Exp Brain Res (2007) 182:199–214
DOI 10.1007/s00221-007-0982-4

RESEARCH ARTICLE

Modality shift effects mimic multisensory interactions: an event-related potential study

Matthias Gondan · Dirk Vorberg · Mark W. Greenlee

Received: 13 October 2006 / Accepted: 8 May 2007 / Published online: 12 June 2007

Abstract A frequent approach to study interactions of the auditory and the visual system is to measure event-related potentials (ERPs) to auditory, visual, and auditory-visual stimuli (A, V, AV). A nonzero result of the AV – (A + V) comparison indicates that the sensory systems interact at a specific processing stage. Two possible biases weaken the conclusions drawn by this approach: first, subtracting two ERPs from one requires that A, V, and AV do not share any common activity. We have shown before (Gondan and Röder in Brain Res 1073–1074:389–397, 2006) that the problem of common activity can be avoided using an additional tactile stimulus (T) and evaluating the ERP difference (T + TAV) – (TA + TV). A second possible confound is the modality shift effect (MSE): for example, the auditory N1 is increased if an auditory stimulus follows a visual stimulus, whereas it is smaller if the modality is unchanged (ipsimodal stimulus). Bimodal stimuli might be affected less by MSEs because at least one component always matches the preceding trial. Consequently, an apparent amplitude modulation of the N1 would be observed in AV. We tested the influence of MSEs on auditory-visual interactions by comparing the results of AV – (A + V) using (a) all stimuli and using (b) only ipsimodal stimuli. (a) and (b) differed around 150 ms, this indicates that AV – (A + V) is indeed affected by the MSE. We then formally and empirically demonstrate that (T + TAV) – (TA + TV) is robust against possible biases due to the MSE.

Keywords Multisensory processes · Event-related potentials · Divided attention · Modality shift effect

Introduction

In everyday perception, the information of the different sensory systems is not processed by independent pathways. This information is rather integrated and processed as a multisensory percept (Welch and Warren 1986), yielding more efficient behavior in many situations. For example, if participants have to make speeded responses to auditory, visual, and bimodal auditory-visual stimuli, faster responses are observed for bimodal stimuli. In order to investigate the neural interactions between sensory systems, a frequent approach is to measure event-related potentials (ERPs) to unimodal and bimodal stimuli, for example, auditory, visual, and simultaneous auditory-visual stimuli (the three ERPs are abbreviated in the following as A, V, and AV, respectively). The arithmetic sum of the ERPs to the unimodal stimuli is then subtracted from the ERP to the bimodal stimulus: AV – (A + V). If the auditory and the visual information is processed in separate pathways, the result should not differ from zero, that is, the bimodal ERP response AV is equivalent to the linear superposition of the two unimodal ERP responses A and V. In contrast, a non-zero result of AV – (A + V) indicates that the sensory systems interact at a particular processing stage. This comparison method has been used to demon-

Electronic supplementary material The online version of this article (doi:10.1007/s00221-007-0982-4) contains supplementary material, which is available to authorized users.

M. Gondan (✉) · M. W. Greenlee
Department of Psychology, University of Regensburg,
93050 Regensburg, Germany
e-mail: matthias.gondan@psychologie.uni-regensburg.de

D. Vorberg
Department of Psychology,
Technical University of Braunschweig,
Braunschweig, Germany

strate interactions between audition and vision (Barth et al. 1995; Fort et al. 2002a, b; Giard and Peronnet 1999; Molholm et al. 2002), audition and touch (Foxe et al. 2000; Gobbelé et al. 2003), and vision and touch (Schürmann et al. 2002). In some of these studies, the ERP to bimodal stimuli differed from the sum of the ERPs to unimodal stimuli as early as 50 ms after stimulus onset. This evidence suggests that the information from the different sensory systems is integrated at very early processing stages (Fort et al. 2002a; Foxe et al. 2000; Giard and Peronnet 1999).

Two potential problems weaken the conclusions drawn by the result of the AV – (A + V) comparison: common activity and modality shift effects. As Teder-Sälejärvi et al. (2002) cautioned, the three ERPs not only reflect perceptual processes, but also contain unspecific activity which is common to the processing of the three different stimuli, for example the P300 wave or the contingent negative variation (CNV, Walter et al. 1964). The CNV is a slow ramp-like negative deflection at frontal and central electrodes which can be observed starting at approximately 1,000 ms before an expected salient stimulus. In AV – (A + V), the CNVs of A and V are subtracted twice from the CNV of AV: the CNV should therefore appear as a slow positivity in the ERP difference. To eliminate the CNV wave, Teder-Sälejärvi et al. suggested a high-pass filter to eliminate the slow components of the three ERPs. Doing so, Teder-Sälejärvi et al. identified the first reliable auditory-visual interaction at approximately 160 ms after stimulus onset, which indicates a rather late processing stage. The authors interpreted a first significant interaction around 100 ms as a residual CNV wave not entirely eliminated by the filter. More generally, it should be noted that the CNV and other possible sources of common activity (e. g. residual activity from previous trials, Talsma and Woldorff 2005) are neither entirely nor selectively eliminated by a high-pass filter.

An alternative ERP comparison which involves the use of an additional tactile stimulus has been suggested by Gondan and Röder (2006). The ERPs for a simple tactile stimulus (T) and for a trimodal stimulus (TAV) are summed and compared to the sum of TA and TV, that is, the ERPs for auditory-tactile and visuo-tactile stimuli: (T + TAV) – (TA + TV). Technically, this comparison is identical to AV – (A + V), but two modifications are made: first, the ERP to a ''null stimulus'' (O) is added to the minuend: (O + AV) – (A + V) (cf. Talsma and Woldorff 2005). This experimental manipulation allows us to control for common activity in the prestimulus baseline (e. g. the CNV). In contrast to the auditory, visual, and bimodal stimuli, the null stimulus does not have a clearly defined onset. Consequently, the assumption that O reflects the common components of A, V, and AV is strictly tenable only before stimulus onset. Therefore, each of the four stimuli is presented together with a tactile stimulus [(O + AV) – (A + V) → (T + TAV) – (TA + TV)]. Since in the resulting ERP difference, two ERPs are subtracted from two others, common activity should be eliminated. Under the assumption that auditory-tactile and visuo-tactile interactions cancel out because they are elicited by both TAV and TA/TV (Table 2), the comparison (T + TAV) – (TA + TV) isolates auditory-visual interactions as does AV – (A + V). This relies on the additional assumption that the trimodal stimulus does not elicit a specific ERP response to a trimodal stimulus (a ''trisensory'' interaction, see Discussion).

The present study will focus on a second potential problem in the AV – (A + V) comparison, the modality shift effect (MSE). In a randomized sequence of stimuli of different modalities, two types of stimuli can be distinguished: in *ipsimodal* stimuli, the modality of the current stimulus is identical to that of the preceding stimulus, e.g. an auditory stimulus following another auditory stimulus. In *crossmodal* stimuli, the modality of the current stimulus is different from the preceding stimulus, e.g. an auditory stimulus following a visual stimulus. If participants have to make speeded responses to auditory, visual, and tactile stimuli presented in random order, responses are usually faster for ipsimodal stimuli than for crossmodal stimuli (Spence et al. 2001). The MSE has most frequently been observed in crossmodal auditory stimuli (Ferstl et al. 1994), primarily in simple reaction time tasks (Cohen and Rist 1992). The exact source of such modality shift effects is a matter of debate; sensory–perceptual facilitation mechanisms (neural ''traces'', expectancy, reviewed in Manuzza 1980) are discussed as well as response-related processes; a more general overview on sequence effects is given by Luce (1986, Chap. 6.6).

The so-called ''neural trace theory'' (Zubin 1975) explains the reaction time difference between ipsimodal and crossmodal stimuli by two mechanisms: a repeated sequence of stimuli of a given modality yields an increase of residual activity in the perceptual system. This residual activity is facilitatory for subsequent stimuli of the same modality, and, as a consequence, evidence for a stimulus of the same modality is reached earlier. This mechanism explains why the reaction time decreases in long sequences of ipsimodal stimuli (Mowrer et al. 1940). A second mechanism is inhibitory: the repeated stimulation in one sensory channel leads to slower responses to stimuli in the other channel. This mechanism explains why the response to a given crossmodal stimulus increases the more stimuli of the other modality precede this stimulus (Sutton and Zubin 1965). The increase of this inhibitory effect was observed only in the visual modality, though. Manuzza (1980) critically reviewed the neural trace theory. A major criticism is that the theory is not specific enough to derive testable predictions (Nuechterlein 1977a, cited in Manuzza 1980).

For example, it is not specified whether the two mechanisms (facilitation of ipsimodal signals, inhibition of crossmodal signals) operate independently of each other. Moreover, the facilitatory mechanism does not seem necessary to explain the observed behavioral effects: fast responses to ipsimodal stimuli (Mowrer et al. 1940) can also be explained by a further decrease of inhibitory traces after repeated ipsimodal stimuli. As will be shown below, it is not necessary to know the exact source of the MSE for the purpose of the present study. We simply assume that the neural activity elicited by modality shifts differs somehow between unimodal and bimodal stimuli, and we will illustrate this using a ''naive'' notion of modality shift effects. If our assumption holds, we demonstrate that in the classical experimental setup used to investigate multisensory interactions, the MSE is able to ''mimic'' interactions between the sensory systems, for example, modulations of ERP components such as the auditory N1.

It has been shown that reaction time gains in bimodal redundant stimuli are partly a consequence of MSEs: in a conventional reaction time analysis, ipsimodal and crossmodal stimuli are pooled, resulting in an increased mean reaction time to unimodal stimuli (due to the MSE which affects the subset of the crossmodal unimodal stimuli). In contrast, the modality of at least one component of a bimodal stimulus always matches the modality of the preceding stimulus. It follows that at least one component of bimodal stimuli will always be ''ipsimodal''; hence it can be assumed that the mean reaction time to bimodal stimuli is affected by the MSE to a lesser extent. Gondan et al. (2004) demonstrated that the MSE can lead to an apparent speeding of response times even in the absence of a coactive multisensory mechanism (see also Miller 1986, p. 338 and Table 4). To eliminate the MSE, they suggested using only ipsimodal trials for the analysis of redundancy gains in reaction times to bimodal stimuli, because ipsimodal trials are free from modality shift effects. If a coactivation effect is still observed, it can be concluded that the redundant information of the two sensory systems is integrated somewhere in the processing pathway. In fact, Gondan et al. (2004) observed coactivation effects in speeded responses to auditory-visual, auditory-tactile, and visuo-tactile stimuli, even if the reaction time analysis was restricted to the subset of ipsimodal stimuli.

