

This article has been published as:

Keitel, C. and M. M. Müller (2016). "Audio-visual synchrony and feature-selective attention co-amplify early visual processing." *Exp Brain Res* **234**(5): 1221-1231

© Springer-Verlag Berlin Heidelberg 2015

TITLE

Audio-visual synchrony and feature-selective attention co-amplify early visual processing

AUTHORS:

Christian Keitel¹, Matthias M Müller²

AFFILIATIONS:

1 – Institute of Neuroscience and Psychology, University of Glasgow, Hillhead Street 58, G12 8QB

Glasgow, UK

2 – Institut für Psychologie, Universität Leipzig, Neumarkt 9-19, 04109 Leipzig, Germany

CORRESPONDING AUTHOR:

Christian Keitel, christian.keitel@glasgow.ac.uk

KEYWORDS:

feature-based attention, selective attention, multisensory integration, audio-visual synchrony, brain oscillations, neural rhythms, steady-state visual evoked potentials (SSVEP), EEG, brain-computer interface (BCI)

ABSTRACT

Our brain relies on neural mechanisms of selective attention and converging sensory processing to efficiently cope with rich and unceasing multisensory inputs. One prominent assumption holds that audio-visual synchrony can act as a strong attractor for spatial attention. Here, we tested for a similar effect of audio-visual synchrony on feature-selective attention. We presented two superimposed Gabor patches that differed in colour and orientation. On each trial participants were cued to selectively attend to one of the two patches. Over time, spatial frequencies of both patches varied sinusoidally at distinct rates (3.14 and 3.63 Hz) giving rise to pulse-like percepts. A simultaneously presented pure tone carried a frequency modulation at the pulse rate of one of the two visual stimuli to introduce audio-visual synchrony. Pulsed stimulation elicited distinct time-locked oscillatory electrophysiological brain responses. These steady-state responses (SSRs) were quantified in the spectral domain to examine individual stimulus processing under conditions of synchronous vs. asynchronous tone presentation and when respective stimuli were attended vs. unattended. We found that both, attending to the colour of a stimulus and its synchrony with the tone, enhanced its processing. Moreover, both gain effects combined linearly for attended in-sync stimuli. Our results suggest that audio-visual synchrony can attract attention to specific stimulus features when stimuli overlap in space.

INTRO

Knowing in advance the colour, shape, or orientation of an object we are looking for will facilitates our visual search. Long-standing conceptualizations of visual attention thus posit that visual objects can be selected based on their constituent features or feature conjunctions (Treisman and Gelade 1980; Wolfe 1994). Corresponding visual search accounts have received broad support from neurophysiological research showing that vast populations of neurons in visual cortices respond preferentially to specific features (reviewed in Maunsell and Treue 2006). More specifically, attending to a preferred feature leads to a neuronal response gain (Treue and Martinez Trujillo 1999). This gain effect enhances the neural representation of a stimulus carrying the attended feature(s). As a consequence, that stimulus experiences a processing advantage as compared with stimuli carrying unattended features. Finally, enhanced neural processing of target stimuli leads to facilitated behavioural performance in visual search tasks (Eimer and Grubert 2014). In summary, tuning the visual system to the feature(s) of the search target aids in determining its location. Search performance however falters when a visual scene comprises objects that carry very similar features because a feature-based selection alone becomes difficult. Van der Burg et al. (2008) demonstrated that in such cases visual search can benefit from informative auditory cues. They presented participants with displays cluttered with small green and red bars of different oblique orientations. During experimental stimulation some randomly chosen bars frequently changed to the opposite colour. The task was to find the one bar that was either vertically or horizontally oriented (irrespective of its colour). In addition to this visual-only condition the authors introduced another condition in which a spatially uninformative tone coincided with the colour flip of the target bar. This manipulation drastically reduced the average time participants needed to locate the target. Moreover, whereas in the visual-only condition search times increased considerably with the number of bars in the display, in the audio-visual condition the display size had a negligible effect on search performance. Van der Burg et al. (2008) interpreted their findings along the following lines: The coincidence of the auditory tone and the task-relevant colour flip – the audio-visual synchrony – led

to a binding of the two unisensory events into a multisensory percept. Compared with the non-target bars this multisensory percept possessed greater saliency and thus automatically attracted attention towards its location. In fact, the suggested underlying auditory-induced pop-out of the target has lent the phenomenon its name, the “pip-and-pop” effect.

Inspired by pip-and-pop and related findings, Talsma et al. (2010) proposed a generalized framework delineating how selective attention and multisensory binding interface in stimulus processing: They suggested that an infrequent salient auditory tone will typically bias visual processing towards a synchronously changing visual stimulus in cluttered scenes. The multisensory interaction between synchronous auditory and visual elements will then guide spatial attention in a stimulus-driven manner.

Whereas ample evidence documents such a pull effect on *spatial* attention in behavioural (Van der Burg et al. 2010) and neuroimaging data (Van der Burg et al. 2011), a similar effect on *feature-selective* attention stands to be demonstrated. This is a non-trivial issue because both forms of attention influence visual processing differently; spatial attention enhances neural stimulus representations based on their positions, i.e. locally. Feature-selective attention instead enhances the representation of specific features globally throughout the entire visual field (Saenz et al. 2002). Moreover, spatial attention likely constitutes a supramodal mechanism (Macaluso and Driver 2003; Busse et al. 2005; Lakatos et al. 2009) while visual feature based selection is necessarily restricted to the visual modality. Nevertheless, results of an earlier behavioural study suggest that feature-based selection can aid in audio-visual synchrony discrimination – albeit to a lesser extent than spatial attention (Fujisaki and Nishida 2008).

In the present study, we tested for effects of synchrony on feature-selective processing of visual stimuli. To this end, participants viewed two superimposed Gabor patches, one red and tilted clockwise and one green and tilted counter-clockwise. Over time, spatial frequencies of both patches varied sinusoidally at distinct rates (*Figure 1A*). These changes gave rise to a pulse-like motion. On each trial participants were cued to selectively attend to the red or green pulsing Gabor patch to

perform a visual detection task. A simultaneously presented pure tone carried a frequency modulation at the pulse rate of one of the two visual stimuli to introduce audio-visual synchrony (*Figure 1A*). This synchrony (or temporal congruency) is widely regarded as a prime requisite for the integration of unisensory input into multisensory percepts (reviewed in Vroomen and Keetels 2010; also see Werner and Noppeney 2011). We thus expected our results to indicate synchrony-induced benefits in neural stimulus processing and in behavioural performance in the detection task.

Pulsing stimuli elicited distinct time-locked oscillatory electrophysiological brain responses. These steady-state responses (SSRs) frequency-tagged the ongoing processing of both Gabor patches and, thus, allowed tracing corresponding stimulus-specific neural activity in the EEG (Regan 1989; Keitel et al. 2014). Crucially, SSRs provide an index of relative attentional allocation to specific stimuli because attention modulates SSR amplitudes in visual (Müller et al. 1998; Müller et al. 2003; Kim et al. 2007), auditory (Ross et al. 2004; Bidet-Caulet et al. 2007; Saupe et al. 2009b) and audio-visual stimulus situations (Saupe et al. 2009a; Keitel et al. 2011; Keitel et al. 2013).

