

Visual stimulus locking of EEG is modulated by temporal congruency of auditory stimuli

Sonja Schall · Cliodhna Quigley · Selim Onat · Peter König

Received: 16 October 2008 / Accepted: 19 May 2009
© Springer-Verlag 2009

Abstract Disparate sensory streams originating from a common underlying event share similar dynamics, and this plays an important part in multisensory integration. Here we investigate audiovisual binding by presenting continuously changing, temporally congruent and incongruent stimuli. Recorded EEG signals are used to quantify spectrotemporal and waveform locking of neural activity to stimulus dynamics. Spectrotemporal analysis reveals locking to visual stimulus dynamics in both a broad alpha and the beta band. The properties of these effects suggest they are a correlate of bottom-up processing in the visual system. Waveform locking reveals two cortically distinct processes that lock to visual stimulus dynamics with differing topographies and time lags relative to the stimuli. Most importantly, these are modulated in strength by the congruency of an accompanying auditory stream. In addition, the waveform locking found at occipital electrodes shows an increase over stimulus duration for visual and congruent audiovisual stimuli. Hence we argue that these effects reflect audiovisual interaction. We thus propose that spectrotemporal and waveform locking reflect different mechanisms involved in the processing of dynamic audiovisual stimuli.

Keywords Audiovisual integration · EEG · Multisensory processing · Spectrotemporal analysis

Abbreviations

EEG Electroencephalogram
MEG Magnetoencephalogram
LFP Local field potential
ERP Event-related potential
fMRI Functional magnetic resonance imaging

Introduction

Multisensory integration makes ecological sense when the incoming signals refer to the same external entity or, more generally, are due to the same underlying physical events. In general, there are three different ways in which signals must correspond for integration to take place: spatially, temporally, and/or semantically (Stein and Meredith 1993; Macaluso and Driver 2005; Calvert 2001; Doehrmann and Naumer 2008).

These correspondences represent a default scheme of integration, and in special cases are influential enough to fool the system into binding streams that do not belong to the same underlying cause. In the well-known ventriloquist illusion, for instance, the concurrent temporal modulation of the puppet's mouth and the ventriloquist's speech leads to a perception of a talking puppet. In spite of the spatial and even semantic discrepancies here, temporal correspondence drives this audiovisual integration (Vroomen and De Gelder 2004; Bonath et al. 2007). A competition in temporal synchronicity between auditory and visual streams, on the other hand, may lead to a visual illusion. In the case of the sound-induced flash illusion (Shams et al. 2000, 2002; Mishra et al. 2007, 2008) an extra illusory flash is perceived when a single flash is interleaved between two beeps. More generally, at a constant visual stimulation rate, the number of perceived flashes increases with a higher auditory stimulation rate (visual illusion) and decreases

S. Schall (✉) · C. Quigley · S. Onat · P. König
Institute of Cognitive Science,
University of Osnabrück, Albrechtstr. 28,
49069 Osnabrück, Germany
e-mail: soschall@uos.de
URL: <http://www.cogsci.uni-osnabrueck.de/NBP/>

when the rate of auditory events is slower (visual suppression) (e.g. Shipley 1964; Noesselt et al. 2008). The perceived temporal pattern of visual events is thus 'adjusted' to match the rate of auditory events. Taken together, these cases nicely illustrate that temporal aspects are essential to the perception of audiovisual events.

Research into the neural substrate of multisensory integration began in earnest after neurophysiological results showed characteristic multimodal response profiles at the level of single cells in superior colliculus in cat (Meredith and Stein 1983; Meredith et al. 1987) and macaque (Wallace et al. 1996). When stimuli were presented in close temporal proximity in two sensory modalities, the response of some collicular cells was found to reach or even exceed the sum of responses to each stimulus delivered in isolation. Imaging studies have since confirmed the importance of superior colliculus in human audiovisual integration (Calvert et al. 2001; Miller and D'Esposito 2005). Several cortical areas have also been implicated in the processing of audiovisual stimuli including the insula (Bushara et al. 2001) and intraparietal sulcus (Calvert et al. 2001). The superior temporal sulcus, in particular, has been consistently shown to be an area of audiovisual convergence and integration (Beauchamp et al. 2004; see Calvert 2001 for a review of earlier work). Importantly, it has been found to be sensitive to the temporal synchrony of audiovisual information, especially audiovisual speech (Calvert et al. 2000; Miller and D'Esposito 2005; Macaluso et al. 2004), but also simpler stimuli (Noesselt et al. 2007; van Atteveldt et al. 2007). In addition, there is a growing body of evidence from ERP and fMRI studies demonstrating that areas traditionally conceived to be unisensory also play a role in the synthesis of audiovisual information (Giard and Peronnet 1999; Molholm et al. 2002; Kayser et al. 2007; Calvert et al. 1999; Miller and D'Esposito 2005; Noesselt et al. 2007; for general reviews on low-level integration see Foxe and Schroeder 2005; Kayser and Logothetis 2007; Driver and Noesselt, 2008). Another line of research follows the hypothesis that cortically, multisensory integration is instantiated by the relative timing of synchronised populations of cortical neurons. This has been supported by human EEG/MEG studies (see Senkowski et al. 2008 for a review).

To date, studies that explicitly investigate the temporal dependence of audiovisual binding, such as those mentioned above, have mainly manipulated the timing of stimulus onset or coincidence in one modality relative to the other. Typically, these experiments employ brief, simple stimuli that are either presented individually (Meredith et al. 1987; Bushara et al. 2001) or in streams (Calvert et al. 2001; Noesselt et al. 2007; Dhamala et al. 2007; Senkowski et al. 2007). However, such static audiovisual events are rarely found under natural conditions, where acoustic

signals mostly emanate continuously from objects in motion, such as rustling leaves or moving cars. Real-world audiovisual events extend over time and are bound less by their simultaneity and more by their common temporal dynamics. For this reason, concentrating only on the isolated temporal coincidence of brief events neglects important temporal aspects inherent in natural events. To capture the full importance of temporal information, it is thus necessary to investigate audiovisual events extended in time.

Indeed, Kayser and König (2004) have recently demonstrated that extended stimulation with natural visual stimuli leads to a continuous modulation of LFP power in cat visual cortex, and that this modulation reflects the dynamics of the presented movies. In addition, rhythmic auditory stimulation has been shown to entrain low-frequency activity in auditory cortex of macaque (Lakatos et al. 2005). The entrainment of human cortical activity to stimulation at a constant frequency is also a well-known phenomenon in the auditory (e.g., Galambos et al. 1981; Bidet-Caulet et al. 2007) and visual modalities (e.g. Regan 1966; Ding et al. 2006); however, it is not clear whether this also holds for stimuli with rich temporal power spectra. One possible role of stimulus locking in temporally dependent multisensory integration may involve an increase in the efficacy of stimuli from a second modality when temporally aligned with neural activity that is entrained to the first modality (Lakatos et al. 2007; Kayser et al. 2008; Schroeder et al. 2008). As such, the extended changes in cortical activity seen in response to extended sensory stimulation are certainly worth closer examination in the context of audiovisual integration.

Here, we are interested in the general relevance of extended, time-varying information for the integration of visual and auditory streams. For this purpose, we paired well-defined visual and auditory signals that changed continuously and irregularly over time according to shared or differing dynamical patterns. Though relatively complex in temporal structure, and in this respect comparable to natural events including audiovisual speech, our stimuli have the advantage of being novel to the experimental subject and free of semantic reference. Using EEG, we employ two measures of stimulus locking to investigate audiovisual interaction. First, as a generalisation of the evoked potential, we examine whether averaged EEG waveforms are locked to stimulus dynamics. Evoked potentials reveal the activity locked to stimulus onset, and here we extend this approach using cross-correlation. Second, following Kayser and König's (2004) study, we investigate the locking of EEG power dynamics at frequencies within and beyond the range of our stimulus dynamics, and thus whether the observations made in cat transfer to man.

We specifically address three research questions. First, whether neural activity in the human cortex locks to the

temporal structure of irregular dynamic stimuli in the visual and auditory modalities. Second and most importantly, we want to investigate whether multisensory integration is reflected in such a mechanism. Finally, we are interested in the time course of stimulus-locking, particularly in response to congruent and incongruent bimodal stimulation.

Methods

Participants

EEG was recorded from 30 healthy university students, who were first screened for normal sight and hearing using standard tests (Landolt C chart and calibrated PC-based audiometer) and gave their informed written consent to participate in the study. Of these, six subjects were discarded prior to data analysis due to noisy data. The remaining 24 subjects (14 female, 23 right-handed) were aged between 19 and 33 years (mean: 23, standard deviation: 3). The experiment was conducted in accordance with the Declaration of Helsinki.

Experimental design

In order to investigate multisensory integration processing, it is possible to contrast responses to unimodal and bimodal stimuli, or alternatively, to compare congruent and incongruent bimodal stimulation. To address both of these contrasts, we presented auditory and visual stimuli in four stimulation conditions: bimodal congruent, bimodal incongruent, unimodal visual and unimodal auditory.

