The spatio-temporal profile of multisensory integration

Johanna Starke[1,2*], Felix Ball[1,2*], Hans-Jochen Heinze[2,3], Toemme Noesselt[1,3]

[1]Department of Biological Psychology, Faculty of Natural Science, Otto-von-Guericke-University Magdeburg, Germany

[2]Department of Neurology, Faculty of Medicine, Otto-von-Guericke-University Magdeburg, Germany

[3]Center for Behavioural Brain Sciences, Otto-von-Guericke-University Magdeburg, Germany

[*]These authors contributed equally

Keywords: multisensory, audiovisual, fMRI, EEG, human

Word Count: 10242

Figures: 8
Tables: 4

Correspondence to

toemme@med.ovgu.de
Department of Biological Psychology
Otto-von-Guericke-University
PO Box 4120
39016 Magdeburg
Germany

Abstract

Task-irrelevant visual stimuli can enhance auditory perception. However, while there is some neurophysiological evidence for mechanisms that underlie the phenomenon, the neural basis of visually-induced effects on auditory perception remains unknown. Combining fMRI and EEG with psychophysical measurements in two independent studies, we identified the neural underpinnings and temporal dynamics of visual-induced auditory enhancement. Lower- and higher-intensity sounds were paired with a non-informative visual stimulus while participants performed an auditory detection task. Behaviourally, visual co-stimulation enhanced auditory sensitivity. Using fMRI, enhanced BOLD-signals were observed in primary auditory cortex for low-intensity audiovisual stimuli which scaled with subject-specific enhancement in perceptual sensitivity. Concordantly, a modulation of event-related potentials could already be observed over frontal electrodes at an early latency (30-80 ms), which again scaled with subject-specific behavioural benefits. Later modulations starting around 280 ms, i.e. in the time range of the P3, did not fit this pattern of brain-behaviour correspondence. Hence, the latency of the corresponding fMRI-EEG brain-behaviour modulation points at an early interplay of visual and auditory signals in low-level auditory cortex, potentially mediated by crosstalk at the level of the thalamus. However, fMRI-signals in primary auditory cortex, auditory thalamus and the P50 for higher-intensity auditory stimuli were also elevated by visual co-stimulation (in the absence of any behavioural effect) suggesting a general, intensity-independent integration mechanism. We propose that this automatic interaction occurs at the level of the thalamus and might signify a first step of audiovisual interplay necessary for visually-induced perceptual enhancement of auditory perception.

Introduction

Real-world events often stimulate more than one sense. By using mutual information from several input modalities the brain can decrease modality-specific noise - caused by the noisiness of the sensory input and of brain responses - and thereby increase the overall robustness of its estimate of the physical world. Evidence for this multisensory interplay (MSI) has been shown with numerous different behavioural paradigms and for different modality-combinations (Chandrasekaran, 2017). Previous behavioural studies have consistently demonstrated an increase in performance for simple audiovisual stimuli (Frassinetti *et al.*, 2002; Jaekl & Soto-Faraco, 2010; Lovelace *et al.*, 2003; Feng *et al.*, 2017; Noesselt *et al.*, 2010) as well as more complex multisensory objects (Beauchamp *et al.*, 2004; Beauchamp, 2005; Doehrmann & Naumer, 2008; Noppeney, 2008). Importantly, these behavioural benefits appear to be more pronounced if unisensory inputs are weak, i.e. less reliable (Nath & Beauchamp, 2011; Rohe & Noppeney, 2015), in accordance with the neural principle of inverse effectiveness (Meredith & Stein, 1983; Stein & Meredith, 1994). On a computational level, the weighting of multisensory inputs by their reliability was first conceptually described as modality-appropriateness (Welch & Warren, 1986), and was later formalized mathematically by maximum likelihood estimation (Ernst & Banks, 2002) and Bayesian causal modelling (Kording, 2014). These methods can successfully predict a range of multisensory phenomena by the weighting of sensory inputs.

On the neural implementation level, multiple structures appear to aid MSI. These include traditional multisensory convergence zones in temporal, parietal and frontal areas. Classical theories on MSI have favoured serial hierarchical models with unimodal processing in putatively unisensory cortex followed by multisensory interplay in higher-order association cortex (Felleman & van Essen, 1991). In clear contradistinction, more recent results from a multitude of imaging studies in humans have provided evidence that putatively unisensory cortex and even primary sensory areas are modulated by MSI (for review see e.g. Driver & Noesselt, 2008). Concordantly, anatomical tracing studies in animals revealed direct connections between V1, A1 and S1 (for gerbils plus a current review of animal studies see Henschke *et al.*, 2015) and even modulations in sensory-specific thalamic nuclei (Komura *et al.*, 2005; Tyll *et al.*, 2011). However, the modulation of fMRI-signals in putatively unisensory cortex – observed in behaving humans – could either be due to integration at the thalamic level, direct cross-talk between sensory-specific cortex or feedback from higher convergence zones (Driver & Noesselt, 2008). These scenarios can only be disentangled if the time course underlying the fMRI-signal modulations is known.

In principle, electrophysiological recordings from animals and humans could shed light on the temporal dynamics but have provided mixed results. Using invasive electrophysiological recordings in anesthetized rats together with pharmacological silencing of V1, Sieben and colleagues (2013) observed an initial superadditive boost of tactile responses in tactile cortex due to visual co-stimulation even after silencing of V1. This finding suggests that initial crossmodal interplay might already occur at the level

of the thalamus. Moreover, animal studies on audiotactile and audiovisual interactions in macaques (Lakatos *et al.*, 2007; Lakatos *et al.*, 2009) have hinted at phase resetting of ongoing brain responses in cortical areas as a potential mechanism governing MSI at the neuronal level. In contrast, only some human event-related potential (ERP)-studies have observed low-latency modulations of auditory evoked potentials (around 50 ms) due to co-stimulation in the tactile (Murray *et al.*, 2005) or visual modality (e.g. Cappe *et al.*, 2010; Giard & Peronnet, 1999) but the exact pattern of results vary: Some studies observed superadditive responses (e.g. Giard & Peronnet, 1999; Molholm *et al.*, 2002) while others observed subadditive effects (Cappe *et al.*, 2010), and some even linked these early ERP effects to unspecific preparatory brain responses (Teder-Sälejärvi *et al.*, 2002). Moreover, EEG-studies investigating the neural correlates of multisensory perception (e.g. the ventriloquist illusion) have often reported modulations at later processing stages (e.g. Bonath *et al.*, 2007; Bruns & Roder, 2010; Busse *et al.*, 2005). On the other hand, using MEG Raij et al. (2010) reported effects in visual cortex due to auditory stimulation already after 50 ms. Concordantly, electrocorticographical (ECoG) results in humans point at a modulation by auditory stimuli in primary visual cortex already after 28 ms (Brang *et al.*, 2015). More recently, some authors have linked this auditory co-activation of visual cortex to an enhancement of visual sensitivity (Feng *et al.*, 2017; van der Burg *et al.*, 2011). Taken together, while there is some evidence for low-latency effects in in non-human animals, the emerging picture in humans is less obvious and the functional relevance of low-latency modulations is still debated.

Some EEG-studies have looked into sound-induced visual benefits (Feng *et al.*, 2017; van der Burg *et al.*, 2011), yet the reverse effect of irrelevant visual stimulation on sound perception has gained considerably less interest. So far, there is some evidence of visually induced auditory enhancement at the behavioural level (Lovelace *et al.*, 2003; Odgaard *et al.*, 2004), but none concerning the temporal dynamics of this effect. However, this topic is of utmost theoretical importance as visually-induced enhancements of auditory perception may be governed by a different mechanism than the sound-induced enhancement of visual perception. In particular, auditory stimuli reach the cortex already after 10 – 15 ms due to the mechanical signal transduction in the cochlea. Visual signals are slower due to the chemical transduction in the retina and reach the visual cortex around 45 – 50 ms (see e.g. Brang and colleagues (2015) for results from invasive recordings in humans). Thus, auditory stimuli could in principle reach the visual cortex earlier than a coincident visual stimulus and act as an advance cue to potentially modulate neural activity there. For the reverse case, this would not be possible. Here, visual stimulation could affect auditory processing only after 40 – 50 ms, a time period when the initial auditory processing has already occurred. Thus, it is at least conceivable that the mechanism in auditory cortex to increase auditory perception for coincident visual and auditory stimuli might differ and may occur at a later processing stage.

Here, we tested this hypothesis in two experiments focusing on fMRI-signal changes in auditory cortex and auditory evoked potentials. In both experiments we used a psychophysical approach to determine if

participants' perceptual sensitivity would increase for near-threshold auditory stimuli when paired with visual stimuli but not for above-threshold auditory stimuli paired with visual stimuli (relative to sound-only conditions). In a first psychophysical-fMRI experiment, we identified the neural basis of the visually-induced enhancement in sound perception in auditory cortex. In a second psychophysical-EEG experiment, the time course of this auditory enhancement was characterised. We hypothesised that visual co-stimulation should enhance the perceptual sensitivity of low-intensity auditory stimuli. Further, fMRI-signals in low-level auditory cortex should be elevated by this visually-induced gain in auditory sensitivity. Finally, mid-latency auditory potentials (P50) should be modulated by the visually-induced enhancement of auditory perception.

Methods

fMRI-experiment

Subjects

In the fMRI-experiment 20 healthy subjects (7 females, age 26.2 ± 3.6 years, range: 21 – 33 years) participated after providing written informed consent. All subjects reported normal hearing, normal or corrected vision, and claimed to be free of any psychiatric or neurological disorders. The experiment was approved by the local ethics commission of the Otto-von-Guericke University.

Apparatus and stimuli

The fMRI-experiment was programmed and presented using Matlab (Version R2011b, The Mathworks, Natick, USA) and the Matlab Toolbox Cogent 2000 (Version 1.3, http://www.vislab.ucl.ac.uk/cogent.php, London, UK). The visual stimuli were projected onto a display via a projector (VC, DLA-G150 CLE, 60 Hz). The subjects saw the display through a mirror mounted on the head coil. Background colour was light grey (RGB: [127,5 127,5 127,5]). A fixation cross was presented in the middle of the display during all experimental conditions. A black and white checkerboard (size: 1 ° x 1 ° visual angel, duration: 100 ms) was used as visual stimulus and presented in the right upper quadrant of the display (5 ° visual angle left and 1 ° visual angle above the fixation cross). Sounds were delivered using a piezo-electric loudspeaker which was connected to a programmable attenuator (Tucker Davis Technologies, Alachua, USA) outside the MRI-room via an electrical shielded cable. The loudspeaker was placed inside the magnet bore above the right upper quadrant of the mirror. Pure sinus tones (frequency: 2000 Hz, duration: 100 ms), which were presented near threshold (3 dB above auditory threshold) and at a high intensity (60 dB above auditory threshold) were individually chosen for each participant and served as targets. In the audiovisual conditions, visual and auditory stimuli were presented simultaneously.

Experimental Design and Experimental Procedure

A 3 x 2 experimental design with the factors audition (factor levels: no sound/low-intensity/high-intensity) and vision (factor levels: visual stimulus present/absent) was used and resulted in six experimental conditions in the main experiment. Low- and high-intensity sounds were individually adjusted to meet predefined accuracy criteria (low-intensity sounds: 50-75 % correct, high-intensity sounds: > 90 % correct).