Similar to the reaction times, modality shifts also affect the ERPs to unimodal stimuli: for example, ERPs to ipsimodal auditory stimuli have smaller N1 amplitudes than ERPs to crossmodal auditory stimuli (Fig. 2/Tone in Cohen and Rist 1992, p. 169). In visual stimuli, the picture is less clear: Fig. 2/Light in Cohen and Rist (1992) shows an increased N100 for crossmodal visual stimuli at Cz. Note however that in the figure, the prestimulus baselines of ipsimodal and crossmodal stimuli differ. If the baseline difference is taken into account in Cohen and Rist (1992), the modality shift effect in visual stimuli seems to vanish or even change its direction. In a randomized sequence of auditory, visual, and bimodal stimuli, the average auditory evoked potential is a mixture of ipsimodal auditory ERPs ($A_{n-1}A_n$; n denotes the current trial) with low N1 amplitudes and crossmodal ERPs with high N1 amplitudes ($V_{n-1}A_n$), resulting in an average ERP with ''intermediate'' N1 amplitude. In bimodal stimuli, either the visual or the auditory component always matches the modality of the preceding stimulus. Therefore, the mean ERP to bimodal stimuli should be less affected by modality shifts and thus have a lower auditory N1 amplitude. The result of AV – (A + V) would suggest an apparent amplitude decrease of the auditory N1. It is also possible that a given amplitude is decreased in crossmodal stimuli (see Fig. 3b, visual N1). Assuming that the bimodal ERP does not contain modality shift-related activity, this would result in an apparent amplitude increase in the bimodal stimulus. Amplitude modulations in AV compared to A and V have in fact been reported (Giard and Peronnet 1999; Molholm et al. 2002; van Wassenhove et al. 2005), as well as in AT compared to A and T (Foxe et al. 2000; Gobbelé et al. 2003; Lütkenhöhner et al. 2002 using MEG; Murray et al. 2005). The goal of the present study was to test whether modality shift-related activity can account for some of the auditory-visual interactions as defined by AV – (A + V).

In Tables 1 and 2, we show that these objections against using the contrast AV – (A + V) are backed by a more elaborate analysis, whereas the contrast (T + TAV) – (TA + TV) avoids such confounds. We assume three additive components of the ERP response to a given crossmodal stimulus. For example, an auditory stimulus which follows a visual stimulus ($V_{n-1}A_n$) elicits a ''raw'' auditory evoked potential (A_0), plus activity related to the shift away from the visual modality (V^-), plus activity related to the shift towards the auditory modality (A^+). The assumption of ''shift away'' components (A^-, V^-) seems counterintuitive at a first glance. Note however, that Spence et al. (2001, p. 330) observed that ''RT costs associated with shifting attention from the tactile modality were greater than those for shifts from either audition or vision''—a behavioral shift away effect. For the present purpose, no assumption is needed about the voltage distribution of the shift components, except that it is non-zero. Based on this assumption, Table 1 demonstrates that even if audition and vision did *not* interact, the ERP contrast AV – (A + V) would differ from zero, because the unimodal ERPs contain activity related to A^+, A^-, V^+, and V^-, whereas the bimodal ERP only contains the positive shift components A^+ and V^+. In contrast, as shown in Table 2, the shift components are balanced in (T + TAV) – (TA + TV): as the minuend (T + TAV) and the subtrahend

Table 1 ERP elicited by auditory, visual, and auditory-visual stimuli in a randomized sequence of stimuli

+AV	Components			–A	Components			–V	Components		
Sequence	Raw	Shift	Interaction	Sequence	Raw	Shift	Interaction	Sequence	Raw	Shift	Interaction
$_{n-1}AAV_n$	A_0, V_0	V^+	A × V	$A_{n-1}A_n$	A_0			$A_{n-1}V_n$	V_0	A^-, V^+	
$V_{n-1}AV_n$	A_0, V_0	A^+	A × V	$V_{n-1}A_n$	A_0	V^-, A^+		$V_{n-1}V_n$	V_0		
$AV_{n-1}AV_n$	A_0, V_0		A × V	$AV_{n-1}A_n$	A_0	V^-		$AV_{n-1}V_n$	V_0	A^-	
Result	3 A_0, 3 V_0	A^+, V^+	3 A × V		3 A_0	2 V^-, A^+			3 V_0	2 A^-, V^+	

A_0 Raw evoked potential, not affected by modality shifts, A^+, V^- hypothetical ERP components elicited by a modality shift away from the visual (V^-) towards the auditory modality (A^+). It is evident that the raw potentials cancel out in AV – A – V: (3 A_0 + 3 V_0) – (3 A_0) – (3 V_0) = 0. In contrast, the shift components are not balanced in AV – A – V: (A^+ + V^+) – (2 V^- + A^+) – (2 A^- + V^+) = – 2 V^- – 2 A^-. Evaluating only the subset of ipsimodal trials $AV_{n-1}AV_n$ – ($A_{n-1}A_n$ + $V_{n-1}V_n$) should avoid this potential source of artifact, thereby isolating the multisensory interaction component A × V

Table 2 ERP elicited by T, TAV, TA and TV in a randomized sequence of stimuli

+T	Components			+TAV	Components		
Sequence	Raw	Shift	Interaction	Sequence	Raw	Shift	Interaction w/o A × V × T
$T_{n-1}T_n$	T_0			$T_{n-1}TAV_n$	T_0, A_0, V_0	A^+, V^+	A × V, A × T, V × T
$TA_{n-1}T_n$	T_0	A^-		$TA_{n-1}TAV_n$	T_0, A_0, V_0	V^+	A × V, A × T, V × T
$TV_{n-1}T_n$	T_0	V^-		$TV_{n-1}TAV_n$	T_0, A_0, V_0	A^+	A × V, A × T, V × T
$TAV_{n-1}T_n$	T_0	A^-, V^-		$TAV_{n-1}TAV_n$	T_0, A_0, V_0		A × V, A × T, V × T
Result	4 T_0	2 A^-, 2 V^-			4 T_0, 4 A_0, 4 V_0	2 A^+, 2 V^+	4 A × V, 4 A × T, 4 V × T

–TA	Components			–TV	Components		
Sequence	Raw	Shift	Interaction	Sequence	Raw	Shift	Interaction
$T_{n-1}TA_n$	T_0, A_0	A^+	A × T	$T_{n-1}TV_n$	T_0, V_0	V^+	V × T
$TA_{n-1}TA_n$	T_0, A_0		A × T	$TA_{n-1}TV_n$	T_0, V_0	A^-, V^+	V × T
$TV_{n-1}TA_n$	T_0, A_0	V^-, A^+	A × T	$TV_{n-1}TV_n$	T_0, V_0		V × T
$TAV_{n-1}TA_n$	T_0, A_0	V^-	A × T	$TAV_{n-1}TV_n$	T_0, V_0	A^-	V × T
Result (cont.)	4 T_0, 4 A_0	2 A^+, 2 V^-	4 A × T		4 T_0, 4 V_0	2 A^-, 2 V^+	4 V × T

T_0, A_0, V_0: raw evoked potential, A^+, V^-: ERP components elicited by modality shifts. In the ERP difference (T + TAV) – (TA + TV), the raw ERPs and the activity elicited by modality shifts cancel out. Under the assumption that trisensory interactions are absent, A × V is isolated

(TA + TV) contain the same number of A^+, V^+, A^-, and V^- components of the same sign, they are eliminated in the ERP difference.

In the present study, we evaluated the traditional analysis of auditory-visual interactions with ERPs in two experimental conditions: in one condition, participants observed auditory, visual, and auditory-visual stimuli. Interactions of audition and vision were isolated using the classical AV – (A + V) comparison. In a first (conventional) analysis, all stimuli were used. We expected a number of auditory-visual interactions in AV – (A + V), some of them visible as a modulation of activity over unisensory areas. In a second analysis, modality shift effects were eliminated using only the ipsimodal stimuli. If the interaction defined by AV – (A + V) is ''contaminated'' by modality shift effects, the results of the two analyses should differ from each other.

In the second condition, tactile, auditory-visuo-tactile, auditory-tactile and visuo-tactile stimuli were presented and auditory-visual interactions were inspected using (T + TAV) – (TA + TV) to control for common activity. As shown in Table 2, (T + TAV) – (TA + TV) should equally be robust with respect to modality shifts. Therefore, we expect the same results, regardless of whether the analysis is based on all stimuli or only on the ipsimodal stimuli. Similar to Gondan and Röder (2006), the assumption that trisensory interactions are absent is tested in the reaction times to trimodal target stimuli.

Method

The experiment was divided into even and odd blocks in which the results of the hypotheses of the present study

were tested in AV – (A + V) and (T + TAV) – (TA + TV), respectively: in the even blocks, a target detection task with auditory, visual, and auditory-visual stimuli was conducted. Participants had to make speeded responses to 10% deviants; 90% of the stimuli were ''standards'' which did not require a response. Auditory-visual interactions were measured using the AV – (A + V) comparison. The results of a first analysis, in which all stimuli were used, were subtracted from the results of a second analysis which was restricted to the subset of ipsimodal stimuli $AV_{n-1}AV_n - (A_{n-1}A_n + V_{n-1}V_n)$ free from modality shift effects.

In the odd blocks, participants detected deviants in a series of tactile, auditory-visuo-tactile, auditory-tactile and visuo-tactile stimuli, and auditory-visual interactions were measured using the ERP difference (T + TAV) – (TA + TV). Again, the results of the analysis, in which all stimuli were used, were subtracted from the analysis in which only ipsimodal stimuli were used.

In both conditions, the hypothesis testing was restricted to the electrodes and intervals at which significant MSEs were observed in either A, V, TA, or TV (''region of interest''-analysis).

Participants

Sixteen students of psychology participated in the study (14 females, 2 males, mean age 24 years, range 20–30 years, two left-handed). All were free of any obvious neurological disorders, had normal hearing and normal or corrected-to-normal vision (based on self-reports). They received partial course credits or payment and gave their written informed consent prior to participation. EEG data from three other participants had to be excluded from the data analysis; two of them had a low signal-to-noise ratio due to high alpha activity at posterior recording sites. In one participant, the visual evoked potential was not visible, suggesting that the participant did not attend to the task.