To engage their attention, participants were required to perform a congruency judgement task. In the bimodal case, subjects were asked to judge the temporal congruency of simultaneously presented auditory and visual stimulus components, thus ensuring that attention would be equally allocated to both modalities. To similarly engage subjects' attention during unimodal presentations, we adjusted the task to a sequential comparison of two consecutively presented unimodal stimuli, meaning that each such trial contained two stimuli. A sequential pair could consist of any combination of the two modalities (V-A, A-V, V-V or A-A). In subsequent EEG analysis, however, each stimulus of these sequentially presented pairs was treated as an independent unimodal stimulus in either the unimodal visual or auditory condition.

We chose a self-paced paradigm in which subjects triggered trial onset via button press (see Fig. 1). In doing so, we aimed to maximise participants' concentration and minimise eye-movements during stimulus presentation. Each trial began with the presentation of a red fixation cross. After a short, random delay, the stimulus appeared on the

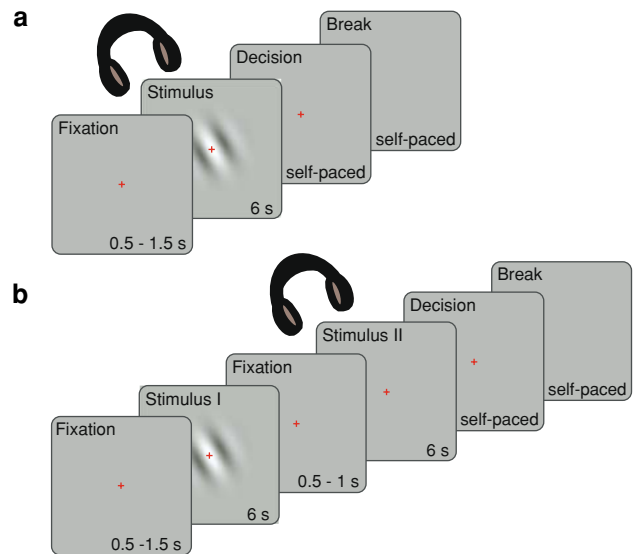


Fig. 1 Schematic diagram of both experimental paradigms. For both paradigms trial onset was self-paced, and subjects indicated their readiness by button press. Each trial began after a random delay with the appearance of a red fixation cross, which subjects were asked to fixate until the end of the trial. **a** Bimodal trials consisted of simultaneously presented visual and auditory stimuli, and were either congruent (stimuli followed the same trajectory) or incongruent (different trajectories). **b** Single modality trials contained two sequentially presented stimuli, separated by a random interval (500–1,500 ms). A pair could entail stimuli from one or both modalities, again either congruent or incongruent. After stimulus presentation, subjects were required to decide whether the trial was congruent or incongruent, answering via button press

screen and/or was heard through headphones. In the case of bimodal stimuli, auditory and visual components were simultaneously presented, after which subjects completed the task by pressing one of two buttons (denoting ‘congruent’ and ‘incongruent’). In the case of sequential trials, two stimuli were shown consecutively, separated by a short, random time interval (500–1,500 ms), and subjects completed the task after the second stimulus. For purposes of behavioural analysis, task performance was evaluated on a trial basis.

Stimuli

Stimuli consisted of a rotating Gabor patch, a frequency-modulated tone, or both presented simultaneously, and lasted for 6 s (see Movies 1 and 2 for examples). Gabor patches were greyscale, had a sinusoidal frequency of 0.33 cycles per degree, subtended approximately 4.5 degrees of visual angle (full width at half-maximum), and were centrally presented on a grey background identical to the mean luminance of the patch. Auditory stimuli had constant amplitude, and were ramped with a 10 ms half-cosine window to avoid clicks or other artefacts. Movies of visual stimuli were sampled at a frame rate of 25 Hz, and

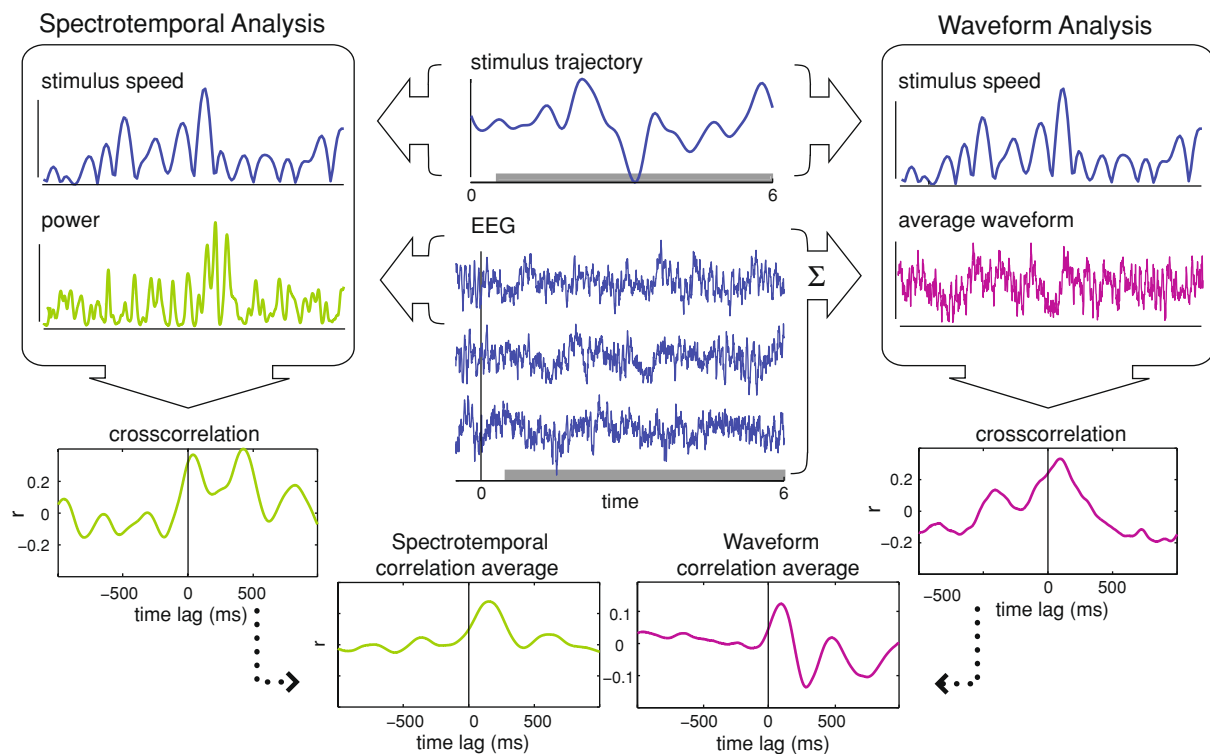


Fig. 2 Two correlation analysis approaches. Spectrotemporal and waveform correlation methods are illustrated on the left and right, respectively. Exemplary stimulus trajectory and single-trial EEG waveforms are shown in the centre, with a grey bar indicating the temporal range used in both analysis approaches. In the case of spectrotemporal analysis, each trial is transformed into a time-frequency representation (see text for details) and then correlated with the speed profile of the corresponding stimulus. The waveform analysis method

involves creating an ERP from all trials corresponding to a single stimulus, and then correlating this waveform directly with the stimulus speed. Peaks at positive time lags indicate that the EEG power or waveform follows the temporal dynamics of the stimulus, in other words that the EEG has locked to the stimulus. In both approaches, correlograms are then averaged over stimuli. Grand averages are created by taking the median over subjects

uncompressed visual and auditory stimuli were merged into an avi file for bimodal presentation using MEncoder (MPlayer version 1.0rc1, <http://www.mplayerhq.hu>).

Both unimodal stimulus components varied along only one feature dimension: orientation of Gabor patches spanned 180 degrees centred around the horizontal axis, and the frequency of tones was modulated between 200 and 300 Hz. This ongoing feature modulation was determined by trajectories that changed smoothly but irregularly over time (see Fig. 2). Visual and auditory feature dimensions were chosen based on a behavioural pilot study (see Appendix for more detail) 32 trajectories were created in Fourier space by assembling a Gaussian power spectrum with a cut-off frequency of 1 Hz (resulting in a cut-off frequency of ~ 4 Hz for stimulus speed) and random phase information. Each trajectory was constructed as part of a set of four mutually orthogonal trajectories, and in total eight sets were created using an iterative optimisation procedure. The Fourier representation was then transformed to the time domain using the inverse fast Fourier transform. Stimuli were presented in congruent and incongruent conditions, where visual and auditory stimuli were modulated by iden-

tical or orthogonal trajectories (see below). This allowed us to precisely control for the degree of incongruence.

Stimuli were organised into four blocks, each consisting of 24 congruent audiovisual, 24 incongruent audiovisual, 24 visual, and 24 auditory stimuli (yielding 48 simultaneous and 24 sequential trials). Each block was created from two sets of orthogonal trajectories, with each trajectory occurring three times in each of the four conditions. Trajectories were balanced over conditions and presentation order of paired stimuli. Block order and stimulus order within blocks was randomised over subjects. Depending on time constraints and subject alertness, participants completed either three or four blocks (equivalent to 288 or 384 stimuli).