The fMRI-experiment comprised an initial auditory threshold determination, a threshold confirmation run with scanner noise, and the main psychophysical-fMRI experiment.

First, the auditory threshold was individually determined via a staircase procedure using intensity steps of 2 dB. Predefined performance level for the low-intensity auditory stimulus was set within the range of 50-70 % accuracy. Further, threshold confirmation runs were implemented to confirm that the auditory threshold was not affected by scanner noise. Here, high- and low-intensity sounds were presented in gaps introduced between volumes (see below for scanning protocol). For threshold confirmation at least three runs with 50 trials were presented. After each run the sound intensity was adjusted to keep the performance level of the low-intensity auditory stimulus within the range of 50-70 % accuracy.

The main experiment consisted of six runs à 180 trials (see Figure 1 for experimental design) in which all conditions were included and presented in random order. The probability of occurrence was chosen so that the ratio between conditions with tone and without as well as the ratio between conditions with visual stimulus and without was balanced (high-intensity sound: 4 %, high-intensity sound & V: 4 %, low-intensity sound: 21 %, low-intensity sound & V: 21 %, V: 25 %, blank: 25 %). Hence, the visual stimulus carried no information about target occurrence.

For the threshold confirmation runs as well as the main experiment the interstimulus interval (ISI) was 2400 ms. Stimuli were presented for 100 ms and after a time period of 200 ms the participants had 1000 ms to respond. A question mark above fixation served as response cue. Participants had to detect the presence or absence of the auditory stimulus (detection task) irrespective of the presence of the irrelevant visual stimulus (only presented during the main experiment). Participants responded with their right index and middle fingers. All stimuli were presented during the 500 ms long scanner noise free intervals. Moreover, null events were included to be able to estimate condition-specific BOLD-responses (Henson, 2007). The respective conditions and the null events were presented in randomized order.

Insert Fig 1 around here

Analysis of behavioural data

The behavioural data was analysed using SPSS (Version 13.0, SPSS Inc, IBM Corporation, Armonk, USA). A 2 x 3 repeated measures ANOVA with two within subject factors (audition: high-intensity sound, low-intensity, no sound; and vision: present/absent) was used to analyse accuracy and was Greenhouse-Geisser-corrected, if appropriate. For post-hoc analyses, paired t-tests were used.

Moreover, signal detection theory was used to distinguish perceptual sensitivity (d') from response bias. Perceptual sensitivity was calculated as follows (Green & Swets, 1966), whereby $\Phi-1$ is a function that converts probabilities into z-values:

$d = \Phi-1\ (hits) - \Phi-1\ (false\ alarms)$

Scanning protocol

A 3-Tesla MRI-scanner (SIEMENS MAGNETOM Trio syngo MR A35, Siemens Medical Systems, Erlangen, Germany) equipped with a Bruker head coil (Bruker-BioSpinMRI, Ettlingen, Germany) were used for MR data acquisition. Participants wore earplugs. Eye movements were monitored with a MRI-compatible camera (Kanowski *et al.*, 2007) and the software PupilTracker (HumanScan, Erlangen, Germany). For the functional measurements gradient-weighted T2*-weighted echo planar imaging (EPI, in-plane resolution: 3 x 3 mm^2, 80 x 80 voxel, slice thickness 3.5 mm, 34 slices, TR: 2000 ms, inter-volume gap: 500 ms, TE: 30 ms, flip angle: 80 °) were used (194 volumes per run, 6 runs). T1-weighted inversion recovery-EPIs with identical distortions as the EPI volumes (IR-EPI, TE: 40 ms TI: 1450 ms, all other parameters identical to the T2-weighted EPIs) were recorded for each participant to determine normalization parameters and for anatomical overlay.

fMRI analysis

The fMRI data was analysed using Statistical Parametric Mapping (SPM 12, Wellcome Trust Centre for Neuroimaging, London, UK) and Matlab (Version R2015b, The Mathworks, Natick, USA). After exclusion of the first 3 volumes to account for saturation effects, pre-processing of the data included slice-time correction, motion correction, co-registering of anatomical IR-EPI on functional EPI volumes, normalisation and smoothing (FWHM 6 mm).

For statistical analysis at the single-subject level all correct trials of all 6 experimental conditions plus a 7^{th} regressor including all incorrect responses and misses were modelled using the hemodynamic

response and its temporal derivate. Further, individual motion parameters were included in the single-subject first-level models. All contrasts of interest of all participants were then used in a flexible factorial design in a group-level random-effects model. To identify the auditory cortex while avoiding double dipping (Kriegeskorte *et al.,* 2009), Neurosynth (Yarkoni *et al.,* 2011) was used to create an independent auditory mask (search term 'auditory', search performed on 17.05.2017; the auditory mask was thresholded at $p < 0.01$, FDR-corrected).

To detect correlations between the subject-specific accuracy gain (AV > A) and the regional BOLD-differences (AV > A), the visually-induced enhancement of accuracy was included in an additional statistical analysis as covariate. For the analysis of the relationship of BOLD-signal changes and changes in perceptual sensitivity we computed the fMRI-contrast (sound paired with vision + no target) vs. (sound + no target with vision) to closely resemble the difference in d'. Subject-specific behavioural gains were checked for outliers (mean +/- 2 standard deviations), and outliers were removed from these analyses to avoid spurious correlations. SPM Anatomy Toolbox (Anatomy 1.5, Research Center Juelich, Juelich, Germany) was used for a detailed labelling of activated areas.

EEG-experiment

The EEG-experiment was identical with the fMRI-experiment except for the following:

Subjects

30 healthy subjects participated after providing written informed consent (22 females, age 25,03 ± 3.99 years, range: 18 – 35 years). Six participants were excluded because their average accuracies did not match the pre-defined criteria (low-intensity sounds: 50-75 % correct, high-intensity sounds: > 90 % correct). One participant was excluded due to strong artefact contamination (> 70 % of trials).

Apparatus and Stimuli

Stimuli were presented using the Psychophysics Toolbox (Brainard, 1997) and Matlab 2012b (Mathworks Inc.). Visual stimuli were presented on a LCD screen (22", SAMSUNG 2233RZ, resolution: 1650 x 1080, refresh rate: 60 Hz) with optimal timing and luminance accuracy for vision researches (Wang & Nikolić, 2011). In the main experiment, participants completed a minimum of 12 runs with 100 trials each. No null events were included.

The checkerboard (size: 0.849 ° x 0.849 ° visual angel) was again presented in the right upper quadrant (visual angle 4.844° horizontal, 4.844° vertical) and below a speaker (distance 7.352 °).

Experimental design

Again, stimuli in the six different conditions were presented for 100 ms. A question mark served as response cue and appeared 800 ms to 1200 ms after stimulus onset. Thus, motor response activity in the analysed EEG time interval (< 500 ms) was reduced, and the onset-jitter reduced motor preparation activity. Participants had one second to respond after which a blank, variable inter-trial-interval was initiated (1300 ms to 1700 ms). Participants' task was again to detect the presence or absence of the auditory stimulus while ignoring any visual co-stimulation. They responded using both thumbs and response-finger mapping was counterbalanced across participants.

EEG Recordings

Participants were seated in a dimly-lit, sound-attenuated chamber. EEG was recorded with a BioSemi Active-Two amplifier system from 64 Ag/AgCl electrodes arranged according to the international 10-10 system, and with two extra, external mastoid electrodes. Moreover, horizontal (2 electrodes) and vertical (1 electrode) electro-oculogram (EOG) was recorded from three additional channels to monitor and later correct for eye movements and blinks. EOG electrodes were placed at the outer canthi of both eyes and infraorbital ridge of the left eye. Two additional electrodes (CMS: Common Mode Sense and DRL: Driven Right Leg) were used as reference and ground (standard for BioSemi system). Signals were sampled at 2056 Hz with 24-bit conversion resolution and 410 Hz low-pass filter (low-pass filter set to approx. 1/5 of the sampling rate). After recording, data were down-sampled to 256 Hz, high-pass filtered at 0.15 Hz, low-pass filtered at 40 Hz, converted to an average reference, and epoched from -500 – 500 ms time-locked to stimulus onset. After removing large artefacts such as electrode drifts and muscle activity, independent-component analysis (ICA) was applied to the raw data to correct for eye blinks and eye movements using the MATLAB toolbox EEGLAB and the extended infomax ICA algorithm (Delorme & Makeig, 2004). Corrected raw data were epoched from -100 – 500 ms (as we focused on auditory event-related potentials), converted to an average mastoid reference and baseline corrected (i.e. -100 – 0 ms) after final inspection of the data.

EEG Analysis

The EEG analysis aimed at identifying the temporal dynamics of visually-induced benefits in auditory perception. For this analysis, we identified common auditory ERP-components in the time range of the P50, N1, and P3. In particular, we computed the differences between the unisensory auditory stimulation and the audiovisual stimulation. Specifically, we subtracted the vision-related ERPs from bimodal ERPs (AV – V) as well as the blank condition from the sound condition (A – blank). These subtractions were

calculated separately for the high- and low-intensity sound conditions. The resulting difference waveforms eliminate the effect of physical visual stimulation and should only reveal non-linear influences on auditory processing due to visual co-stimulation.

Results

Behavioural Results

Participants of the fMRI-experiment performed better in the high-intensity sound conditions than in the low-intensity sound conditions, as expected (high-intensity: 98.6 %; low-intensity: 73.5 %). Moreover, there was also an increase in accuracy for the low-intensity sound condition if paired with vision relative to the low-intensity sound condition alone (77.1 % and 69.9 %, respectively). Even performance in the no-target condition was slightly increased if the no-target was paired with vision (non-target plus sound: 90.9 % vs non-target alone 89.9 %), indicating that visual co-stimulation did not induce a response bias. A repeated measures ANOVA confirmed these observations and revealed main effects of vision and audition (audition [high-intensity/low-intensity/no sound]: $F(1,19) = 42.3$, $p = 2.17*10^{-10}$; vision [present/absent]: $F(1,19) = 20.3$, $p = 2.4*10^{-4}$). Most importantly, the interaction term also reached significance ($F(2,38) = 8.1$, $p = 1.15*10^{-3}$). Post-hoc t-tests (Bonferroni corrected [pBF] when appropriate) confirmed a significant difference between low-intensity auditory stimuli with and without sounds ($T(19) = 4.6$; $pBF = 5.94*10^{-4}$) with 16 out of 20 subjects showing this effect, which was not found for high-intensity conditions and no-target conditions (pBF = .51 and pBF = 1, respectively).

To confirm that this increase in performance was indeed due to an increase in perceptual sensitivity, a complementary t-test was calculated for the d' of the low-intensity conditions. Note that we did not include high-intensity conditions because a majority of participants (n = 11) performed at ceiling there. For the low-intensity condition a significant increase in sensitivity was found if the tone was paired with vision ($T(19) = 2.9$, $p = .009$).