Frequency-tagging studies have also investigated effects of audio-visual synchrony on SSRs (Jenkins et al. 2011; Giani et al. 2012). Nozaradan et al. (2012) demonstrated enhanced amplitudes (and inter-trial phase coherence) of SSRs driven by an auditory and a visual stimulus when both obeyed a synchronous presentation. They established synchrony by creating a “visual beat” – a centrally presented white rectangle periodically moving left and right – and an auditory beat, both presented at a temporal rate of ~ 2 Hz. A contrast condition abolished synchrony by presenting beats at slightly different rates and led to smaller SSR amplitudes. The authors found their results well in line with the notion of an audio-visual synchrony related attentional gain.

Our design enabled a closer investigation of this notion. Employing a related but extended paradigm that featured concurrent in sync and out-of-sync visual stimulus presentations allowed us to disentangle, quantify and compare the magnitude of gain effects of feature-selective attention and the “pull” exerted by audio-visual synchrony.

[Insert Figure 1 here]

METHODS

Participants

We collected data from thirteen participants (8 women), aged 20 to 31 years, with normal or corrected-to-normal vision and normal hearing. Participants gave informed written consent prior to experiments. None reported a history of neurological diseases or injury. The experiment was conducted in accordance with the Declaration of Helsinki and the guidelines of the ethics committee of the University of Leipzig.

Stimulation

Stimuli were presented on a 19-inch cathode ray tube screen positioned 0.8 m in front of participants. The screen was set to a refresh rate of 85 frames per second and a resolution of 1024 x 768 pixel (width x height). Visual experimental stimulation consisted of two centrally presented superimposed Gabor patches (5° of visual angle), one coloured green and tilted 45° counter-clockwise and the other one coloured red and tilted 45° clockwise. Shades of green were delivered via the green channel of the screen (RGB colour space) and shades of red via the red channel. This enabled an additive combination of colour channels where stimuli overlapped. Although mixing red and green channels yielded local yellow spots (such as the centre of the stimulation, see *Figure 2A* inset) it abolished effects of depth perception, i.e. one of the patches being perceived in front of the other.

Participants individually adjusted the maximum luminance of green and red colours by means of heterochromatic flicker photometry (Wagner and Boynton 1972) against a grey background (RGB: 128,128,128; luminance = 41 cd/m²) prior to the experiment. During stimulation Gabor patches were presented against a black background (RGB: 0,0,0; luminance < 0.1 cd/m²). A small grey circle (.4° degrees of visual angle, RGB: 128,128,128) in the centre of the stimulation served as fixation point.

Both Gabor stimuli underwent two independent periodic changes in the course of a trial: (1) Green patch presentation followed a cycle of 4 on-frames and 2 off-frames (2/1 on/off-ratio) resulting in a 17 Hz flicker. The red patch flickered at a rate of 14.2 Hz achieved by repetitive cycles of 3 on-frames and 2 off-frames (3/2 on/off-ratio). (2) In addition to the flicker, the spatial frequency of the Gabor patches oscillated between a maximum of 2 Hz/° and a minimum of 1 Hz/° at a rate of 3.14 Hz for the green patch and 3.62 Hz for the red patch. Periodic spatial frequency changes gave the impression of alternating contractions and relaxations that led to the percept of pulsing Gabor patches over time (*Figure 1A*). Pulse frequencies were chosen based on pilot experiments that served to determine a trade-off frequency range in which pulsing was readily perceptible, yet, still allowed driving near-sinusoidal frequency-following brain responses. In brief, both Gabor patches were concurrently flickering and pulsing, each at its distinct frequencies (*Figure 1B*).

In addition to the visual stimuli we presented a tone with a centre frequency of 440 Hz binaurally via headphones. The frequency of the tone was rhythmically modulated following sinusoidal excursions from the centre frequency (10% maximum excursion = ± 44 Hz). On each trial the modulation rate exactly matched the pulse rate of one of the two Gabor patches. Common rhythmic changes over time resulted in sustained audio-visual synchrony (see e.g. Schall et al. 2009).

Prior to the experiment, we employed the method of limits (Leek 2001) to approximate individual hearing thresholds using one of the experimental stimuli, a 3.14-Hz frequency modulated tone (see e.g. Herrmann et al. 2014). In our implementation, participants listened to a series of 10 tone sequences with a maximum duration of 15 s per sequence. Tone intensity changed during each sequence while alternating between log-linear decreases and increases across sequences.

Participants were instructed to indicate by button press when they stopped or started hearing respective tones. Cross-referencing button response times with tone intensity functions yielded individual estimates of psychophysical hearing thresholds, i.e. sensation levels (SL). In the experiment, acoustical stimulation was presented at an intensity of 35 dB (above) SL.

Procedure and Task

Participants were seated comfortably in an acoustically dampened and electromagnetically shielded chamber and directed gaze towards the fixation ring on the computer screen. At the beginning of each trial, participants were cued to attend to the colour Green or Red. To this end, a green or red circle appeared inside the fixation ring for 500 ms. Subsequently, the two flickering and pulsing Gabor patches and the pulsing tone were presented for 3500 ms. At the end of each trial, the fixation ring remained on screen for an extra 700 ms allowing participants to blink before the next trial started (*Figure 1C*).

Participants were instructed to respond to occasional brief fadings of the colour-cued Gabor patch (= targets) while ignoring similar events in the other patch (= distracters). For that purpose, Gabor patch luminance faded out to a minimum of 50%, then back in within a 300 ms interval. Targets and distracters occurred in 50% of trials and up to 3 times in one trial with a minimum interval of 800 ms between subsequent onsets. Responses were recorded as space-bar presses on a standard keyboard. The responding hand was changed halfway through the experiment with the starting hand counterbalanced across participants.

We manipulated the two factors *attended* Gabor patch (green vs. red) and audio-visual *synchrony* between attended Gabor patch and tone (synchronous vs. asynchronous) in a fully balanced design. Trials of the resulting four conditions – [green attended, synchronous], [green attended, asynchronous], [red attended, synchronous], [red attended, asynchronous] – were presented in a pseudo-randomized order. Note that the tone was always in sync with one of the two Gabor patches. Therefore, in the two conditions in which the tone was out of sync with the attended Gabor patch, it was in sync with the unattended patch.

In total we presented 600 trials (= 150 trials per condition) divided into 10 blocks (~5 min each).

Before the experiment, participants performed training for at least one block. After each training and experimental block they received feedback upon average hit rate and reaction time.

