a; Boenke et al., 2007b; Boenke et al., in prep.).

1.3 Figures and Tables

Table 1: In the left column the experimental conditions are shown [Low intensity (VL) and high intensity (VH) and their durations]. In the right column the corresponding means (±STE) of the estimated point of subjective simultaneity (PSE) are given.

Exp Cond	PSE ± STE
VH 9 ms	50 ±15
VH 40 ms	36 ± 11
VH 500 ms	32 ± 13
VL 9 ms	84 ± 19
VL 40 ms	72 ± 15
VL 500 ms	69 ± 12

Table 1. Means of the estimated PSE for all durations and intensities (±STE).

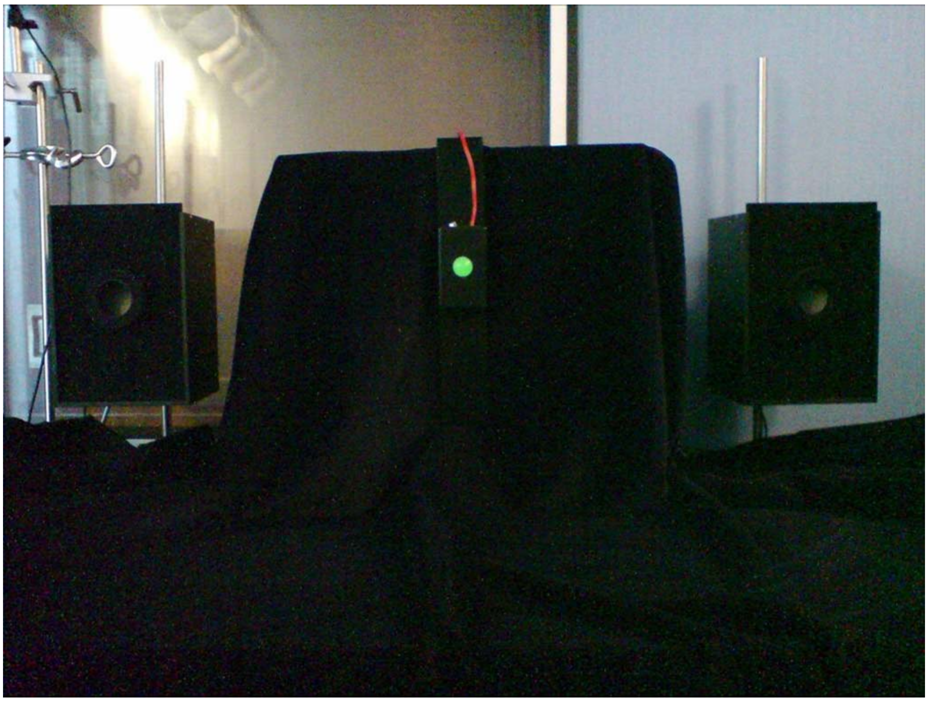

Figure 1: Experimental set up with the fixation help in the center (green LED) and two boxes with a light and sound source inside.

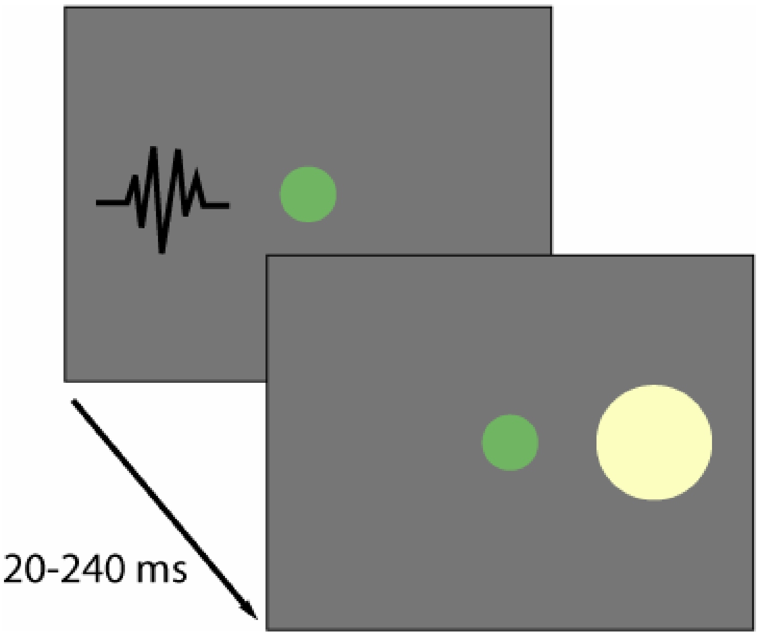

Figure 2: AV stimulus pairs were separated by a 20-240 ms SOA.

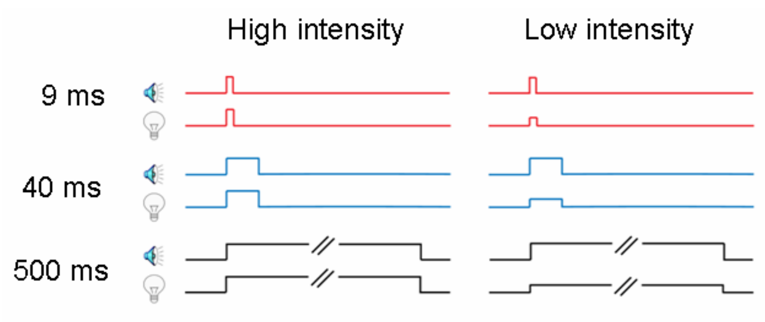

Figure 3: Orthogonal design of the experiment. AV stimulus pairs were separated by SOAs ranging from 20-240ms.