The pattern of results for the EEG-experiment was virtually identical: High-intensity sounds were perceived more accurately as low-intensity sounds, as expected (99.5 % vs. 58.6 %); and low-intensity sounds were perceived more accurately if paired with visual stimuli vs. without (55.3 % vs 61.9 %). Here, even the no-target trials showed an increase in accuracy if paired with a visual stimulus (93.0 % and 94.2 %). For accuracy we again observed main effects of audition ($F(1.6,34.8) = 415.3$, $p = 1.24*10^{-23}$), vision ($F(1,22) = 18.9$, $p = 2.56*10^{-4}$) and most importantly an interaction of vision and audition ($F(1.3,27.8) = 18.0$, $p = 8.3 * 10^{-5}$). Accordingly, post-hoc t-tests indicated that the visually-induced benefit was most pronounced for low-intensity sounds with vision but reduced in the no-target condition and high-intensity sound condition (low-intensity: $T(22) = 4.7$, $pBF = 3.51*10^{-4}$; no-target: $T(22) = 2.3$, pBF = .093; high-intensity: pBF = 1). Concordantly, an analysis of perceptual sensitivity revealed that low-intensity sounds were perceived better when paired with vision ($T(22) = 4.7$, $p = 1.11*10^{-4}$).

Finally, a direct comparison of accuracies across experiments again revealed main effects of sound ($F(1.5,62.2) = 256.1$, $p = 6.09*10^{-28}$) and vision ($F(1,41) = 39.5$, $p = 1.69*10^{-7}$). Auditory intensity interacted with experiments ($F(1.5,62.2) = 20.6$, $p = 1*10^{-6}$), which was due to a higher overall performance in the low-intensity conditions in the fMRI- than the EEG-experiment. Post-hoc t-tests indicate that overall performance was enhanced in the low-intensity sound condition in the fMRI-experiment relative to the EEG-experiment (two-sample t-tests for low-intensity: $T(41) = 4.8$, $pBF = 6*10^{*-5}$; high-intensity: $pBF = .63$; no-target: $pBF = .39$). Most importantly, the crucial interaction of vision and audition did NOT interact with experiment ($p = .75$) but again revealed a significant increase for low-intensity sounds if paired with vision across both experiments ($F(1.5,60.8) = 23.2$, $p = 5.03*10^{-7}$; post-hoc paired t-tests for low-intensity: $T(42) = 6.6$, $pBF = 15*10^{-8}$; high-intensity: $pBF = .78$; no-target: $pBF = .24$). Likewise, the increase of perceptual sensitivity for the low-intensity conditions across experiment [repeated measures ANOVA with between subject factor experiment (EEG/fMRI)] revealed a main effect for vision on perceptual sensitivity ($F(1,41) = 18.3$, $p = 1.1*10^{-4}$) which did not interact with the experimental set-up (EEG vs. fMRI: $F(1,41) = 2.8$, $p = .105$). Together, this pattern of results indicates that participants of both experiments successfully used the visual information to increase perceptual sensitivity.

fMRI Results

For the fMRI-results, we first compared sound with no-sound conditions and vision present with vision absent conditions as manipulation checks. As expected, for the contrast [vision > no-vision conditions] enhanced fMRI-signals were found in visual cortex mostly contralateral to the stimulated hemifield (Fig. 2a, Table 1a, $p < 0.05$, FWE-corrected, $k > 50$). For the contrast [sound > no-sound] conditions low-level auditory areas and bilateral medial geniculate bodies (MGB) of the thalamus expressed enhanced fMRI-signals (see Fig. 2b, Table 1b, $p < 0.05$, FWE-corrected, $k > 50$).

Insert Fig 2 around here

We then compared the effects of sound intensity in auditory regions (see Fig. 3a for auditory mask). High-intensity sounds (relative to low-intensity sounds) activated low-level auditory cortex ($p < .05$, FWE-corrected) and even auditory thalamus (Fig. 3b, Table 2a; note that we restricted our analysis to auditory areas independently identified by posterior probability maps provided by Neurosynth, Yarkoni *et al.*, 2011). We next tested whether low-intensity auditory stimulation (with and without sound) would lead to increased fMRI-signals thereby identifying loudness-sensitive auditory areas. As predicted, we observed an increase in left-hemispheric superior temporal gyrus (STG) extending into the superior temporal sulcus (STS), albeit at a lower statistical threshold (Fig. 3c, $p < .001$, uncorrected at voxel level

but FWE-corrected at cluster level within the auditory mask). Note that low-intensity auditory stimuli generally evoke smaller fMRI-responses (Behler & Uppenkamp, 2016).

Insert Figure 3 around here

The critical comparison tested for nonlinear enhancements of audiovisual stimulation relative to unisensory auditory and visual stimulation. We observed a modulation of the STS adjacent to loudness-sensitive temporal cortex by visual co-stimulation in the left hemisphere (Fig. 4a, Table 3a) and even the auditory thalamus (Table 3a). Unexpectedly, for the high-intensity sound condition paired with vision we also found enhanced responses in bilateral low-level auditory cortex and the thalamus (Fig. 4b, Table 3b).

Insert Figure 4 around here

Finally, we tested for a direct link between fMRI-signal modulations and participants' behavioural benefits. To this end, we calculated for each participant the difference between the audiovisual and unisensory conditions for the two measures, the BOLD response and accuracy/d' ([AV-A]), and tested for a correlation of both. Indeed, we observed a significant relationship of accuracy gain with BOLD-increase in bilateral low-level auditory cortex (Fig. 5a, Table 4a). Automatic labelling with the SPM Anatomy toolbox indicated that this effect already occurs in TE 1.0 and TE 1.1, i.e. includes primary auditory cortex. Concordantly, a modulation of BOLD-responses in virtually identical areas was observed for the correlation of BOLD signal differences and perceptual sensitivity (Fig. 5b, Table 4b).

Insert Figure 5 around here

EEG Results

For the ERP-results the high-intensity sound conditions were analysed first, as robust ERP-components could be expected for these more salient stimuli (Hegerl & Juckel, 1993). Components and time-intervals were identified by visual inspection (of the unisensory and multisensory difference waves) which was based on common locations and time ranges for these components reported in the literature

(Hillyard, 1993). At frontal electrodes a P50 component was observed around 30 – 70 ms poststimulus, followed by a subsequent N1 component around 80 – 140 ms over central electrodes, and a subsequent positive deflection again over central electrodes in the time range of the P3, i.e. starting around 180 ms and peaking around 300 ms (see Fig. 6 upper part for topographies and Fig. 7, left side for ERP time course). A statistical comparison of mean voltage in these time windows for the contrast (AV minus V) vs. (A minus blank) revealed a significant enhancement of the P50 component over frontal electrodes for the audiovisual relative to unisensory auditory condition ($T(22) = 2.71$, $p = .013$). Moreover, significant modulations were found for the P3 component, with an enhanced positive deflection for audiovisual stimuli over central electrodes in the time range of 180 – 240 ms ($T(22) = 4.69$, $p = 1.11*10^{-4}$) and an inversion of this effect ([A - blank] > [AV - V]) at central and parietal electrodes in the range of 250 – 400 ms ($T(22) = -2.13$, $p < .045$). The effect for the N1 component was non-significant ($T(22) = .17$, $p = .866$).

Insert Figure 6 around here

Turning to the low-intensity stimuli, similar components were observed albeit at a longer latency, possibly due to the lower-intensity of the auditory stimulus (see Fig. 6 lower part for topographies and Fig. 7, right side for ERP time course). A statistical comparison of the P50 (30 – 100 ms), N1 (90 – 240 ms) and P3-components (280 – 400 ms) only revealed a significant modulation in the time range of the P3 (i.e. starting around 280 ms) with a suppression of audiovisual relative to auditory stimuli ($T(22) = -2.47$, $p = .022$). P50 and N1 did not reach significance ($p = .972$ and $p = .233$, respectively).

Insert Figure 7 around here

One of the reasons for the reduction of effects might be that low-intensity stimuli do not only shift the peak latency to a later point in time, but the variability of peak latencies across participants may increase with reduced intensity as well. We therefore identified subject-specific peaks within the time range of the auditory ERP-components and used the peak voltage (plus a window of ±10 ms for early and short components, and ±30 ms for the longer P3) for a more in-depth analysis. However, again a significant modulation was only observed for the late time range of the P3 over centro-parietal electrodes ($T(22) = -2.45$, $p = .023$) with a suppression of audiovisual relative to auditory stimuli. P50 and N1 did not reach significance ($p = .297$ and $p = .441$, respectively).

Finally, to directly test for a link of the EEG-signal with behaviour, we correlated the audiovisual increase in perceptual sensitivity (i.e. the d' gain: $d'_{AV\ low} - d'_{A\ low}$) point-by-point with the difference EEG-signal (i.e. $EEG_{(AV\ low\ -\ V)} - EEG_{(A\ low\ -\ blank)}$) for all EEG channels separately (see Fig. 8, top). The point-wise correlation was corrected by using a cluster cut-off criterion (significant regression coefficient ($p < .01$) in 5 successive time points, see e.g. Myers *et al.* (2014) for a similar approach). This analysis revealed a significant effect over right-sided frontal electrodes (see Fig. 8, top) in the time range of the P3 (rho = .736, $p = 1*10^{-4}$). While the aforementioned approach provides an unbiased analysis of the data it has at least one shortcoming: if components are not precisely localized in time across participants, correlations are likely to be non-significant. Thus, correlations of behaviour with lower-latency components (such as the P50) might be obscured if the latencies of these components vary across participants. This potential problem may less affect later components as these occur over a wider time range (slow components).

As activity in the primary auditory cortex (see fMRI results) indicated that MSI may occur rather early in the ERP, we additionally analysed the P50 amplitude (i.e. the multisensory benefit: $EEG_{(AV\ low\ -\ V)} - EEG_{(A\ low\ -\ blank)}$) and correlated it with the audiovisual increase in perceptual sensitivity (i.e. the d' gain: $d'_{AV\ low} - d'_{A\ low}$; see Fig. 8, bottom). Here, we used the subject-specific peaks of the P50 to reduce the variability in latency across participants. Indeed, this analysis revealed a correspondence of behaviour with electrophysiology for the P50 similar to the observed fMRI-behaviour relationship (rho = .38; $p = 3.78*10^{-2}$). Hence, this effect appears to closely resemble the correlation of fMRI-responses and behaviour in low-level auditory cortex for which we observed a significant relationship of subject-specific fMRI-signals with behaviour.

Put Figure 8 around here

Discussion

In this study we tested whether visual stimuli would enhance auditory perceptual sensitivity, and characterized the neural underpinnings and temporal dynamics of this effect. In two independent experiments, we consistently observed a robust increase in perceptual sensitivity for lower-intensity sounds paired with visual stimulation. Concordantly, with fMRI a difference in fMRI-signal was observed in auditory STS. Importantly, a direct link between subject-specific increase in perceptual sensitivity and BOLD-signal was observed in low-level auditory regions including TE 1.0. For EEG, modulations of low-intensity sounds plus vision were only observed in a later time range starting around 280 ms over central electrodes when analysing ERP mean-amplitudes. Moreover, a correlation of ERPs and perceptual sensitivity was also observed in the late ERP time range. However, this ERP-behaviour relationship showed smaller behavioural gains for increased ERP-amplitudes unlike the relationship

found with fMRI. Most importantly, for subject-specific amplitude modulations we also observed an interplay of ERP-signals and perceptual sensitivity at an early time range (P50) closely resembling the effects observed with fMRI in low-level auditory cortex including A1 (TE 1.0).