Behavioural data recording and analyses

Responses were considered a 'hit' when a button press occurred between 200 to 1000 ms after target onset. We further defined correct rejections as omitted responses to distracter stimuli. Based on these data, we calculated the response accuracy as the ratio of correct responses (number of hits and correct rejections) to the total number of targets and distracters for each condition and participant. Accuracies were subjected to a two-way repeated measures analysis of variances (ANOVA) with factors of *attention* (green vs. red) and *synchrony* (synchronous vs. asynchronous). Reaction times were analysed accordingly.

Note that, due to technical issues, behavioural data of one participant were not recorded. Results of task performance analyses reported below are thus based on the data of the 12 remaining participants.

Electrophysiological data recording

EEG was recorded from 64 scalp electrodes that were mounted in an elastic cap using a BioSemi ActiveTwo system (BioSemi, Amsterdam, Netherlands) set to a sampling rate of 256 Hz. Lateral eye movements were monitored with a bipolar outer canthus montage (horizontal electrooculogram). Vertical eye movements and blinks were monitored with a bipolar montage positioned below and above the right eye (vertical electrooculogram). From continuous data, we extracted epochs of 3500 ms starting at audio-visual stimulus onset were. In further preprocessing, we excluded (1) epochs that corresponded to trials containing transient targets and distracters (luminance fadings) as well as (2) epochs with horizontal and vertical eye movements exceeding $25 \mu\text{V}$ ($= 2.5^\circ$ of visual angle) or containing blinks. To correct for additional artefacts, such as single noisy electrodes, we applied the 'fully automated statistical thresholding for EEG artefact rejection' (FASTER, Nolan et al. 2010). This procedure corrected or discarded epochs with residual artefacts based on statistical parameters of the data. Artefact correction employed a spherical-spline-based channel interpolation. For each participant FASTER interpolated up to 2 electrodes (median = 1) across recordings and an average of

up to 4.3 electrodes (minimum = 2.1, median = 3.3) per epoch. Note that epochs with more than 12 artefact-contaminated electrodes were excluded from further analysis. In total, we discarded an average of 14% of epochs per participant and condition. Subsequently, data were re-referenced to average reference and averaged across epochs for each condition and participant, separately. Basic data processing steps such as extraction of epochs from continuous recordings, re-referencing and plotting of scalp iso-contour voltage maps made use of EEGLAB (Delorme and Makeig 2004) in combination with custom routines written in MATLAB (The Mathworks, Natick, MA).

Electrophysiological data analyses

Averaged artefact-free epochs were truncated to 3000 ms segments that started 500 ms after audio-visual stimulation onset. The first 500 ms were omitted in order to exclude event-related potentials to stimulus onset from spectral analyses of EEG time series. From de-trended (i.e. linear trend removed) 3-s segments we quantified individual amplitude spectra by means of Fourier transforms. Peaks in condition-averaged power (= squared amplitude) spectra pooled across all 64 scalp electrodes corresponded to the pulse and flicker frequencies of the stimulation (*Figure 2A*) and thus illustrated that our stimulation was effective in driving distinct SSRs. Notably, spectra revealed strong harmonic responses at twice the pulse frequencies (6.28 and 7.24 Hz). We included these pulse-driven harmonics in further analyses because fundamental and harmonic responses have been repeatedly found to reflect different aspects of stimulus processing (Pastor et al. 2007; Kim et al. 2011; Porcu et al. 2013).

Grand-average topographical distribution of pulse-driven 3.14, 3.62, 6.28, 7.24, as well as flicker-driven 14.2 and 17 Hz SSR power averaged over conditions showed similar focal maxima at parieto-occipital electrode sites (see Fig. 2B). For each participant and condition, SSR amplitudes were averaged across the 5 best (maximum amplitude) electrodes within this cluster and divided by the individual mean amplitude across conditions to control for differences in absolute amplitude between participants. Thus normalized amplitudes allowed pooling across pulse frequency following

(‘pulse 1F’) 3.14 Hz and 3.62 Hz, pulse frequency harmonic (‘pulse 2F’) 6.28 and 7.24 Hz as well as flicker frequency following (‘flicker 1F’) 14.2 and 17 Hz SSRs, respectively. Note that collapsing across frequencies necessarily entailed collapsing across SSRs driven by green and red Gabor patches. This step was justified by the fact that we were interested in the net effect of feature-based attention on visual stimulus processing rather than a comparison between attention effects on red and green Gabor patches. Behavioural performance in attend-green and attend-red conditions was comparable (see Results section below) thus supporting our approach.

Normalized collapsed amplitudes of pulse 1F, pulse 2F and flicker 1F SSRs were subjected to two-way repeated measures analysis of variances (ANOVAs) with factors of *attention* (attended vs. unattended) and *synchrony* (synchronous vs. asynchronous). Note the conversion of the factor *attention* from [green vs. red] to [attended vs. unattended] that is due to the collapsing across green- and red-Gabor-driven SSRs.

[Insert Figure 2 here]

RESULTS

Behavioural data

Participants performed comparably accurate in responding to luminance fadings while they attended to green vs. red Gabor patches (main effect *attention* to colour: $F(1,11) < 1$). However, accuracy was slightly but systematically reduced when the tone pulsed in congruence with the attended Gabor patch (main effect *synchrony*: $F(1,11) = 22.21$, $p < 0.001$, $\eta^2 = 0.11$, also see Table 1). The interaction of both factors was insignificant ($F(1,11) < 1$). On average, participants responded 606 ms (± 16 ms [SEM]) after target presentation. Reaction times were similar in all conditions (F 's < 3.14 , p 's > 0.10 , also see Table 1).

EEG data

The topographical distribution of SSR power averaged across experimental conditions revealed maxima at occipital electrode sites for all stimulation frequencies (Fig. 2B). Results described below are based on SSR amplitudes averaged across 5 maximum-power parieto-occipital electrodes that were selected from SSR amplitude topographies for each frequency and participant, separately. This approach captured typical small variations in the exact location of amplitude maxima between participants. (Note that all statistical effects reported below were robust against excluding the one participant with missing behavioural data.)

Pulse 1F

Amplitudes of 3.14- and 3.62-Hz SSRs showed strong modulation by attention, but only little influence of audio-visual synchrony (*Figure 2C*, top chart). Indeed, amplitudes were systematically greater for attended than for unattended Gabor patches (main effect attention: $F(1,12) = 7.37$, $p < 0.05$, $\eta^2 = 0.22$; also see *Figure 2D*, top chart) but remained comparable during synchronous vs. asynchronous tone presentation (main effect audiovisual synchrony: $F(1,12) = 1.68$, $p = 0.22$, $\eta^2 = 0.04$). The factors of attention and audiovisual synchrony did not interact ($F(1,12) < 1$).

