

Steady-state evoked potentials as an index of multisensory temporal binding

Sylvie Nozaradan^{a,b}, Isabelle Peretz^b, André Mouraux^{a,*}

^a Institute of Neuroscience (Ions), Université catholique de Louvain (UCL), Belgium

^b International Laboratory for Brain, Music and Sound Research (BRAMS), Université de Montréal (UdeM), Canada

ARTICLE INFO

Article history:

Received 6 July 2011

Revised 10 November 2011

Accepted 22 November 2011

Available online 1 December 2011

Keywords:

EEG

Multisensory integration

Rhythm perception

Neuronal entrainment

Steady-state evoked potentials

ABSTRACT

Temporal congruency promotes perceptual binding of multisensory inputs. Here, we used EEG frequency-tagging to track cortical activities elicited by auditory and visual inputs separately, in the form of steady-state evoked potentials (SS-EPs). We tested whether SS-EPs could reveal a dynamic coupling of cortical activities related to the binding of auditory and visual inputs conveying synchronous vs. non-synchronous temporal periodicities, or *beats*. The temporally congruent audiovisual condition elicited markedly enhanced auditory and visual SS-EPs, as compared to the incongruent condition. Furthermore, an increased inter-trial phase coherence of both SS-EPs was observed in that condition. Taken together, these observations indicate that temporal congruency enhances the processing of multisensory inputs at sensory-specific stages of cortical processing, possibly through a dynamic binding by synchrony of the elicited activities and/or improved dynamic attending. Moreover, we show that EEG frequency-tagging with SS-EPs constitutes an effective tool to explore the neural dynamics of multisensory integration in the human brain.

© 2011 Elsevier Inc. All rights reserved.

Introduction

Building coherent representations of the external world requires integrating and merging information concurrently sampled through our different senses (Gibson, 1966; Spence and Driver, 2004). Most events occurring in the environment concomitantly activate afferents from different sensory modalities. For example, the perception of an explosion simultaneously emitting light, noise, vibrations and heat requires the integration of combined visual, auditory and somatosensory inputs. Because the information conveyed by these different sensory modalities is often complementary, cross-modal integration of these inputs may provide information about the environment that is absent in any one modality presented in isolation and, hence, cross-modal integration may improve behavior (Adrian, 1949; Elliott et al., 2010; Stein and Meredith, 1993).

Temporal congruency facilitates cross-modal integration (Bertelson, 1999; Fujisaki and Nishida, 2005; Petrini et al., 2009; Sekuler et al., 1997; Vatakis and Spence, 2006; Vroomen and Keetels, 2010; Welch and Warren, 1980; Zampini et al., 2003). Multisensory perception may result from a process of *binding by synchrony* of the cortical responses to sensory inputs sharing similar temporal dynamics (Kayser, 2009; Luo et al., 2010; Schroeder et al., 2008; Senkowski et al., 2008). Support for this hypothesis can be found in the electrophysiological recordings performed in the sensory cortices of monkeys where congruent multisensory inputs elicit an increased phase coherence of neuronal oscillatory

activity within the activated sensory cortices, as compared to incongruent multisensory inputs (Kayser and Logothetis, 2007; Kayser et al., 2008; Senkowski et al., 2007). Similarly, in humans, electroencephalographic (EEG) recordings reveal that the congruency of combined auditory and visual stimulation enhances the magnitude of stimulus-induced EEG oscillations across both auditory and visual cortices (Luo et al., 2010; Schall et al., 2009; Schroeder et al., 2008). However, because of the unavoidable temporal overlap between the neural responses to concurrent streams of sensory input, disentangling the neural activities related to each sensory stream, although critical to study multisensory integration, is difficult (Besle et al., 2009). Hence, current knowledge of how the human brain extracts, integrates and exploits the temporal dynamics of sensory input remains, at present, poorly understood.

Frequency-tagging using EEG steady-state evoked potentials (SS-EPs) could overcome this limitation, and thus may constitute a mean to study, non-invasively, multisensory integration in humans (Regan, 1989; Regan and Heron, 1969). SS-EPs are elicited by the continuous presentation of a sensory stimulus in which a given feature is modulated periodically at a given frequency. SS-EPs appear as an increase in the EEG frequency spectrum peaking specifically at the frequency of stimulation (Regan, 1989). Therefore, different SS-EP frequencies can be used to tag the different sensory inputs constituting a multimodal stimulus and, thereby, isolate the neural activity related specifically to each stream of sensory input (Morgan et al., 1996; Regan, 1989; Tononi et al., 1998). This frequency-tagging approach has already been used to characterize the neural activity triggered by intermodal interactions of selective attention, using simultaneous auditory and visual inputs and comparing the magnitude of SS-EPs obtained in unisensory vs. multisensory conditions, according to the

* Corresponding author at: Institute of Neurosciences, Université catholique de Louvain, 53, Avenue Mounier – UCL 53.75, B-1200 Bruxelles, Belgium.

E-mail address: andre.mouraux@uclouvain.be (A. Mouraux).

focus of attention (de Jong et al., 2010; Saupé et al., 2009; Talsma et al., 2006; Toffanin et al., 2009).

Here, we used the same frequency-tagging technique to investigate a different phenomenon. We aimed to explore, using SS-EPs, the cortical processes involved in the binding of temporally congruent multisensory inputs, occurring in the context of temporal periodicities. This context was chosen because the temporal dynamics of the eliciting stimuli is particularly important for beat perception in music (Large, 2008; London, 2004; van Noorden and Moelants, 1999). Although beats are preferentially conveyed by auditory input (Glenberg et al., 1989; Grahn et al., 2011; Patel et al., 2005; Repp, 2005), beat perception often co-occurs with visual movements such as when dancing or watching a conductor directing an orchestra (Repp, 2006). Furthermore, it has been hypothesized that beat perception is subtended by a periodic modulation of the responsiveness of neuronal populations frequency-tuned to the temporal periodicity extracted from the external stimulus (Jones and Boltz, 1989; Large, 2008; Large and Jones, 1999; Large and Kolen, 1994; Nozaradan et al., 2011; van Noorden and Moelants, 1999). Beat perception induced by multimodal sensory inputs may thus constitute a unique opportunity to study the mechanisms involved in the multimodal integration of temporally regular events.