Our behavioural results are in good agreement with earlier behavioural studies reporting enhanced detectability for sounds if paired with visual co-stimulation (Lovelace *et al.*, 2003; Odgaard *et al.*, 2004). Moreover, the pattern of results also fit well with the large body of evidence reporting beneficial effects of non-informative co-stimulation in a task-irrelevant modality on perceptual sensitivity in an attended modality (Feng *et al.*, 2017; Frassinetti *et al.*, 2002; Hoefer *et al.*, 2013; Jaekl & Soto-Faraco, 2010; Lovelace *et al.*, 2003; Noesselt *et al.*, 2010). Hence, the increase in perceptual sensitivity supports the notion of a general multisensory mechanism which enhances sensory representations by task-irrelevant co-stimulation in a second modality.

The neural basis of visually-induced enhancements of low-intensity sounds was observed in auditory STS close to loudness-sensitive auditory regions in the left STG/STS, contralateral to the visual stimulus. Here, the fMRI-signal for low-intensity sounds was elevated if sounds were paired with a visual stimulus. In accord, several previous imaging studies have linked the STS to audiovisual integration (e.g. Hein & Knight, 2008; Noesselt *et al.*, 2010; Noesselt *et al.*, 2012; Stevenson & James, 2009; Venezia *et al.*, 2017; Watson *et al.*, 2014; Werner & Noppeney, 2010). Note, however, that many of these previous imaging studies have used audiovisual speech-stimuli, which might not be comparable to the simple stimuli used here. In particular, during speech, mouth movements precede vocalisation by about 100 – 150 ms (Schroeder *et al.*, 2008). Hence, changes in the visual speech signal can serve as a temporal cue to inform the auditory modality that auditory information is imminent, unlike here. In our two experiments, non-semantic visual and auditory information were presented synchronously. Here, we corroborated previous imaging studies on audiovisual interplay with non-speech stimuli (Noesselt *et al.*, 2010; Werner & Noppeney, 2010). These studies reported modulations in STS for non-semantic audiovisual stimuli, and directly linked them to subject-specific changes in behaviour, including sound-induced visual contrast enhancements (Noesselt *et al.*, 2010). We extend these findings to the paradigm of visually-induced increases in auditory sensitivity for simple stimuli.

Most importantly, a direct link between brain responses and behaviour was already observed in primary auditory cortex (TE 1.0). This pattern of results lends further support to the notion that increases in perceptual sensitivity are indeed related to modulations of sensory representations and that this modulation can be caused by a visual stimulus. Accordingly, previous unisensory auditory studies have linked modulations in low-level auditory cortex with perceived loudness in addition to an increase in physical sound pressure level (e.g. Behler & Uppenkamp (2016)). Moreover, several imaging studies on audiovisual interplay reported modulations of neural signalling in low-level auditory cortex (including A1) in animals (Brosch *et al.*, 2005; Kayser *et al.*, 2008; Kayser *et al.*, 2009) and humans

(Lehmann *et al.*, 2006; Noesselt *et al.*, 2007; Noesselt *et al.*, 2010). In our study we demonstrate that this increase in BOLD-response in auditory regions is related to enhanced auditory performance.

Remarkably, we also observed some indication of a direct link between event-related potentials and behaviour in the time range of the auditory P50 paralleling the effects found in the fMRI-experiment. These findings are in line with electrophysiological effects due to multisensory stimulation in a similar time window (Cappe *et al.*, 2010; Giard & Peronnet, 1999; Mercier *et al.*, 2013; Murray *et al.*, 2005; Senkowski *et al.*, 2011; Talsma *et al.*, 2007). However, previous effects within this time range were either observed in no-go trials (Cappe *et al.*, 2010) or resulted in speeded reaction times (Mercier *et al.*, 2013; Senkowski *et al.*, 2011) but not an increase in perceptual sensitivity as found here. Although we observed a medium sized ERP-behaviour relationship (rho = .38), the group-mean P50 for the low-intensity sounds was not enhanced by visual stimulation. This could be due to the smaller P50 elicited by low-intensity sounds, i.e. a poorer signal-to-noise ratio (SNR), which is unavoidable when working with near-threshold stimuli. Thus, a lower SNR, may reduce the chance of identifying differential effects across conditions, unlike the higher-intensity sounds which elicited a robust P50.

Notably, higher-intensity sounds plus vision also enhanced both, average fMRI-signals in low-level auditory cortex and the P50. However, these imaging and electrophysiological findings were not mirrored by a behavioural benefit as accuracy was always at ceiling. In sum, for low-intensity sounds, we observed that an increase in audiovisual P50 scaled with behavioural performance but found no overall enhancement of the P50 by vision; and we found the exact opposite for the high-intensity sound conditions in both experiments, i.e. an increase in mean BOLD-response and mean amplitude of the P50. A possible explanation to reconcile these findings could be that the influence of visual input on processes in auditory cortex occurs automatically. In case of low-intensity sounds, the increase in amplitude of the audiovisual P50 might be masked by noise (see also Picton et al., 1997; Rapin et al., 1966 & Tucker et al., 2001) while it is observed more readily at higher intensity levels. At the behavioural level, this automatic neuronal interplay can only affect the behavioural outcome for near-threshold stimuli as performance is well below ceiling. In contrast, no behavioural benefits can be observed for higher-intensity sounds as performance is already at ceiling there.

Given the latency of this modulation, the functional neuroanatomy underlying such automatic interplay may be found at the level of the thalamus which receives visual and auditory input (Henschke *et al.*, 2015; Komura *et al.*, 2005; Noesselt *et al.*, 2010; Tyll *et al.*, 2011). Indeed, we observed a robust modulation of the auditory thalamus for auditory stimuli paired with vision, especially for higher-intensity sounds paired with vision. Thus the enhancement of the audiovisual P50, which was observed for higher-intensity sounds paired with vision, might be due to interactions within the thalamus.

One functional role the thalamus has been implicated in is gatekeeping; and an electrophysiological marker for gatekeeping is indeed the P50 as this component is consistently reduced for redundant auditory stimuli (Smith *et al.*, 1994). In our audiovisual context, however, the size of the P50

enhancement scales with elevated behavioural performance. This may suggest that concurrent audiovisual stimulation acts like inverse gatekeeping, thereby leading to a bottom-up driven increase in perceptual sensitivity. Possible related effects have recently been reported for an auditory detection task (Sadaghiani *et al.*, 2009), where higher pre-stimulus activity in the dorsal attentional network effectively reduced performance, whereas higher activity in the default mode network increased behavioural performance. This may also explain the inverse relationship of behavioural benefit and ERP-enhancement for the P3. Recall, that we observed a decrease in the mean amplitude of audiovisual P3 over central electrodes and that this effect scaled inversely with performance. While some have linked modulations of the P3 to auditory sensitivity of near threshold auditory stimuli (Hillyard *et al.*, 1971) most authors have related this component to several post-perceptual phenomena, including memory update and awareness (Railo *et al.*, 2011). However, it is not obvious how this modulation of a late and putatively cognitive component could fit the pattern of behavioural results, as behavioural performance suggested that a modulation of sensory representations should occur. Hence, one would expect to observe a modulation of sensory ERPs. If the late P3 modulation would be related to a change in neural signalling in auditory cortex, this could be due to feedback signals from higher cortical regions which may be more commonly associated with the P3 (Railo *et al.*, 2011). This large effect in higher cortical areas might also mask concurrent but smaller sized modulations in auditory cortex.

However, this alternative explanation does not account for the direction of the brain-behaviour relationship found with fMRI. There, we observed a coupling of increased fMRI-signals with increased behavioural performance for the AV relative to the A condition. For the P3-component, this relationship was reversed. One could now argue that ERP results and fMRI results do not always fully match, as fMRI accumulates all signals of a given trial, whereas ERPS reflect signal processes within a trial. But we observed a similar pattern of brain-behaviour-coupling for fMRI in low-level auditory cortex and the P50. Thus, the P3 modulation might rather be due to cognitive processes which may in fact hinder audiovisual integration, similar to the effects observed in an unisensory auditory detection task (Sadaghiani *et al.*, 2009).

In sum, the comparison of results of Experiments 1 and 2 revealed remarkable similarities across experiments. For high-intensity sounds we observed increased amplitudes of the P50 for audiovisual stimuli in the ERP-experiment together with enhanced BOLD-responses for audiovisual stimuli in primary auditory cortex in the fMRI-Experiment. For low-intensity sounds audiovisual stimulation enhanced fMRI-signals in low-level auditory cortex. This effect was not mirrored by a similar increase in audiovisual ERPs. It should be noted, however, that accuracy for the low-intensity conditions were also lower in the ERP- relative to the fMRI-experiment. It may be the case that this lower overall accuracy in the ERP-experiment further decreased the signal to noise ratio, thereby reducing the amplitude differences for low-latency responses to auditory and audiovisual stimuli. Nevertheless, the behavioural benefit for low-intensity sounds when paired with sounds was observed in both experiments

and was paralleled by significant brain-behaviour correlations in both the fMRI- and EEG-experiment. While the fMRI-experiment revealed a positive correlation of behavioural gain and fMRI-signal in low-level auditory cortex, the EEG-experiment provided evidence for a positive correlation in the time range of the P50 but also for a negative correlation in the time range of the P3. Although we focused in this study on the auditory cortex, we decided to test the fMRI-signal for a negative correlation of fMRI-signal and behaviour outside auditory cortex. Indeed, a negative relationship of fMRI-signal was observed in right-hemispheric superior and middle frontal gyrus (at $p < 0.03$, $k > 50$; SFG: $k = 276$, $T = 4.17$; $p = 3.41*10^{-4}$; MNI = 22/-2/54; MFG: $k = 202$; $T = 4.04$; $p = 4.29*10^{-4}$; MNI = 44/22/34). This finding supports the suggestion that P3 activity and its correlation with behaviour was linked to activity in higher cortical areas. Taken together, we observed a striking similarity between fMRI- and ERP-results. Both experiments consistently revealed a positive correlation of multisensory behavioural benefit with neural signals in auditory cortex in line with our hypotheses. Hence, our results suggest that subject-specific visually-induced enhancements in auditory perception are consistently linked to elevated BOLD-responses in low-level auditory cortex and early modulations in the time range of the P50.

The underlying small circuit mechanisms supporting visually-induced changes in auditory processing still need to be elucidated. Schroeder & Lakatos (2009) have suggested that phase resetting of ongoing oscillatory activity could serve as a general underlying mechanism. This mechanism would enhance sensory processing by a signal from the second modality which occurs just before the task-relevant stimulus reaches the cortex. Further, the authors provided evidence for phase resets in primary visual cortex due to auditory co-stimulation and in primary auditory cortex due to tactile co-stimulation (Lakatos *et al.*, 2007; Lakatos *et al.*, 2008). In both cases, the second modality can influence the task-relevant modality as signal transduction times are similar or even faster for the second modality. In our paradigm auditory stimuli coincided with visual stimuli, which would lead to auditory information reaching the cortex before visual information (Brang *et al.*, 2015). Hence, a phase reset of auditory cortex might not necessarily help ongoing sensory processing. Others however, have pointed at more complex relationships between frequency bands and at specific time points which may increase or reduce sensory processing (Naue *et al.*, 2011). Future invasive electrophysiological studies in auditory cortex are needed to comprehend the exact neural mechanisms underlying our fMRI and ERP effects.