Presentation and recording

Subjects were comfortably seated approximately 1 m from the presentation screen in a dimly lit recording room. Stimuli were presented on a 21-in. monitor (Samsung SyncMaster 1100 DF) with a refresh rate of 100 Hz. The presentation software (Neurobehavioural Systems, version

10.3) was run on a Macintosh Pro (Apple Computers). Sounds were delivered via in-ear headphones (EAR-Tone 3A, Etymotic Research Inc.) at a comfortable volume chosen by the subject. During trials, subjects were asked to remain still and keep their gaze fixated on a centred, red fixation cross.

EEG was recorded from 28 sintered Ag/AgCl ring electrodes mounted in a cap (EasyCap GmbH, Herrsching-Breitbrunn, Germany) according to the standard 10/20 system, referenced to linked mastoids. Four additional electrodes were used to capture vertical and horizontal electro-oculograms (bipolar montages). The upper threshold for impedances was 5 kOhm. Two of the 30 standard electrodes (FT8 and FT7) were exchanged for diodes to capture visual and auditory stimulus onset. Signal amplification, filtering and digitisation were carried out by a 32-channel amplifier system (Synamps1, Neuroscan, Compumedics, TX, USA). The data were digitised at 500 Hz and online bandpass-filtered within 0.3 and 100 Hz to prevent aliasing. The digital signal was recorded by a PC-workstation (Intel Pentium 4, 2.41 GHz, 1 GB RAM).

The diode signals revealed that the visual stimulus preceded the auditory by 25 ms—a small delay compared to the cut-off frequency of the trajectories—due to the stimulation software used.

Data analysis

Behavioural analysis

Behavioural analysis served two purposes: first, to determine whether subjects were attentive during the experiment; second, to ensure that subjects could perceptually distinguish bimodal congruent from bimodal incongruent stimuli. Task performance was evaluated in terms of trials, and simultaneous and sequential stimulus presentations were dealt with separately.

To determine whether subjects performed above chance level, an exact Binomial test was performed for each subject. We furthermore used signal detection methods to quantify subjects' sensitivity and bias in discriminating between congruent and incongruent bimodal stimuli (see Wickens 2002 for more detail). The sensitivity measure d' was computed from hit (correct identification of congruent stimulus) and false alarm rates. Assuming that both rates come from standard normal distributions, d' estimates the distance between the two distributions in units of standard deviations—the further apart the two distributions, the easier the congruency can be detected, and the higher the d' . The response bias estimate ($\log \beta$) informs us whether the observer favours a 'congruent' or 'incongruent' response, with a negative value indicating a bias to congruence, and a positive value indicating an incongruence bias.

Pre-processing and artefact removal

All data analysis was carried out using MatLab (The MathWorks, Natick, MA), using the EEGLab toolbox (Delorme and Makeig 2004) and custom-made functions. Before artefact removal, all data were bandpass-filtered between 0.5 and 100 Hz and epoched. Eye-movement artefacts were eliminated using Independent Component Analysis (Jung et al. 2000). Other muscle artefacts were detected by visual inspection, and whole trials were rejected. All trials were baseline corrected relative to 500 ms of pre-stimulus activity.

Stimulus locking

Two approaches were used to quantify stimulus locking both involving cross-correlations between EEG activity and the magnitude of change of the stimulus (see Fig. 2). All correlation coefficients were normalised so that the values lie between -1 and 1 , with 0 indicating the absence of a correlation. Correlation coefficients are calculated for all temporal offsets between stimulus and response. Significant coefficients at positive time-lags indicate that the stimulus leads the cortical responses, that is, the brain response is locked to the stimulus dynamics. Due to the finite autocorrelation of the stimulus, a locking might even occur at negative lags.

Our first analysis approach was concerned with the question of whether the time course of stimulus-induced power changes in the EEG correlates with changes in stimulus speed (spectrotemporal analysis). This approach allowed us to investigate whether the power in any frequency band of the measured EEG signals locks to the ongoing stimulus.

A second analysis approach (waveform analysis) was used to examine whether the EEG waveform itself locks to the temporal profile of the stimulus. This approach was used to investigate whether there is a stable phase-relationship between evoked EEG response and stimulus. A significant correlation coefficient indicates that the phase of the EEG signal locks to the ongoing stimulus. In contrast to the spectrotemporal approach, this analysis only reveals locking of frequencies common to both stimulus and EEG. A detailed description of each approach is given below.

In the spectrotemporal analysis approach, spectral power changes in response to visual speed were determined on a trial-by-trial basis (Fig. 2, left panels). For each trial, we estimated the Short-time Fourier Transform using a window length of 80 ms shifted with an overlap of 78 ms. Each data segment was windowed (Hamming window) and zero-padded (to 256 samples \triangleq 512 ms). Thus, we achieved a nominal temporal resolution of 2 ms and frequency resolution of approximately 2 Hz. The amplitude of the Short-time Fourier Transform was squared to obtain spectral

power, and then z-transformed with respect to baseline power. Due to the power increase evoked by stimulus onset, we removed the first 500 ms of each trial which is uninformative for the present purpose and investigated only the stationary response. The power of individual frequencies was then either summed across our frequency range of interest (20–35 Hz) corresponding to previous results (Kayser and König 2004), or kept at its full resolution. Each frequency band was then separately cross-correlated with the speed of the stimulus, given by the absolute values of the trajectory's first derivative. The resulting cross-correlations were then averaged using Fisher's z-transform over all trials within each condition, yielding one cross-correlogram per subject, condition, frequency band and channel. Bimodal incongruent trials resulted in two cross-correlograms due to the difference in visual and auditory stimulation trajectories.

The waveform analysis approach evaluated whether stimulus speed is reflected in the EEG waveform by cross-correlating stimulus speed with stimulus-specific event-related potentials (ERPs). An ERP was computed by averaging the EEG over those trials in which the same stimulus (i.e. identical trajectories) was shown for the modality in question (Fig. 2, right panels). Due to the limited number of trajectory repetitions throughout the experiment, ERPs were constructed from a maximum of nine waveforms (please note that the number of trials is substantially larger). Stimulus-specific ERPs were then cross-correlated with the speed profile of the corresponding trajectory. The stimulus average for each subject, condition, and channel was obtained, again using Fisher's z-transform. As above, the grand average was calculated using the median.

Our first research question asked whether there is any evidence for stimulus-locking to visual and auditory stimulus dynamics, in other words how the cortex responds to the presence of a time-varying stimulus. To address this, for each modality we averaged over all conditions containing input to that modality, e.g. unimodal visual and both bimodal conditions were averaged to investigate visual stimulus-locking. This increased the amount of data used in our analysis, thus stabilising effects that were otherwise small and less consistent between subjects. In addition, we can thus estimate the modality-specific, automatic, bottom-up response irrespective of task factors or information present in other modalities. After evaluating the presence or absence of stimulus locking, the original conditions were used to explore the second question, namely whether locking is modulated by multisensory interaction. Finally, to address the third research question, we estimated changes in stimulus locking over the duration of a trial. This was done using a time-dependent cross-correlation function. Cross-correlations were computed within a 500-ms shifting

time window (250 ms overlap), thus yielding 23 cross-correlograms per trial. Averaging over trials, conditions and subjects was performed as above.

Evaluating statistical significance

Significance of the resulting correlation coefficients was tested using bootstrap techniques. To test for significant peaks within a subject and for a given condition, we computed 1,000 cross-correlations composed of averaged cross-correlations between non-matching EEG data and stimuli. From this distribution, the 95 and 99% confidence intervals were estimated. To test significance of the grand average, a distribution of 1,000 medians was generated to determine confidence intervals.

To test for significant condition differences, permutation testing was used. For each subject, 1,000 surrogate conditions were created by pooling and randomly redrawing trials from both conditions without replacement. A distribution of condition differences was obtained by subtracting the resulting 1,000 grand averages (median) of both conditions.

Results

Here we first evaluate subjects' task performance in order to determine their alertness during the experiment and ability to distinguish audiovisual congruent from incongruent stimuli. Next we address the issue of whether stimulus locking can be found in the human brain. Finally, we compare congruent and incongruent bimodal conditions to investigate whether stimulus locking is involved in the crossmodal processing of temporal information.

Task performance

All subjects performed well above chance in the task ($p < 0.01$, exact binomial test), with an average performance of 76% ($\pm 8\%$ standard deviation). This result indicates that subjects were attentive during the experiment and understood the task. Dividing data according to simultaneous audiovisual and sequential single modality trials, subjects' performance was better for the simultaneous ($87 \pm 11\%$) than the sequential paradigm ($65 \pm 7\%$). All subjects performed above chance in simultaneous trials (23 with $p < 0.01$, 1 with $p < 0.05$). In sequential trials the majority of subjects remained significantly above chance (14 with $p < 0.01$, 5 with $p < 0.05$), while 5 of the 24 were not able to significantly discriminate congruent from incongruent trials. The differences in performance are most likely related to the difficulty of the task—it is intuitively more difficult to compare two temporally extended patterns when

they are serially presented than when they are simultaneously available.