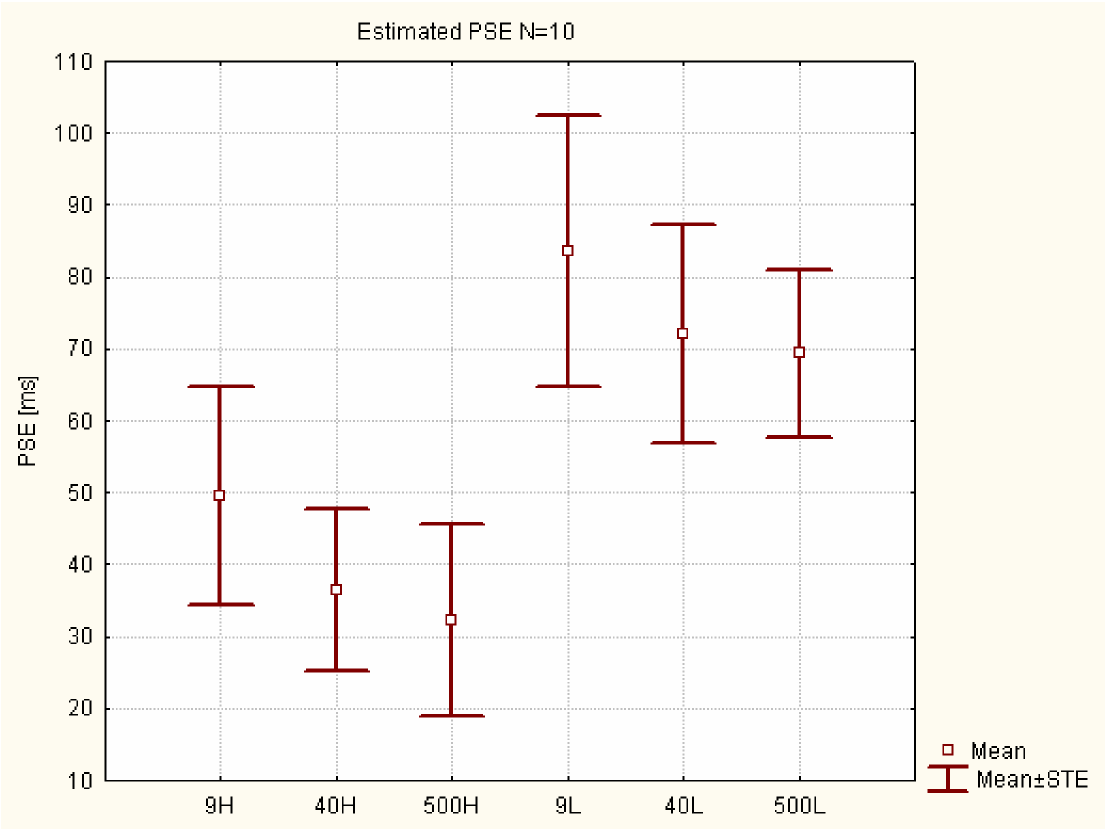

Figure 4. Estimated PSEs (mean ± STE) for all six experimental conditions. A clear effect of the relative intensities (L: low intensity, H: high intensity of the visual stimulus) on the PSE is revealed. Also, for a fixed intensity relationship the PSE significantly decreased with increasing stimulus duration.

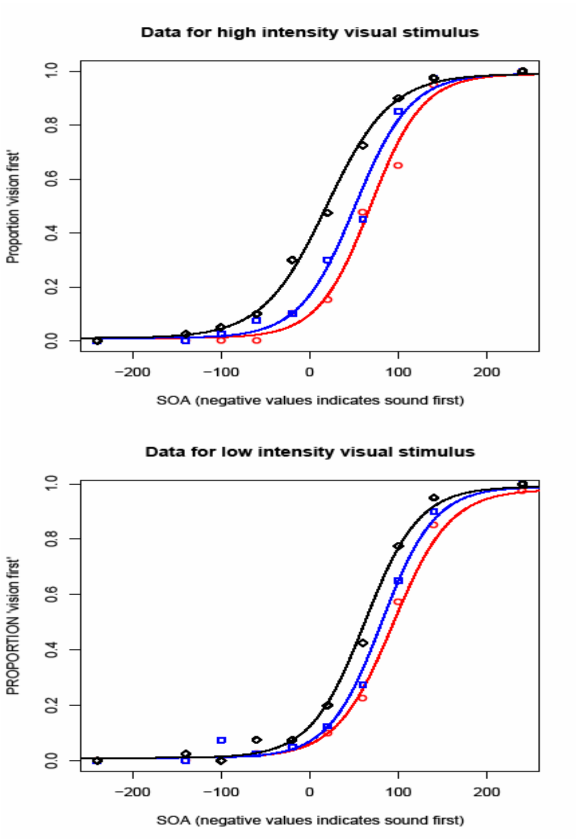

Figure 5: Exemplary estimated psychometric functions for one participant (participant A in Figure 6). The top and bottom panels depict the data for the high intensity condition and low intensity condition, respectively. In each panel data are given for stimulus duration 9ms (red), 40ms (blue), and 500ms (black).