Together, our results point at a modulation of auditory perception by visual stimulation that can be linked to a modulation of brain responses in auditory cortex. These changes in neural processing already start at around 40 ms, which is consistent with a mechanism operating at the level of the thalamus and/or direct crosstalk between sensory-specific cortex and we provide fMRI-evidence for an involvement of the thalamus. However, this modulation was also observed for higher-intensity sounds with behavioural performance already at ceiling. This suggests that this early modulation may reflect an automatic,

bottom-up driven intersensory mechanism which can boost perception in particular contexts independent of stimulus intensity.

Acknowledgements

This work was funded by DFG-SFB-TRR31/TPA8. We thank Rebecca Burnside for revising the manuscript.

Conflict of interest

The authors have no conflicts of interest concerning this work.
Author contributions

Author contribution statement

We hereby declare that all authors (JS, FB and TN) met the four authorship criteria as per the recommendation of ICMJE. All authors made substantial contributions to the conception or design of the work (JS, FB and TN); or the acquisition (JS and FB), analysis (JS, FB and TN), or interpretation of data for the work (JS, FB and TN). All authors contributed drafting and revising the manuscript critically for important intellectual content, approved the final version of the manuscript, and agreed to be accountable for all aspects of the work.

Data accessibility

Requests for supporting data and material should be sent to the corresponding author.

References

Beauchamp, M.S. (2005) *See me, hear me, touch me. Multisensory integration in lateral occipital-temporal cortex. Current opinion in neurobiology*, **15**, 145–153.

Beauchamp, M.S., Lee, K.E., Argall, B.D. & Martin, A. (2004) *Integration of auditory and visual information about objects in superior temporal sulcus. Neuron*, **41**, 809–823.

Behler, O. & Uppenkamp, S. (2016) *The representation of level and loudness in the central auditory system for unilateral stimulation. NeuroImage*, **139**, 176–188.

Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H.-J. & Hillyard, S.A. (2007) *Neural basis of the ventriloquist illusion. Curr. Biol.*, **17**, 1697–1703.

Brainard, D.H. (1997) *The Psychophysics Toolbox. Spatial Vision*, **10**, 433–436.

Brang, D., Towle, V.L., Suzuki, S., Hillyard, S.A., Di Tusa, S., Dai, Z., Tao, J., Wu, S. & Grabowecky, M. (2015) *Peripheral sounds rapidly activate visual cortex: evidence from electrocorticography. Journal of Neurophysiology*, **114**, 3023–3028.

Brosch, M., Selezneva, E. & Scheich, H. (2005) *Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. The Journal of neuroscience : the official journal of the Society for Neuroscience*, **25**, 6797–6806.

Bruns, P. & Roder, B. (2010) *Tactile capture of auditory localization: an event-related potential study. The European journal of neuroscience*, **31**, 1844–1857.

Busse, L., Roberts, K.C., Crist, R.E., Weissman, D.H. & Woldorff, M.G. (2005) *The spread of attention across modalities and space in a multisensory object. Proceedings of the National Academy of Sciences of the United States of America*, **102**, 18751–18756.

Cappe, C., Thut, G., Romei, V. & Murray, M.M. (2010) *Auditory-visual multisensory interactions in humans: timing, topography, directionality, and sources. The Journal of neuroscience : the official journal of the Society for Neuroscience*, **30**, 12572–12580.

Chandrasekaran, C. (2017) *Computational principles and models of multisensory integration. Current opinion in neurobiology*, **43**, 25–34.

Delorme, A. & Makeig, S. (2004) *EEGLAB. An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods*, **134**, 9–21.

Doehrmann, O. & Naumer, M.J. (2008) *Semantics and the multisensory brain. How meaning modulates processes of audio-visual integration. Brain research*, **1242**, 136–150.

Driver, J. & Noesselt, T. (2008) *Multisensory Interplay Reveals Crossmodal Influences on 'Sensory-Specific' Brain Regions, Neural Responses, and Judgments. Neuron*, **57**, 11–23.

Ernst, M.O. & Banks, M.S. (2002) *Humans integrate visual and haptic information in a statistically optimal fashion. Nature*, **415**, 429–433.

Felleman, D.J. & van Essen, D.C. (1991) *Distributed hierarchical processing in the primate cerebral cortex. Cerebral cortex (New York, N.Y. : 1991)*, **1**, 1–47.

Feng, W., Stormer, V.S., Martinez, A., McDonald, J.J. & Hillyard, S.A. (2017) *Involuntary orienting of attention to a sound desynchronizes the occipital alpha rhythm and improves visual perception. NeuroImage*, **150**, 318–328.

Frassinetti, F., Bolognini, N. & Ladavas, E. (2002) *Enhancement of visual perception by crossmodal visuo-auditory interaction. Experimental brain research*, **147**, 332–343.

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

Green, D.M. & Swets, J.A. (1966) *Signal detection theory and psychophysics*. Wiley, New York.

Hegerl, U. & Juckel, G. (1993) *Intensity dependence of auditory evoked potentials as an indicator of central serotonergic neurotransmission. A new hypothesis. Biological Psychiatry*, **33**, 173–187.

Hein, G. & Knight, R.T. (2008) *Superior temporal sulcus—It's my area. Or is it? Journal of Cognitive Neuroscience*, **20**, 2125–2136.

Henschke, J.U., Noesselt, T., Scheich, H. & Budinger, E. (2015) *Possible anatomical pathways for short-latency multisensory integration processes in primary sensory cortices. Brain structure & function*, **220**, 955–977.

Henson, R. (2007) Efficient experimental design for fMRI. In , *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, pp. 193–210.

Hillyard, S.A. (1993) *Electrical and magnetic brain recordings: contributions to cognitive neuroscience. Current opinion in neurobiology*, **3**, 217–224.

Hillyard, S.A., Squires, K.C., Bauer, J.W. & Lindsay, P.H. (1971) *Evoked potential correlates of auditory signal detection. Science*, **172**, 1357–1360.

Hoefer, M., Tyll, S., Kanowski, M., Brosch, M., Schoenfeld, M.A., Heinze, H.J. & Noesselt, T. (2013) *Tactile stimulation and hemispheric asymmetries modulate auditory perception and neural responses in primary auditory cortex. NeuroImage*, **79**, 371–382.

Jaekl, P.M. & Soto-Faraco, S. (2010) *Audiovisual contrast enhancement is articulated primarily via the M-pathway. Brain research*, **1366**, 85–92.

Kanowski, M., Rieger, J.W., Noesselt, T., Tempelmann, C. & Hinrichs, H. (2007) *Endoscopic eye tracking system for fMRI. J. Neurosci. Methods*, **160**, 10–15.

Kayser, C., Petkov, C.I. & Logothetis, N.K. (2008) *Visual modulation of neurons in auditory cortex. Cerebral cortex (New York, N.Y. : 1991)*, **18**, 1560–1574.

Kayser, C., Petkov, C.I. & Logothetis, N.K. (2009) *Multisensory interactions in primate auditory cortex. FMRI and electrophysiology. Hearing research*, **258**, 80–88.

Komura, Y., Tamura, R., Uwano, T., Nishijo, H. & Ono, T. (2005) *Auditory thalamus integrates visual inputs into behavioral gains. Nature neuroscience*, **8**, 1203–1209.

Kording, K.P. (2014) *Bayesian statistics: relevant for the brain? Current opinion in neurobiology*, **25**, 130–133.

Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S. & Baker, C.I. (2009) *Circular analysis in systems neuroscience. The dangers of double dipping. Nature neuroscience*, **12**, 535–540.

Lakatos, P., Chen, C.-M., O'Connell, M.N., Mills, A. & Schroeder, C.E. (2007) *Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex. Neuron*, **53**, 279–292.

Lakatos, P., Karmos, G., Mehta, A.D., Ulbert, I. & Schroeder, C.E. (2008) *Entrainment of neuronal oscillations as a mechanism of attentional selection. Science (New York, N.Y.)*, **320**, 110–113.

Lakatos, P., O'Connell, M.N., Barczak, A., Mills, A., Javitt, D.C. & Schroeder, C.E. (2009) *The leading sense: supramodal control of neurophysiological context by attention. Neuron*, **64**, 419–430.

Lehmann, C., Herdener, M., Esposito, F., Hubl, D., di Salle, F., Scheffler, K., Bach, D.R., Federspiel, A., Kretz, R., Dierks, T. & Seifritz, E. (2006) *Differential patterns of multisensory interactions in core and belt areas of human auditory cortex. NeuroImage*, **31**, 294–300.

Lovelace, C.T., Stein, B.E. & Wallace, M.T. (2003) *An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. Brain Res Cogn Brain Res*, **17**, 447–453.

Mercier, M.R., Foxe, J.J., Fiebelkorn, I.C., Butler, J.S., Schwartz, T.H. & Molholm, S. (2013) *Auditory-driven phase reset in visual cortex. Human electrocorticography reveals mechanisms of early multisensory integration. NeuroImage*, **79**, 19–29.

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

Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E. & Foxe, J.J. (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.

Murray, M.M., Molholm, S., Michel, C.M., Heslenfeld, D.J., Ritter, W., Javitt, D.C., Schroeder, C.E. & Foxe, J.J. (2005) *Grabbing your ear. Rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. Cerebral cortex (New York, N.Y. : 1991)*, **15**, 963–974.

Myers, N.E., Stokes, M.G., Walther, L. & Nobre, A.C. (2014) *Oscillatory brain state predicts variability in working memory. The Journal of neuroscience : the official journal of the Society for Neuroscience*, **34**, 7735–7743.

Nath, A.R. & Beauchamp, M.S. (2011) *Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. The Journal of neuroscience : the official journal of the Society for Neuroscience*, **31**, 1704–1714.

Naue, N., Rach, S., Strüber, D., Huster, R.J., Zaehle, T., Körner, U. & Herrmann, C.S. (2011) *Auditory event-related response in visual cortex modulates subsequent visual responses in humans. Journal of Neuroscience*, **31**, 7729–7736.

Noesselt, T., Bergmann, D., Heinze, H.-J., Munte, T. & Spence, C. (2012) *Coding of multisensory temporal patterns in human superior temporal sulcus. Frontiers in integrative neuroscience*, **6**, 64.

Noesselt, T., Rieger, J.W., Schoenfeld, M.A., Kanowski, M., Hinrichs, H., Heinze, H.-J. & Driver, J. (2007) *Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. The Journal of neuroscience : the official journal of the Society for Neuroscience*, **27**, 11431–11441.

Noesselt, T., Tyll, S., Boehler, C.N., Budinger, E., Heinze, H.-J. & Driver, J. (2010) *Sound-induced enhancement of low-intensity vision. Multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity. The Journal of neuroscience : the official journal of the Society for Neuroscience*, **30**, 13609–13623.

Noppeney, U. (2008) *The neural systems of tool and action semantics. A perspective from functional imaging. Journal of physiology, Paris*, **102**, 40–49.

Odgaard, E.C., Arieh, Y. & Marks, L.E. (2004) *Brighter noise: sensory enhancement of perceived loudness by concurrent visual stimulation. Cogn Affect Behav Neurosci*, **4**, 127–132.

Picton, T. W., Woods, D. L., Baribeau-Braun, J., & Healey, T. M. (1977). *Evoked potential audiometry*. *Journal of Otolaryngology*, *6*(2), 90-119.