Pulse 2F

In addition to the strong attention effect on pulse 1F SSRs, 6.28- and 7.24-Hz (i.e. harmonic) SSR amplitudes further modulated with audio-visual synchrony (*Figure 2C*, centre spectrum). Statistical analyses confirmed that amplitudes were significantly greater when the corresponding Gabor patch was attended compared with when it was unattended (main effect attention: $F(1,12) = 8.51$, $p < 0.05$, $\eta^2 = 0.15$) and significantly reduced when the tone pulsed asynchronously compared with when it pulsed in synchrony with visual stimulation (main effect audio-visual synchrony: $F(1,12) = 5.58$, $p < 0.05$, $\eta^2 = 0.18$). We found no indication that the attentional gain depended on audio-visual synchrony (interaction attention x synchrony: $F(1,12) = 2.30$, $p = 0.16$, $\eta^2 = 0.01$). Thus, gain effects of colour-selective attention and audio-visual synchrony modulated SSR amplitudes additively.

A post-hoc comparison (paired two-tailed t-test) of attentional (attended minus unattended) and synchrony-based modulation (synchronous minus asynchronous) of normalized SSR amplitudes showed that both effects were of similar magnitude ($t(12) = -0.19$, $p = 0.85$).

We further tested whether pulse 2F SSR amplitudes depended on pulse 1F amplitudes. To this end, we evaluated amplitude correlations of 1F with 2F SSRs driven by Green (3.14/6.28 Hz) and Red Gabor patches (3.62/7.25 Hz), separately, by means of Spearman's rho (ρ). Both correlations were negligible (Green: $\rho = 0.05$, $p = 0.88$; Red: $\rho = -0.25$, $p = 0.42$), thus, indicating that 1F and 2F SSR amplitudes were independent.

Flicker 1F

As can be seen in spectra in *Figure 2A* and *C*, flicker stimulation at 14.2 and 17 Hz elicited SSRs with smallest amplitudes. Decreasing amplitudes with increasing frequency is a common finding (see e.g. Andersen et al. 2008; Porcu et al. 2014) although, here, the substantial overlap of Gabor patches, the small amount of off-frames as well as the low average luminance of the stimuli (after adjusting for isoluminance) might have further contributed to low luminance flicker-driven SSR amplitudes.

A repeated measures ANOVA carried out on normalized amplitudes collapsed across 14.2- and 17-Hz SSRs indicated comparable processing whether participants attended a Gabor patch or not (main effect attention: $F(1,12) < 1$). However, asynchronous tone presentation led to greater amplitudes than synchronous presentation (main effect synchrony: $F(1,12) = 6.77$, $p < 0.05$, $\eta^2 = 0.09$; also see *Figure 2D*). No interaction between factors was found ($F(1,12) < 1$).

DISCUSSION

To date, the vast majority of research into the interplay of audio-visual interactions and attention has considered effects of spatial attention on transiently presented synchronous or asynchronous sensory events. Here, we investigated the sustained effect of audio-visual synchrony on early cortical processing of visual stimuli that were attended based on their features rather than location. We

presented attended and unattended stimuli concurrently and for several seconds, thus, mimicking natural viewing conditions more closely.

During the experiment participants viewed two superimposed Gabor patches of different orientations, one red and one green, to enforce attentional selection based on the stimulus feature “colour”. We established audio-visual synchrony between one of the two Gabor patches and a continuous tone by “pulsing” both stimuli at the same temporal rate. Pulsed stimulation (as well as visual stimulus flicker) drove robust SSRs that indexed the processing of each individual Gabor patch. Therefore, we were able to directly compare the processing of an attended in-sync stimulus with that of attended out-of-sync, unattended in-sync as well as unattended out-of-sync stimuli.

Synchrony-related gain in pulse-driven SSRs

Previous experiments have demonstrated the property of synchronous audio-visual stimuli to attract attention towards their location (Van der Burg et al. 2008; Van der Burg et al. 2011). Here, we demonstrate for the first time that audio-visual synchrony can also attract attention towards specific stimulus features when spatial selection is hardly possible.

In our case, visual stimuli experienced sustained gain effects when presented in sync with the auditory tone. These effects were statistically robust on pulse 2F (harmonic) SSRs but not on pulse 1F (fundamental) SSRs. This difference likely relates to findings that harmonics constitute genuine neural responses that code complementary rather than redundant stimulus representations (Pastor et al. 2007; Jenkins et al. 2011; Kim et al. 2011). Further corroborating these findings, we found individual fundamental and harmonic SSR amplitudes to be uncorrelated.

With regard to our experimental stimulation, pulse 1F responses could have encoded the rhythmic stimulus modulation in its entirety. Instead, strong transients as generated by most drastic stimulus changes, i.e. when Gabor patches expanded or contracted fastest, may have contributed to pulse 2F responses. Naturally, transients occurred at exactly twice the stimulation frequencies (i.e. at 6.28/7.24 Hz) during maximum up- and down-slopes of the employed continuous sinusoidal stimulus

modulation. Recent research has emphasised the prevailing role of such sensory transients over continuous cues of audio-visual synchrony in multisensory integration (Van der Burg et al. 2010; Werner and Noppeney 2011; Van der Burg et al. 2014) and, thus, supports greater susceptibility of pulse 2F SSRs to corresponding gain effects.

Observing multisensory interactions during synchronous presentation of audio-visual transients at rates > 6 Hz, however, collides with a series of elegant studies that established a speed limit of < 4 Hz for the perception of synchrony (Fujisaki and Nishida 2005; Fujisaki and Nishida 2009). Several factors may account for this discrepancy: Fujisaki and Nishida (2005; 2009) measured explicit judgements of synchrony while varying the temporal lag between one auditory and one visual stream presented at identical temporal rates. In our experiment, synchrony was task-irrelevant and asynchrony was established through stimulus modulation at distinct and incommensurable temporal rates. Most importantly, we presented synchronous and asynchronous stimuli concurrently and participants were only required to attend to the visual aspects of the stimulation.

Fujisaki and Nishida (2005) argued that decreasing synchrony judgments depended on the inability of sensory systems to individuate salient temporal features from rhythmic stimulus streams with increasing frequency. In their framework, it is this feature extraction process that affords subsequent multisensory integration. We suggest that our findings can be reconciled with the 4-Hz speed limit by assuming that attention may be able to facilitate the postulated feature extraction process. In fact, a recent study found that spatial attention can indeed increase the temporal precision of multisensory integration (i.e. decrease the likelihood of integrating sensory inputs that coincide just because they are resented in rapid succession) when participants did not have to judge the simultaneity of auditory and visual events explicitly but rather focussed on visual information only (Donohue et al. 2015). Not only is this situation comparable to our paradigm and may explain why we find effects of synchrony while stimulating > 6 Hz but, when relating it to our results, it also suggests that feature-selective attention has a similar influence on the temporal precision of multisensory integration.

Reversed effect of synchrony on flicker-driven SSRs

Much to our surprise, audio-visual synchrony modulated flicker 1F responses in the opposite direction: Although pulse 2F SSRs indicated that in sync stimuli gained a processing advantage over out-of-sync stimuli, flicker 1F SSRs suggested a contrary processing disadvantage. We suggest that this paradox can be solved by assuming that our concurrent flicker and pulse stimulation at distinct (and incommensurable) rates led to a perceptual dissociation of the visual stimulus into two distinct streams over time - a phenomenon well described as stream segregation in auditory perception (reviewed in Winkler et al. 2012).