Stimuli and procedure

The experiment followed a typical oddball design with 90% ''standard'' stimuli and 10% ''target'' stimuli. The entire experiment was divided into 18 blocks of about 5 min stimulation each, yielding a total duration of about 2 h, breaks included. In the even blocks, unimodal auditory, visual, and bimodal auditory-visual stimuli were presented in random order (A, V, AV, each 25%, plus 25% ''gaps'' in which no stimulus was presented, O[1]). Auditory standards were bursts of white noise (20 ms, 65 dBA) emitted by a loudspeaker at a distance of 80 cm, which was located straight ahead of the participant. Visual standards were light flashes (20 ms), emitted by a group of four LEDs (60 mcd) mounted into the housing of the loudspeaker and visible through the front grid. Target stimuli were auditory, visual, or auditory-visual double stimuli of 20 ms each, presented with a gap of 100 ms: A-gap-A, V-gap-V, AV-gap-AV. Each standard was presented 405 times; each target was presented 45 times during the experiment. Subjects responded to the target stimuli by pressing a button with the left hand.

In the odd blocks, tactile, auditory-tactile, visuo-tactile, and auditory-visuo-tactile stimuli were used. The tactile impulse was delivered above threshold to the right index finger (small metallic post, diameter 0.2 mm) by a custom-made mechanical stimulator. A faint white background noise had to be continuously presented during the entire session by a second loudspeaker to mask any sounds emitted by the mechanical stimulator. Participants had again to respond to rare (10%) target stimuli in which a standard stimulus was presented twice in rapid succession.

Participants were seated in a comfortable chair and were asked to fixate the loudspeaker with their gaze and to respond to the target stimuli as quickly as possible. No response was required for the standard stimuli. The inter-stimulus-interval varied between 1,300 and 1,700 ms. The experiment took place in a dimly lit, electrically and acoustically shielded room (Industrial Acoustics).

EEG recording

The EEG was recorded from 62 equally distant scalp electrodes (non-polarizable Ag/AgCl electrodes) mounted into an elastic cap (Easy Cap, FMS). FCz served as the reference in the recordings. The electrode impedance was kept at 10 kΩ or below by preparing the skin with ''Abralyt 2000'' (FMS, Herrsching, Germany) and isopropyl alcohol. The band pass of the amplifiers (BrainAmp MR plus, MesMed, Munich, Germany) was set from 0.1 to 100 Hz, the sampling rate was 500 Hz. Horizontal eye movements were monitored with EOG at AF7 and AF8, vertical eye movements and eye blinks were measured with an electrode placed under the left eye. The EOG channels served for offline rejection of trials with eye artifacts. Segments with ocular activity larger than 50 μV between 100 ms before and 400 ms after stimulus onset were rejected.

ERP analysis

Only the standard stimuli (non-targets) were used for the ERP analysis. ERPs were averaged separately for each stimulus condition, baseline-corrected to the mean activity 100 to 0 ms preceding stimulus onset, and referenced

[1] The purpose of the null stimuli O was to evaluate the comparison suggested by Talsma and Woldorff (2005), (O + AV) – (A + V).

offline to the mean voltage of both mastoids. Modality shift effects were investigated by subtracting the ERP time course for crossmodal stimuli from the ERP time course for ipsimodal stimuli, that is, $V_{n-1}A_n - A_{n-1}A_n$ for auditory stimuli, and $A_{n-1}V_n - V_{n-1}V_n$ for visual stimuli in the even blocks. In the odd blocks, MSEs were defined as $TV_{n-1}TA_n - TA_{n-1}TA_n$ and $TA_{n-1}TV_n - TV_{n-1}TV_n$, accordingly. A modality shift effect was considered reliable if the ipsimodal and the crossmodal curves differed significantly ($P < 0.05$) for at least 10 ms at one of the highlighted channels in Fig. 1. This was done using a point-wise one sample t test (which is numerically identical to a test of paired samples, Whitley and Ball 2002).

Our main hypothesis was that the AV – (A + V) comparison does not only reflect multisensory interactions (MSI), but is contaminated by activity related to modality shifts, especially the shift away components (SAC, Table 1). In a more formal notation, this can be written as: AV – (A + V) = MSI + SAC. Restricting the analysis to only ipsimodal stimuli should eliminate all activity related to MSEs: $AV_{n-1}AV_n - (A_{n-1}A_n + V_{n-1}V_n)$ = MSI. Under these assumptions, it is evident that a non-zero SAC will yield a difference between the two analyses. Therefore, the hypothesis is tested using the double difference [AV – (A + V)] – $[AV_{n-1}AV_n - (A_{n-1}A_n + V_{n-1}V_n)]$ = [MSI + SAC] – [MSI] = SAC. As shown in Table 1, the activity reflected by SAC in this comparison is a subset of the brain activity observed in modality shifts in unimodal auditory and visual stimuli. Consequently, significant areas of the [AV – (A + V)] – $[AV_{n-1}AV_n - (A_{n-1}A_n + V_{n-1}V_n)]$ difference should be within the ''region of interest'' (ROI) defined by the union of $V_{n-1}A_n - A_{n-1}A_n \neq 0$ and $A_{n-1}V_n - V_{n-1}V_n \neq 0$. Such a region of interest approach is able to reduce type I errors while preserving power: at electrodes and intervals at which a modality shift effect is not observed, SAC should not differ from zero and it is not plausible that the two analysis methods yield different results. As the modality shift effects obtained for A/V (even blocks) and TA/TV (odd blocks) did not perfectly overlap, the regions of interest actually used for both main analyses were pooled, that is, they represented the union of the electrodes and intervals at which significant MSEs were observed in A, V, TA, and TV.

To test whether the criteria for the main analysis are sufficiently strict to avoid false positive results, we calculated a t_{max} distribution (Blair and Karninsky 1993) using 10,000 permutations. In each permutation, the sign of the ERP difference [AV – (A + V)] – $[AV_{n-1}AV_n - (A_{n-1}A_n + V_{n-1}V_n)]$ was selected at random for each participant, and we calculated the t values at each sampling point which fell into the region of interest. We then chose the maximum absolute t value which met our analysis criterion ($P < 0.05$ for 10 ms). The 95th percentile of this distribution was selected as the critical t value. Since the same restrictions were used in the permutations and in the main analysis, the probability is 0.05 that any absolute t value in the main analysis is above the critical t value if the null hypothesis holds.

The analysis of multisensory interactions, that is, the intervals during which AV – (A + V) is different from zero or (T + TAV) – (TA + TV) is different from zero, was not restricted to specific electrodes and intervals. As we did not have specific hypotheses concerning the outcome of AV – (A + V) or (T + TAV) – (TA + TV), this analysis is rather descriptive in nature (criterion: point-wise two-tailed t test, $P < 0.05$ for at least 10 ms).

Reaction time analysis

Several models have been suggested to explain the reaction time gain in redundant stimuli (Diederich and Colonius 2005; Miller and Ulrich 2003; Schwarz 1994), of which the two most prominent are the race model and the coactivation model (Miller 1982). According to the race model, processing of the two stimuli occurs in separate channels, and a response is triggered as soon as the faster of the two channels has finished processing. As a consequence, the probability for a fast response is increased if two stimuli are presented instead of one (''statistical facilitation'', Raab 1962). The maximal redundancy gain obtained by statistical facilitation has an upper limit which is described

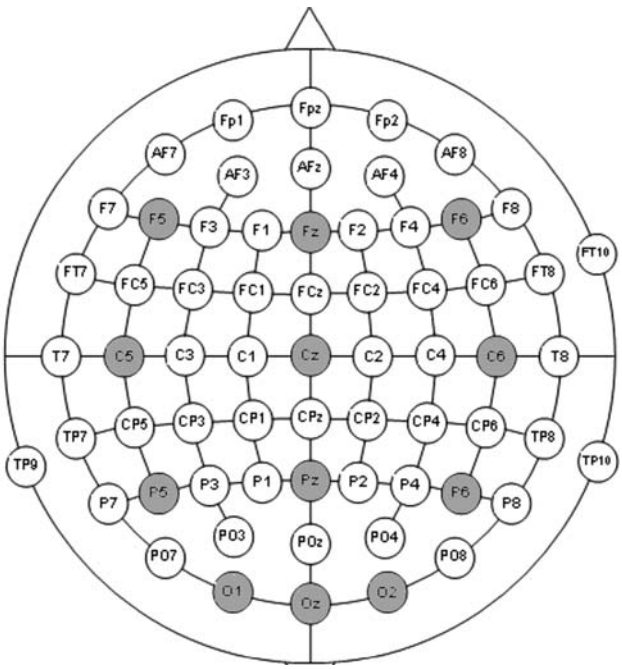

Fig. 1 Electrode montage. EEG data were re-referenced offline to the mean voltage of TP9 and TP10. Analysis was restricted to the highlighted channels

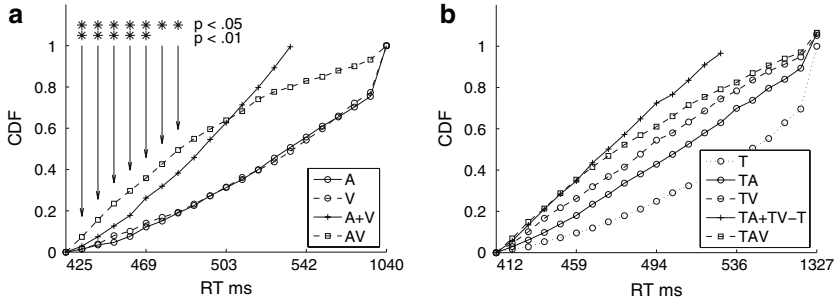

Fig. 2 **a** Race model test, auditory-visual stimuli (Eq. 1). Higher amounts of fast responses to auditory-visual stimuli (*squares*) than the sum of the probabilities of fast responses to auditory and visual stimuli (+) indicate a violation of the race inequality. *Stars* indicate where this violation is significant (binomial test, $p = q = 0.5$). **b** Race model test, trimodal stimuli (Eq. 2). In this condition, F_{TAV} did not exceed $F_{TA} + F_{TV} - F_T$. Redundancy gains in the trimodal stimulus can entirely be explained by a race between the two coactivation components TA and TV, and the three single channel racers T, A, and V

by the race inequality (Miller 1982). If this upper limit is violated at some t, coactive processing is usually concluded, implying that the information of the auditory and the visual system is integrated. This is tested using the cumulative response time distributions F_{AV}, F_A, and F_V:

$$F_{AV}(t) \leq F_A(t) + F_V(t). \quad (1)$$

In the odd blocks, tactile, auditory-visuo-tactile, auditory-tactile and visuo-tactile stimuli were presented. To test for auditory-visual coactivation in the responses to trimodal stimuli, a modified race model is necessary which explicitly allows for auditory-tactile and visuo-tactile coactivation effects. The upper limit in Eq. 2 is formally derived in the Appendix. If Eq. 2 is violated, auditory-visual or trisensory coactive mechanisms are candidates for the observed redundancy gains.

$$F_{TAV}(t) \leq F_{TA}(t) + F_{TV}(t) - F_T(t). \quad (2)$$

To estimate the cumulative reaction time distributions, 20 bins of equal size (5% of all reaction times) were defined separately for each participant (Corballis 2002). $F_{AV} - F_A - F_V$ (Eq. 1) and $F_{TAV} - F_{TA} - F_{TV} + F_T$ (Eq. 2) were tested using a sign test for each bin in the lower percentile range.