Raij, T., Ahveninen, J., Lin, F.H., Witzel, T., Jääskeläinen, I.P., Letham, B., Israeli, E., Sahyoun, C., Vasios, C., Stufflebeam, S., Hämäläinen, M. & Belliveau, J.W. (2010) *Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. European Journal of Neuroscience*, **31**, 1772–1782.

Rapin, I., Schimmel, H., Tourk, L. M., Krasnegor, N. A., & Pollak, C. (1966). *Evoked responses to clicks and tones of varying intensity in waking adults. Electroencephalography and clinical neurophysiology*, *21*(4), 335-344.

Railo, H., Koivisto, M. & Revonsuo, A. (2011) *Tracking the processes behind conscious perception. A review of event-related potential correlates of visual consciousness. Consciousness and cognition*, **20**, 972–983.

Rohe, T. & Noppeney, U. (2015) *Sensory reliability shapes perceptual inference via two mechanisms. Journal of vision*, **15**, 22.

Sadaghiani, S., Hesselmann, G. & Kleinschmidt, A. (2009) *Distributed and antagonistic contributions of ongoing activity fluctuations to auditory stimulus detection. The Journal of neuroscience : the official journal of the Society for Neuroscience*, **29**, 13410–13417.

Schroeder, C.E. & Lakatos, P. (2009) *Low-frequency neuronal oscillations as instruments of sensory selection. Trends in neurosciences*, **32**, 9–18.

Schroeder, C.E., Lakatos, P., Kajikawa, Y., Partan, S. & Puce, A. (2008) *Neuronal oscillations and visual amplification of speech. Trends in cognitive sciences*, **12**, 106–113.

Senkowski, D., Saint-Amour, D., Höfle, M. & Foxe, J.J. (2011) *Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. NeuroImage*, **56**, 2200–2208.

Sieben, K., Roder, B. & Hanganu-Opatz, I.L. (2013) *Oscillatory entrainment of primary somatosensory cortex encodes visual control of tactile processing. The Journal of neuroscience : the official journal of the Society for Neuroscience*, **33**, 5736–5749.

Smith, D.A., Boutros, N.N. & Schwarzkopf, S.B. (1994) *Reliability of P50 auditory event-related potential indices of sensory gating. Psychophysiology*, **31**, 495–502.

Stein, B.E. & Meredith, M.A. (1994) *The merging of the senses*. MIT Press, Cambridge, Mass.

Stevenson, R.A. & James, T.W. (2009) *Audiovisual integration in human superior temporal sulcus. Inverse effectiveness and the neural processing of speech and object recognition. NeuroImage*, **44**, 1210–1223.

Talsma, D., Doty, T.J. & Woldorff, M.G. (2007) *Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? Cerebral cortex (New York, N.Y. : 1991)*, **17**, 679–690.

Teder-Sälejärvi, W.A., McDonald, J.J., Di Russo, F. & Hillyard, S.A. (2002) *An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. Cognitive Brain Research*, **14**, 106–114.

Tucker, D. A., Dietrich, S., McPherson, D. L., & Salamat, M. T. (2001). *Effect of stimulus intensity level on auditory middle latency response brain maps in human adults. Journal of the American Academy of Audiology*, *12*(5), 223-232.

Tyll, S., Budinger, E. & Noesselt, T. (2011) *Thalamic influences on multisensory integration. Communicative & integrative biology*, **4**, 378–381.

van der Burg, E., Talsma, D., Olivers, C.N.L., Hickey, C. & Theeuwes, J. (2011) *Early multisensory interactions affect the competition among multiple visual objects. NeuroImage*, **55**, 1208–1218.

Venezia, J.H., Vaden, K.I., Rong, F., Maddox, D., Saberi, K. & Hickok, G. (2017) *Auditory, visual and audiovisual speech processing streams in superior temporal sulcus. Frontiers in Human Neuroscience*, **11**.

Wang, P. & Nikolić, D. (2011) *An LCD Monitor with Sufficiently Precise Timing for Research in Vision. Frontiers in Human Neuroscience*, **5**, 85.

Watson, R., Latinus, M., Charest, I., Crabbe, F. & Belin, P. (2014) *People-selectivity, audiovisual integration and heteromodality in the superior temporal sulcus. Cortex; a journal devoted to the study of the nervous system and behavior*, **50**, 125–136.

Welch, R.B. & Warren, D. (1986) Intersensory interactions. In Boff, K.R. (ed), *Handbook of perception and human performance.* Wiley, New York, pp. 1–36.

Werner, S. & Noppeney, U. (2010) *Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. Cerebral cortex (New York, N.Y. : 1991)*, **20**, 1829–1842.

Yarkoni, T., Poldrack, R.A., Nichols, T.E., van Essen, D.C. & Wager, T.D. (2011) *Large-scale automated synthesis of human functional neuroimaging data. Nature methods*, **8**, 665–670.

Table captions

Table 1. Main Effects of Vision and Sound. Table 1a: Local maxima for the comparison vision > no vision ($p < .05$, FWE-corrected, $k > 50$). Table 1b: Local maxima for the comparison sound > no-sound conditions ($p < .05$, FWE-corrected, $k > 50$). Table shows for each significant cluster the number of contiguous voxels, T-, p-value and MNI-coordinates of local maxima, plus anatomical regions (from left to right). Cortical regions are named in accordance with the SPM anatomy toolbox, percentage corresponds to likelihood of voxel belonging to a particular functional region (NA = no well-defined anatomical label). Abbreviations: L=Left, R = Right; ACC = Anterior Cingulate Cortex, IFG = Inferior Forntal Gyrus, IPL = Intraparietal Lobule, IPS = Intraparietal Sulcus, LOG = Lateral Occipital Gyrus, MGB = Medial Geniculate Body, MOG = Medial Occipital Gyrus, MTG = Medial Temporal Gyrus, POS = Parieto-Occipital Sulcus, SMG = Supramarginal Gyrus, STG = Superior Temporal Gyrus, TOS = Transversal Occipital Sulcus; MNI = Montreal Neurological Institute.

Table 2. Main effect of Loudness in Auditory Areas. Table 2a: Local maxima for the comparison High-intensity > Low-intensity sound conditions ($p < .05$, FWE-corrected). Table 2b: Local maxima for the comparison Low-intensity > No sound conditions ($p<.001$, Small Volume-cluster corrected (SVC) within auditory mask). Table shows for each significant cluster the number of contiguous voxels, T-, p- value and MNI-coordinates of local maxima, plus anatomical regions (from left to right). Cortical regions are named in accordance with the SPM anatomy toolbox (NA = no well-defined anatomical label), percentage corresponds to likelihood of voxel belonging to a particular functional region. Abbreviations: L = Left, R = Right; MGB = Medial Geniculate Body, STG = Superior Temporal Gyrus; MNI = Montreal Neurological Institute; * = not cluster-corrected.

Table 3. Superadditive Effects of Audiovisual relative to Visual and Auditory stimulation. Table 3a: Local maxima for the comparison Low-intensity sound with vision > [Low-intensity sound only and vision only] ($p < .01$, $k > 50$). Table 3b: High-intensity sound with vision > [High-intensity sound only and vision only] ($p<.05$, FWE-corrected). Table shows for each significant cluster the number of contiguous voxels, T-, p- value and MNI-coordinates of local maxima, plus anatomical regions (from left to right). Cortical regions are named in accordance with the SPM anatomy toolbox (NA = no well-defined anatomical label), percentage corresponds to likelihood of voxel belonging to a particular functional region. Abbreviations: L= Left, R=Right; MGB=Medial Geniculate Body, MTG = Medial Temporal Gyrus, STG = Superior Temporal Gyrus, STS = Superior Temporal Sulcus; MNI = Montreal Neurological Institute; * = not cluster-corrected.

Table 4. fMRI-Signal Behaviour Correlation in Auditory Areas. Table 4a: Local maxima for the fMRI-accuracy correlation of difference of low-intensity sounds with vision minus low-intensity sound alone ($p < .03$, $k > 50$). Table 4b: Local maxima for the fMRI- d' correlation of difference of [low-intensity sounds with vision +blank] minus [low-intensity sound only + vision only] ($p < .03$, $k > 50$, $n = 19$ due to one behavioural outlier). Table shows for each significant cluster the number of contiguous voxels, T-, p- value and MNI-coordinates of local maxima, plus anatomical regions (from left to right). Cortical regions are named in accordance with the SPM anatomy toolbox (NA = no well-defined anatomical label), percentage corresponds to likelihood of voxel belonging to a particular functional region. Abbreviations: L = Left, R = Right; MTG = Medial Temporal Gyrus, STG = Superior Temporal Gyrus; MNI = Montreal Neurological Institute.

Figure Captions

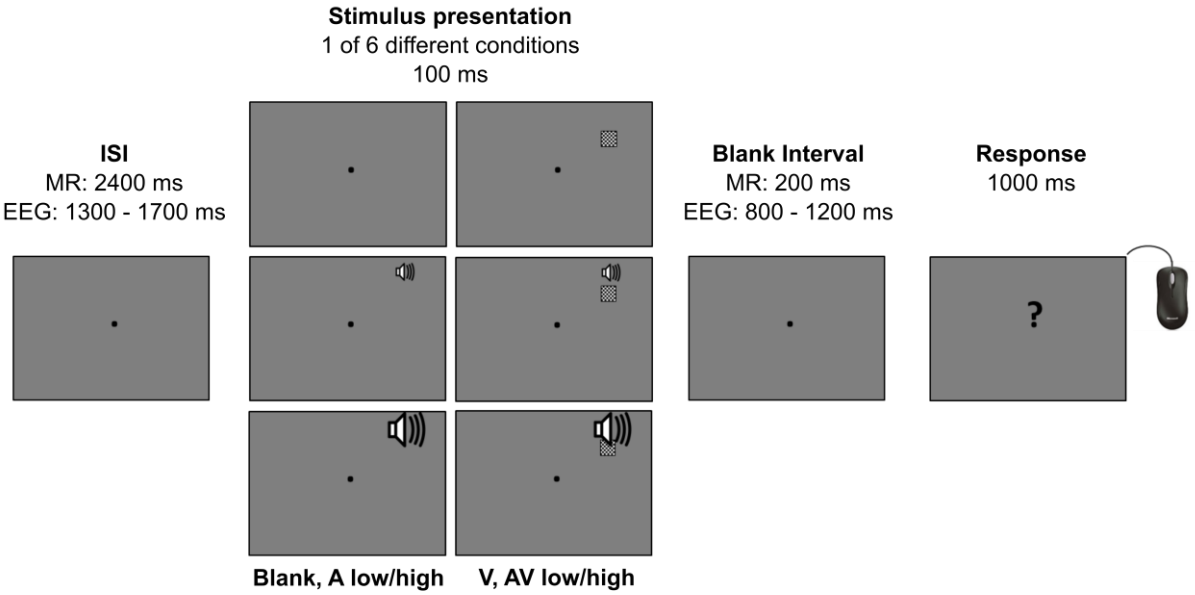

Figure 1. Experimental design of fMRI and ERP-study. Different timing between MR and EEG experiments are denoted above each subpart of the trial. Either unimodal auditory stimuli (low or high intensity) or concurrent audiovisual stimuli (sound plus checkerboard) were presented. In the remaining two conditions no stimulus or just the checkerboard were shown. A question mark always served as response cue. Participants had to report whether a sound was present. Note that the visual stimulus provided no information whether a sound had occurred (see main text for details).