The EEG response to auditory stimulation was tagged using a periodic modulation of loudness at 11 Hz (auditory SS-EP), whereas the EEG response to visual stimulation was tagged using a periodic modulation of luminance at 10 Hz (visual SS-EP) (Fig. 1). Independently of these periodic modulations, the auditory and visual stimulation contained temporal regularities, in the form of a periodic amplitude modulation of the sound and a periodic movement of the visual object, at rates corresponding to the ecological range of musical tempo perception and production (see Supplementary material for downloadable audiovisual examples). Across the two modalities, these “musical beats” were either synchronous or non synchronous, i.e. temporally congruent or incongruent (Fig. 1). This approach allowed us to compare the cortical activity related to processing auditory and visual streams of sensory input in a condition in which the temporal congruency of the auditory and visual beats is expected to promote a unified, temporally-bound, audiovisual percept and a condition in which the temporal incongruency of the auditory and visual beats is expected to prevent such an integration to take place.

Materials and methods

Participants

Twelve healthy volunteers (4 females, 11 right-handed, aged 22 to 37 years, mean 31 ± 5 years) took part in the study after providing

written informed consent. They all had musical experience as amateur listeners or dancers, and no history of hearing, neurological or psychological disorders. The study was approved by the local Ethics Committee.

Audiovisual stimulation

The auditory and visual stimuli were generated using the PsychToolbox extensions (Brainard, 1997) running under Matlab 6.5 (The MathWork, USA).

Auditory stimuli

The auditory stimuli were presented binaurally through earphones at a comfortable hearing level (BeyerDynamic DT 990 PRO, Germany). Each auditory stimulus lasted 33 s, and consisted of a 333.3 Hz pure tone in which an auditory beat was induced by modulating periodically the amplitude envelope of the tone, using an asymmetrical Hanning function (22 ms rise time and 394 ms fall time, amplitude modulation between 0 and 1) (Fig. 1). The frequency of the beat was either 2.1 Hz (126 bpm, 476 ms period) or 2.4 Hz (144 bpm, 416 ms period), thus lying within an ecological range of musical tempo perception and production (Drake and Botte, 1993; Repp, 2005), and shown to be suitable to induce beat perception (Nozaradan et al., 2011). Finally, the obtained sound was amplitude-modulated at 70% using an 11 Hz sinusoidal function. This additional amplitude-modulation of the tone was used to tag the neural responses elicited by the auditory stimuli, which appeared as an auditory SS-EP at 11 Hz (Galambos et al., 1981; Regan, 1989).

Visual stimuli

The visual stimuli were presented using a CRT computer monitor (resolution: 1024×768 , refresh rate: 100 Hz) positioned 50 cm in front of the participant's eyes. The stimulus consisted of a 1.5° visual angle white square displayed on a dark grey background, moving horizontally from left to right and from right to left following a sinusoidal function (Fig. 1). The amplitude of the horizontal movement was 3° visual angle, which allowed subjects to track the moving square without performing ocular movements. The frequency of the visual beat induced by the back and forth horizontal movement of the square was 2.1 Hz (126 bpm, 476 ms period) or 2.4 Hz (144 bpm, 416 ms period), like the frequency of the auditory beat. The luminance of the moving object was modulated using a 5 Hz sinusoidal function oscillating between 0.3 and 1, thus periodically changing the luminance of the moving square from dark to light and from light to dark at a rate of 10 Hz. This modulation of luminance was used to tag the neural

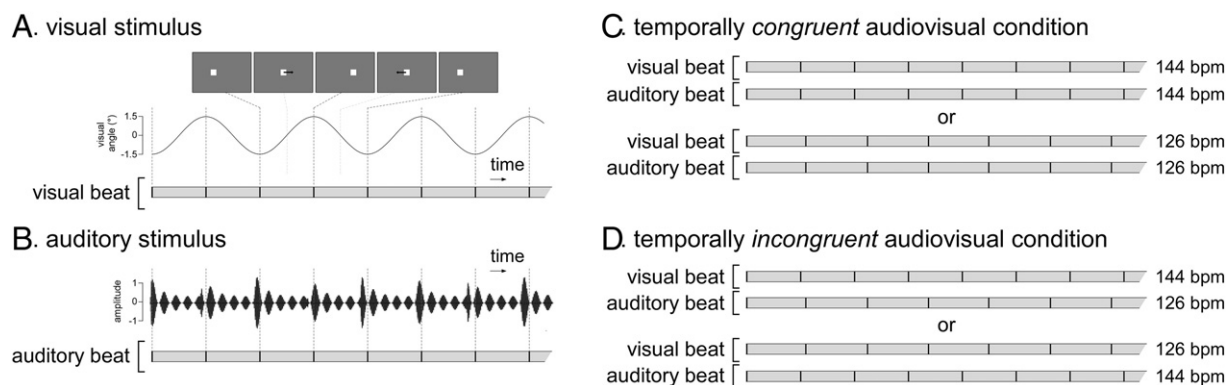


Fig. 1. Audiovisual stimulation. A: Visual stimulus. A white square presented on a grey background moved periodically from left to right and from right to left. This sinusoidal periodic movement was used to induce a *visual beat* with a frequency of 2.1 or 2.4 Hz (126 or 144 beats per minute, bpm). B: Auditory stimulus. The loudness of a 333.3 Hz pure tone was periodically modulated using an asymmetrical Hanning function, such as to induce an *auditory beat*. Such as the visual beat, the frequency of the auditory beat was either 2.1 or 2.4 Hz (126 or 144 bpm). C and D: Auditory and visual stimuli were presented simultaneously. In half of the trials, the beats carried by the auditory and visual stimuli were synchronous, thus generating a temporally congruent multimodal percept. In the other half of the trials, the beats carried by each of the two streams of sensory input were plesiochronous, thus preventing the emergence of a unified multimodal percept. See Supplementary material for downloadable audiovisual examples.

responses elicited by the visual stimuli, which elicits a visual SS-EP at 10 Hz (Kim et al., 2007; Saupe et al., 2009; Talsma et al., 2006). Like the auditory stimuli, each visual stimulus lasted 33 s.

Audiovisual stimulation

The auditory and visual beats were presented simultaneously, thus constituting audiovisual stimuli of 33 s duration. In half of the trials, the beats carried by the auditory (A) and visual (V) stimuli were temporally congruent, their periodicities being synchronous (A: 2.1 Hz with V: 2.1 Hz, or A: 2.4 Hz with V: 2.4 Hz). In the other half of the trials, the beats carried by each of the two streams of sensory input were temporally incongruent, their periodicities being pleo-synchronous (A: 2.1 Hz with V: 2.4 Hz, or A: 2.4 Hz with V: 2.1 Hz). In the temporally congruent audiovisual condition, the synchrony between the auditory and visual beat was expected to generate a temporally-binded multimodal percept. In contrast, in the temporally incongruent audiovisual condition, the lack of synchrony between the two beats was expected to prevent the emergence of a unified multimodal percept. This perceptual difference was assessed after the recording session, by asking the participants to judge the temporal synchrony or non-synchrony of 10 additionally-presented temporally congruent and incongruent audiovisual stimuli, through a forced choice. Two 30 s audiovisual clips of temporally congruent and incongruent stimulation are available online as a supplementary material.