Results

Reaction times

Reaction times, omission rates and false alarm rates to auditory, visual and auditory-visual target stimuli (even blocks) are shown in Table 3. The reaction time distributions for auditory, visual, and bimodal stimuli are shown in Fig. 2a. False alarms (that is, ''standards'' to which participants gave a response) and misses were below 10% on average (<4 per condition) and were not further analyzed. Mean reaction times for auditory-visual stimuli were below the mean reaction times for auditory and visual stimuli (bimodal < unimodal:$F(1,15) = 64.9$, $P < 0.001$). Figure 2a shows that the amount of fast responses to bimodal stimuli F_{AV} was higher than $F_A + F_V$ (Eq. 1), hence the redundancy gain was higher than predicted by the race model.

Mean reaction times to tactile, auditory-visuo-tactile, auditory-tactile and visuo-tactile target stimuli (odd blocks) are shown in Table 3. Mean reaction times to trimodal targets were lower than to auditory-tactile or visuo-tactile stimuli (trimodal < bimodal: $F(1,15) = 8.04$, $P < 0.05$), while reaction times to unimodal tactile target stimuli were highest (bimodal < unimodal: $F(1,15) = 13.7$, $P < 0.01$). The redundancy gain in responses to trimodal stimuli did not violate Eq. 2 (Fig. 2b), that is, they were in accordance with a model which did not allow auditory-visual or trisensory coactivation.

Table 3 Mean reaction times, omission and false alarm rates in the even and odd blocks

Condition	RT	SE	OR	FA
Even blocks				
A	551	21	3.6	0.0
V	545	21	5.5	0.6
AV	493	19	2.1	0.1
Odd blocks				
T	566	23	8.2	0.2
TAV	499	24	0.8	2.9
TA	528	23	3.0	0.6
TV	506	21	3.7	1.1

RT Mean reaction time in ms, *SE* standard error, *OR%* omission rate in percent, *FA%* false alarm rate to non-targets in percent

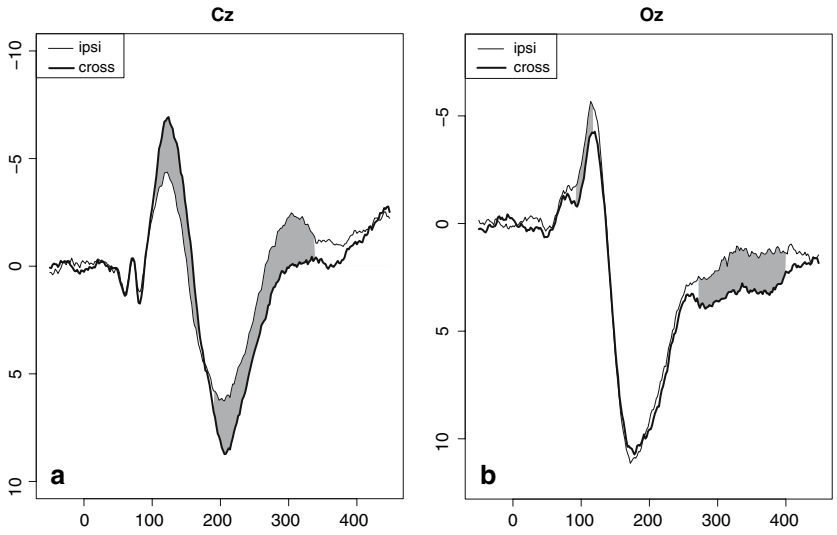

Fig. 3 Auditory evoked potential (**a** *Cz*) and visual evoked potential (**b** *Oz*). The *bold curves* show the average (approximately 100 samples per participant) which was formed using only crossmodal stimuli. The *thin curves* indicate the average based on ipsimodal stimuli only (approximately 100 samples). Intervals during which the ipsimodal and crossmodal time courses differ significantly are marked in gray (two-tailed t test, $P < 0.05$ for 10 ms)

ERP results: AV – (A + V)

The auditory evoked potential (AEP, Fig. 3a) showed characteristic deflections at Cz (Pa: 60 ms, P1: 80 ms, N1: 120 ms, P2: 200 ms). Modality shifts caused increased N1 and P2 amplitudes (the intervals during which the ipsimodal and crossmodal ERPs differ are marked in gray). The visual evoked potential (VEP, Fig. 3b) showed a P1–N1–P2 deflection at occipital recording sites, around 100, 120 and 180 ms, respectively. The ERPs to ipsimodal and crossmodal visual stimuli first differed around 100 ms after stimulus onset, with a more positive time course between P1 and N1 in the crossmodal condition (Fig. 3b, gray areas).

Multisensory interactions, as indicated by the conventional AV – (A + V) comparison, are shown in Fig. 4 (bold lines). A first significant interaction emerged at Cz, starting at approximately 90 ms after stimulus onset (positive), followed by a negative deflection at around 130 ms. A broad fronto-central interaction (positive) was observed between 150 and 200 ms. Later interactions were not analyzed. The analysis was repeated using only the subset of ipsimodal stimuli: $AV_{n-1}AV_n - (A_{n-1}A_n + V_{n-1}V_n)$. The resulting ERP difference is shown in Fig. 4 (thin lines). Up to 120 ms after stimulus onset, both analyses yield similar results. Thus, controlling for modality shift effects did not eliminate the early onset of the auditory-visual interaction observed in AV – (A + V).

The main hypothesis of the present study is tested by the direct subtraction of AV – (A + V) and $AV_{n-1}AV_n - (A_{n-1}A_n + V_{n-1}V_n)$ within the region of interest defined by significant modality shift effects: starting around 120 ms after stimulus onset, the direct comparison of the two analyses within this region of interest yields a significant difference (Fig. 4, gray areas). Upon visual inspection of the two ERP differences, one sees that AV – (A + V) indicates a positivity at Cz and Fz, whereas $AV_{n-1}AV_n - (A_{n-1}A_n + V_{n-1}V_n)$ does not. Note however, that the number of trials which enter AV – (A + V) is, by definition, higher than in $AV_{n-1}AV_n - (A_{n-1}A_n + V_{n-1}V_n)$. Therefore, the significance patterns of AV – (A + V) and $AV_{n-1}AV_n - (A_{n-1}A_n + V_{n-1}V_n)$ only provide qualitative hints for the interpretation of the result.

The permutation test in which the results of the two analyses were directly contrasted yielded a critical absolute t value of 4.132. Thirteen of the observed absolute t values met this criterion (F5, Fz, F6, C5, C6, between 116 and 162 ms).

ERP results: (O + AV) – (A + V)

The results obtained by the (O + AV) – (A + V) comparison (Talsma and Woldorff 2005) mostly confirm the observations obtained by AV – (A + V): an early, though not significant, interaction was observed 90 ms after stimulus onset at Cz in both analyses (ipsimodal stimuli, all stimuli). Around 120 ms, the result of the ERP comparison depends on whether only ipsimodal stimuli or all stimuli are used. This is reflected by a significant difference between the two curves (marked in gray, Fig. 1 in online supplementary material).

The critical t value estimated by the permutation test was 4.301. Twelve observed absolute t values met this criterion (F5, Fz, F6, C5, C6, again between 116 and 162 ms).

ERP results: (T + TAV) – (TA + TV)

Tactile stimuli elicited somatosensory evoked potentials (SEPs, Fig. 5a) over central recording sites of the left