In the case of bimodal trials, we further investigated subjects' ability to detect congruence and whether any bias was involved in their congruency judgement. This was necessary before proceeding with the contrast between EEG responses to bimodal congruent and incongruent stimuli. 22 out of 24 subjects had d' estimates greater than 1, with 19 of these exceeding a d' of 2. The subject average was 2.92, indicating that the difference between congruent and incongruent bimodal stimuli was clearly perceptually detectable. Interestingly, almost all subjects (21 out of 24) had a negative $\log \beta$ estimate revealing a bias to respond 'congruent' with a mean $\log \beta$ of -0.8 (± 1 standard deviation). Maximum and minimum bias estimates were 1.65 and -2.6132 , respectively.

As mentioned above, each stimulus of sequential pairs is treated as an independent stimulus in either the unimodal visual or unimodal auditory condition. Thus, we are not concerned with the behavioural results for those conditions. In all further analysis, all stimuli are used to calculate results, regardless of whether they appeared in correctly or incorrectly answered trials.

Stimulus locking: can we find it in human EEG?

We first had to establish whether EEG entrains to the dynamics of the presented stimuli. As mentioned in the Methods section, we quantify the locking for each individual input signal by correlating it with the measured EEG from any trial in which the modality of interest was stimulated with that input, thus averaging over conditions. Separate conditions are compared in the next section.

Visual locking

The spectrotemporal analysis approach quantifies the amount of correspondence between the magnitude of stimulus change and the spectrotemporal power of the EEG within a given frequency range. To evaluate whether our data show stimulus locking, we first concentrate on the 20–35 Hz frequency band reported by Kayser and König (2004). Here we examine the effect at the population level, averaging first over all visual conditions (bimodal congruent and incongruent and unimodal visual) and then over subjects. As can be seen in Fig. 3a, we find stimulus locking to visual stimuli within this frequency range at occipital electrodes (O1, OZ, O2), peaking at a lag of 92 ms with a maximal correlation coefficient of 0.01.

Looking now at the entire range of frequencies that are available, again at the population level, we see at occipital electrodes that in addition to this positive correlation between visual stimulus dynamics and EEG power

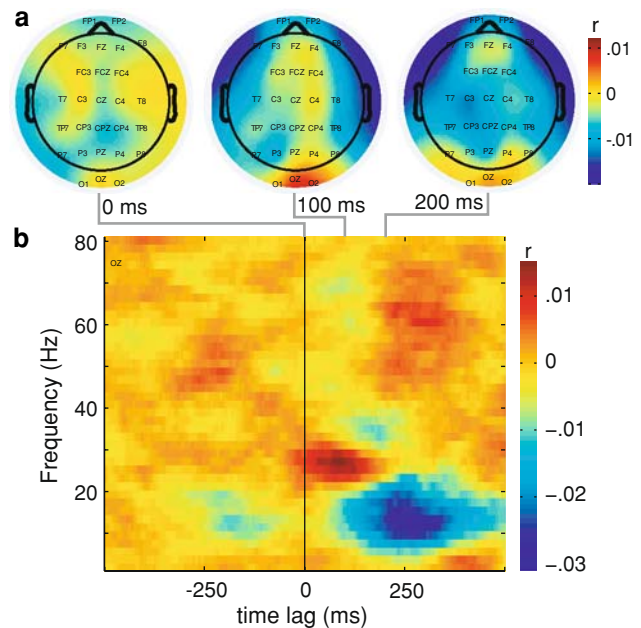


Fig. 3 Stimulus locking of EEG power to visual stimuli. **a** Each plot shows the topographic distribution of the grand average of the spectrotemporal analysis correlation results (averaged over bimodal congruent, incongruent and unimodal visual conditions and over all subjects) within the 20–35 Hz band, at selected time lags beginning at 0 ms and ending at 200 ms lag. Dots represent locations of labelled electrodes, with locations below head centre drawn outside the cartoon head. Colour codes for the magnitude of the correlation coefficient, with warm colours indicating positive and cold colours negative correlation, and values are linearly interpolated between recording sites for visualisation purposes. **b** Grand average spectrotemporal analysis correlograms (as in **a**) for individual frequencies are shown for electrode OZ. Each row represents the results for a single frequency, with frequencies given on the y-axis and correlation lags on the x-axis. Colour codes for magnitude of correlation coefficient

($r = 0.015$ at OZ), there is also a negative correlation found in lower frequencies at a later correlation lag ($r = -0.03$ at OZ, see Fig. 3b). Although these effect sizes are small, they are clearly different to baseline, and significance testing using permutation tests reveals that these peaks are indeed highly significant ($p < 0.01$). Thus, we report two ranges of stimulus-locking: a positive peak centred at approximately 27 Hz occurring at 80 ms lag after visual stimulation; and a strong negative correlation between 8 and 20 Hz beginning at approximately 180 ms lag. The difference in direction of these correlations suggests that there are at least two frequency-specific stimulus-locking phenomena in response to visual stimulation.

These results are supported by an analysis of individual subjects. In 10 out of 24 subjects, a significant peak was found in the beta range for at least one occipital recording site. The clearest beta correlation was not found in the same occipital channel for all participants, with some displaying a strongly lateralised effect. The anti-correlation found in lower frequencies showed less variability across subjects,

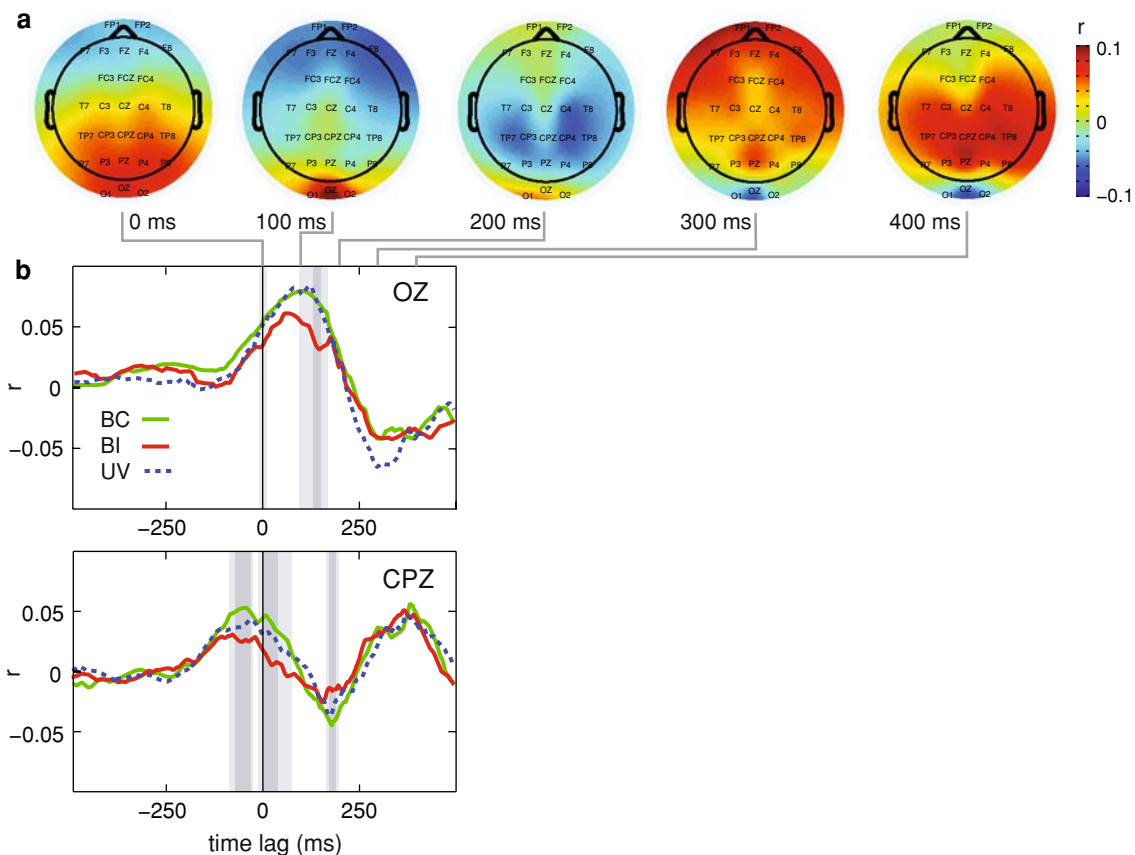


Fig. 4 EEG waveforms lock to visual input at multiple sites. **a** *Head-plots* show the topographic distribution of the grand average correlation coefficients between ERP waveform (averaged over bimodal congruent and incongruent and unimodal visual conditions) and visual stimulus speed at selected time lags between 0 and 400 ms. Correlation magnitude is colour coded as in Fig. 3a, b) Condition comparison for electrodes OZ (*upper panel*) and CPZ (*lower panel*). Waveform anal-

ysis correlograms for bimodal congruent (BC, *green line*), bimodal incongruent (BI, *red*) and unimodal visual (UV, *dashed blue*) are shown with time lags on the *x*- and correlation coefficients on the *y*-axis. Time lags showing significant differences between bimodal incongruent and bimodal congruent conditions are highlighted in *light grey* for $p < 0.05$ and *darker grey* for $p < 0.01$

and EEG power of 20 out of 24 subjects was found to significantly anti-correlate with visual stimuli at an occipital electrode.