Experimental design

The experiment consisted of two identical blocks of 12 trials, each consisting of a 3 s foreperiod followed by the presentation of the audiovisual stimulus (6 temporally congruent and 6 temporally incongruent trials, presented in random order). The rate of presentation of the trials was self-paced: participants were asked to initiate the upcoming trial by pressing a key on a computer keyboard. During the trial, participants were asked to detect the occurrence of very short-lasting (4 ms) interruptions in the auditory stimulus (perceived as a short cracking of the sound) and changes in the color of the moving square (briefly changed to red during 10 ms). These changes were inserted at a random position in ten additional trials interspersed within the two blocks. These ten trials were excluded from further analyses. The change detection task ensured that subjects maintained their attention on both the auditory and the visual streams during recording.

In eight subjects, two additional blocks were recorded in which auditory and visual stimuli were presented in isolation (unisensory auditory and visual conditions). The participants performed the same detection task as in the multisensory condition. These additional recordings were used to compare the amplitude of the auditory and visual SS-EPs elicited in a unisensory vs. multisensory context (see Supplementary Fig. 1).

EEG recording

Subjects were comfortably seated in a chair with their head resting on a support. They were instructed to relax, to avoid any head or body movement during the recording, and to keep their eyes fixated on the visual screen. The electroencephalogram (EEG) was recorded using 64 Ag-AgCl electrodes placed on the scalp according to the International 10/10 system (Waveguard64 cap, Cephalon A/S, Denmark). Vertical and horizontal eye movements were monitored using four additional electrodes placed on the outer canthus of each eye and in the inferior and superior areas of the right orbit. Electrode impedances were kept below 10 k Ω . The signals were recorded using an average reference, amplified and digitized using a sampling rate of 1000 Hz (64-channel high-speed amplifier, Advanced Neuro Technology, The Netherlands).

EEG analysis

Continuous EEG recordings were referenced offline to the left and right mastoids and filtered using a 0.5 Hz high-pass Butterworth zero-phase filter to remove very slow drifts in the recorded signals. EEG epochs lasting 32 s were then obtained by segmenting the recordings from +1 s to +33 s relative to the onset of the audiovisual stimuli, thus yielding 12 epochs for each subject and condition (temporally congruent vs. incongruent audiovisual conditions). Such long-lasting epochs were chosen to improve the signal-to-noise ratio as (1) previous studies have shown that SS-EPs require several cycles of stimulation to be steadily entrained, (2) when installed, the phase and amplitude of SS-EPs are considered to remain relatively constant over time and (3) the high frequency resolution of the obtained frequency spectra concentrates the SS-EP signal within very narrow frequency bands (Regan, 1989). Furthermore, the EEG recorded during the first second of stimulation was excluded from further analyses to discard the transient auditory and visual evoked potentials triggered by the onset of the stimulus (Nozaradan et al., 2011; Saupe et al., 2009). Artifacts produced by eye blinks or eye movements were removed using a validated method based on an Independent Component Analysis (Jung et al., 2000). EEG epochs were then submitted to two distinct analysis procedures to estimate SS-EP amplitude: across-trial averaging in the time domain and across-trial averaging in the frequency domain.

SS-EP amplitude: across-trial averaging in the time domain

For each subject and condition, EEG epochs were averaged across trials in the time domain. Across-trial averaging in the time domain is expected to cancel out or at least markedly reduce the contribution of EEG signals that are not phase-locked to the stimulation train (Mouraux and Iannetti, 2008). The obtained average waveforms were then transformed in the frequency domain using a discrete Fourier transform (Frigo and Johnson, 1998), yielding a frequency spectrum ranging from 0 to 500 Hz with a resolution of 0.031 Hz (Bach and Meigen, 1999). Amplitude spectra were then obtained using the modulus of the complex Fourier coefficients (Fig. 2).

SS-EP amplitude: across-trial averaging in the frequency domain

For each subject and condition, EEG epochs were transformed in the frequency domain using a discrete Fourier transform (Frigo and Johnson, 1998), yielding a frequency spectrum ranging from 0 to 500 Hz with a resolution of 0.031 Hz (Bach and Meigen, 1999). Single-trial amplitude spectra were then obtained using the modulus of the complex Fourier coefficients. The obtained spectra were then averaged across trials, yielding an average spectrum for each subject and condition. Unlike across-trial averaging of amplitude in the time domain, across-trial averaging of amplitude in the frequency domain is expected to preserve the contribution of EEG signals that are not aligned across trials (Lachaux et al., 2000; Mouraux and Iannetti, 2008; Pfurtscheller and Lopes da Silva, 1999; Regan, 1989) (Fig. 2).

Comparing the results obtained using these two approaches allowed assessing whether observed differences in SS-EP amplitude are due to an actual difference in the magnitude of the neuronal entrainment, or to a difference of phase coherence across trials (Lachaux et al., 2000; Pfurtscheller and Lopes da Silva, 1999).

Assuming additive noise, within the obtained frequency spectra, signal amplitude may be expected to correspond to the sum of (1) EEG activity induced by the auditory and visual stimuli, i.e. the auditory and visual SS-EPs and (2) unrelated background noise due, for example, to spontaneous EEG activity, muscle activity or eye movements. Therefore, to obtain estimates of the magnitude of the auditory and visual SS-EPs, the contribution of this noise was removed by subtracting, at each bin of the frequency spectra, the average amplitude measured at neighboring frequency bins (two frequency bins ranging from -0.15 to -0.09 Hz and two frequency bins ranging