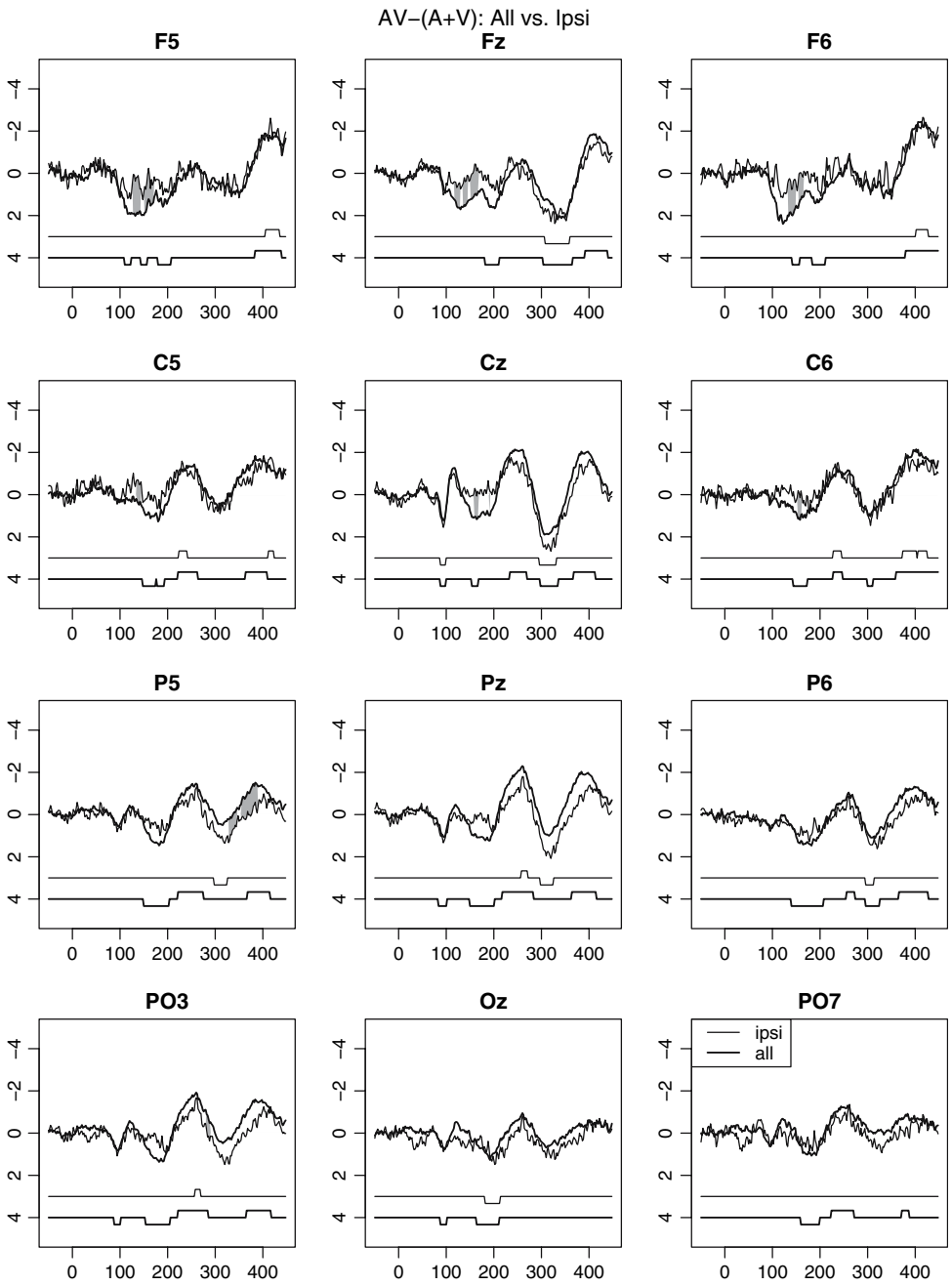

Fig. 4 Auditory-visual interactions according to AV – (A + V). The *bold line* shows the ERP difference for all stimuli. The *thin line* shows the ERP difference for the ipsimodal stimuli. *Gray areas* indicate the intervals during which the two analyses yield different results at $P < 0.05$ for at least 10 ms (restricted to the intervals during which modality shift effects were observed in unimodal stimuli). In line with our main hypothesis, the results of the two analysis methods differ. Intervals during which the result of AV – (A + V) differs from zero are marked by the rectangular curve ($P < 0.05$, 10 ms, *thin* ipsimodal, *bold* all). Note that these significance patterns are only a qualitative hint, they cannot be directly compared, since the number of trials in 'ipsimodal' is only about one-fourth of the number of trials in 'all'

hemisphere, contralateral to the stimulated hand. The amplitude of the SEPs was generally low, presumably due to the low intensity of the tactile stimuli. Consequently, the ERPs for auditory-tactile and visuo-tactile stimuli closely resembled the ERPs for auditory and visual stimuli in the even blocks (Fig. 5b, c). Again, modality shifts caused increased amplitudes of the auditory N1 and P2 (marked in gray), and a more positive time course during P1 and N1 (n. s.).

Multisensory interactions, as indicated by the (T + TAV) – (TA + TV) comparison, are shown in Fig. 6 (bold curves). As in AV – (A + V), a first significant interaction was visible around 90 ms (Oz, positivity), followed directly by a negative deflection (120 ms, PO3), and a broader positivity at central and parietal recording sites. In contrast to the AV – (A + V) difference, the ERP difference fell back to zero around 350 ms after stimulus onset.

Eliminating potential MSEs by using only ipsimodal trials for the analysis yielded similar results (thin curves). This was confirmed by the direct comparison of the results of (T + TAV) – (TA + TV) and the results of

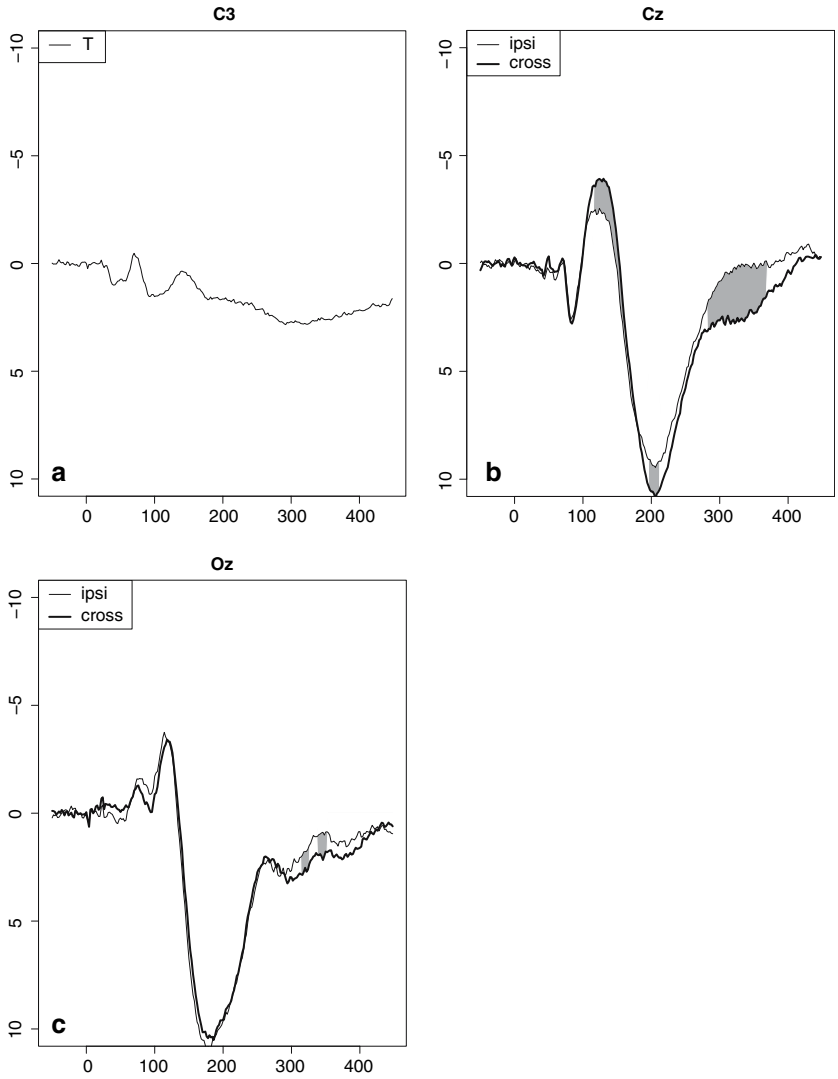

Fig. 5 Somatosensory (**a** *Cz*), auditory-tactile (**b** *Cz*), and visuo-tactile evoked potential (**c** *Oz*). The *thin curves* show the average (approximately 100 samples per participant) which was formed using only ipsimodal stimuli. The *bold curves* indicate the average based on crossmodal stimuli (approximately 100 samples). Intervals during which the ipsimodal and crossmodal time courses differ significantly are marked in *gray* (two-tailed *t* test, $P < 0.05$ for 10 ms)

$(T_{n-1}T_n + TAV_{n-1}TAV_n) - (TA_{n-1}TA_n + TV_{n-1}TV_n)$. In contrast to AV – (A + V), but in line with our hypothesis that the results of (T + TAV) – (TA + TV) are not affected by the MSE (Table 2), restricting the analysis to the subset of ipsimodal stimuli did not change the ERP results. The permutation test yielded a critical absolute *t* value of 4.158. None of the observed absolute *t* values met this criterion.

Discussion

The basic paradigm for the study of multisensory interactions in event-related potentials (ERPs) relies on the measurement of the ERPs to unimodal and bimodal stimuli and compares the bimodal ERP (e.g. AV) to the sum of the two unimodal ERPs (A + V). This additive model assumes that AV contains ERP activity related to the auditory stimulus, plus activity related to the visual stimulus, as well as activity related to the interaction of the auditory and the visual system in the bimodal situation: AV = A + V + A × V. In order to guarantee equivalent task requirements, auditory, visual, and bimodal stimuli have to be presented in randomized order (Besle et al. 2004; van Wassenhove et al. 2005). However, in such sequences, modality shift effects (MSEs, e.g. Spence et al. 2001) need to be controlled for, because they primarily affect the unimodal stimuli; in bimodal stimuli, at least one stimulus component always matches the preceding stimulus (Gondan et al. 2004; Miller 1986). As a consequence, the mean ERP to unimodal stimuli might differ from the unisensory component of the bimodal ERP. This would result in an apparent multisensory interaction which might be observed in AV – (A + V), even if audition and vision do not interact (see Table 1 for a detailed analysis of ERP activity related to modality shifts in the ERP difference). Such an interaction would ''mimic'' a modulation of unisensory activity in the ERP to the bimodal stimulus, which has been