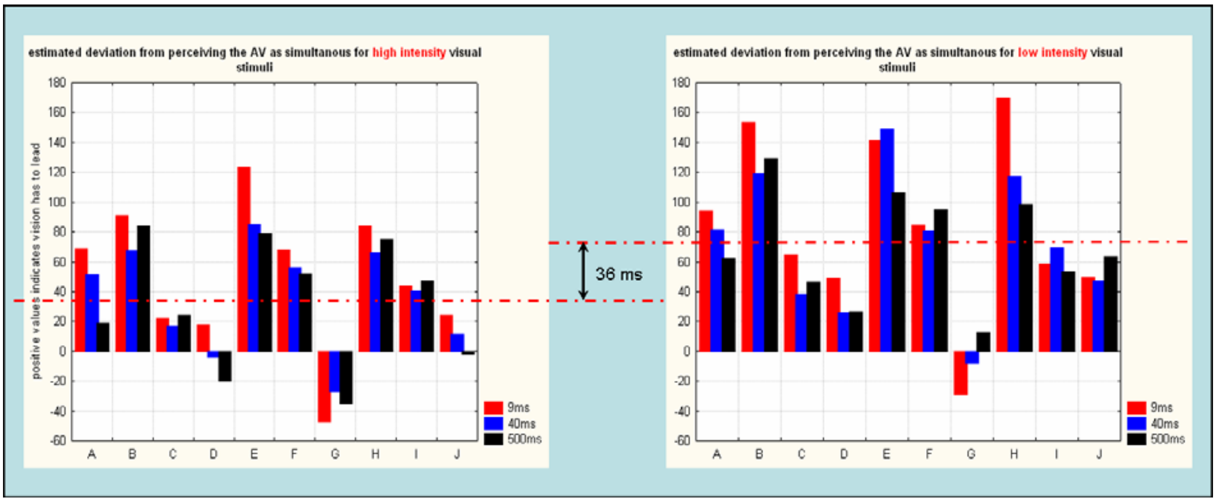

Figure 6. Estimated PSEs for individual participants. Left and right panels show the PSEs in the high-intensity and low intensity condition, respectively. The red dotted lines visualize the mean intensity effect over participants (36ms) indicating that the necessary visual lead for perceived simultaneity is smaller in the high-intensity condition. While for most participants the visual stimulus had to lead the auditory stimulus (PSE>0) an opposite pattern was found for subject G.

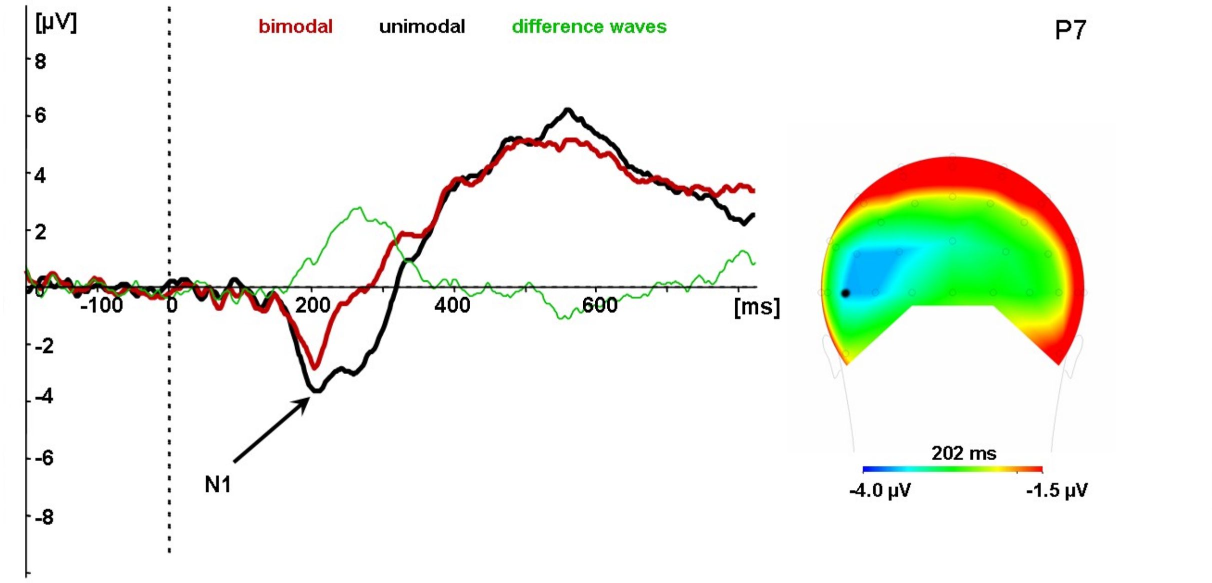

Figure 7. Grand mean local maximum of activation at electrode P7. The left panel shows ERP traces for the bimodal (AV+VA red) and unimodal (AA+VV black) conditions as well as the difference ((AV+VA)-(AA+VV) green). The right panel shows the voltage map at the time of maximum local activation.