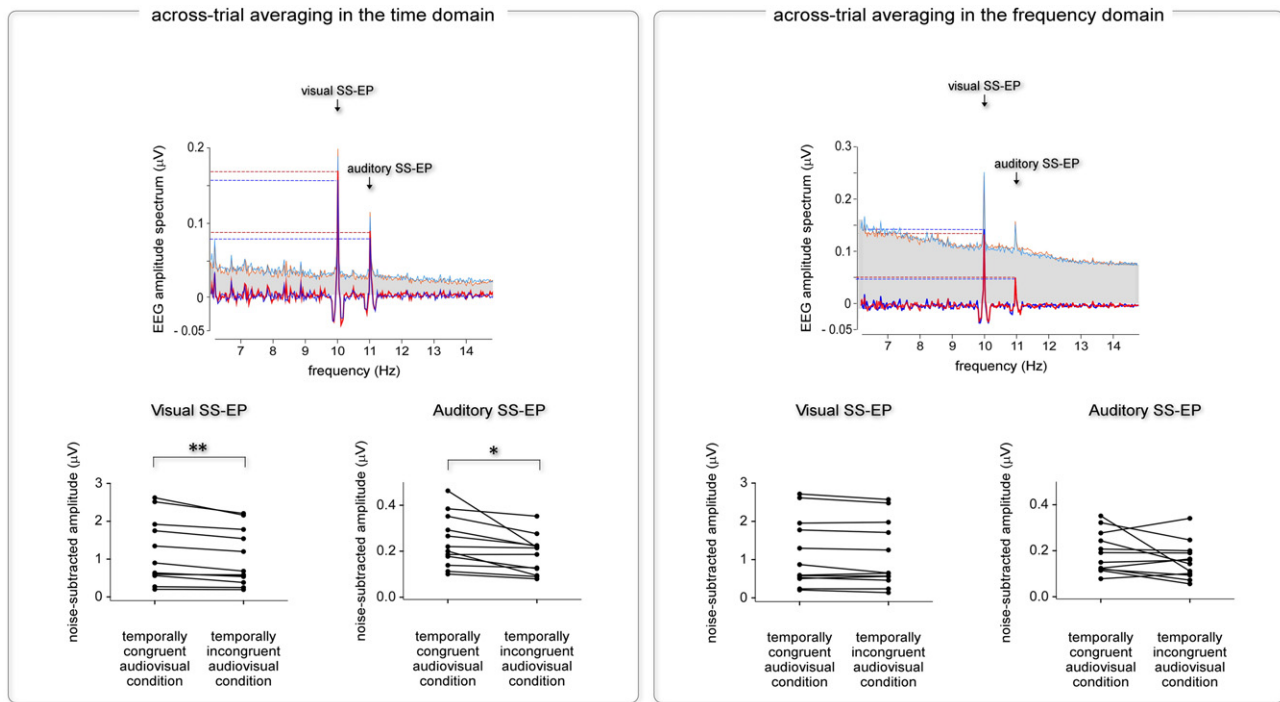


Fig. 2. The upper parts of the panels represent the auditory and visual SS-EPs elicited by temporally congruent (red) and incongruent (blue) audiovisual stimuli. The left panels display the frequency spectra obtained after averaging trials in the time domain. The right panels display the frequency spectra obtained after averaging trials in the frequency domain (group-level average of the signals recorded across the 64 scalp channels). The noise-subtracted spectra are shown using dark-colored lines, whereas the original non-subtracted spectra are shown using light-colored lines. The bottom parts of the panels represent single-subject noise-subtracted auditory and visual SS-EP amplitudes obtained in the temporally congruent and incongruent audiovisual conditions, after averaging trials in the time domain and in the frequency domain, respectively.

from +0.09 to +0.15 Hz, relative to each frequency bin). This procedure is justified by the fact that, in the absence of an SS-EP, the amplitude at a given frequency bin should be similar to the amplitude of the mean of the surrounding frequency bins (Mouraux et al., 2011; Nozaradan et al., 2011). Hence, in the absence of an SS-EP, the noise-subtracted amplitude should tend towards zero. Even if our analyses relied on a comparison of amplitude across two conditions, the subtraction procedure is important because the magnitude of the background noise is not equally distributed across scalp channels. For example, high-frequency noise related to electromyographic activity may be expected to be more present at temporal electrodes, whereas low-frequency noise related to eye-blink artifacts may be expected to be more present at frontal electrodes. Hence, although undermined by some assumptions (additive noise), and possibly leading to an underestimation of the actual SS-EP amplitude, the subtraction procedure is expected to yield more meaningful topographical maps of SS-EP amplitude (Fig. 2).

Finally, the magnitude of visual and auditory SS-EPs was estimated by averaging the signal amplitude measured at the three frequency bins centered on the corresponding target frequencies (10 Hz visual SS-EP: bins ranging from 9.968 to 10.031 Hz; 11 Hz auditory SS-EP: bins ranging from 10.968 to 11.031 Hz), thereby considering a possible spectral leakage due to the fact that the Fourier transform did not estimate signal amplitude at the exact SS-EP frequency.

All these EEG processing steps were carried out using Analyzer 1.05 (Brain Products, Germany), Letswave (Mouraux and Iannetti, 2008), Matlab (The MathWorks, USA) and EEGLAB (<http://scn.ucsd.edu>).

Statistical analyses

For each participant and analysis procedure (across-trial averaging in the time domain and across-trial averaging in the frequency domain), the obtained frequency spectra were averaged in the two

conditions. Then, for each participant and SS-EP, an electrode of interest was then determined by selecting, within these averaged spectra, the electrode exhibiting the maximum amplitude at the corresponding SS-EP frequency (Fig. 3). This selection procedure was applied to take into account the across-subject variability of SS-EP scalp topographies, and the widespread scalp topography of auditory SS-EPs (Johnson et al., 1988) (Fig. 3). Importantly, because the electrode of interest was determined based on the scalp distribution of the spectrum averaged across the two conditions, the procedure did not bias our results towards finding a difference between the two conditions.

Then, a two-way repeated-measures ANOVA was performed to compare the SS-EP amplitude estimates obtained after averaging the trials in the time domain and measured (1) in each of the two experimental conditions (temporally congruent vs. incongruent), and (2) for each of the two sensory modalities (visual vs. auditory). A second two-way repeated-measures ANOVA was performed to compare the SS-EPs amplitudes estimates obtained after averaging the trials in the frequency domain, (1) across experimental conditions and (2) across modalities. Paired-sample *t* tests were used to perform *post hoc* pairwise comparisons of the magnitude of the SS-EPs measured between the two experimental conditions. Significance level was set at $p < 0.05$.

Inter-trial phase coherence

Phase coherence across trials of the visual and auditory SS-EPs was assessed in each of the two conditions (temporally congruent and incongruent), as follows (Cottereau et al., 2011). At the electrode displaying maximum SS-EP amplitude, phase values at the SS-EP frequency were extracted from the Fourier transform of each entire single trial (by computing the *argument* of the complex Fourier coefficient). Using the Circular Statistics Toolbox (Berens, 2009), the across-trial standard deviation of phase angles was then computed, for each subject, experimental condition (temporally congruent vs.