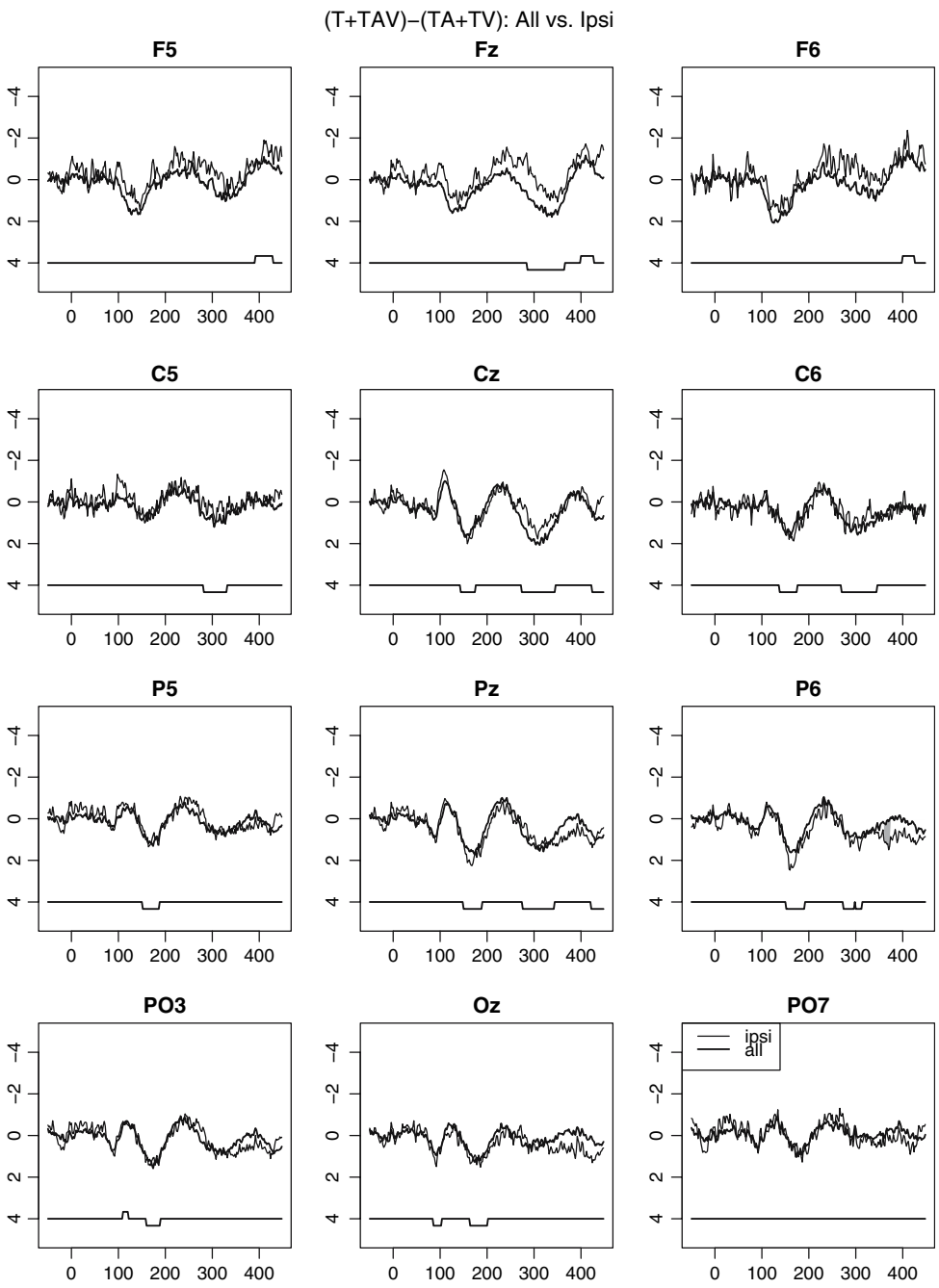

Fig. 6 Interactions of the auditory and visual system according to (T + TAV) – (TA + TV). The *bold line* shows the ERP difference for the entire set of stimuli. The *thin line* shows the ERP difference for ipsimodal stimuli. In line with our hypothesis, the two analyses yield similar results, except for a small interval around 360 ms (P6). Intervals during which the result of (T + TAV) – (TA + TV) differs from zero are marked by the rectangular curve ($P < 0.05$, 10 ms)

reported in several studies (cf. Foxe et al. 2000; Giard and Peronnet 1999; Gobbelé et al. 2003; Lütkenhöhner et al. 2002; Molholm et al. 2002; van Wassenhove et al. 2005).

Indeed, modality shifts had a significant influence on the evoked potentials to unimodal stimuli, at least for auditory ERPs: N1 and P2 of the ERP to ipsimodal stimuli were significantly smaller than N1 and P2 of the ERP to crossmodal stimuli (Figs. 3a, 6b). As the available literature on modality shift effects primarily focuses on the difference between behavioral MSE measures in normals and schizophrenics, we can provide only a preliminary discussion of this finding. The neural trace theory (Zubin 1975) assumes that residual activity in the auditory system accounts for the speeding of response times for ipsimodal stimuli; as a consequence, evidence for subsequent stimuli of the same modality is reached earlier. The present ERP results for auditory stimuli do not contradict this interpretation; because residual activity does not need to be time-locked to the onset of the stimulus. Therefore, it is well possible that the ERP amplitude in ipsimodal stimuli is lower than in crossmodal stimuli, reflecting the lower amount of ''work'' needed to process the stimulus. In contrast, the visual N1 was more positive after a modality shift (Fig. 3b), that is, the amplitude of the ERP component was increased in

ipsimodal stimuli. We have already noted that this finding does not contradict the study of Cohen and Rist (1992), because a baseline correction has not been applied there (Fig. 2 in Cohen and Rist 1992). The increased N1 amplitude in ipsimodal stimuli compared to crossmodal stimuli observed in the present study can be accommodated with an attentional interpretation of the visual MSE: if a participant is actively attending to a visual stimulus, response times are shortened (Posner 1980), and the ERP amplitude is increased (Mangun et al. 1993). Taken together, the behavioral and ERP findings suggest different mechanisms responsible for the MSE in auditory and visual stimuli: in ipsimodal auditory stimuli, ''neural traces'' account for faster responses and lower ERP amplitudes; in ipsimodal visual stimuli, the previous stimulus of the same modality causes the participant to attend the visual modality, consequently, response are faster, and ERP amplitudes are increased. Of course, this post hoc explanation needs further empirical testing.

We eliminated MSE-related activity in AV – (A + V) by performing an additional ERP analysis in which only ipsimodal stimuli were used. During the first 130 ms after stimulus onset, the different analyses yielded similar results, with and without controlling for MSEs. Around approximately 150 ms however, the two time courses differ significantly, and eliminating the MSE equally seems to diminish the positive deflection of the AV – (A + V) difference wave. This occurred simultaneously with the auditory N1, which has been shown to be affected by modality shifts (Figs. 3a, 5b). This suggests that the MSE at least partially accounts for multisensory interactions defined by AV – (A + V), challenging the validity of the method in the interval between 0 and 200 ms after stimulus onset (cf. Besle et al. 2004). We should underline, though, that the relatively early onset of the auditory-visual interaction around 90 ms seems to be a robust finding, unrelated to the MSE.

Another problem of the AV – (A + V) comparison has already been outlined in the Introduction: if two ERPs are subtracted from one, unspecific common activity (CNV, P300, motor-related processes) are subtracted twice from one ERP (Teder-Sälejärvi et al. 2002). Common activity, therefore, can lead to a non-zero AV – (A + V) even if audition and vision do not interact. A solution for this problem is to include an additional zero stimulus and to present the stimuli together with a tactile stimulus, thus evaluating the expression (T + TAV) – (TA + TV). The underlying additive model is outlined in Table 2: at the expense of a potential artifact due to trisensory interactions in the trisensory stimulus, (T + TAV) – (TA + TV) should isolate auditory-visual interactions similar to AV – (A + V). Moreover, common activity is eliminated because two ERPs are subtracted from two other ERPs. Finally, Table 2 demonstrates that in this ERP comparison, unlike AV – (A + V), modality shift effects are actually balanced: to test our assumption and to further evaluate the (T + TAV) – (TA + TV) method, we included a second condition in which participants observed a series of tactile, trimodal, auditory-tactile and visuo-tactile stimuli. Modality shift effects were investigated in TA and TV, and multisensory interactions were analyzed using (a) the entire set of stimuli and (b) the subset of ipsimodal stimuli. Since in (T + TAV) – (TA + TV), the MSEs are canceled out (Table 2), we expected the two analyses (a) and (b) to yield the same results. In line with this hypothesis, Fig. 6 shows that controlling for modality shifts did not alter the main ERP finding in (T + TAV) – (TA + TV).

As already stated in the Introduction, the (T + TAV) – (TA + TV) comparison formally requires that trisensory interactions are zero. This is an assumption which may hold or may not hold. Using an adapted race model test for reaction times to trimodal stimuli, Gondan and Röder (2006) did not find evidence for coactivation effects specific for the trimodal stimulus. Evidence for trimodal cells has been reported in the superior colliculus (Wallace and Stein 2001), in primate parietal cortex (auditory-visual-vestibular: Schlack et al. 2005), in primate superior temporal gyrus (Hikosaka et al. 1988; Schroeder and Foxe 2002), and in human temporo-parietal junction (Matsuhashi et al. 2004). In contrast, Wallace et al. (2004) report only a very low number of trisensory neurons in rat cortex (Tables 1, 2, p. 2169). Although the existence of trisensory neurons is a necessary condition for trisensory interactions, this does not imply that these neurons respond to trimodal stimuli in a way specific to trisensory stimuli. We have argued elsewhere (Gondan and Röder 2006) that a trimodal stimulus is already highly 'amplified' due to auditory-visual, auditory-tactile, and visuo-tactile integration mechanisms, a system which exclusively integrates trimodal events seems of little use, especially due to the enormous complexity of the calculations needed to map the different spatial representations onto each other. It should be noted, however, that the problem of trisensory interactions is far from settled, as a systematic study of trisensory interactions has not yet been undertaken. We should iterate, however, that the AV – (A + V) method relies on two strong assumptions, as well: the first assumption is that common activity is zero; the second assumption is that modality shifting effects can be neglected. As we have demonstrated in the present study and in Gondan and Röder (2006), these two assumptions might be violated in a ''standard'' experimental setup.

In the even blocks of the experimental session, participants had to detect targets in a sequence of auditory, visual, and auditory-visual stimuli. Figure 2a shows the race model test for the reaction times to auditory-visual target stimuli. Significant violations of Eq. 1 were observed in this session, indicating that the information of the two sensory channels is integrated at some particular processing stage.

In the odd blocks, participants had to detect targets in a sequence of tactile, auditory-tactile, visuo-tactile and trimodal stimuli. The adapted race model test to evaluate auditory-visual and trisensory coactivation is derived in Eq. 2. As shown in Fig. 2b, the reaction time distribution for the trimodal stimulus were in line with a model in which auditory-visual and trisensory coactive effects were not allowed. This is in contrast to earlier findings (Diederich and Colonius 2005; Gondan and Röder 2006).