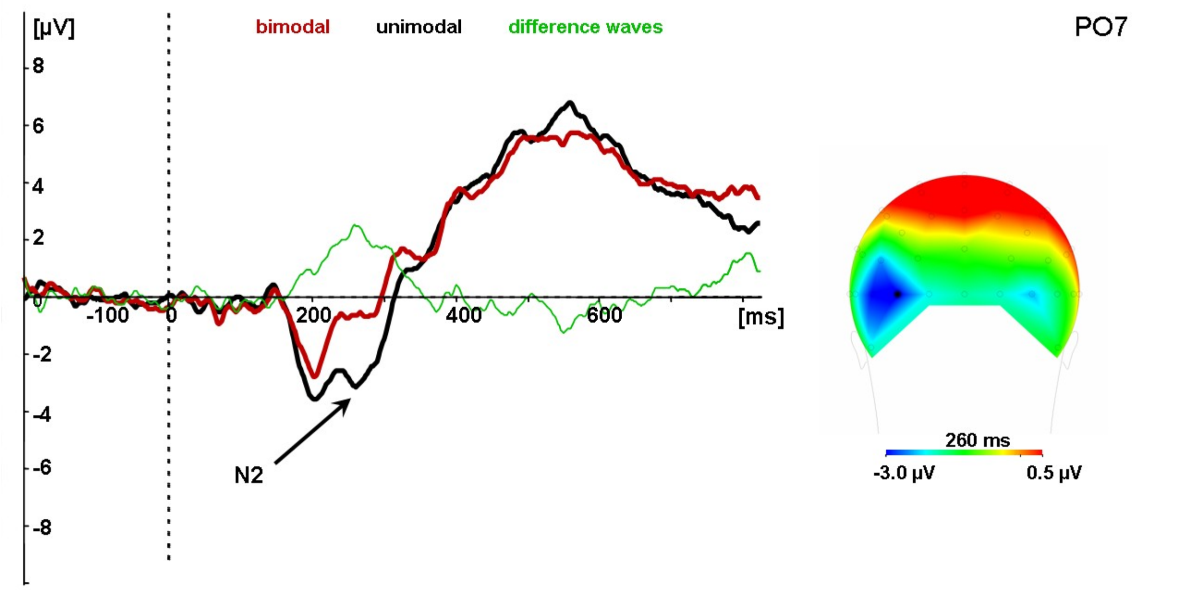

Figure 8. Grand mean local maximum activation at electrode PO7. Figure organization and labels as in Figure 7. The neighboring electrode PO8 (not shown) reached maximum activation slightly later.

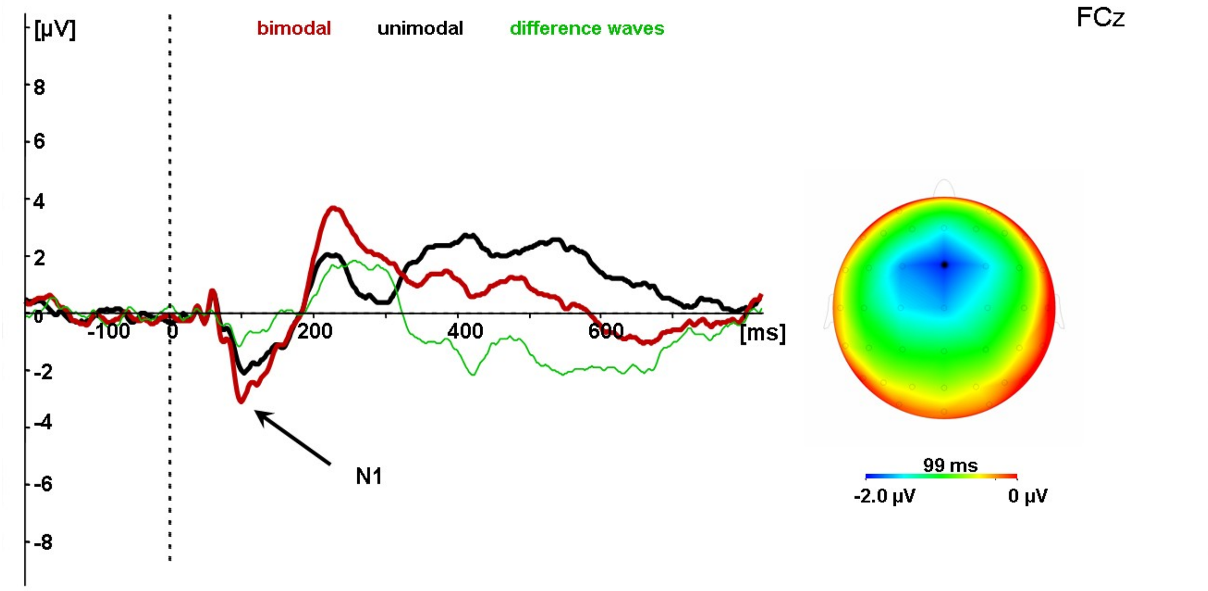

Figure 9. Grand mean local maximum activation at electrode FCz. Figure organization and labels as in Figure 7. Here the auditory-evoked ERP component N1 is shown. With respect to the comparison (AV+VA)-(AA+VV), this component did not reach significance (based on the 5 measured participants).