To illustrate this argument, imagine observing passing cars through lines of trees on a train ride. The motion of the train will lead to intermittent occlusions of the cars. Despite the spatial superposition, the visual system is able to interpolate occluded parts of the cars based on their trajectory and thus retains a sense of cars and trees being separate visual objects (Yi et al. 2008; Franconeri et al. 2012; Atmaca et al. 2013). In our example, the moving cars correspond to the pulsing stimuli and the trees resemble transient occlusions by flicker-off frames. Accordingly, the visual system may have interpolated Gabor pulsation during flicker-off frames. This process may have effectively segregated visual input into concurrent Gabor and “occluder” streams. Due to distinct flicker frequencies each occluder stream was stimulus-specific because one Gabor patch could have been visible while the other one was occluded.

This visual stream segregation hypothesis resolves the paradoxical effect of audio-visual synchrony on flicker 1F SSR amplitudes: Synchrony (i.e. the associated pull of attention) biased processing towards the respective Gabor stream. Consequentially, this bias reduced the processing of the corresponding occluder stream. Similar effects have been described recently for attention shifts between features tagged with different flicker frequencies (concurrent rotation and colour changes) that constitute one visual object (Müller 2014, p. 132).

Synchrony-related effects on behaviour

At first glance our results suggest juxtaposed effects of audio-visual synchrony on visual processing and behavioural performance. More specifically, whereas *synchrony* led to a visual processing gain we found a slight but systematic counter-intuitive behavioural effect: Participants performed better when they attended *asynchronous* stimuli. Recall, however, that the task was to detect transient fadings of the cued Gabor patch. In line with our visual stream segregation hypothesis (as laid out above), these fadings, effectively a reduction in patch luminance, could be similarly conceived of as (partial) stimulus occlusions. The postulated Gabor motion interpolation during occlusions might have been aided by synchronous auditory input. Thus, in-sync stimulation may have counteracted Gabor occlusion more strongly and consequentially reduced fading detection performance in synchronous relative to asynchronous conditions. Assuming these assumptions hold, the task employed here was inadequate to reveal effects of audio-visual synchrony.

We emphasize, however, that the main purpose of the task in this study was to control whether participants allocated their attention to the cued colour on each trial. As a consequence, audio-visual (a)synchrony was entirely task-irrelevant. Employing an alternative task design that included any form of judgment about synchrony might have revealed expected multisensory benefits (Van der Burg et al. 2010).

Feature-selective gain effects

Attending to a specific colour enhances the neural representation of stimuli carrying that colour (Wolfe 1994; Maunsell and Treue 2006). This gain mechanism has been shown to subserve goal-directed selective processing in cluttered visual scenes with co-localized, overlapping stimuli.

Employing frequency-tagged superimposed moving red and blue dot clouds, Müller et al. (2006) specifically demonstrated feature-selective attentional modulation of early cortical visual as indexed by stimulus flicker-driven SSRs. The present results complement previous findings: Attended Gabor

patches underwent enhanced processing relative to when they remained unattended as indicated by pulse-driven fundamental and harmonic SSRs.

Interestingly though, we found feature-selective attention effects only on SSRs driven by stimulus pulsation. SSRs driven by stimulus on-off flicker remained unaffected. This result contrasts with above mentioned studies. Recall however that we optimised our experimental stimulation towards establishing audio-visual synchrony. To this end, we employed an atypical stimulus flicker with long on- and short off-phases that drove less pronounced SSRs. The low amplitude of these signals (relative to pulse-driven SSRs) might thus have concealed the expected effects of feature-selective attention.

Although speculative, our visual stream segregation hypothesis provides a plausible alternative explanation for low flicker-driven SSR amplitudes and their lacking modulation by feature-selective attention: The behavioural task required participants to attend to the Gabor patches. This might have led to a general enhancement of Gabor stream processing and a parallel suppression of occluder stream processing that resulted in lower flicker-driven amplitudes. Moreover, the segregation from the Gabor stream rendered the occluders colourless and thus unreceptive to effects of feature-selective attention.

Synchrony and Attention co-amplify visual processing

Our study focused on characterizing gain effects of audio-visual synchrony and feature-selective attention. Stimulus frequency-tagging elicited 3 prominent SSRs per stimulus, one of which indicated concurrent modulation by synchrony and attention (pulse 2F SSRs). The following discussion mainly focuses on the results obtained from this response.

Importantly, pulse 2F SSR amplitudes indicated that visual stimulus processing experienced greatest gain when a given Gabor patch was attended and in sync with the tone, intermediate gain when the Gabor patch was attended but asynchronous or vice versa, and lowest when it was unattended and asynchronous (*Figure 2D*). We found both gain effects to be of similar magnitude. Moreover, neither

pulse 2F SSR amplitudes nor the other two responses implied that one gain effect depended on the other. Put differently, audio-visual synchrony pulling attention away from one stimulus had comparably detrimental effects on its processing regardless of whether it was attended or not. Vice versa, attended and unattended in-sync stimuli showed similar processing increases indicating that feature-selective and audio-visual synchrony-related gains occurred independently.

Our results are in line with the notion that both, goal-directed and cross-modal stimulus-driven influences can modulate early visual processing concurrently and in parallel (Talsma et al. 2010). Goal-directed feature-selective influences are likely conveyed and relayed to occipital cortex via a fronto-parietal network similar to that of spatial attention (Snyder and Foxe 2010). Stimulus-driven auditory influences on early visual processing, instead, may have been conveyed via direct cortico-cortical connections or relayed subcortically (Lakatos et al. 2009). In our case, the goal-directed bias was introduced by cueing participants to attend to a Gabor patch with a specific colour while audio-visual synchrony constituted the stimulus-driven bias (Van der Burg et al. 2008; Nozaradan et al. 2012).

An interesting question for future research would be whether audio-visual synchrony and spatial attention combine gain effects in a similar additive fashion.

Conclusion

We studied how the perception of audio-visual synchrony and feature-selective attention modulate stimulus representations in early visual cortex. Both mechanisms led to comparable processing gains. Moreover, their effects combined linearly suggesting that audio-visual synchrony and feature-selective attention can act in parallel to influence neural stimulus representations. Our results add to the growing literature on the interplay of attention and multisensory integration (reviewed in Talsma et al. 2010) and may have practical implications for the design of multisensory brain computer interfaces (An et al. 2014).

Acknowledgments

Work was supported by the Deutsche Forschungsgemeinschaft (grant no. MU972/21-1). Data presented here were recorded at the Institut für Psychologie, Universität Leipzig. The authors appreciate the assistance of Renate Zahn and Emanuele Porcu in data collection. The experimental stimulation was realized using Cogent 2000 developed by the Cogent 2000 team at the Functional Imaging Laboratory and the Institute of Cognitive Neuroscience and Cogent Graphics developed by John Romaya at the Laboratory of Neurobiology at the Wellcome Department of Imaging Neuroscience.