We first note that the (T + TAV) – (TA + TV) method assumes that trisensory interactions are absent (Gondan and Röder 2006); therefore, the result shown in Fig. 2b is in line with this assumption. We also note that Eq. 2 is very conservative in detecting auditory-visual coactivation. One reason for the different findings might be the modified stimulus protocol used in the present experiment, compared to Gondan and Röder (2006): in the previous study, participants had to detect target stimuli which were delivered in all modalities, T, A, V, TA, TV, AV, and TAV. In contrast, the present experiment was split into blocks: even blocks (A, V, AV) and odd blocks (T, TA, TV, TAV): in order to control for sequence effects like the MSE, the number of trial replications has to be increased considerably. If N different conditions are used and each stimulus is presented M times, a given ipsimodal sequence occurs only about M/N times during the experiment. Since the present study had $N = 8$ different stimulus conditions, it would have been necessary to increase the total number of trials by the factor 8 in order to get reliable ERP waveforms. Therefore, we decided to split the entire study into two parts with four conditions each. Within each experimental block, the stimulus sequence was randomized. Doing so, the total number of trials had only to be increased by the factor 4, thereby reducing the duration of the entire session to approximately 120 min. In addition, this experimental manipulation enabled us to simulate both the stimulus protocol and the modality shift effects of a typical AV – (A + V) experiment (Table 1). Likewise, Table 2 shows that MSEs were already balanced in the odd blocks if T, TAV, TA and TV enter the analysis.

Since the main interest of the present study was to investigate auditory-visual interactions, auditory and visual stimuli were delivered from a centrally located loudspeaker, in close spatial proximity (Meredith and Stein 1987) and in the focus of the participant's attention. The tactile stimulus was presented at the right index finger, separated from the location of the auditory and the visual stimulus. In doing so, we tried to avoid interactions of the auditory or the visual system with the tactile system, although spatial coincidence might not be required for multisensory interactions (Murray et al. 2005). One reason why the participants gained less by the auditory-visual stimuli might be that, in the odd blocks, every stimulus had a tactile component. As a consequence, the tactile stimulus component included all necessary information to decide whether a stimulus was a target or a non-target. Therefore, it might be argued that participants paid less attention to the central source of auditory and visual stimuli, but rather directed their attention on their index finger, because the tactile stimulus was the most relevant of the three different stimuli. Moreover, participants had problems detecting targets in the tactile modality (8.2% omissions, Table 3); this might have increased the amount of attention towards the tactile modality, as well. As a consequence, in the present study, participants might have concentrated less on the central loudspeaker than in Gondan and Röder (2006). For the efficient integration of redundant features in purely visual targets, spatial attention seems necessary (Feintuch and Cohen 2002). If this principle also applies for auditory-visual coactivation (e.g. Alsius et al. 2005), this might explain why, in the present study, an auditory-visual coactivation effect was not found in the trimodal stimuli.

As already stated, the onset of the first auditory-visual interaction occurred at about 90 ms after stimulus onset, and this effect seems to be robust with respect to MSEs and common activity. The latency of this interaction replicates earlier findings (Fig. 4 in Gondan and Röder 2006: different topography, significant negativity at T8; Fig. 4 in Talsma and Woldorff 2005: unattended condition, only qualitative results for occipital sites; Fig. 4 in Teder-Sälejärvi et al. 2002: same topography). Although significant, Teder-Sälejärvi et al. did not further analyze the interaction around 100 ms because it closely resembled the prestimulus slow wave which they tried to eliminate using a high-pass filter. The present results confirm this finding. However, evidence for even earlier interactions around 40 ms after stimulus onset (e.g. Giard and Peronnet 1999; Fort et al. 2002a; Molholm et al. 2002) is not provided by our data, neither by AV – (A + V) shown in Fig. 4, nor by (T + TAV – (TA + TV) shown in Fig. 6. Of course, common activity and MSEs were not controlled for in the latter studies, but in the AV – (A + V) comparison shown in Fig. 4 they are not controlled either. Therefore, it is not plausible to conclude that common activity or MSE-related problems exclusively account for the different findings, although the influence of common activity and MSEs need not to be constant across experiments.[2]

[2] A possibly crucial methodological distinction between the present study, Gondan and Röder, Talsma and Woldorff and Teder-Sälejärvi et al. on one hand and Fort et al. (2002a), Giard and Peronnet and Molholm et al. on the other hand is the choice of the reference electrode used in the EEG recordings: in the former studies, the earlobes or the mastoids served as the reference, and the first interactions were observed over Cz. In contrast, in Fort et al., Giard and Peronnet and Molholm et al., the nose served as the reference, and the first interactions were observed over posterior regions.

Several brain regions are candidates for the multisensory interactions observed when comparing the ERP responses to unimodal and bimodal stimuli. Recently, Molholm et al. (2006) provided direct evidence for audio-visual interactions in the superior parietal lobule: the authors recorded evoked potentials directly on the surface of the brain in three patients undergoing epileptic surgery. Responses to auditory-visual stimuli were consistently observed to deviate from the summed responses starting at around 120 ms after stimulus onset. Though not significant in the present study, the negative deflection visible at Cz and Pz around 120 ms (Figs. 4, 6) could reflect this activity. It should be noted that the superior parietal lobule might not be the first target of auditory-visual convergence, because the onset of the interaction observed in Molholm et al. occurs relatively late. In fact, the negative deflection observed in the present study (Figs. 4, 6) which might arise from the superior parietal lobule immediately follows a positivity around 90 ms. The latency of this first positive deflection resembles the latencies reported by Ghazanfar et al. (2005). Ghazanfar et al. recorded local field potentials (LFPs) in the core and belt regions of auditory cortex when rhesus monkeys were attending to short movie clips of vocalizing conspecifics. Starting at around 90 ms after the onset of the vocalization, the audio-visual LFP response differed from the LFP response to an auditory vocalization. Although the comparison of latencies seems problematic across different species, regions around the auditory cortex are likely candidates for early feed-forward multisensory interactions (Calvert et al. 1997; Schroeder et al. 2004).

In summary, our findings suggest that modality shift effects partly account for some of the multisensory interactions observed in simple target detection tasks with auditory, visual, and auditory-visual stimuli. In more complex tasks, MSEs might even have a greater influence (Cohen and Rist 1992; Rist and Cohen 1987). Together with the problems related to common activity in A, V, and AV, the findings of the present study question the validity of the AV – (A + V) method in a randomized stimulus protocol. Therefore, researchers should consider testing for MSEs in their data and performing additional analyses as outlined in the present study, e.g. by repeating the analysis using only ipsimodal stimuli which are free from MSEs, or by using the modified ERP analysis (T + TAV) – (TA + TV) in which both unspecific common activity and the MSE are balanced. Although the present study focused on ERPs and auditory-visual interactions for very simple stimuli, the conclusions drawn here can be readily generalized to any combination of sensory stimuli, including more complex and/or meaningful stimuli (e.g. Beauchamp et al. 2004; Molholm et al. 2004), and to any method in which the additive model is used, such as advanced ERP analysis techniques (spectral analysis, e.g. Sakowitz et al. 2005; inverse solutions, e. g. Murray et al. 2005), or functional magnetic resonance imaging (e.g. Calvert et al. 2001).

Appendix

According to the race model (Miller 1982), a response is triggered as soon as the faster of the two sensory channels has finished processing: $T_{AV} = \min(T_A, T_V)$ (with an implicitly included motor execution time unrelated to the sensory decision which can be omitted). A crucial assumption is that sensory processing of a stimulus does not depend on whether it occurs in the unimodal or in the bimodal context (''context independence'', Colonius 1990). This assumption allows to relate the response time distribution for bimodal stimuli to those for unimodal stimuli:

$$\begin{aligned} P\{T_{AV} < t\} &= P[\min(T_A, T_V) < t] \\ &= P[\{T_A < t\} \cup \{T_V < t\}] \\ &= P\{T_A < t\} + P\{T_V < t\} \\ &\quad - P[\{T_A < t\} \cap \{T_V < t\}] \end{aligned}$$

The conjunction term $P[\{T_A < t\} \cap \{T_V < t\}]$ cannot be estimated without the additional assumption that T_A and T_V are stochastically independent—which is probably wrong, because the two channels might compete for ressources (Colonius 1990). Dropping it yields the well known upper bound (Miller-inequality, Miller 1982).

$$F_{AV}(t) \leq F_A(t) + F_V(t),$$

which holds for all t.

Demonstrating coactivation in trimodal stimulus trials requires an extension of the race model test to three stimuli, for which several upper bounds have been proposed, none of which is uniformly stricter than the others (see Colonius and Vorberg 1994). One of them is a straightforward extension of Miller's inequality:

$$F_{TAV}(t) \leq F_T(t) + F_A(t) + F_V(t).$$

Rejection of this inequality, however, leaves open the question as to the source of the coactivation effect. For example, it is well plausible that the bimodal stimulus TA (implicitly included in TAV) elicits a redundancy gain, because some brain region [TA] profits from redundant auditory-tactile information. The output of [TA], T_{TA}, might then compete in a race with the visual channel. Assuming that T_{TA} and T_V are stochastically equal to the hidden processing times of the auditory-tactile and the visual component in TAV, the model outlined would predict

that $T_{TAV} = \min(T_{TA}, T_V)$, which implies a different upper bound:

$$F_{TAV}(t) \leq F_{TA}(t) + F_V(t).$$

Violation of this upper bound would imply that a race between the information provided by [TA] and by the visual channel cannot explain the redundancy gain observed in trimodal stimuli. However, the problem basically remains: what if fast reactions to trimodal stimuli were more frequent than predicted because there is an additional bimodal coactivation from, e.g. [TV]? Here we sketch a new approach for testing redundancy gains in trimodal stimuli which explicitly allows lower order coactivation effects. These lower order coactivation effects are conceived of as one or more additional runners in the race, which become active if all their constituent stimuli are present. Thus, assuming coactivation in [TA], T_{TAV} is determined by the winner in a race that includes four rather than two runners, $T_{TAV} = \min(T_T, T_A, T_V, T_{TA})$. The present purpose is to test for auditory-visual coactivation effects in trimodal stimuli. Consequently, the model to be tested allows both auditory-tactile and visuo-tactile effects, thus $T_{TAV} = \min(T_T, T_A, T_V, T_{TA}, T_{TV})$:

$$\begin{aligned} P\{T_{TAV} < t\} &= P[\min(T_T, T_A, T_V, T_{TA}, T_{TV}) < t] \\ &= P[\{T_T < t\} \cup \{T_A < t\} \cup \{T_V < t\} \\ &\quad \cup \{T_{TA} < t\} \cup \{T_{TV} < t\}] \\ &= P[(\{T_A < t\} \cup \{T_{TA} < t\}) \cup (\{T_V < t\} \\ &\quad \cup \{T_{TV} < t\}) \cup \{T_T < t\}] \\ &= P[B_1 \cup B_2 \cup B_3] \text{ with} \\ B_1 &= \{T_A < t\} \cup \{T_{TA} < t\}, \\ B_2 &= \{T_V < t\} \cup \{T_{TV} < t\}, B_3 = \{T_T < t\} \end{aligned}$$