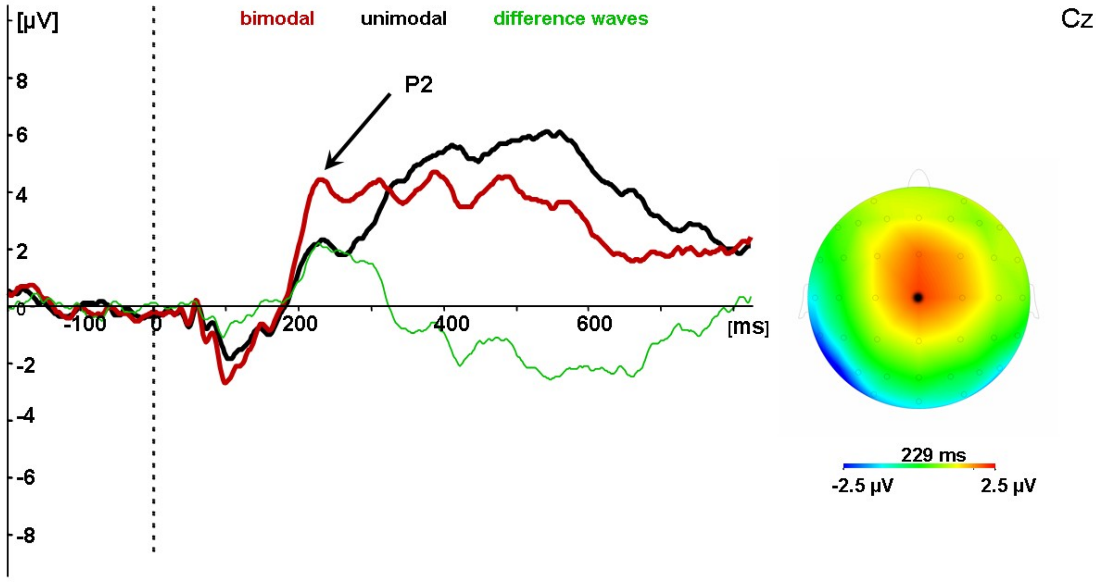

Figure 10. Grand mean local maximum activation at electrode Cz. Figure organization and labels as in Figure 7.

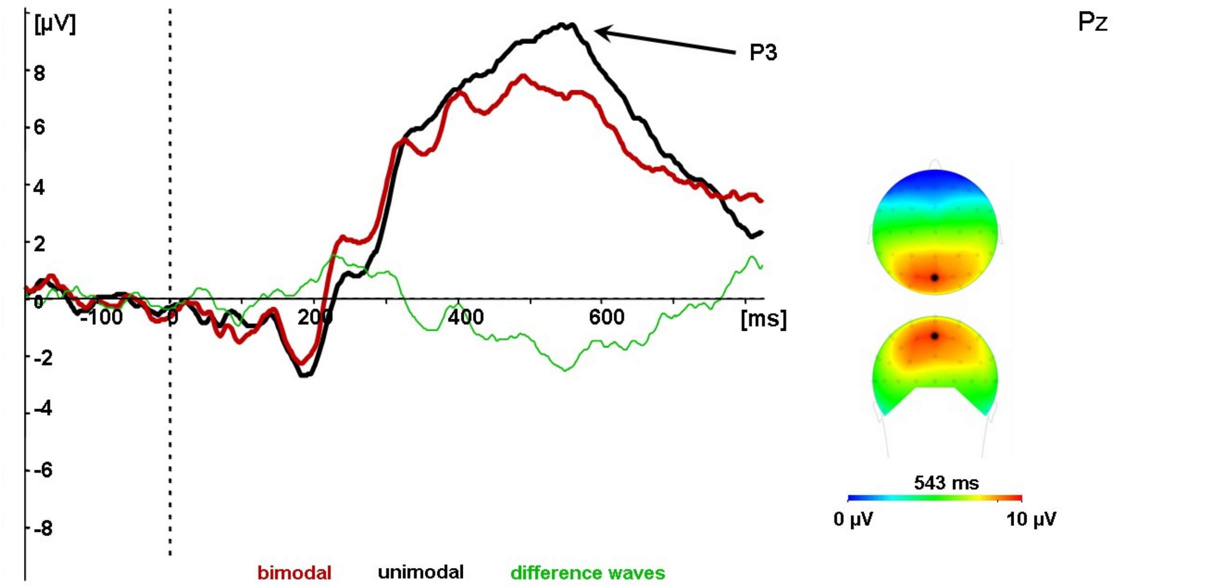

Figure 11. Grand mean local maximum activation at electrode Pz. Figure organization and labels as in Figure 7. The suppressed activation at classical P3 sites for the bimodal condition could reflect a higher degree of uncertainty of perceptual categorization which was also reported by the subjects for these stimuli.