Conflict of interest: None declared.

REFERENCES

- An X, Hohne J, Ming D, Blankertz B (2014) Exploring combinations of auditory and visual stimuli for gaze-independent brain-computer interfaces. *PLoS One* 9:e111070 doi: 10.1371/journal.pone.0111070
- Andersen SK, Hillyard SA, Müller MM (2008) Attention facilitates multiple stimulus features in parallel in human visual cortex. *Curr Biol* 18:1006-1009
- Atmaca S, Stadler W, Keitel A, Ott DV, Lepsien J, Prinz W (2013) Prediction processes during multiple object tracking (MOT): involvement of dorsal and ventral premotor cortices. *Brain Behav* 3:683-700 doi: 10.1002/brb3.180
- Bidet-Caulet A, Fischer C, Besle J, Aguera PE, Giard MH, Bertrand O (2007) Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. *J Neurosci* 27:9252-9261
- Busse L, Roberts KC, Crist RE, Weissman DH, Woldorff MG (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 doi: DOI 10.1073/pnas.0507704102
- 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

- Donohue SE, Green JJ, Woldorff MG (2015) The effects of attention on the temporal integration of multisensory stimuli. *Front Integr Neurosci* 9:32 doi: 10.3389/fnint.2015.00032
- Eimer M, Grubert A (2014) The gradual emergence of spatially selective target processing in visual search: From feature-specific to object-based attentional control. *J Exp Psychol Hum Percept Perform* 40:1819-1831 doi: 10.1037/a0037387
- Franconeri SL, Pylyshyn ZW, Scholl BJ (2012) A simple proximity heuristic allows tracking of multiple objects through occlusion. *Atten Percept Psychophys* 74:691-702 doi: 10.3758/s13414-011-0265-9
- Fujisaki W, Nishida S (2005) Temporal frequency characteristics of synchrony-asynchrony discrimination of audio-visual signals. *Exp Brain Res* 166:455-464 doi: 10.1007/s00221-005-2385-8
- Fujisaki W, Nishida S (2008) Top-down feature-based selection of matching features for audio-visual synchrony discrimination. *Neurosci Lett* 433:225-230 doi: 10.1016/j.neulet.2008.01.031
- Fujisaki W, Nishida S (2009) Audio-tactile superiority over visuo-tactile and audio-visual combinations in the temporal resolution of synchrony perception. *Exp Brain Res* 198:245-259 doi: 10.1007/s00221-009-1870-x
- Giani AS, Ortiz E, Belardinelli P, Kleiner M, Preissl H, Noppeney U (2012) Steady-state responses in MEG demonstrate information integration within but not across the auditory and visual senses. *Neuroimage* 60:1478-1489 doi: 10.1016/j.neuroimage.2012.01.114
- Herrmann B, Schlichting N, Obleser J (2014) Dynamic range adaptation to spectral stimulus statistics in human auditory cortex. *J Neurosci* 34:327-331 doi: 10.1523/JNEUROSCI.3974-13.2014
- Jarmasz J, Hollands JG (2009) Confidence intervals in repeated-measures designs: The number of observations principle. *Can J Exp Psychol* 63:124-138 doi: 10.1037/a0014164
- Jenkins J, 3rd, Rhone AE, Idsardi WJ, Simon JZ, Poeppel D (2011) The elicitation of audiovisual steady-state responses: multi-sensory signal congruity and phase effects. *Brain Topogr* 24:134-148 doi: 10.1007/s10548-011-0174-1
- Keitel C, Maess B, Schroger E, Muller MM (2013) Early visual and auditory processing rely on modality-specific attentional resources. *Neuroimage* 70:240-249 doi: DOI 10.1016/j.neuroimage.2012.12.046
- Keitel C, Quigley C, Ruhnau P (2014) Stimulus-Driven Brain Oscillations in the Alpha Range: Entrainment of Intrinsic Rhythms or Frequency-Following Response? *Journal of Neuroscience* 34:10137-10140 doi: Doi 10.1523/Jneurosci.1904-14.2014
- Keitel C, Schröger E, Saupe K, Müller MM (2011) Sustained selective intermodal attention modulates processing of language-like stimuli. *Exp Brain Res*:1-7

- Kim YJ, Grabowecky M, Paller KA, Muthu K, Suzuki S (2007) Attention induces synchronization-based response gain in steady-state visual evoked potentials. *Nat Neurosci* 10:117-125
- Kim YJ, Grabowecky M, Paller KA, Suzuki S (2011) Differential roles of frequency-following and frequency-doubling visual responses revealed by evoked neural harmonics. *J Cogn Neurosci* 23:1875-1886 doi: 10.1162/jocn.2010.21536
- Lakatos P, O'Connell MN, Barczak A, Mills A, Javitt DC, Schroeder CE (2009) The leading sense: supramodal control of neurophysiological context by attention. *Neuron* 64:419-430 doi: 10.1016/j.neuron.2009.10.014
- Leek MR (2001) Adaptive procedures in psychophysical research. *Percept Psychophys* 63:1279-1292
- Macaluso E, Driver J (2003) Multimodal spatial representations in the human parietal cortex: evidence from functional imaging. *Adv Neurol* 93:219-233
- Maunsell JH, Treue S (2006) Feature-based attention in visual cortex. *Trends Neurosci* 29:317-322 doi: 10.1016/j.tins.2006.04.001
- Müller MM (2014) Neural Mechanisms of Feature-Based Attention. In: Mangun GR (ed) *Cognitive Electrophysiology of Attention: Signals of the Mind*. Elsevier, pp 123-135
- Müller MM, Andersen S, Trujillo NJ, Valdes-Sosa P, Malinowski P, Hillyard SA (2006) Feature-selective attention enhances color signals in early visual areas of the human brain. *Proc Natl Acad Sci U S A* 103:14250-14254
- Müller MM, Malinowski P, Gruber T, Hillyard SA (2003) Sustained division of the attentional spotlight. *Nature* 424:309-312
- Müller MM, Picton TW, Valdes-Sosa P, Riera J, Teder-Salejarvi WA, Hillyard SA (1998) Effects of spatial selective attention on the steady-state visual evoked potential in the 20-28 Hz range. *Brain Res Cogn Brain Res* 6:249-261
- Nolan H, Whelan R, Reilly RB (2010) FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection. *J Neurosci Methods* 192:152-162 doi: 10.1016/j.jneumeth.2010.07.015
- Nozaradan S, Peretz I, Mouraux A (2012) Steady-state evoked potentials as an index of multisensory temporal binding. *Neuroimage* 60:21-28 doi: 10.1016/j.neuroimage.2011.11.065
- Pastor MA, Valencia M, Artieda J, Alegre M, Masdeu JC (2007) Topography of cortical activation differs for fundamental and harmonic frequencies of the steady-state visual-evoked responses. An EEG and PET H215O study. *Cereb Cortex* 17:1899-1905 doi: 10.1093/cercor/bhl098
- Porcu E, Keitel C, Muller MM (2013) Concurrent visual and tactile steady-state evoked potentials index allocation of inter-modal attention: a frequency-tagging study. *Neurosci Lett* 556:113-117 doi: 10.1016/j.neulet.2013.09.068