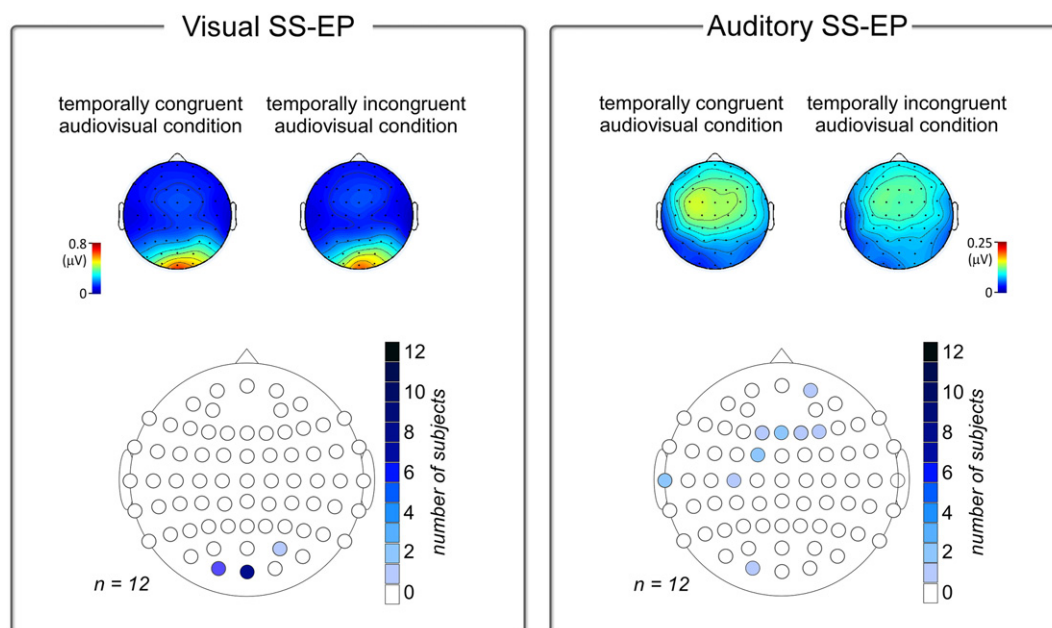


Fig. 3. Topographical distribution of visual and auditory SS-EPs. The upper parts represent the group-level average topographical maps of the magnitude of the SS-EPs elicited by 10-Hz visual stimulation and 11-Hz auditory stimulation, obtained after averaging trials in the time domain. The bottom parts of the figure represent the location of the scalp electrodes displaying maximal auditory and visual SS-EP amplitude across subjects. Note that the visual SS-EP was predominant at occipital electrodes, whereas the auditory SS-EP was, in most subjects, predominant over fronto-central electrodes.

incongruent) and SS-EP (visual vs. auditory). A two-way repeated-measures ANOVA was then used to compare these measures of across-trial phase coherence in each of the conditions (temporally congruent vs. incongruent) and each sensory modality (visual and auditory SS-EPs). Paired-sample *t* tests were used to perform *post hoc* pairwise comparisons of the inter-trial phase coherence values between the two experimental conditions. Significance level was set at $p < 0.05$.

Results

Behavioral results

During the recording, participants performed the detection task with a median score of 8.5/10 (interquartile range: 8–10). Detection performance for discontinuities in the visual stimulus was not significantly different than detection performance for discontinuities in the auditory stimulus (one sample *t* test: $t = 1.45$, $p = 0.18$). After the recording, all participants discriminated the temporally congruent and incongruent audiovisual trials with a perfect score.

Electrophysiological results

Visual SS-EP

A marked increase of EEG signal amplitude was recorded in all participants at the frequency of 10 Hz, corresponding to the SS-EP elicited by the visual stimuli (Fig. 2). In both the temporally congruent and incongruent audiovisual conditions, the scalp topography of this response was maximal over occipital regions (Fig. 3). When averaging the trials in the time domain, the magnitude of the visual SS-EP was $1.16 \pm 0.25 \mu\text{V}$ in the temporally congruent condition and $1.01 \pm 0.21 \mu\text{V}$ in the temporally incongruent condition (mean $\pm$ sem) (Fig. 2). On average, SS-EP magnitude was 13.7 ± 1.0 times greater than the mean background activity (ranging from 6.6 to 18.1 across individual subjects). When averaging the trials in the frequency domain, the magnitude of the visual SS-EP was $1.16 \pm 0.26 \mu\text{V}$ in the temporally congruent condition and $1.10 \pm 0.25 \mu\text{V}$ in the temporally incongruent condition. On average, SS-EP

magnitude was 6.3 ± 0.8 times greater than the mean background activity (ranging from 3.5 to 11.9 across individual subjects).

Auditory SS-EP

A marked increase of EEG signal amplitude was recorded in all participants at the frequency of 11 Hz, corresponding to the SS-EP elicited by the auditory stimuli (Fig. 2). In both the temporally congruent and incongruent audiovisual conditions, the scalp topography of this response was widespread over both hemispheres, but predominant over fronto-central regions (Fig. 3). When averaging the trials in the time domain, the magnitude of the auditory SS-EP was $0.24 \pm 0.03 \mu\text{V}$ in the temporally congruent condition and $0.18 \pm 0.05 \mu\text{V}$ in the temporally incongruent condition (Fig. 2). On average, SS-EP magnitude was 5.9 ± 0.8 times greater than the mean background activity (ranging from 2.1 to 9.5 across individual subjects). When averaging the trials in the frequency domain, the magnitude of the auditory SS-EP was of $0.19 \pm 0.03 \mu\text{V}$ in the temporally congruent condition and $0.16 \pm 0.02 \mu\text{V}$ in the temporally incongruent condition. On average, SS-EP magnitude was 2.1 ± 0.2 times greater than the mean background activity (ranging from 1.2 to 3.6 across individual subjects).

Comparison of the visual and auditory SS-EPs elicited by the temporally congruent and incongruent audiovisual stimuli

The repeated-measures ANOVA used to compare the magnitude of visual and auditory SS-EPs obtained after averaging the trials in the time domain across experimental conditions (temporally congruent vs. incongruent) and modalities (visual and auditory) revealed a significant main effect of the type of stimulus ($F = 14.87$, $\eta^2 = 0.57$, $p = 0.003$). Regardless of the experimental condition, the magnitude of the SS-EPs elicited by 10-Hz visual stimulation was significantly greater than the magnitude of the SS-EPs elicited by 11-Hz auditory stimulation. Regardless of the type of stimulus, there was a significant main effect of temporal congruency ($F = 19.77$, $\eta^2 = 0.64$, $p = 0.001$). There was a significant interaction between the factors 'temporal congruency' and 'modality' ($F = 5.47$, $\eta^2 = 0.33$, $p = 0.039$). Post-hoc pairwise comparisons showed that the magnitude of the SS-EPs obtained when averaging trials in the time domain were significantly

greater in the temporally congruent vs. incongruent conditions (visual SS-EP: $t = 3.89$, $p = 0.0025$; auditory SS-EP: $t = 2.94$, $p = 0.0135$).