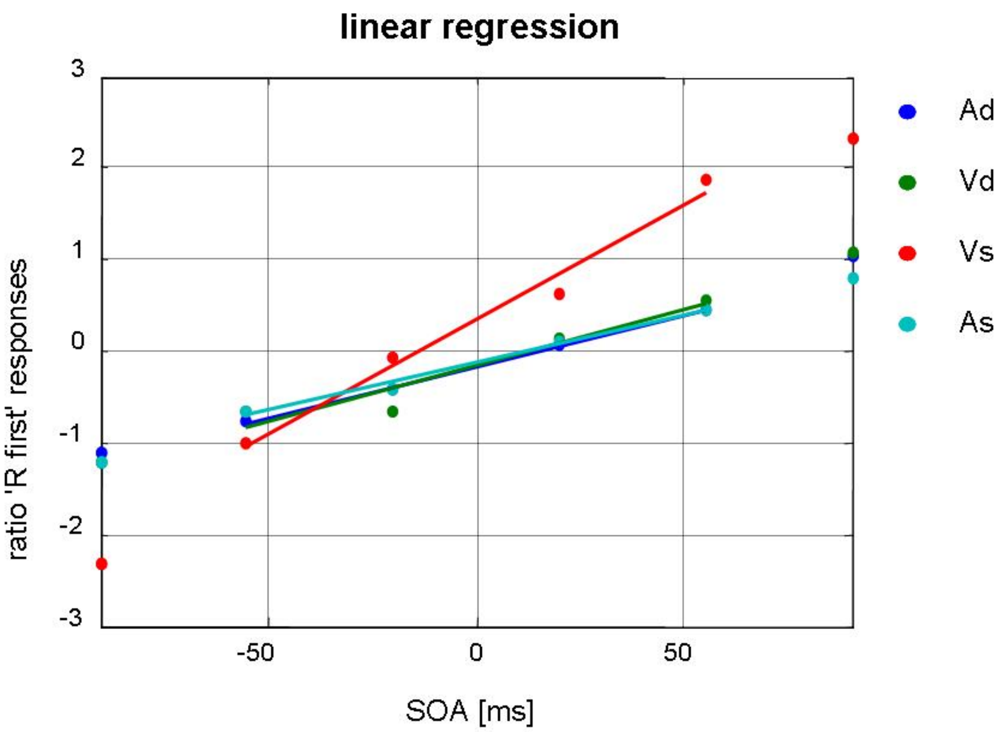

Figure 12. Mean proportion of "right-first" responses converted to equivalent z-scores as a function of stimulus onset asynchrony (SOA) shown together with linear (affine) fits to the z-transformed scores in each condition. Negative SOA values indicate that the stimulus presented from the left side was first. The labels correspond to "experiment 2" in Zampini et al. (2003), i.e. "A" and "V" indicate the modality of the stimulus presented on the left side, "s" and "d" indicates whether the two stimuli were presented from the same or opposite ("different") position. For the limited data set (5 participants) condition Vs differed significantly from all others with respect to JND (regressed slope) but not PSE.

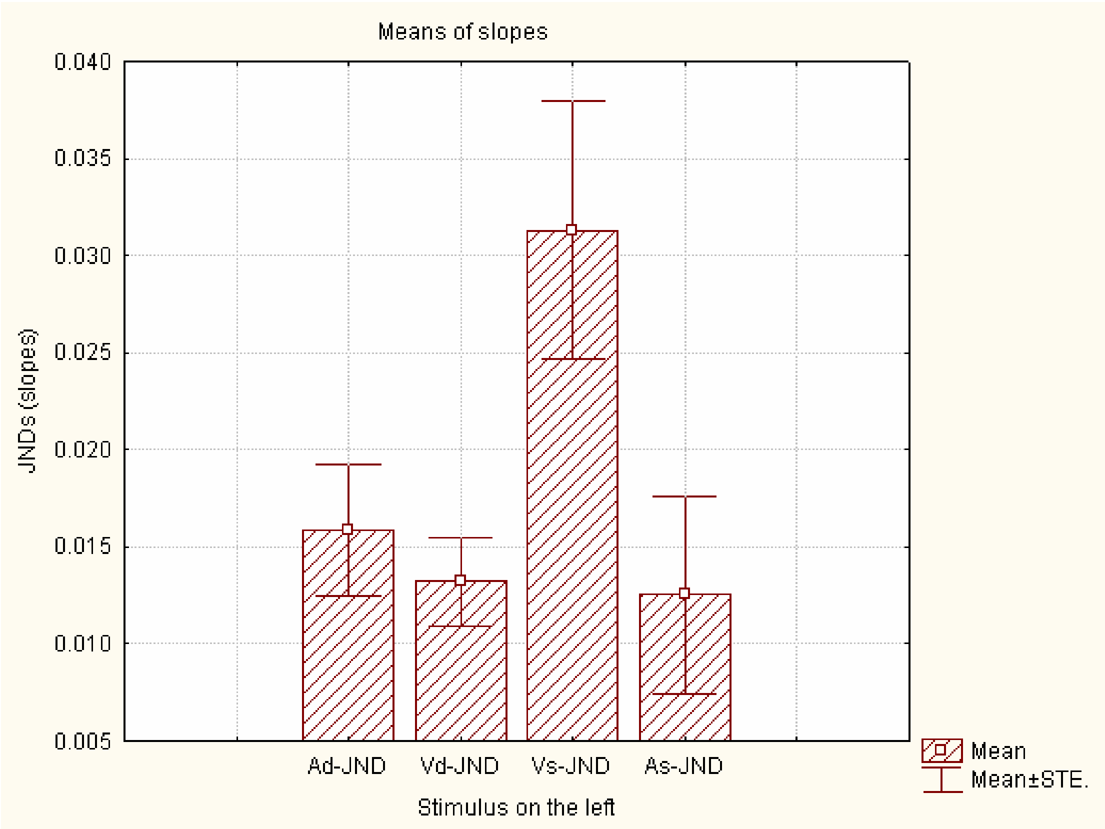

Figure 13. Just noticeable differences for the 5 participants and the 4 stimulus pairings. The JND was significantly higher in the Vs condition compared to all others. This pattern is similar to the results by Zampini et al. (2003).