- Porcu E, Keitel C, Muller MM (2014) Visual, auditory and tactile stimuli compete for early sensory processing capacities within but not between senses. *Neuroimage* 97:224-235 doi: 10.1016/j.neuroimage.2014.04.024
- Regan D (1989) *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine*. Elsevier, New York
- Ross B, Picton TW, Herdman AT, Pantev C (2004) The effect of attention on the auditory steady-state response. *Neurol Clin Neurophysiol* 2004:22
- Saenz M, Buracas GT, Boynton GM (2002) Global effects of feature-based attention in human visual cortex. *Nat Neurosci* 5:631-632 doi: 10.1038/nn876
- Saupe K, Schröger E, Andersen SK, Müller MM (2009a) Neural mechanisms of intermodal sustained selective attention with concurrently presented auditory and visual stimuli. *Frontiers in Human Neuroscience* 3:-
- Saupe K, Widmann A, Bendixen A, Müller MM, Schröger E (2009b) Effects of intermodal attention on the auditory steady-state response and the event-related potential. *Psychophysiology* 46:321-327
- Schall S, Quigley C, Onat S, König P (2009) Visual stimulus locking of EEG is modulated by temporal congruency of auditory stimuli. *Exp Brain Res* 198:137-151
- Snyder AC, Foxe JJ (2010) Anticipatory attentional suppression of visual features indexed by oscillatory alpha-band power increases: a high-density electrical mapping study. *J Neurosci* 30:4024-4032 doi: 10.1523/jneurosci.5684-09.2010
- Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG (2010) The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences* 14:400-410
- Treisman AM, Gelade G (1980) A feature-integration theory of attention. *Cogn Psychol* 12:97-136
- Treue S, Martinez Trujillo JC (1999) Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399:575-579
- Van der Burg E, Cass J, Alais D (2014) Window of audio-visual simultaneity is unaffected by spatio-temporal visual clutter. *Sci Rep* 4:5098 doi: 10.1038/srep05098
- Van der Burg E, Cass J, Olivers CN, Theeuwes J, Alais D (2010) Efficient visual search from synchronized auditory signals requires transient audiovisual events. *PLoS One* 5:e10664 doi: 10.1371/journal.pone.0010664
- Van der Burg E, Olivers CN, Bronkhorst AW, Theeuwes J (2008) Pip and pop: nonspatial auditory signals improve spatial visual search. *J Exp Psychol Hum Percept Perform* 34:1053-1065

- Van der Burg E, Talsma D, Olivers CN, Hickey C, Theeuwes J (2011) Early multisensory interactions affect the competition among multiple visual objects. *Neuroimage* 55:1208-1218 doi: 10.1016/j.neuroimage.2010.12.068
- Vroomen J, Keetels M (2010) Perception of intersensory synchrony: a tutorial review. *Atten Percept Psychophys* 72:871-884 doi: 10.3758/APP.72.4.871
- Wagner G, Boynton RM (1972) Comparison of four methods of heterochromatic photometry. *J Opt Soc Am* 62:1508-1515
- Werner S, Noppeney U (2011) The contributions of transient and sustained response codes to audiovisual integration. *Cereb Cortex* 21:920-931
- Winkler I, Denham S, Mill R, Bohm TM, Bendixen A (2012) Multistability in auditory stream segregation: a predictive coding view. *Philos Trans R Soc Lond B Biol Sci* 367:1001-1012 doi: 10.1098/rstb.2011.0359
- Wolfe JM (1994) Guided Search 2.0 - a Revised Model of Visual-Search. *Psychonomic Bulletin & Review* 1:202-238 doi: Doi 10.3758/Bf03200774
- Yi DJ, Turk-Browne NB, Flombaum JI, Kim MS, Scholl BJ, Chun MM (2008) Spatiotemporal object continuity in human ventral visual cortex. *Proc Natl Acad Sci U S A* 105:8840-8845 doi: 10.1073/pnas.0802525105