Visual stimulus speed is reflected not only in the power profile of single EEG trials but also in the EEG waveform itself, as can be seen from the waveform analysis results (Fig. 4a). The grand average cross-correlogram between stimulus speed and the stimulus-specific ERP waveform, averaged over all visual conditions, reveals peaks at various recording sites across the scalp. The strongest effects are found at occipital sites at 78 ms lag with a magnitude of correlation of 0.12. An extensive cluster of electrodes in the centro-parietal region shows a more complex pattern of locking, with a first positive peak at approximately -60 ms lag followed by a negative peak at around 180 ms and a second positive peak at approximately 390 ms lag. The strongest effect within this cluster is seen at CPZ ($r = 0.09$). Given that EEG at these sites follows the stimulus with differing delays, occipital and centro-parietal clusters are likely to reflect distinct underlying cortical sources.

Contrasting the two analytic approaches, we see that the waveform analysis provides a stronger, consistent measure of stimulus locking. For almost all subjects (23 of 24), EEG recorded from OZ correlates with stimulus speed ($p < 0.01$ for 22 subjects, $p < 0.05$ for one subject). Significant peaks occur between zero and 300 ms lag (median r : 0.1372 for significant subjects, 0.1353 for all subjects) for the average of all visual conditions. Stimulus locking in the centro-parietal region is similarly stable, with 20 subjects showing a significant effect (19 with $p < 0.01$, 1 with $p < 0.05$) at CPZ between 200 and 460 ms lag, (median r : 0.1229 for significant subjects, 0.1100 for all subjects). Thus, phase-locking, as captured by waveform analysis, is more reliable than the power-locking quantified by our spectrotemporal approach, making it more useful for investigating condition differences.

Auditory locking

As a next step, we asked whether stimulus locking can also be found in the auditory system. Here, we did not have an

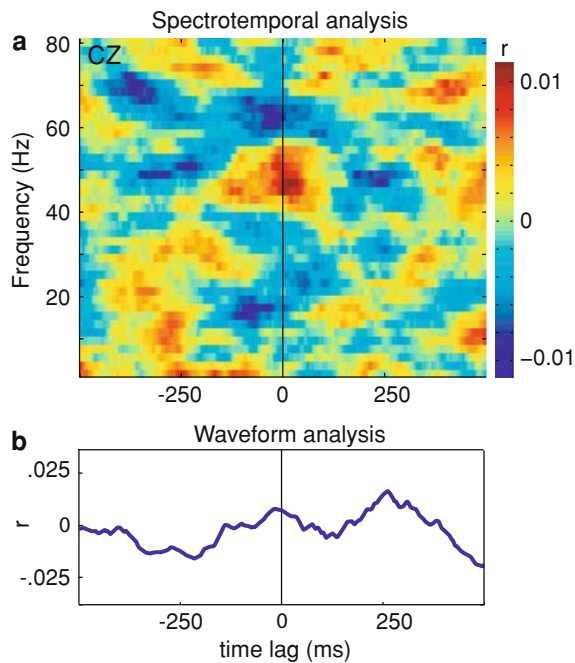


Fig. 5 No stimulus locking to auditory input. Exemplary unimodal auditory results are depicted for electrode CZ. **a** Spectrotemporal analysis cross-correlograms are shown in rows for individual frequency bands (y-axis). Colours represent correlation coefficients. **b** The waveform analysis cross-correlation is represented with time lags on the x- and correlation coefficients on the y-axis

initial frequency band or site to guide our investigation. Representative results for both analytic approaches are depicted in Fig. 5 for electrode CZ (chosen because its auditory ERP showed the largest evoked potential of all channels shortly (~ 100 ms) after stimulus onset) for the unimodal auditory condition. Although the spectrotemporal analysis results for this recording site suggest there may be stimulus-locking between 45 and 55 Hz, this correlation is not salient with respect to other lags and frequencies, and is furthermore not found at neighbouring channels, nor is it consistent across auditory conditions. Waveform analysis cross-correlograms also show no salient peaks, with very low correlation coefficients over all time lags. As in the visual case, we also looked at the average over all auditory conditions, since we assumed that averaging over more trials might uncover small effects. However, the obtained results are similarly uninformative about auditory stimulus-locking. In addition, no consistent effects were seen on the individual subject level for either analytic approach.

Is there evidence for multisensory interactions?

To address the question of whether stimulus locking is subject to multisensory interactions, we contrasted different stimulation conditions. As we did not find any evidence for auditory stimulus locking, we concentrate in the following

on visual stimulus locking and how it is modulated by the presence of matching or mismatching auditory input.

Despite the absence of auditory stimulus locking, the waveform analysis results revealed that incongruent auditory information modulates visual stimulus locking. This crossmodal effect is indicated by diminished peak correlation coefficients in bimodal trials containing incongruent auditory input, compared to bimodal congruent trials. Occipitally, the maximum correlation of the bimodal incongruent grand average ($r = 0.0612$) is almost 25% smaller than the peak of the bimodal congruent condition ($r = 0.0795$).¹ Time lags between 98 and 168 ms have statistically less locking in incongruent than congruent bimodal trials ($p < 0.05$). The unimodal visual condition, in contrast, shows the same correlation as the bimodal congruent condition ($p > 0.05$). For centro-parietal EEG, the biggest condition differences are found for the first of the two peaks mentioned above. For time bins between -90 and -30 , and -10 and 72 ms of lag, bimodal incongruent coefficients are significantly smaller than for the bimodal congruent or unimodal visual conditions ($p < 0.05$, see Fig. 4b). Likewise, the magnitude of the subsequent anti-correlation is lower for bimodal incongruent than both other conditions for lags between 168 and 198 ms ($p < 0.05$). It seems then that incoherent auditory input suppresses visual stimulus locking during multimodal stimulation.

Applying the spectrotemporal analysis method, no condition differences were found for the stimulus-locking effects mentioned above. However, the unimodal visual, bimodal-congruent and bimodal-incongruent condition all showed significant peaks in the frequency ranges and time-lags reported for the average condition reported above: peak cross-correlation values for the positive peak within the limits of 0–150 ms lag and 20–30 Hz were 0.018, 0.020, and 0.17, respectively; for the negative peak they were -0.035 , -0.034 , and -0.035 within lag limits 150–500 ms and 8–20 Hz. Thus, spectrotemporal stimulus locking to visual stimulus dynamics is rather similar between different conditions and seems to be a bottom-up driven cortical response independent of input to other modalities.

Do the effects change over time?

To examine the progression of stimulus locking over the duration of stimulus presentation, we evaluated cross-correlations as a function of stimulus time. To this end, correlation coefficients were determined within a shifting time

¹ Peak correlations for single conditions are all smaller than for the condition average reported earlier, as the single condition waveform is only constructed from a maximum of 3 trials instead of 9 trials in the condition average case.

window (see “Methods”). Before investigating unimodal visual, bimodal congruent and bimodal incongruent conditions separately, we again examined their condition average.

Figure 6a shows the waveform analysis results for selected occipital and central electrodes. At occipital recording sites, stimulus locking increases with progressing

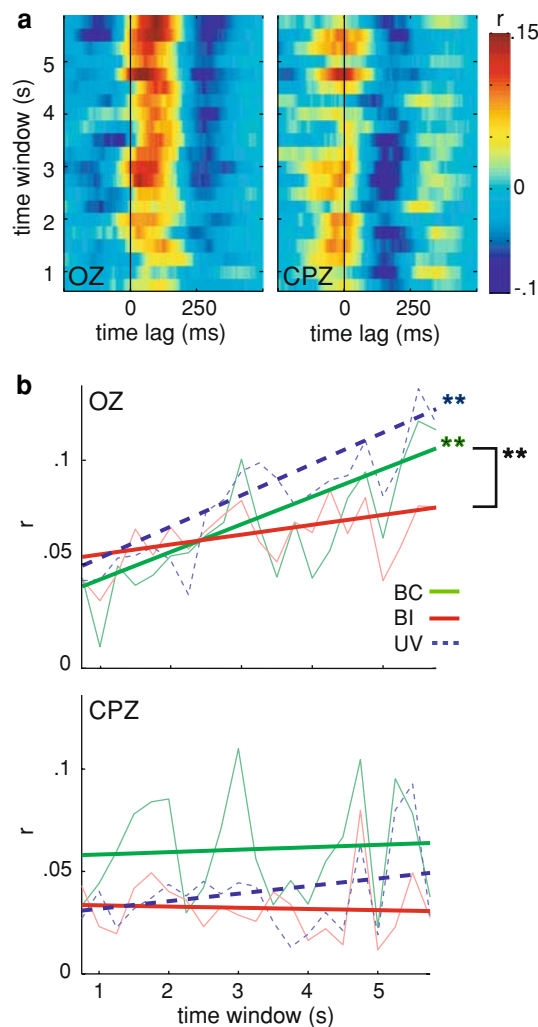


Fig. 6 Temporal progression of stimulus locking. **a** Each row shows one grand average waveform analysis cross-correlogram (waveforms calculated from average of bimodal congruent and incongruent and unimodal visual conditions), which was computed within a 500 ms shifting time window with 250 ms overlap. From bottom to top, the window is shifted from stimulus onset to the end of stimulus presentation. Correlation coefficients are colour-coded. **b** The maximum waveform analysis correlation coefficients (y-axis) are plotted as a function of time for OZ (left panel) and CPZ (right panel). Peaks were determined within a region of interest (0–250 ms lag) for each time window from the grand average of each condition. The resulting data points are shown in light-coloured lines for bimodal congruent (BC, green), bimodal incongruent (BI, red) and unimodal visual (UV, dashed blue). The according linear fits for each condition are shown in darker thick lines in the corresponding colour scheme. Stars at line ends indicate significant linear trends ($p < 0.05$) within a condition, and starred brackets indicate significantly different slopes between two conditions