2. Conclusions

It is currently debated which parameters of auditory and visual stimuli influence the perception of temporal order when such stimuli are presented in close temporal proximity. Previous research has demonstrated that the relative spatial locations and relative intensity relationships have an influence on (1) the perceived temporal order and on (2) the necessary temporal order to achieve perception of simultaneity. Here, using an orthogonal design embedded in an audiovisual temporal-order judgment task, we studied the influence of relative intensities and stimulus duration on temporal-order judgement. We could replicate the known intensity effect and discovered an often overlooked additional confounding stimulus parameter, stimulus duration. Identical increase of duration of auditory and visual stimuli (all other parameters held constant) led to a shift of PSE to smaller values (in the extreme case to negative values).
We have also begun to combine a classical temporal-order judgment paradigm with the simultaneous recording of scalp EEG. A preliminary analysis of ERPs revealed audiovisual interaction affects brain activity at both early and late processing stages. The fact that an N2pc component could be demonstrated in our experiment, indicates the possibility for further investigation of how processing of visual stimulus might be modulated in a top-down fashion by processes evoked from stimuli of another sensory modality (e.g. audition).

3. Outlook

With respect to our psychophysical studies we have planned two further experiments. The first one is designed to determine if the described shift of PSE to smaller values with increasing stimulus duration, seen in most subjects, is a general feature of audiovisual stimulus processing or, alternatively, whether this pattern reflects some mechanism by which the delay between temporally separated stimuli is compensated towards PSE values closer to zero - a theoretical possibility for which the data from subject G provides some indication. Such a hypothetical mechanism could employ longer stimulus durations quite generally to perceptually "attract" temporally separated stimuli in time, or it could specifically "accelerate" the processing of the sensory modality which under the specific parameter constellation is less rapidly perceived.
A second experiment addresses the current debate about discordant results from temporal-order judgements and reaction time measurements, respectively, for the hypothesized underlying physiological processes of audiovisual integration. The aim here will be to determine which physiological processes correlate better with temporal-order judgement and which better with determinants of reaction time.
With respect to our electrophysiological studies we will continue measurement and analysis of the presented preliminary data. We will place additional focus on how the predictability of correct responses based on observation of the contralateral N2pc component, known from visual search studies, is influenced by audiovisual interaction.

4. References

Lars T Boenke, Matthias Deliano, and Frank W Ohl. Influence of stimuli duration in audiovisual temporal order judgement. *In prep*.

Lars T Boenke, Matthias Deliano, and Frank W Ohl. Interstimulus interaction in an audio-visual temporal order judgment task: The role of offset of the preceding stimulus and stimulus length on temporal perception. *Proceedings of the 8th Annual Meeting of the International Multisensory Research Forum (IMRF)*: Sydney. 2007a.

Lars T Boenke, Matthias Deliano, and Frank W Ohl. Role of stimulus length, intensity, and offset-onset interaction in audiovisual temporal order judgement. [Abstract]. In J Grainger, F-X Alario, B Burle, and N Janssen (Eds.), *Proceedings of the 15th meeting of the European Society for Cognitive Psychology*: p. 107. Marseille: ESCoP. 2007b.

Khallafalla O Bushara, Jordan Grafman, and Mark Hallett. Neural correlates of auditory-visual stimulus onset asynchrony detection. *J Neurosci* **21(1)**: 300-304. 2001.

Bjoern Bonath, Toemme Noesselt, Antigona Martinez, Jyoti Mishra, Kati Schwiecker, Hans-Jochen Heinze, and Steven A Hillyard. Neural basis of the ventriloquist illusion. *Curr Biol* **17(19)**:1697-1703. 2007.

Larry Cahill, Frank Ohl, and Henning Scheich (1996). Alteration of auditory cortex activity with a visual stimulus through conditioning: a 2-deoxyglucose analysis. *Neurobiol Learn Mem.* **65(3)**: 213-22.

Gemma A Calvert and Thomas Thesen. Multisensory integration: methodological approaches and emerging principles in the human brain. *J Physiol Paris* **98(1-3)**: 191-205. 2004.

John Driver and Toemme Noesselt. Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron* **57(1)**:11-23. 2008

Sigmund Exner. Experimentelle Untersuchung der einfachsten psychischen Prozesse. III. Abhandlung. Der persönlichen Gleichung zweiter Teil [Experimental examination of the most simple psychological processes. III. Treatise Second part of the personal equation]. *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Thiere* **7**: 601-660. 1875.

Waka Fujisaki, Shinsuke Shimojo, Makio Kashino, and Shin'ya Nishida. Recalibration of audiovisual simultaneity. *Nat Neurosci* **7(7)**: 773-778. 2004.

JW Gebhard and GH Mowbray. On discriminating the rate of visual flicker and auditory flutter. *Am J Psychol* **72**: 521-529. 1959.

Matthias Gondan, Birgit Niederhaus, Frank Rösler, and Brigitte Röder. Multisensory processing in the redundant-target effect: a behavioral and event-related potential study. *Percept Psychophys* **67(4)**:713-726. 2005.

James VM Hanson, James Heron and David Whitaker. Recalibration of perceived time across sensory modalities. Exp Brain Res [Epub ahead of print]. 2008.