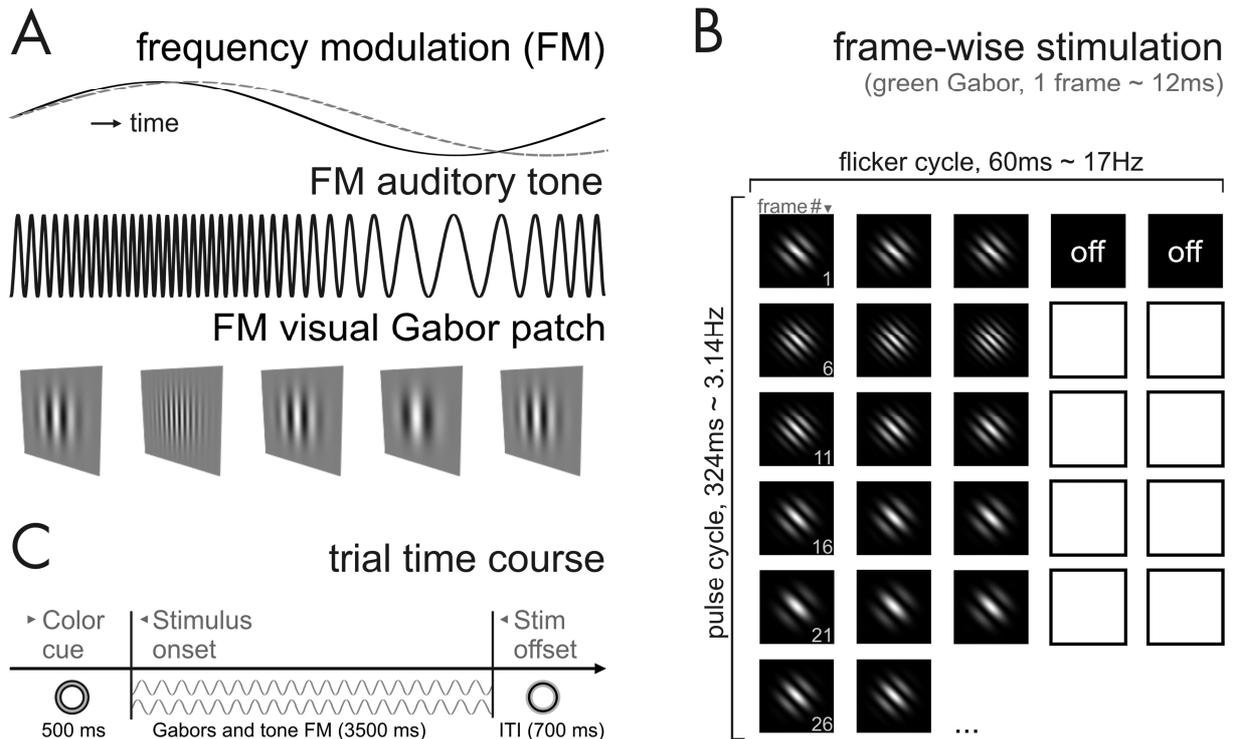


Figure 1 Schematics of audio-visual stimulation. (A) A common frequency modulation (FM, solid black line) of the pitch of the auditory tone and of the spatial frequency of one of the two Gabor patches led to a synchronous pulsing audio-visual percept. Concurrently, the spatial frequency of the other Gabor patch modulated at a slightly different frequency (dashed grey line), thus, rendering it asynchronous to the tone. Note that for illustrative purposes, only one monochrome upright Gabor patch is shown here whereas two patches were presented in colour, tilted and superimposed during the experiment (see *Figure 2A*, inset). (B) Frame-by-frame visual stimulation for the green Gabor patch. The illustration shows the first 27 frames of each trial. Note the emphasis on the on-off cycles leading to a 17 Hz flicker along the horizontal axis (white boxes = further off frames) and one full cycle of the spatial frequency modulation leading to a 3.14 Hz “pulsation” along the vertical axis. (C) Schematic trial time course (see Methods for a detailed description). Annuli indicate transient changes at fixation. Grey sinusoids represent concurrent FM of Gabor patches and the tone.

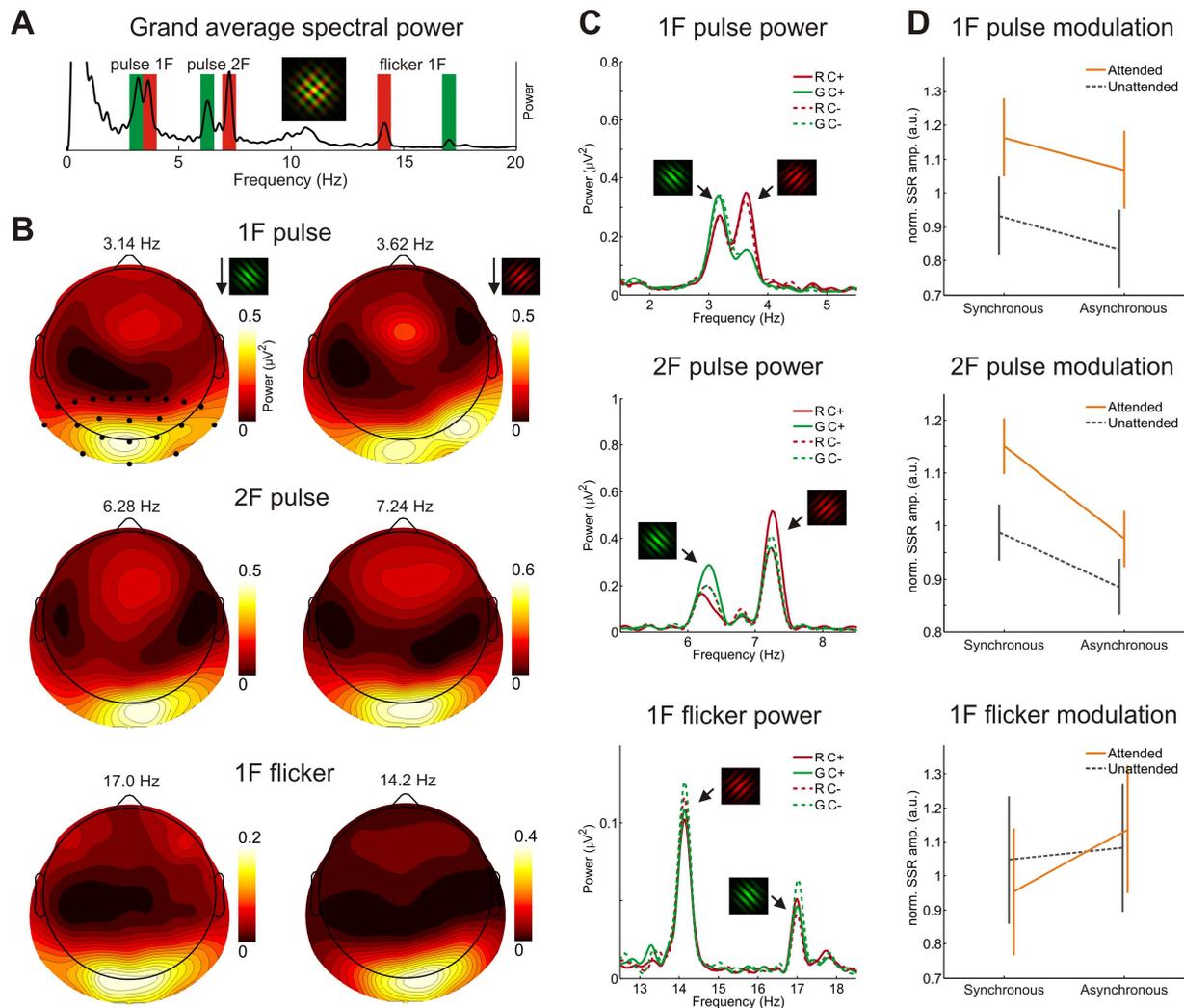


Figure 2 Results of SSR analyses. (A) Grand average power spectrum (in μV^2) averaged across conditions and electrodes. Peaks correspond to pulse frequency following (pulse 1F), pulse frequency doubling (pulse 2F) and flicker frequency following (flicker 1F) SSRs driven by the two superimposed Gabor patches (inset). (B) Grand average scalp iso-contour voltage maps of SSR power (in μV^2) averaged across conditions for each peak in A. Black dots in top left map indicate parieto-occipital electrodes used for SSR analyses. (C) Condition-resolved grand average power spectra (in μV^2) for each pair of stimulation frequencies (red solid line, R C+: red attended, synchronous; green solid line, G C+: green attended, synchronous; red dashed line, R C-: red attended, asynchronous; green dashed line, G C-: green attended, asynchronous). (D) Interaction plots of normalized SSVEP amplitudes pooled across frequencies. Lines connect group means of SSR amplitudes driven during synchronous vs. asynchronous stimulation (x-axis) when the corresponding stimulus was attended (orange solid line) or unattended (grey dashed line). Error bars correspond to 95%-within-subject confidence intervals (Jarmasz and Hollands 2009). Normalized amplitudes are scaled in arbitrary units (a.u.).

Table 1 Average behavioural performance in the visual fading detection task (N = 12).

Attended Colour		Green		Red	
		S+	S-	S+	S-
Proportion	<i>M</i>	81.7 %	83.8 %	81.9 %	83.8 %
correct (%)	$\pm SEM$	2.3 %	2.3 %	2.2 %	2.0 %
Reaction	<i>M</i>	596	610	609	610
time (ms)	$\pm SEM$	17	18	15	17

M = mean, *SEM* = standard error of the mean. *S+* = synchronous, *S-* = asynchronous.