stimulus time—immediately after stimulus onset, correlation coefficients are very small and only start to increase after approximately 2 s of visual stimulation. In contrast, the magnitude of stimulus locking at central sites is not dependent on time. Locking is roughly constant, with no apparent systematic change. Overall, we see different time courses of visual stimulus locking at occipital and centro-parietal electrodes.

To quantify the temporal progression and obtain a better comparison across conditions, we extracted the maximum correlation coefficient for each time window. These time series were then fitted using linear regression. Results are shown for bimodal congruent and incongruent and unimodal visual conditions at selected occipital and central electrodes (Fig. 6b). At the occipital site, there is a significant linear increase for the bimodal congruent (regression coefficient = 0.0027, $r^2 = 0.5$; $p < 0.01$) and unimodal visual condition (regression coefficient = 0.0034, $r^2 = 0.75$; $p < 0.01$). In contrast, locking in the bimodal incongruent condition does not follow any linear trend (regression coefficient = 0.0007, $r^2 = 0.12$; $p > 0.05$). As expected from these data, a comparison of bimodal congruent and incongruent slopes yields highly significant results ($p < 0.01$), while there is no statistical difference between bimodal congruent and unimodal visual stimuli. At our representative central site, there are no increases in locking over time in any condition ($p > 0.05$). There are, however, condition differences regarding the magnitude of stimulus locking: although all conditions show significant stimulus locking ($p < 0.01$), peak correlations are significantly higher for bimodal congruent than incongruent and unimodal visual stimuli ($p < 0.01$). In summary, our results suggest two distinct multisensory interactions at work. At central recording sites we see a sensitivity to congruence between auditory and visual input streams, and at occipital sites a further time-dependent facilitation of visual stimulus-locking is evident that is suppressed when conflicting auditory information is present.

In the case of the spectrotemporal analysis approach, we first defined frequency bands of interest, guided by the visual stimulus locking results reported above (8–20 Hz and 20–30 Hz). The sum of spectrotemporal power in these bands was then correlated with stimulus speed using the shifting time window method just described, and peak correlation values (minimum in the case of the anti-correlation, and maximum for the correlation) were extracted from lag limits defined from the earlier results (200–300 and 0–150 ms, respectively). Finally, these time series were fitted using linear regression. No linear trend was seen for the 8–20 Hz band at occipital sites, but effect sizes were in the same range for bimodal congruent, incongruent, and unimodal visual conditions. In the case of the beta correlation, no linear trend was found for bimodal congruent and

incongruent conditions at occipital sites ($p > 0.05$); however, locking of spectrotemporal power to unimodal visual stimuli was seen to increase over time in OZ ($p < 0.05$, $r^2 = 0.27$) and O1 ($p < 0.01$, $r^2 = 0.47$).

Discussion

We measured EEG in 24 subjects while they viewed rotating Gabor patches and listened to frequency-modulated tones. Our results reveal two forms of stimulus locking of EEG to the temporal dynamics of visual stimuli, evident from changes in EEG power over time and from the temporal structure of the EEG waveform. Auditory stimuli presented alone do not lead to any measurable, structured changes in EEG power or waveform. However, analysis of bimodal trials shows that visual stimulus locking is modulated by the temporal congruence of simultaneous auditory input—locking of EEG waveforms to visual stimuli was reduced when the auditory input mismatched the temporal profile of the visual input. Furthermore, this multisensory interaction became even clearer when the time course of waveform locking was examined. Under congruent bimodal stimulation conditions, visual stimulus locking increased steadily over the duration of the trial, while the incongruent bimodal condition showed no such effect. Thus we suggest that stimulus locking is a suitable tool for studying and characterising the multisensory processing of dynamically changing auditory and visual stimuli.

In order to explore the importance of extended dynamic modulation of auditory and visual stimuli, we required a relatively long trial duration. This raised the need for a task to ensure participants' attention was maintained. We decided on a task that required attention to both stimulus components in parallel—to decide whether the auditory and visual signals matched—rather than directing the subject to focus on a single sensory modality. As a result, the task could not be directly transferred to single unimodal components of the audiovisual stimuli, and was instead applied to sequentially presented pairs of components from either modality. The simultaneous and sequential comparison of stimulus dynamics constitutes two different tasks and necessitates a careful comparison of unimodal and bimodal conditions. However, the task allows us to be confident of subjects' attentiveness during the experiment.

Here, as a first step, we report results for signals measured at electrodes, i.e. in sensor space. An EEG signal is a combination of signals from different cortical sources and noise. It would be desirable to further process the data in order to translate our findings into source space by performing a source localisation that would attempt to isolate the location and activity of the cortical sources involved. However, this requires additional assumptions and currently

there is no available method that is generally accepted and free of problems (Nunez and Srinivasan 2005). An additional issue is that algorithms for source localization make assumptions about the interaction or independence of different signals. Work is still under way to develop synchrony measures in source space (e.g. Marzetti et al. 2008). Although our results do not rely on a spatial interpretation, we do make some tentative assumptions here: that EEG measured at occipital channels captures activation in early visual areas proximal to these measurement sites, whereas effects we find at centro-parietal electrodes reflect cortical processing at some higher stages. A more detailed investigation of the cortical sources of the stimulus locking phenomena described here must be left to future work.

We did not see any evidence for stimulus locking to auditory stimuli, or to the auditory component of audiovisual stimuli. This is surprising, as it is commonly assumed that the temporal reliability of the auditory system is higher than the visual. As a consequence, its contribution in optimal sensory fusion is high for temporal estimates (e.g. Bresciani et al. 2008) and small for spatial estimates (e.g. Körding et al. 2007). In addition, the modulation of brain responses to changes in amplitude or frequency of simple auditory stimuli is a well-established phenomenon. Auditory entrainment has been optimally found in response to 40 Hz auditory stimulation (Galambos et al. 1981), but has also been reported at lower frequencies (e.g. Ding et al. 2006). A systematic examination of the steady-state response to sinusoidal frequency modulation by Picton et al. (1987) found that for lower modulation frequencies, responses were most reliable between 3 and 7 Hz. The amplitude and phase of responses to tones modulated at the rates of change used in our study were not found to be significantly reliable, and thus it may simply be the case that the auditory stimuli used here change too slowly. Indeed, preliminary results of current work with 3 Hz frequency modulation have shown a tendency to a locking of 40 Hz power to the dynamics of the auditory stimulus.

We applied two analytic approaches to investigate how the dynamics of the stimulus are reflected in the power and phase of the measured EEG signals, respectively. The first approach, spectrotemporal analysis, correlates power changes with stimulus speed, and can reveal locking of EEG power changes to stimulus dynamics at frequencies beyond the rate of change of the stimulus. The second approach, waveform analysis, correlates ERP waveforms with stimulus speed, and thus mirrors a stable phase-relationship between EEG signals and the time-course of the stimulus. Due to the use of correlation, the entrainment we observe for the EEG waveform must result from the phase-locking of frequencies within the range contained in the stimulus speed. Previously, locking of LFP power to dynamic visual stimulation has been investigated in cat

visual cortex (Kayser and König 2004) using the same spectrotemporal locking analysis. In addition, entrainment or phase-locking of neural responses to regularly repeating stimuli has been extensively studied (e.g. Rager and Singer 1998) and this steady-state cortical response has been used to investigate many other phenomena, especially attention (e.g. Müller et al. 2003). Our analytic approaches are in line with both kinds of locking mechanisms, but are applied here to explore stimulus locking of human EEG to continuously, irregularly changing visual stimuli.

Our spectrotemporal approach investigated locking to all frequencies between 0 and 100 Hz. Within this range, different frequency bands have been defined, including delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (>30 Hz) and these bands have been suggested to reflect different functionalities (see Steriade et al. 1990 for an early review). We found two instances of spectrotemporal locking that are centred at distinct frequencies (20–30 and 8–20 Hz) and furthermore differ in correlation lag and direction of effect. As such, we suggest that we are dealing with two distinct spectrotemporal locking processes with different functional significance.