The repeated-measures ANOVA used to compare the magnitude of visual and auditory SS-EPs obtained after averaging the trials in the frequency domain across experimental conditions (temporally congruent vs. incongruent) and modalities (visual and auditory) revealed a significant main effect of the type of stimulus ($F = 15.25$, $\eta^2 = 0.58$, $p = 0.002$). Regardless of the experimental condition, the magnitude of the SS-EPs elicited by 10-Hz visual stimulation was significantly greater than the magnitude of the SS-EPs elicited by 11-Hz auditory stimulation. Regardless of the type of stimulus, there was a significant main effect of temporal congruency ($F = 6.72$, $\eta^2 = 0.38$, $p = 0.025$). There was no significant interaction between the factors 'temporal congruency' and 'modality' ($F = 0.3$, $\eta^2 = 0.026$, $p = 0.6$). Post-hoc pairwise comparisons showed that the magnitude of the SS-EPs obtained when averaging trials in the frequency domain were not significantly greater in the temporally congruent vs. incongruent conditions (visual SS-EP: $t = 2.16$, $p = 0.053$; auditory SS-EP: $t = 1.51$, $p = 0.16$).

Comparison of the visual and auditory SS-EPs elicited in unisensory and multisensory conditions

The magnitude of the visual and auditory SS-EPs recorded in the unisensory condition were significantly smaller than the magnitude of the visual and auditory SS-EPs recorded in the multisensory conditions, both after across-trial averaging in the time domain and after across-trial averaging in the frequency domain (see Supplementary Fig. 1).

Inter-trial phase coherence

The mean intertrial coherence values obtained were 0.30 ± 0.14 radians and 0.66 ± 0.24 radians (mean and standard deviation of the mean) for the visual SS-EP, and were of 0.60 ± 0.23 and 0.79 ± 0.19 radians (mean and standard deviation of the mean) for the auditory SS-EPs, in the temporally congruent and incongruent conditions respectively. The repeated-measures ANOVA used to compare the estimated standard deviation of phase across trials of the visual and auditory SS-EPs recorded in the temporally congruent and incongruent experimental conditions revealed a significant main effect of temporal congruency ($F = 22.21$, $\eta^2 = 0.61$, $p = 0.001$). The estimated standard deviation of phase across trials was significantly greater in the temporally incongruent audiovisual condition as compared to the temporally congruent audiovisual condition, both for the visual SS-EP ($t = 3.65$, $p = 0.0038$) and the auditory SS-EP ($t = 2.88$, $p = 0.0149$), thus indicating that temporal congruency of the visual and auditory stimuli enhanced the phase coherence across trials of the elicited visual and auditory SS-EPs. There was also a significant main effect of sensory modality ($F = 18.21$, $\eta^2 = 0.62$, $p = 0.001$), but no interaction between the two factors ($F = 2.22$, $\eta^2 = 0.17$, $p = 0.164$).

Discussion

The aim of the present study was to explore the neural mechanisms underlying the multimodal integration of sensory inputs sharing similar temporal dynamics. For this purpose, a novel EEG approach – frequency tagging using steady-state evoked potentials (SS-EPs) – was used to isolate the neural activities related to the processing of simultaneously presented auditory and visual stimuli. The audiovisual stimuli carried temporal periodicities, or *beats*, that were either synchronous (leading to a unified perception of beat) or non-synchronous (not leading to a unified audiovisual beat percept). Importantly, the auditory and visual streams of sensory input were identical in the two conditions, which thus differed only by the relative temporal congruency of their temporal dynamics.

Both the EEG responses to the auditory stimulus (i.e. the auditory SS-EP) and the EEG responses to the visual stimulus (i.e. the visual SS-EP) were significantly enhanced when the auditory and visual beats were congruent, as compared to when the auditory and visual beats were incongruent. In addition, the phase of the elicited auditory and visual SS-EPs remained significantly more constant across trials in the temporally congruent vs. incongruent audiovisual conditions (Lopes da Silva, 2006; Shah et al., 2004).

The finding that temporal congruency between the auditory and visual beats enhanced the phase coherence of the SS-EPs elicited by each of the two streams of sensory input indicates that, from trial to trial, there was less variability in the timing of the elicited SS-EPs when both sensory systems were synchronized to a common beat frequency. This interpretation is consistent with previous studies showing that *cross-modal congruency*, i.e. the congruency of the information conveyed by different sensory modalities in term of their spatial, temporal or semantic aspects, can induce an enhancement of the phase coherence of oscillatory activities across cortical areas (Kayser et al., 2008, 2010; Luo et al., 2010; Schall et al., 2009; Schroeder et al., 2008; Senkowski et al., 2008; Talsma et al., 2010). In fact, this enhancement of synchrony across cortical areas has been proposed as the actual mechanism underlying the binding of congruent multisensory inputs into a unified coherent percept (reviewed in Senkowski et al., 2008). The enhancement of across-trial phase coherence of auditory and visual SS-EPs that was observed in the temporally congruent audiovisual condition of the present study could thus be interpreted as revealing an increased binding of auditory and visual cortical processing, contributing to the emergence of a coherent and unified audiovisual percept, as compared to when auditory and visual beats were incongruent.

The finding that temporal congruency between the auditory and visual beats enhanced the inter-trial phase coherence of both auditory and visual SS-EPs is also consistent with the results of studies having shown that *selective attention* can enhance the phase coherence of neuronal oscillatory activities (Iversen et al., 2008; Kayser, 2009; Kim et al., 2007; Lakatos et al., 2009; Senkowski et al., 2005; Zhang et al., 2010). More specifically, this enhancement could be interpreted as resulting from a beat-induced process of *dynamic attending* (Jones and Boltz, 1989), occurring in the temporally congruent audiovisual condition. Indeed, the cortical processing of temporal periodicities such as musical beats has been proposed to be related to the induction of a dynamic modulation of attention, frequency-tuned to the beat, and subtended by an entrainment of the neuronal populations responding to the beat (Large, 2008; Large and Jones, 1999; Nozaradan et al., in press). As previous studies have shown that the responsiveness to external stimuli of oscillating neurons is dependent on the phase of these oscillations (Busch et al., 2009; Fries, 2005), neuronal entrainment at beat frequency is expected to generate a periodic modulation of neuronal responsiveness, which could explain the beat-induced periodic modulation of selective attention across time. In the temporally congruent audiovisual condition of the present study, the improved coherence of the elicited auditory and visual SS-EPs could thus be due to temporal synchronization and, thereby, enhancement of auditory and visual dynamic attending processes. Alternatively, in the temporally incongruent audiovisual condition, the reduced coherence of the auditory and visual SS-EPs could be explained by the lack of synchrony of the auditory and visual beats, preventing the formation of a well-structured periodic dynamic attending.