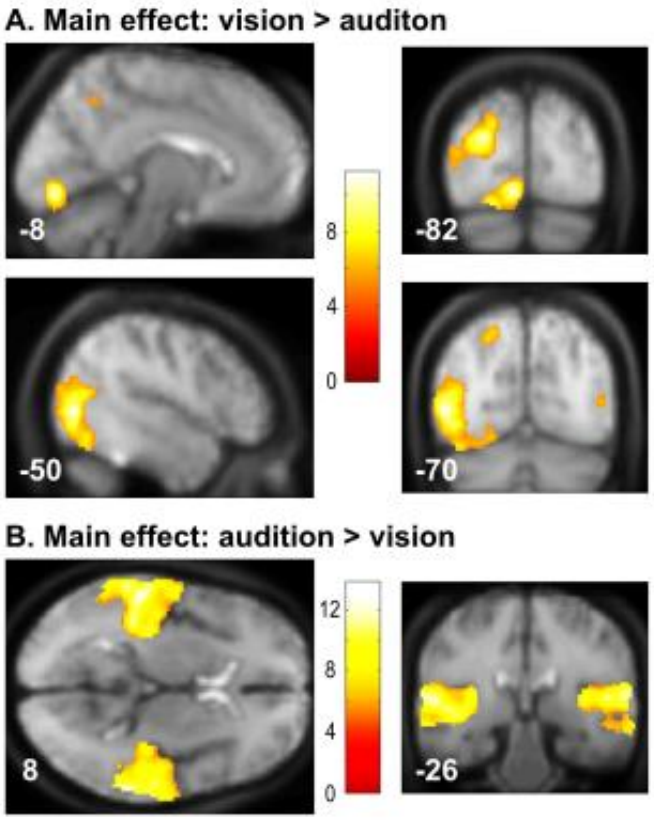

Figure 2. Main effects of vision and sound. Fig. 2a depicts the main effect of vision > no-vision ($p < .05$ FWE-corrected) in medial (upper row) and lateral occipital regions (lower row). Fig. 2b depicts the main effect of sound > no-sound ($p < .05$ FWE-corrected) in superior temporal regions. Activation maps are overlaid on the group-mean anatomical image. Neurological convention is used. Numbers at the bottom of each slice denote MNI-coordinates. Colour coding indicate T-value.

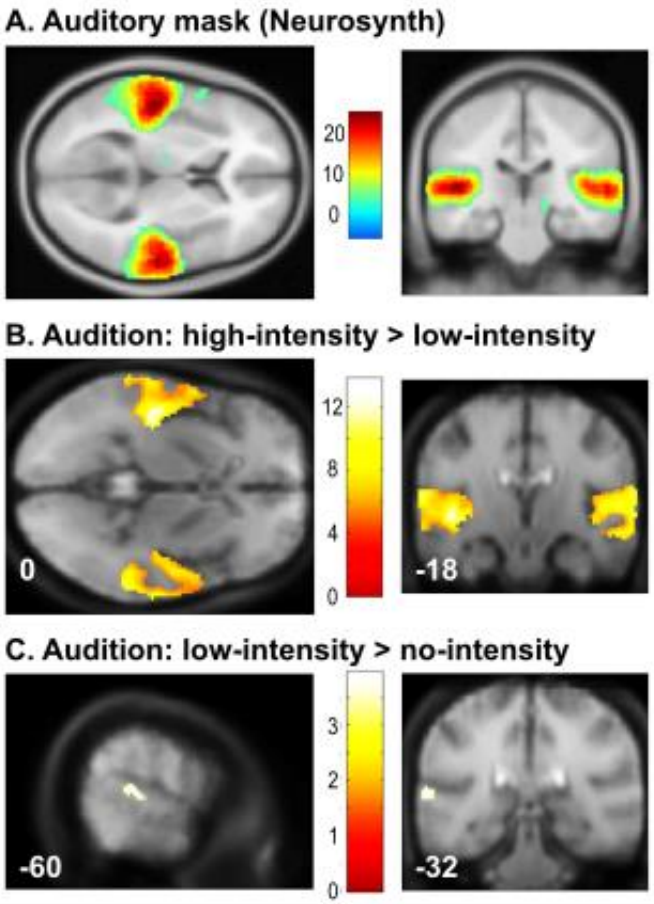

Figure 3. Main effect of Loudness in Auditory Areas. Fig. 3a depicts the auditory mask in superior temporal cortex (created with Neurosynth; posterior-probability map; search term 'auditory'; p<.01 FDR-corrected) which was used for all subsequent analyses (see Fig. 3b,c Fig. 4, Fig. 5 and Tab. 2-4). Fig. 3b depicts the effect of high-intensity > low-intensity sounds (p<.05 FWE-corrected) in superior temporal cortex. Fig. 3c depicts the effect of low-intensity sound > no-sound (p < .001 FWE-cluster-corrected) in left superior temporal gyrus. Activation maps are overlaid on the group-mean anatomical image. Neurological convention is used. Numbers at the bottom of each slice denote MNI-coordinates. Colour coding indicates T-value.

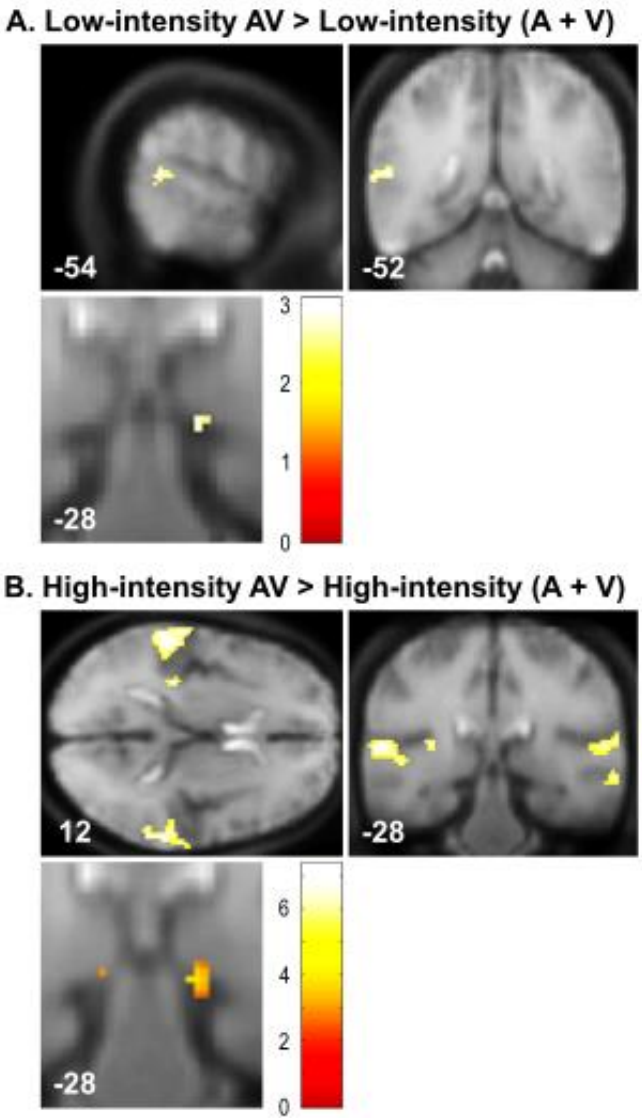

Figure 4. Superadditive Effects of Audiovisual relative to Visual and Auditory stimulation. Fig. 4a: depicts the effect of Low-intensity sound with vision > [Low-intensity sound only and vision only] ($p < .01$, $k > 50$) in left-hemispheric superior temporal gyrus/sulcus (top) and auditory thalamus (below). Fig. 4b depicts the effect of high-intensity sound with vision > [high-intensity sound only and vision only] ($p < .05$, FWE-corrected) in bilateral superior temporal gyrus including TE 1.0 (top) and auditory thalamus (below). Activation maps are overlaid on the group-mean anatomical image. Neurological convention is used. Numbers at the bottom of each slice denote MNI-coordinates. Colour coding indicates T-value.

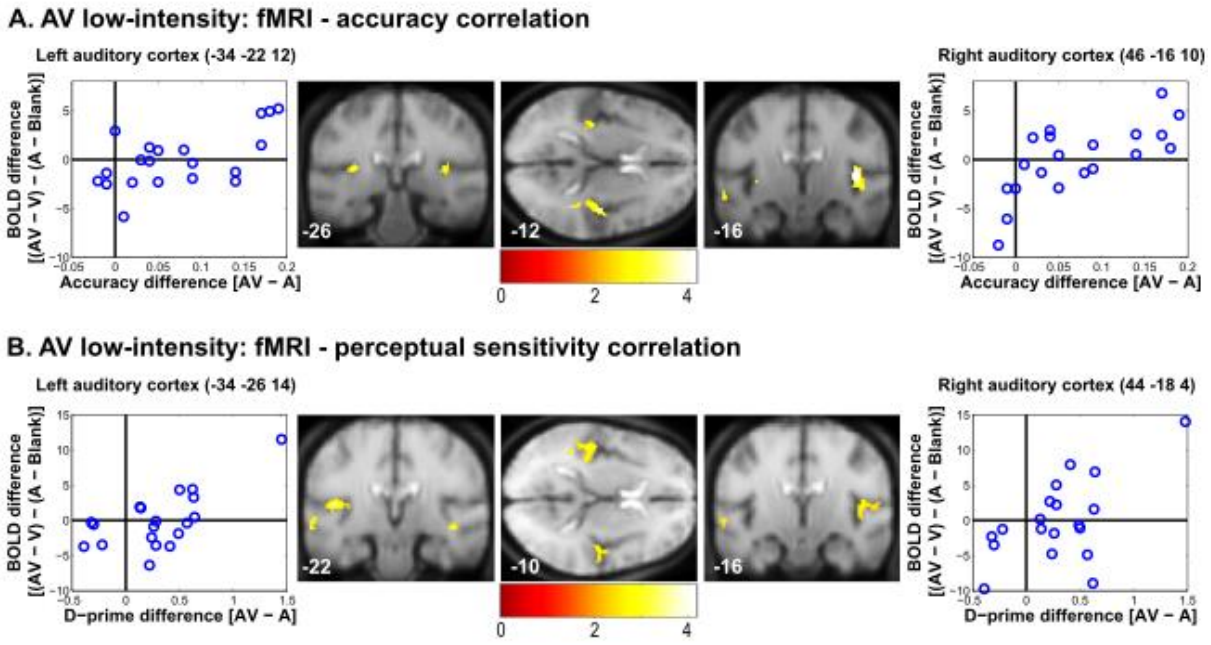

Figure 5. fMRI-Signal Behaviour Correlation in Auditory Areas. Fig. 5a depicts the effects for the fMRI-accuracy correlation of difference of low-intensity sounds with vision minus low-intensity sound only ($p < .03$, $k > 50$) in low-level auditory cortex. Fig. 5b depicts the effects for the fMRI-d' correlation of difference of [low-intensity sounds with vision + blank] minus [low-intensity sound only plus vision only] ($p < .03$, $k > 50$, $n = 19$ due to one behavioural outlier) in low-level auditory cortex. Scatter plots next to brain slices show brain-behaviour correlation of local maxima in left and right-hemispheric low-level auditory cortex. Activation maps are overlaid on the group-mean anatomical image. Neurological convention is used. Numbers at the bottom of each slice denote MNI-coordinates. Colour coding indicates T-value.