The first effect found was that induced power changes in the beta (20–30 Hz) range lock to the speed profile of visual input with a lag of 92 ms. This is in agreement with previous results that found locking between 20 and 35 Hz at around 100 ms lag in cat visual cortex (Kayser and König 2004). Curiously, the frequency showing the strongest locking in our results is close to 25 Hz, which is exactly the frame rate at which our movies were created, but is far from the refresh rate of the monitor (100 Hz). However, it is not clear whether the correspondence with the frame rate is simply coincidental, which brings into question the functional significance of the 20–30 Hz band. There is a crucial difference between our study and the earlier work in cats—humans have a comparatively lower flicker fusion frequency than cat, so the 25-Hz frame rate can be considered to be adequate for human viewers. Although we have clearly found evidence for spectrotemporal locking to irregular visual stimulation, we cannot generalise this finding to more natural conditions.

The second effect revealed by the spectrotemporal analysis was an anti-correlation in a broad alpha band (8–20 Hz). The human occipital alpha rhythm has traditionally been associated with a state of cortical rest, as put forward by the idling hypothesis. More recent studies have extended the idling hypothesis, proposing that a strong alpha rhythm characterises a dominance of “top-down” processing in the absence of any external stimulation (von Stein et al. 2000; see Palva and Palva 2007 for a general review). Although the results here span a frequency band too broad to compare to the classical alpha band (8–13 Hz), our results fit with the idling hypothesis account, as stimulation strength

anti-correlates with alpha power, with stronger input leading to reduced alpha oscillations. The anti-correlation found here does not show any modulation by auditory input, so we assume that the spectrotemporal locking we have observed is a purely visual phenomenon.

Overall, the effect sizes found using the spectrotemporal approach were small. These correlations are calculated using total power, which includes both induced (not phase-locked to stimulus onset) and evoked (phase-locked to stimulus onset) components (Tallon-Baudry and Bertrand 1999). Evoked power constitutes only a small part of total power—here approximately 15%—so if the stimulus-locked power changes are indeed consistent in their phase with respect to the stimulus dynamics, then the stimulus-locking effect may be hidden in the total power. In comparison, Kayser and König's (2004) results were indeed larger, with correlations of approximately 0.13 compared to our 0.01, but were based on intra-cranial LFP measurements that do not suffer from the spatial smearing inherent in EEG.

The second approach we employed, waveform analysis, correlated ERP waveforms with stimulus speed. Our results revealed locking to visual stimulus dynamics, distributed across many measurement sites. An examination of the characteristics of the correlation results suggests at least two distinct locking mechanisms: one measured occipitally with a positive correlation at a lag of 78 ms, interpreted as an early visual process; and the other appearing at centro-parietal sites as positive locking at lags of –60 and 390 ms, interpreted as later stages of visual processing. A positive correlation indicates that the phase of the EEG waveform is aligned to the dynamics of the stimulus at a given time lag; however, the relative magnitude of both signals also contributes to this measure of locking and we cannot isolate the role of phase alignment from the role of amplitude changes. As mentioned above, we can furthermore specify that entrained frequencies must lie in the range contained in the speed profiles of our stimuli (0.5–4 Hz). Although steady-state responses are typically evoked using stimuli presented at higher, regular stimulation frequencies, entrainment has also been observed for frequencies in the delta range (Ding et al. 2006). Less regular stimulus trains jittered in time within this frequency range have also been found to have an entrainment effect on visual and auditory cortex, which has been shown to be the result of an instantaneous phase reset of the ongoing oscillatory activity (Lakatos et al. 2008; Lakatos et al. 2005). Hence, we assume that in the case of our much more irregular stimuli, faster visual input leads to a stronger phase-reset across a large population of neurons, and thus to stronger EEG amplitude that is phase-aligned with the stimulus. Furthermore, we found that entrainment to visual stimuli increased over stimulus duration at occipital sites, indicating that phase alignment to our stimuli is a gradual process—there is no instantaneous phase reset.

The role of ongoing neural oscillations has been emphasised in multisensory research, in particular frequencies in the gamma range (see Senkowski et al. 2008 for a recent review). The entrainment of lower frequency oscillations, which are of particular interest here due to our choice of stimuli, has also been implicated in multisensory processes. In auditory cortex, the phase of all ongoing neuronal oscillations has been proposed to be reset by tactile or visual input, with this general phase reset hypothesised to allow more effective processing of subsequent auditory signals arriving at an optimal phase of this reset activity (Lakatos et al. 2007; Kayser et al. 2008). Such a mechanism has also been proposed to be useful for audiovisual speech, with the regularities of the visual input allowing a facilitation of temporally matched auditory input in noisy backgrounds (Schroeder et al. 2008). In conclusion, crossmodal resetting of cortical activity in a “unimodal” area may allow prioritised processing of temporally corresponding stimuli of the preferred modality, which would be of great importance for multisensory processing.

Our waveform analysis results for audiovisual conditions can be related to these findings. In general, we found stronger visual locking to congruent than incongruent audiovisual stimuli. This difference in entrainment to visual dynamics must be due to an effect of the auditory stimulus dynamics. One possible interpretation is that congruently timed auditory stimulation enhances the phase reset of visually entrained neurons. Furthermore, an examination of the time course of the visual entrainment of waveforms suggested two locking mechanisms, which seem to be differently involved in multisensory processing when the entrainment to purely visual stimuli is taken into account. The visual process measured at occipital sites shows a continual increase in visual entrainment over the duration of unimodal visual and congruent audiovisual stimulation, but a constant level of entrainment for mismatched audiovisual inputs. As such, incongruent auditory input seems to counteract visual entrainment, and the difference in entrainment between congruent and incongruent conditions unfolds gradually and slowly. The time course involved, in the magnitude of seconds, is a strong indication that this process is initially driven by bottom-up visual input, with information regarding multisensory congruence arriving later, most likely via feedback from higher areas. Along these lines, Noesselt et al. (2007) recently showed that modulations in primary unisensory areas in response to temporally aligned audiovisual patterns were a result of feedback from superior temporal sulcus. The visual locking found here at centroparietal sites shows no change over time, with congruent audiovisual input showing a clear advantage for locking over both incongruent and unimodal visual stimulation. Thus, this process depends on both input streams; however, it is unclear whether this mechanism reflects an evaluation of congruence or is itself the result of

a match/mismatch evaluation from elsewhere. Source analysis methods, and a direct evaluation of the directional interaction between the two waveform locking mechanisms, may help elucidate this in future work.

The behaviour of a dynamical system can be characterised by examining the relationship between its output and a known input. Often-used inputs are of two different kinds: impulses or continuous stimuli. In the case of a linear system, these two inputs yield equivalent results. To date, sensory processing has been analysed with an emphasis on this impulse-response concept, specifically multimodal interaction (see Calvert and Thesen 2004 for a general review of methodological approaches in multisensory research). Here we want to fully exploit the high temporal resolution of EEG to examine the nature of a temporal process, and it is rather the second kind of input that is of interest. The present analysis amounts to an investigation of the linearity of the audiovisual interactions. Indeed the observation of a time-dependant increase in coupling in the bimodal congruent condition and unimodal visual is difficult to reconcile with a purely linear system. The characteristics of the stimuli and the analysis methods used in this study are better suited to such investigations of non-linear system behaviour than traditional impulse response approaches.

The integration of two sensory sources critically depends on the compatibility of the temporal dynamics of both information streams. We used stimulus locking to extended stimuli with complex temporal profiles as a tool to investigate multisensory interactions. Spectrotemporal locking revealed a bottom-up, purely visual effect. In contrast, waveform locking showed modulation by auditory congruency and a differential effect over time. We thus propose that spectrotemporal and waveform locking reflect different mechanisms involved in the processing of dynamic audiovisual stimuli, and that these analysis approaches may prove useful in future research.

Acknowledgments This work was supported by Perception on Purpose, a European Commission specific targeted research project (FP6-IST-2004-027268). We thank Lars Hausfeld for programming of the experiment and help with data acquisition, and Hans-Peter Frey for assistance with experimental design and set-up and anonymous reviewers for their constructive comments.

Appendix 1

Behavioural pilot study

A behavioural pilot experiment was conducted to determine whether and how stimulus parameters affect the perceptual detectability of audiovisual congruence.

We tested different visual and auditory feature dimensions, as well as slower and faster modulation frequencies.

Otherwise, audiovisual stimuli were constructed as described for the main experiment. The Gabor patch could be modulated in one of 6 dimensions: size (100–200 pixels), contrast (0.1–0.5), orientation (180 degrees around horizontal axis), frequency (~0.2–1 cycles/degree), phase (0–360 degrees) or colour saturation (0.1–0.5 along the red–green axis in DKL colour space). The tone was either amplitude- (0.1–1) or frequency- (Carrier Frequency \pm 40 Hz) modulated. In the latter case, the carrier frequency was chosen randomly from a uniform distribution (100–500 Hz). Modulation frequencies were bound by either 0.7, 1.0 or 1.3 Hz. The combinations of parameter triplets (visual, auditory, cutoff) were fully balanced.