The cortical areas involved in the processing of multisensory integration and perceptual binding remain a matter of ongoing debate. In the classical hierarchical view of sensory processing, multimodal integration is considered as a higher-order process that occurs only after sensory information has undergone preliminary processing through modality-specific cortical structures (Jones and Powell, 1970; Stein and Meredith, 1993). However, this view has been progressively

abandoned in the light of the growing amount of evidence showing cross-sensory influences occurring already at the earliest stages of sensory processing (Kayser and Logothetis, 2007). Primary sensory cortices, as well as subcortical nuclei, have been shown to take part in multimodal binding, through a modulation of their neuronal activity (Foxe and Schroeder, 2005; Ghazanfar and Schroeder, 2006; Krauslis et al., 2004; Stein and Wallace, 1996). In the present experiment, temporally congruent vs. incongruent audiovisual stimulations appeared to modulate the cortical activity triggered by auditory and visual stimulation within primary sensory areas, as suggested by the occipital scalp topography obtained for the visual SS-EPs elicited in the two conditions, as well as the fronto-central scalp topography of the auditory SS-EPs elicited in the two conditions (Fig. 3).

Taken together, our findings provide support to the notion that perception of temporally congruent multisensory input is subtended by an enhancement of the neural activities related to processing each of the different streams of sensory input, occurring already at the sensory-specific stages of cortical processing. Moreover, we show that EEG frequency-tagging with SS-EPs constitutes a highly effective tool to explore the neural dynamics of multisensory binding in the human brain.

Supplementary materials related to this article can be found online at doi:10.1016/j.neuroimage.2011.11.065.

Acknowledgments

S.N. is supported by the Fund for Scientific Research of the French-speaking community of Belgium (F.R.S.-FNRS). I.P. is supported by the Natural Sciences and Engineering Research Council of Canada, the Canadian Institutes of Health Research and by a Canada Research Chair. We thank the two anonymous reviewers for their constructive comments to the manuscript.

References

- Adrian, E.D., 1949. The Sherrington lectures: 1. Sensory Integration. University of Liverpool Press.
- Bach, M., Meigen, T., 1999. Do's and don'ts in Fourier analysis of steady-state potentials. *Doc. Ophthalmol.* 99, 69–82.
- Berens, P., 2009. CircStat: a Matlab toolbox for circular statistics. *J. Stat. Softw.* 31 (10).
- Bertelson, P., 1999. Ventiloquism: a case of cross-modal perceptual grouping. In: Aschersleben, G., Bachmann, T., Musseler, J. (Eds.), *Cognitive Contributions to the Perception of Spatial and Temporal Events*. Elsevier, Amsterdam, pp. 347–362.
- Besle, J., Bertrand, O., Giard, M.H., 2009. Electrophysiological (EEG, sEEG, MEG) evidence for multiple audiovisual interactions in the human auditory cortex. *Hear. Res.* 258, 143–145.
- Brainard, D.H., 1997. The psychophysics toolbox. *Spat. Vis.* 10, 433–436.
- Busch, N.A., Dubois, J., VanRullen, R., 2009. The phase of ongoing EEG oscillations predicts visual perception. *J. Neurosci.* 29, 7869–7876.
- Cottareau, B.R., McKee, S.P., Ales, J.M., Norcia, A.M., 2011. Disparity-tuned population responses from human visual cortex. *J. Neurosci.* 31, 954–965.
- de Jong, R., Toffanin, P., Harbers, M., 2010. Dynamic crossmodal links revealed by steady-state responses in auditory-visual divided attention. *Int. J. Psychophysiol.* 75, 3–15.
- Drake, C., Botte, M.C., 1993. Tempo sensitivity in auditory sequences: evidence for a multiple look model. *Percept. Psychophys.* 54, 277–286.
- Elliott, M.T., Wing, A.M., Welchman, A.E., 2010. Multisensory cues improve sensorimotor synchronization. *Eur. J. Neurosci.* 31, 1828–1835.
- Foxe, J.J., Schroeder, C.E., 2005. The case for feedforward multisensory convergence during early cortical processing. *Neuroreport* 16, 419–423.
- Fries, P., 2005. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* 9, 474–480.
- Frigo, M., Johnson, S.G., 1998. FFTW: an adaptive software architecture for the FFT. *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing*, 3, pp. 1381–1384.
- Fujisaki, W., Nishida, S., 2005. Temporal frequency characteristics of synchrony–asynchrony discrimination of audio–visual signals. *Exp. Brain Res.* 166, 455–464.
- Galambos, R., Makeig, S., Talmachoff, P.J., 1981. A 40-Hz auditory potential recorded from the human scalp. *Proc. Natl. Acad. Sci. U. S. A.* 78, 2643–2647.
- Ghazanfar, A.A., Schroeder, C.E., 2006. Is neocortex essentially multisensory? *Trends Cogn. Sci.* 10, 278–285.
- Gibson, J.J., 1966. The problem of temporal order in stimulation and perception. *J. Psychol.* 62, 141–149.
- Glenberg, A.M., Mann, S., Altman, L., Forman, T., 1989. Modality effects in the coding and reproduction of rhythms. *Mem. Cogn.* 17, 373–383.
- Grahn, J.A., Henry, M.J., McAuley, J.D., 2011. fMRI investigation of cross-modal interactions in beat perception: audition primes vision, but not vice versa. *Neuroimage* 54, 1231–1243.
- Iversen, J.R., Patel, A.D., Chen, Y., Gally, J.A., Edelman, G.M., 2008. Selective attention to one of two competing auditory rhythms modulates phase of brain responses to the attended rhythm. *Cognitive Neuroscience Society Annual Meeting*, San Francisco.
- Johnson, B.W., Weinberg, H., Ribary, U., Cheyne, D.O., Ancill, R., 1988. Topographic distribution of the 40 Hz auditory evoked-related potential in normal and aged subjects. *Brain Topogr.* 1, 117–122.
- Jones, M.R., Boltz, M., 1989. Dynamic attending and responses to time. *Psychol. Rev.* 96, 459–491.
- Jones, E.G., Powell, T.P., 1970. An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain* 93, 793–820.
- Jung, T.P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., Sejnowski, T.J., 2000. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin. Neurophysiol.* 111, 1745–1758.
- Kayser, C., 2009. Phase resetting as a mechanism for supramodal attentional control. *Neuron* 64, 300–302.
- Kayser, C., Logothetis, N.K., 2007. Do early sensory cortices integrate cross-modal information? *Brain Struct. Funct.* 212, 121–132.
- Kayser, C., Petkov, C.I., Logothetis, N.K., 2008. Visual modulation of neurons in auditory cortex. *Cereb. Cortex* 18, 560–574.
- Kayser, C., Logothetis, N.K., Panzeri, S., 2010. Visual enhancement of the information representation in auditory cortex. *Curr. Biol.* 20, 19–24.
- Kim, Y.J., Grabowecy, M., Paller, K.A., Muthu, K., Suzuki, S., 2007. Attention induces synchronization-based response gain in steady-state visual evoked potentials. *Nat. Neurosci.* 10, 117–125.
- Krauslis, R.J., Liston, D., Carello, C.D., 2004. Target selection and the superior colliculus: goals, choices and hypotheses. *Vision Res.* 44, 1445–1451.
- Lachaux, J.P., Rodriguez, E., Martinerie, J., Adam, C., Hasboun, D., Varela, F.J., 2000. A quantitative study of gamma-band activity in human intracranial recordings triggered by visual stimuli. *Eur. J. Neurosci.* 12, 2608–2622.
- Lakatos, P., O'Connell, M.N., Barczak, A., Mills, A., Javitt, D.C., Schroeder, C.E., 2009. The leading sense: supramodal control of neurophysiological context by attention. *Neuron* 64, 419–430.
- Large, E.W., 2008. Resonating to musical rhythm: theory and experiment. In: Grondin, Simon (Ed.), *The Psychology of Time*. Emerald, West Yorkshire.
- Large, E.W., Jones, M.R., 1999. The dynamics of attending: how we track time varying events. *Psychol. Rev.* 106, 119–159.
- Large, E.W., Kolen, J.F., 1994. Resonance and the perception of musical meter. *Connect. Sci.* 6, 177–208.
- London, J., 2004. *Hearing in Time: Psychological Aspects of Musical Meter*. Oxford University Press Inc.
- Lopes da Silva, F.H., 2006. Chapter 1 event-related neural activities: what about phase? *Prog. Brain Res.* 159, 3–17.
- Luo, H., Liu, Z., Poeppel, D., 2010. Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biol.* 10 :8 (8), e1000445.
- Morgan, S.T., Hansen, J.C., Hillyard, S.A., 1996. Selective attention to stimulus location modulates the steady-state visual evoked potential. *Proc. Natl. Acad. Sci. U. S. A.* 93, 4770–4774.
- Mouraux, A., Iannetti, G.D., 2008. Across-trial averaging of event-related EEG responses and beyond. *Magn. Reson. Imaging* 26, 1041–1054.
- Mouraux, A., Iannetti, G.D., Colon, E., Nozaradan, S., Legrain, V., Plaghki, L., 2011. Nociceptive steady-state evoked potentials elicited by rapid periodic thermal stimulation of cutaneous nociceptors. *J. Neurosci.* 31, 6079–6087.
- Nozaradan, S., Peretz, I., Missal, M., Mouraux, A., 2011. Tagging the neuronal entrainment to beat and meter. *J. Neurosci.* 31, 10234–10240.
- Patel, A.D., Iversen, J.R., Chen, Y., Repp, B.H., 2005. The influence of metricity and modality on synchronization with a beat. *Exp. Brain Res.* 163, 226–238.
- Petrini, K., Russell, M., Pollick, F., 2009. When knowing can replace seeing in audiovisual integration of actions. *Cognition* 110, 432–439.
- Pfurtscheller, G., Lopes da Silva, F.H., 1999. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* 110, 1842–1857.
- Regan, D.M., 1989. *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine*. Elsevier, New York.
- Regan, D.M., Heron, J.R., 1969. Clinical investigation of lesions of the visual pathway: a new objective technique. *J. Neurol. Neurosurg. Psychiatry* 32, 479–483.
- Repp, B.H., 2005. Sensorimotor synchronization: a review of the tapping literature. *Psychon. Bull. Rev.* 12, 969–992.
- Repp, B.H., 2006. Musical synchronization. *Music, Motor Control and the Brain*. In: Altenmüller, E., Wiesendanger, M., Kesselring, J. (Eds.), Oxford University Press, pp. 55–76. Chapter 4.
- Saupe, K., Schröger, E., Andersen, S.K., Müller, M.M., 2009. Neural mechanisms of inter-modal sustained selective attention with concurrently presented auditory and visual stimuli. *Front. Hum. Neurosci.* 3, 58.
- 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.
- Schroeder, C.E., Lakatos, P., Kajikawa, Y., Partan, S., Puce, A., 2008. Neuronal oscillations and visual amplification of speech. *Trends Cogn. Sci.* 12, 106–113.
- Sekuler, R., Sekuler, A.B., Lau, R., 1997. Sound alters visual motion perception. *Nature* 385, 308.
- Senkowski, D., Talsma, D., Herrmann, C.S., Woldorff, M.G., 2005. Multisensory processing and oscillatory gamma responses: effects of spatial selective attention. *Exp. Brain Res.* 166, 411–426.