Applying Lemma 1 from Colonius and Vorberg (1994),

$$\begin{aligned} P[B_1 \cup B_2 \cup B_3] &\leq P(B_1 \cup B_3) + P(B_2 \cup B_3) - P(B_3) \\ &= P[\{T_A < t\} \cup \{T_T < t\} \cup \{T_{TA} < t\}] \\ &\quad + P[\{T_V < t\} \cup \{T_T < t\} \cup \{T_{TV} < t\}] \\ &\quad - P\{T_T < t\} \end{aligned}$$

By the assumptions, $P[\{T_{TA} < t\} \cup \{T_T < t\} \cup \{T_A < t\}] = F_{TA}(t)$ and $P[\{T_{TV} < t \cup \{T_T < t\} \cup \{T_V < t\}] = F_{TV}(t)$; this yields the upper bound:

$$F_{TAV}(t) \leq F_{TA}(t) + F_{TV}(t) - F_T(t)$$

References

Alsius A, Navarra J, Campbell R, Soto-Faraco S (2005) Audiovisual integration of speech falters under high attention demands. Curr Biol 15:839–843

Barth DS, Goldberg N, Brett B, Di S (1995) The spatiotemporal organization of auditory, visual, and auditory-visual evoked potentials in rat cortex. Brain Res 678:177–190

Beauchamp MS, Lee KE, Argall BD, Martin A (2004) Integration of auditory and visual information about objects in superior temporal sulcus. Neuron 41:809–823

Besle J, Fort A, Giard MH (2004) Interest and validity of the additive model in electrophysiological studies of multisensory interactions. Cogn Process 5:189–192

Blair RC, Karninski W (1993) An alternative method for significance testing of waveform difference potentials. Psychophysiology 30:518–524

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SCR, McGuire PK, Woodruff PWR, Iversen SD, David AS (1997).Activation of auditory cortex during silent lipreading. Science 276:593–596

Calvert GA, Hansen PC, Iversen SD, Brammer M (2001) Detection of audio-visual integration sites by application of electrophysiological criteria to the BOLD effect. Neuroimage 14:427–438

Cohen R, Rist F (1992) The modality shift effect. Further explorations at the crossroads. Ann N Y Acad Sci 658:163–181

Colonius H (1990) Possibly dependent probability summation of reaction time. J Math Psychol 34:253–275

Colonius H, Vorberg D (1994) Distribution inequalities for parallel models with unlimited capacity. J Math Psychol 38:35–58

Corballis MC (2002) Hemispheric interactions in simple reaction time. Neuropsychologia 40:423–434

Diederich A, Colonius H (2005) Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. Percept Psychophys 66:1388–1404

Feintuch U, Cohen A (2002) Visual attention and coactivation of response decisions for features from different dimensions. Psychol Sci 13:361–369

Ferstl R, Hanewinkel R, Krag P (1994) Is the modality shift effect specific for schizophrenia patients? Schizophr Bull 20:367–373

Fort A, Delpuech C, Pernier J, Giard MH (2002a) Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. Cereb Cortex 12:1031–1039

Fort A, Delpuech C, Pernier J, Giard MH (2002b) Early auditory-visual interactions in human cortex during nonredundant target identification. Brain Res Cogn Brain Res 14:20–30

Foxe JJ, Morocz IA, Murray MM, Higgins BA, Javitt DC, Schroeder CE (2000) Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. Brain Res Cogn Brain Res 10:77–83

Ghazanfar AA, Maier JX, Hofmann KL, Logothetis NK (2005) Multisensory integration of dynamic faces and voices in Rhesus monkey auditory cortex. J Neurosci 25:5004–5012

Giard MH, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J Cogn Neurosci 11:473–490

Gobbelé R, Schürmann M, Forss N, Juottonen K, Buchner H, Hari R (2003) Activation of the human posterior parietal and temporo-parietal cortices during audiotactile interaction. Neuroimage 20:503–511

Gondan M, Röder B (2006) A new method for detecting interactions between the senses in event-related potentials. Brain Res 1073–1074:389–397

Gondan M, Lange K, Rösler F, Röder B (2004) The redundant target effect is affected by modality switch costs. Psychon Bull Rev 11:307–313

Hikosaka K, Iwai E, Saito HA, Tanaka K (1988) Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey. J Neurophysiol 60:1615–1637

Luce RD (1986) Response times. Their role in inferring elementary mental organization. Oxford University Press, New York

Lütkenhöhner B, Lammertmann C, Simões C, Hari R (2002) Magnetoencephalographic correlates of audiotactile interaction. Neuroimage 15:509–522

Mangun GR, Hillyard SA, Luck SJ (1993). Electrocortical substrates of visual selective attention. In: Meyer DE, Kornblum S (eds) Attention and performance XIV, MIT Press, Cambridge, pp 219–243

Manuzza S (1980) Cross-modal reaction time and schizophrenic attentional deficit: a critical review. Schizophr Bull 6:654–675

Matsuhashi M, Ikeda A, Ohara S, Matsumoto R, Yamamoto J, Takayama M, Satow T, Begum T, Usui K, Nagamine T, Mikuni N, Takahashi J, Miyamoto S, Fukuyama H, Shibasaki H (2004) Multisensory convergence at human temporo-parietal junction—epicortical recording of evoked responses. Clin Neurophysiol 115:1145–1160

Meredith MA, Stein BE (1987) Spatial factors determine the activity of multisensory neurons in the cat's superior colliculus. Brain Res 420:162–166

Miller J (1982) Divided attention: evidence for coactivation with redundant signals. Cognit Psychol 14:247–279

Miller J (1986) Timecourse of coactivation in bimodal divided attention. Percept Psychophys 40:331–343

Miller J, Ulrich R (2003) Simple reaction time and statistical facilitation: a parallel grains model. Cognit Psychol 46:101–151

Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. Brain Res Cogn Brain Res 14:115–128

Molholm S, Ritter W, Javitt DC, Foxe JJ (2004) Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. Cereb Cortex 14:452–465

Molholm S, Sehatpour P, Mehta AD, Shpaner M, Gomez-Ramirez M, Ortigue S, Dyke JP, Schwartz TH, Foxe JJ (2006) Audio-visual multisensory integration in superior parietal lobule revealed by human intracranial recordings. J Neurophysiol 96:721–729

Mowrer OH, Rayman NN, Bliss EL (1940) Preparatory set (expectancy)—an experimental demonstration of its 'central' locus. J Exp Psychol 26:357–372

Murray MM, Molholm S, Michel CM, Heslenfeld DJ, Ritter W, Javitt DC, Schroeder CE, Foxe JJ (2005) Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. Cereb Cortex 5:963–974

Posner MI (1980) Orienting of attention. Q J Exp Psychol 32:3–25

Raab DH (1962) Statistical facilitation of simple reaction times. Trans N Y Acad Sci 24:574–590

Rist F, Cohen R (1987) Effects of modality shift on event-related potentials and reaction times of chronic schizophrenics. In: Johnson JR, Rohrbaugh JW, Parasuraman R (eds) Current trends in event-related potential research. Electroencephalogr Clin Neurophysiol Suppl 40:738–745

Sakowitz OW, Quiroga RQ, Schürmann M, Başar E (2005) Spatio-temporal frequency characteristics of intersensory components in audiovisually evoked potentials. Brain Res Cogn Brain Res 23:316–326

Schlack A, Strebing-D'Angelo S, Hartung K, Hoffmann KP, Bremmer F (2005) Multisensory space representations in the Macaque ventral intraparietal area. J Neurosci 25:4616–4625

Schroeder CE, Foxe JJ (2002) The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. Brain Res Cogn Brain Res 14:187–198

Schroeder CE, Molholm S, Lakatos P, Ritter W, Foxe JJ (2004) Human–simian correspondence in the early cortical processing of multisensory cues. Cogn Process 5:140–151

Schürmann M, Kolev V, Menzel K, Yordanova J (2002) Spatial coincidence modulates interaction between visual and somatosensory evoked potentials. Neuroreport 13:779–783

Schwarz W (1994) Diffusion, superposition, and the redundant-targets effect. J Math Psychol 38:504–520

Spence C, Nicholls MER, Driver J (2001) The cost of expecting events in the wrong sensory modality. Percept Psychophys 63:330–336

Sutton S, Zubin J (1965) Effect of sequence on reaction time in schizophrenia. In: Welford AT, Birren JE (eds) Behavior, aging, and the nervous system. Thomas, Springfield, pp 562–597

Talsma D, Woldorff MG (2005) Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. J Cogn Neurosci 17:1098–1114

Teder-Sälejärvi WA, McDonald JJ, Di Russo F, Hillyard SA (2002) An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. Brain Res Cogn Brain Res 14:106–114

Wallace MT, Stein BE (2001) Sensory and multisensory responses in the newborn monkey superior colliculus. J Neurosci 21:8886–8894

Wallace MT, Ramachandran R, Stein BE (2004) A revised view of sensory cortical parcellation. Proc Natl Acad Sci 101:2167–2172

Walter WG, Cooper R, Aldridge VJ, McCallum WC, Winter AL (1964) The contingent negative variation. Nature 203:380–384

van Wassenhove V, Grant KW, Poeppel D (2005) Visual speech speeds up the neural processing of auditory speech. Proc Natl Acad Sci 102:1181–1186

Welch RB, Warren DH (1986) Intersensory interactions. In: Boff KR, Kaufman L, Thomas JP (eds) Handbook of perception and human performance. Wiley, New York, pp 25/1–25/36

Whitley E, Ball J (2002) Statistics review 5: comparison of means. Crit Care 6:424–428

Zubin J (1975) The problem of attention in schizophrenia. In: Kietzmann ML, Sutton S, Zubin J (eds) Experimental approaches to psychopathology. Academic, New York, pp 139–166