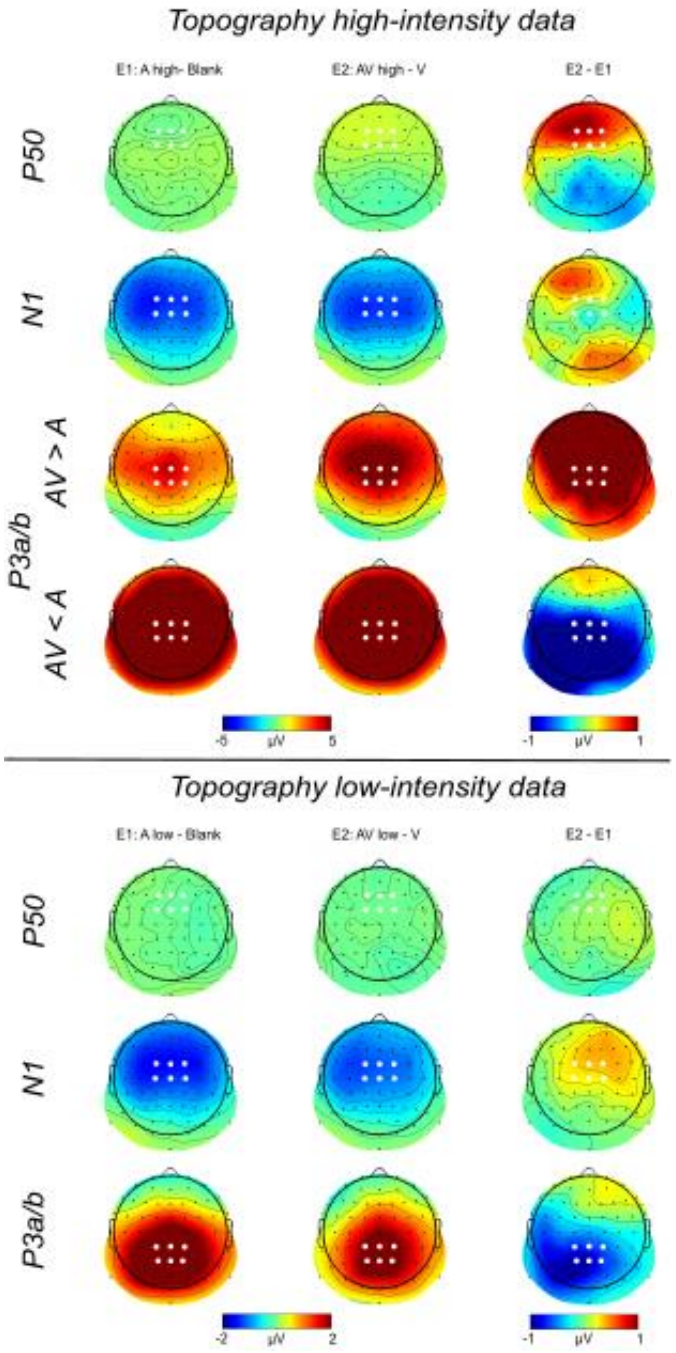

Figure 6. **Topographies of Low- (bottom) and High-Intensity (top) Auditory Evoked Potentials (component names displayed on left side).** The left column shows the ERP difference waveforms for A – blank (E1), The middle column shows the difference for AV – V (E2), and the right column shows the difference E2 – E1. Electrodes of Interest (EOIs) from which waveforms were derived are highlighted in white.

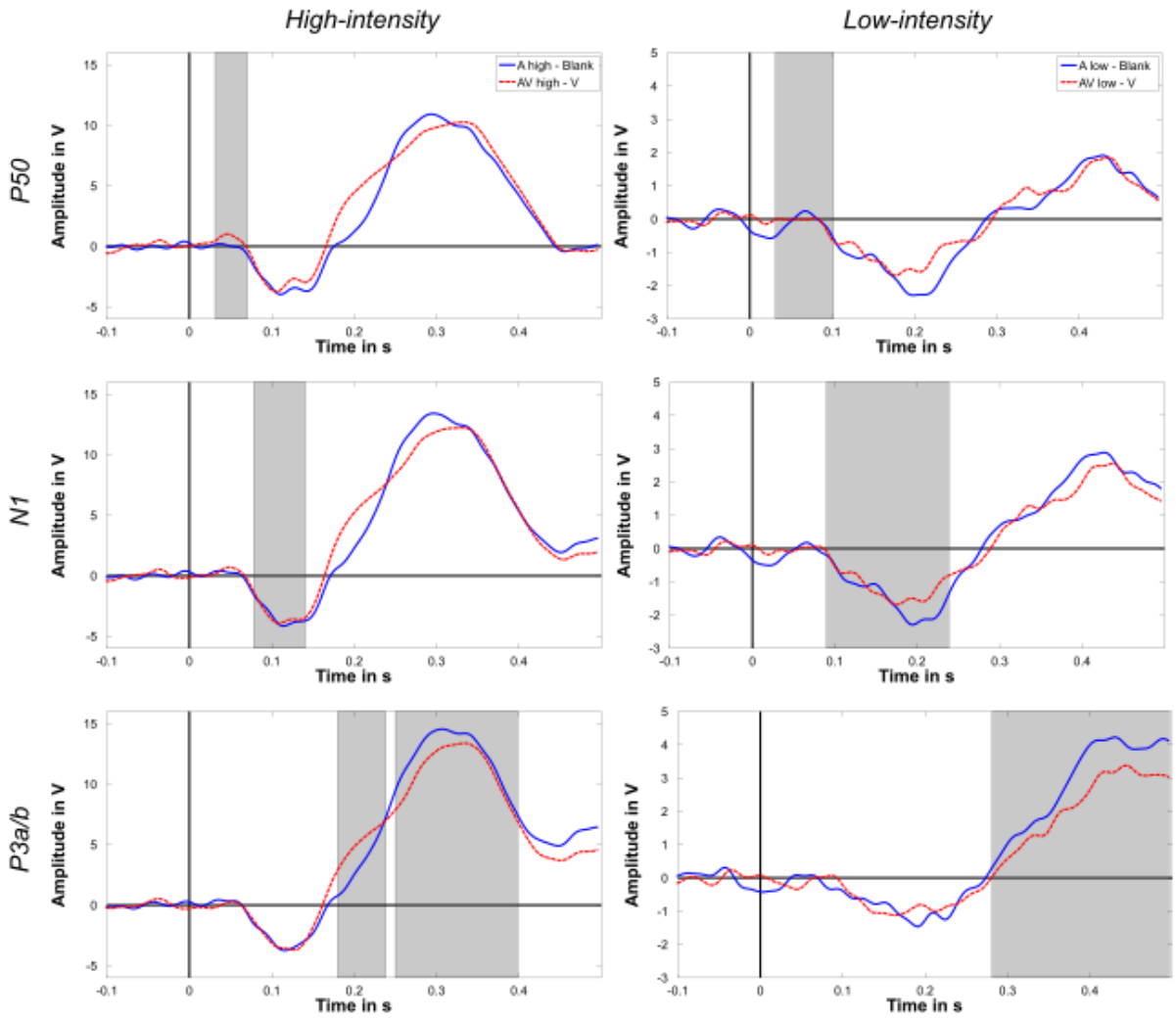

Figure 7. Time course of Low- (right) and High-Intensity (left) ERPs (component names displayed on left side). The time ranges used for statistical analyses are highlighted in grey. The left column shows ERPs for high intensity trials and the right side shows ERPs for low intensity trials.

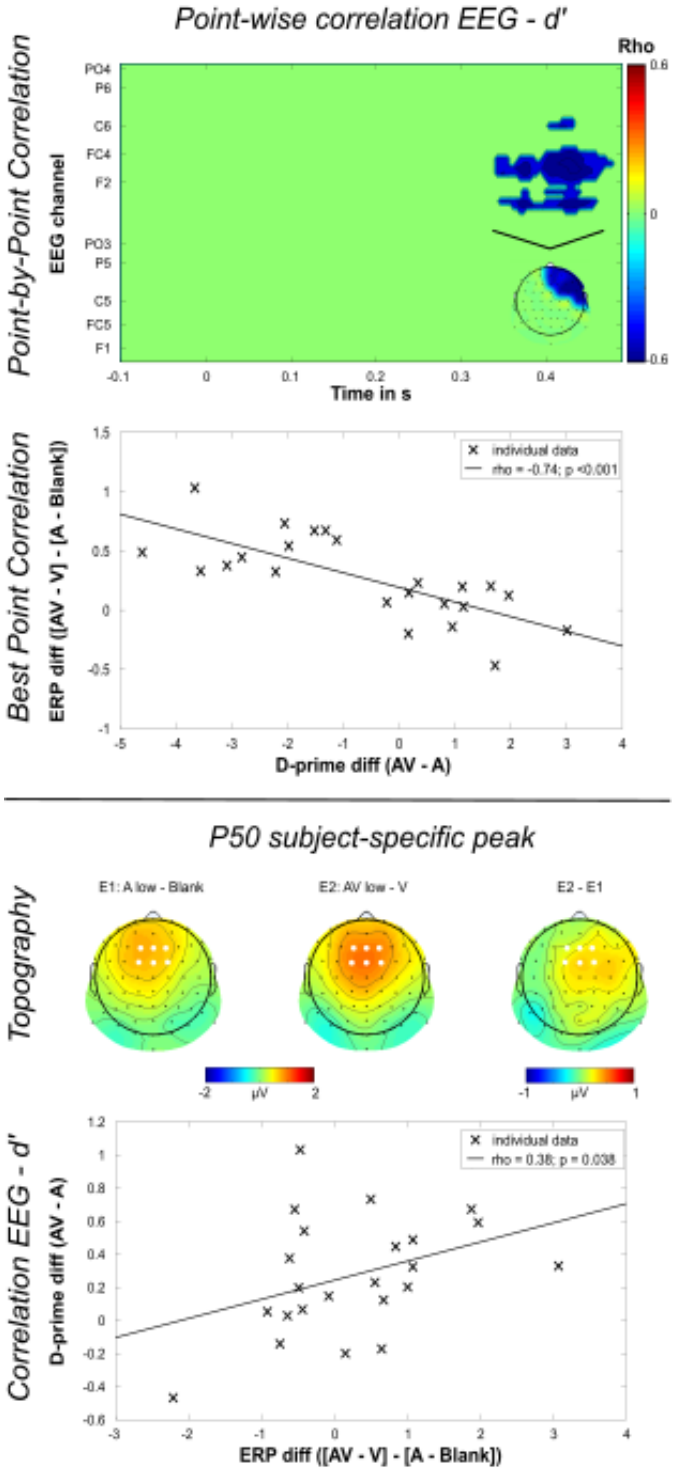

Figure 8. Channel-wise Point-by-Point Regression with Behaviour (Top). All non-significant (rho close to zero) data points shown in green. Furthermore, we show the topography of the significant cluster. The scatter-plot shows the individual improvement in d-prime (AV vs. A) plotted against the individual mean amplitude shift within the cluster (AV vs. A). **Topography and scatter plot (Bottom)** for the correlation of behavioural performance gain (AV vs. A) and P50 increase (AV vs. A).