Additionally, each combination was shown an equal number of times in congruent and incongruent stimuli. Movies were created out of 6 trajectories per modulation frequency, resulting in a total number of 432 movies. Button presses were recorded from nine naive subjects. The whole stimulus set was divided and balanced between pairs of subjects, i.e. each subject saw 216 movies of a single balanced set. Paradigm, trial organisation and instruction were identical to the main experiment (simultaneous trials). Unimodal stimuli were not shown.

On average, subjects responded correctly on 74.9% of all trials (standard deviation: 9.3%). The analysis of responses indicated that each subject performed well above chance (exact Binomial test, $p < 0.01$). A comparison of different visual features, auditory features and modulation frequencies revealed no significant differences (χ^2 , $p < 0.05$).

We conclude that subjects are capable of distinguishing audiovisual congruent from incongruent stimuli. Furthermore, sensitivity to temporal congruence does not seem to be dependent on the specific low-level features through which the temporal structure is conveyed.

References

- Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A (2004) Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nat Neurosci* 7(11):1190–1192
- Bidet-Caulet A, Fischer C, Bauchet F, Aguera P, Olivier B (2007) Neural substrate of concurrent sound perception: direct electrophysiological recordings from human auditory cortex. *Front Hum Neurosci* 1(5) doi:10.3389/neuro.09.005.2007
- Bonath B, Noesselt T, Martinez A, Mishra J, Schwiecker K, Heinze HJ, Hillyard SA (2007) Neural basis of the ventriloquist illusion. *Curr Biol* 17(19):1697–1703
- Bresciani J-P, Dammeier F, Ernst MO (2008) Tri-modal integration of visual, tactile and auditory signals for the perception of sequences of events. *Brain Res Bull* 75(6):753–760
- Bushara KO, Grafman J, Hallett M (2001) Neural correlates of auditory-visual stimulus onset asynchrony detection. *J Neurosci* 21(1):300–304
- Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex* 11(12):1110–1123
- Calvert GA, Thesen T (2004) Multisensory integration: methodological approaches and emerging principles in the human brain. *J Physiol Paris* 98(1–3):191–205
- Calvert GA, Brammer MJ, Bullmore ET, Campbell R, Iversen SD, David AS (1999) Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport* 10(12):2619–2623
- Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* 10(11):649–657
- Calvert GA, Hansen PC, Iversen SD, Brammer MJ (2001) Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the bold effect. *Neuroimage* 14(2):427–438
- 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(1):9–21
- Dhamala M, Assisi CG, Jirsa VK, Steinberg FL, Kelso JAS (2007) Multisensory integration for timing engages different brain networks. *Neuroimage* 34(2):764–773
- Ding J, Sperling G, Srinivasan R (2006) Attentional modulation of SSEVP power depends on the network tagged by the flicker frequency. *Cereb Cortex* 16(7):1016–1029
- Doehrmann O, Naumer M (2008) Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. *Brain Res* 1242:136–150
- Driver J, Noesselt T (2008) Multisensory interplay reveals crossmodal influences on ‘sensory-specific’ brain regions, neural responses, and judgments. *Neuron* 57(1):11–23
- Foxe JJ, Schroeder CE (2005) The case for feedforward multisensory convergence during early cortical processing. *Neuroreport* 16(5):419–423
- Galambos R, Makeig S, Talmachoff PJ (1981) A 40-Hz auditory potential recorded from the human scalp. *Proc Natl Acad Sci USA* 78(4):2643–2647
- Giard MH, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci* 11(5):473–490
- Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ (2000) Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin Neurophysiol* 111(10):1745–1758
- Kayser C, König P (2004) Stimulus locking and feature selectivity prevail in complementary frequency ranges of V1 local field potentials. *Eur J Neurosci* 19(2):485–489
- Kayser C, Logothetis NK (2007) Do early sensory cortices integrate cross-modal information? *Brain Struct Funct* 212(2):121–132
- Kayser C, Petkov CI, Augath M, Logothetis NK (2007) Functional imaging reveals visual modulation of specific fields in auditory cortex. *J Neurosci* 27(8):1824–1835
- Kayser C, Petkov CI, Logothetis NK (2008) Visual modulation of neurons in auditory cortex. *Cereb Cortex* 18(7):1560–1574
- Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L (2007) Causal inference in multisensory perception. *PLoS One* 2(9):e943
- Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE (2005) An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J Neurophysiol* 94(3):1904–1911
- Lakatos P, Chen C-M, O’Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53(2):279–292
- Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320(5872):110–113

- Macaluso E, Driver J (2005) Multisensory spatial interactions: a window onto functional integration in the human brain. *Trends Neurosci* 28(5):264–271
- Macaluso E, George N, Dolan R, Spence C, Driver J (2004) Spatial and temporal factors during processing of audiovisual speech: a PET study. *Neuroimage* 21(2):725–732
- Marzetti L, Del Gratta C, Nolte G (2008) Understanding brain connectivity from EEG data by identifying systems composed of interacting sources. *Neuroimage* 42(1):87–98
- Meredith MA, Stein BE (1983) Interactions among converging sensory inputs in the superior colliculus. *Science* 221(4608):389–391
- Meredith MA, Nemitz JW, Stein BE (1987) Determinants of multisensory integration in superior colliculus neurons I temporal factors. *J Neurosci* 7(10):3215–3229
- Miller LM, D'Esposito M (2005) Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *J Neurosci* 25(25):5884–5893
- Mishra J, Martinez A, Sejnowski TJ, Hillyard SA (2007) Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *J Neurosci* 27(15):4120–4131
- Mishra J, Martinez A, Hillyard SA (2008) Cortical processes underlying sound-induced flash fusion. *Brain Res* 1242:102–115
- 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(1):115–128
- Müller MM, Malinowski P, Gruber T, Hillyard SA (2003) Sustained division of the attentional spotlight. *Nature* 424(6946):309–312
- Noesselt T, Rieger JW, Schoenfeld MA, Kanowski M, Hinrichs H, Heinze H, Driver J (2007) Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *J Neurosci* 27(42):11431–11441
- Noesselt T, Bonath B, Boehler CN, Schoenfeld MA, Heinze HJ (2008) On perceived synchrony-neural dynamics of audiovisual illusions and suppressions. *Brain Res* 1220:132–141
- Nunez PL, Srinivasan R (2005) *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, New York
- Palva S, Palva JM (2007) New vistas for alpha-frequency band oscillations. *Trends Neurosci* 30(4):150–158
- Picton TW, Skinner CR, Champagne SC, Kellett AJ, Maiste AC (1987) Potentials evoked by the sinusoidal modulation of the amplitude or frequency of a tone. *J Acoust Soc Am* 82(1):165–178
- Rager G, Singer W (1998) The response of cat visual cortex to flicker stimuli of variable frequency. *Eur J NeuroSci* 10(5):1856–1877
- Regan D (1966) Some characteristics of average steady-state and transient responses evoked by modulated light. *Electroceph clin Neurophysiol* 20:238–248
- Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A (2008) Neuronal oscillations and visual amplification of speech. *Trends Cogn Sci* 12(3):106–113
- Senkowski D, Talsma D, Grigutsch M, Herrmann CS, Woldorff MG (2007) Good times for multisensory integration: effects of the precision of temporal synchrony as revealed by gamma-band oscillations. *Neuropsychologia* 45(3):561–571
- Senkowski D, Schneider TR, Foxe JJ, Engel AK (2008) Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci* 31(8):401–409
- Shams L, Kamitani Y, Shimojo S (2000) Illusions: what you see is what you hear. *Nature* 408(6814):788
- Shams L, Kamitani Y, Shimojo S (2002) Visual illusion induced by sound. *Brain Res Cogn Brain Res* 14(1):147–152
- Shipley T (1964) Auditory flutter-driving of visual flicker. *Science* 145:1328–1330
- Stein BE, Meredith MA (1993) *The merging of the senses*. MIT Press, Cambridge
- Steriade M, Gloor P, Llinas RR, Lopes da Silva FH, Mesulam MM (1990) Basic mechanisms of cerebral rhythmic activities. *Electroceph Clin Neurophysiol* 76(6):481–508
- Tallon-Baudry C, Bertrand O (1999) Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn Sci* 3(4):151–162
- van Atteveldt NM, Formisano E, Blomert L, Goebel R (2007) The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cereb Cortex* 17(4):962–974
- von Stein A, Chiang C, König P (2000) Top-down processing mediated by interareal synchronization. *Proc Natl Acad Sci USA* 97(26):14748–14753
- Vroomen J, De Gelder B (2004) Perceptual effects of cross-modal stimulation: ventriloquism and the freezing phenomenon. In: Calvert GA, Spence C, Stein BE (eds) *Handbook of multisensory processes*. MIT Press, Cambridge, pp 141–150
- Wallace MT, Wilkinson LK, Stein BE (1996) Representation and integration of multiple sensory inputs in primate superior colliculus. *J Neurophysiol* 76(2):1246–1266
- Wickens DT (2002) *Elementary signal detection theory*. Oxford University Press, New York