Vanessa Harrar and Lawrence R. Harris. The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity. *Exp Brain Res* [Epub ahead of print]. 2008.

Piotr Jaśkowski. Simple reaction time and perception of temporal order: dissociations and hypotheses. *Perceptual and Motor Skills* **82**: 707-730. 1996.

Piotr Jaśkowski, Feliks Jaroszyk, and Dorota Hojan-Jezierska. Temporal-order judgments and reaction time for stimuli of different modalities. *Psychological Research* **52**: 35-38. 1990.

Alan Johnston and Shin'ya Nishida. Time perception: Brain time or event time? *Current Biology* 11: 427-430. 2001.

Christoph Kayser and Nikos K. Logothetis. Do early sensory cortices integrate cross-modal information? *Brain Struct Funct* **212:** 121-132.

Malte Kuss, Frank Jäkel, and Felix A Wichmann. Bayesian inference for psychometric functions. *Journal of Vision*: 478-492. 2005.

Harry McGurk and John MacDonald. Hearing lips and seeing voices. *Nature* **264(5588)**:746-748. 1976.

Raphaël V Meylan and Micah M Murray. Auditory-visual multisensory interactions attenuate subsequent visual responses in humans. *Neuroimage* **35(1)**: 244-254. 2007.

Jyoti Mishra, Antigona Martinez, Terrence J Sejnowski, and Steven A Hillyard. Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *J Neurosci* **27(15)**: 4120-31. 2007.

Sophie Molholm, Walter Ritter, Micah M Murray, Daniel C Javitt, Charles E Schroeder, and John J Foxe. Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res Cogn Brain Res* **14(1)**: 115-28. 2002.

Micah M Murray, John J Foxe, Beth A Higgins, Daniel C Javitt, and Charles E Schroeder. Visuo-spatial neural response interactions in early cortical processing during a simple reaction time task: a high-density electrical mapping study. *Neuropsychologia* **39(8)**: 828-44. 2001.

Odmar Neumann and Michael Niepel. Timing of „perception“ and perception of „time“. In: C Kärnbach , E Schröger, H Müller (eds). Psychophysics Beyond Sensation: Laws and Invariants of Human Cognition. Erlbaum, Mahwah NJ: pp.245-269. 2004.

Odmar Neumann, Robert Koch, Michael Niepel, and Thomas Tappe. Reaktionszeit und zeitliches Reihenfolgeurteil: Übereinstimmung oder Dissoziation? [Reaction time and temporal order judgment: Correspondance or Dissociation]. *Zeitschrift für Experimentelle und Angewandte Psychologie* **39**: 621-645. 1992.

Michael Niepel and Odmar Neumann. A dissociation between temporal order judgment and reaction time: Evidence for direct parameter specification. *Unpublished manuscript.* 2002.

MI Posner, CR Snyder, and BJ Davidson. Attention and the detection of signals. *J Exp Psychol* **2**: 160-174. 1980.

Gregg H Recanzone. Auditory influences on visual temporal rate perception. *J Neurophysiol* **89(2)**: 1078-1093. 2003.

J Rutschmann and R Link. Perception of temporal order of stimuli differing in sense mode and simple reaction time. *Perceptual and Motor Skills* **18**: 345-352. 1964.

Ladan Shams, Yukiyasu Kamitani, and Shinsuke Shimojo. Visual illusion induced by sound. *Brain Res Cogn Brain Res* **14(1)**:147-152. 2002.

Ladan Shams, Yukiyasu Kamitani, and Shinsuke Shimojo. Illusions. What you see is what you hear. *Nature* **408(6814)**: 788. 2000.

WF Smith. The relative quickness of visual and auditory perception. *J Exp Psychol* **16**: 239-257. 1933.

Charles Spence, David I Shore, and Raymond M Klein. Multisensory prior entry. *J Exp Psychol Gen* 130: 799-832. 2001.

Yoichi Sugita and Yôiti Suzuki. Audiovisual perception: Implicit estimation of sound-arrival time. *Nature* **421(6926)**: 911. 2003.

Wolfgang A Teder-Sälejärvi, John J McDonald, Francesco Di Russo, and Steven A Hillyard. An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Brain Res Cogn Brain Res* **14(1)**: 106-14. 2002.

Wolfgang A Teder-Sälejärvi, Francesco Di Russo, John J McDonald, and Steven A Hillyard. Effects of spatial congruity on audio-visual multimodal integration. *J Cogn Neurosci* **17(9)**:1396-1409. 2005.

Edward B Titchener. Lectures of the elementary psychology of feeling and attention. *New York: Macmillan*. 1908.

J Vibell, C Klinge, M Zampini, C Spence, and AC Nobre. Temporal order is coded temporally in the brain: early event-related potential latency shifts underlying prior entry in a cross-modal temporal order judgment task. *J Cogn Neurosci* **19(1)**: 109-20. 2007.

Jean Vroomen, Mirjam Keetels, Beatrice de Gelder, and Paul Bertelson. Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cogn Brain Res* **22**: 32-35. 2004.

RB Welch, LD DuttonHurt, and DH Warren. Contributions of audition and vision to temporal rate perception. *Percept Psychophys* **39(4)**: 294-300. 1986.

Massimiliano Zampini, David I Shore, and Charles Spence. Audioviusal temporal order judgments. *Exp Brain Res* **152**: 198-210. 2003.