- Senkowski, D., Talsma, D., Grigutsch, M., Herrmann, C.S., Woldorff, M.G., 2007. Good times for multisensory integration: effects of the precision of temporal synchrony as revealed by gamma-band oscillations. *Neuropsychologia* 45, 561–571.
- Senkowski, D., Schneider, T.R., Foxe, J.J., Engel, A.K., 2008. Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci.* 31, 401–409.
- Shah, A.S., Bressler, S.L., Knuth, K.H., Ding, M., Mehta, A.D., Ulbert, I., Schroeder, C.E., 2004. Neural dynamics and the fundamental mechanisms of event-related brain potentials. *Cereb. Cortex* 14, 476–483.
- Spence, C., Driver, J. (Eds.), 2004. *Crossmodal Space and Crossmodal Attention*. Oxford University Press, Oxford.
- Stein, B.E., Meredith, M.A., 1993. *Merging of the Senses*. MIT Press, Cambridge.
- Stein, B.E., Wallace, M.T., 1996. Comparisons of cross-modality integration in midbrain and cortex. *Prog. Brain Res.* 112, 289–299.
- Talsma, D., Doty, T.J., Strowd, R., Woldorff, M.G., 2006. Attentional capacity for processing concurrent stimuli is larger across sensory modalities than within a modality. *Psychophysiology* 43, 541–549.
- Talsma, D., Senkowski, D., Soto-Faraco, S., Woldorff, M.G., 2010. The multifaceted interplay between attention and multisensory integration. *Trends Cogn. Sci.* 14, 400–410.
- Toffanin, P., de Jong, R., Johnson, A., Martens, S., 2009. Using frequency tagging to quantify attentional deployment in a visual divided attention task. *Int. J. Psychophysiol.* 72, 289–298.
- Tononi, G., Srinivasan, R., Russell, D.P., Edelman, G.M., 1998. Investigating neural correlates of conscious perception by frequency-tagged neuromagnetic responses. *Proc. Natl. Acad. Sci. U. S. A.* 95, 3198–3203.
- van Noorden, L., Moelants, D., 1999. Resonance in the perception of musical pulse. *J. New Music Res.* 28, 43–66.
- Vatakis, A., Spence, C., 2006. Audiovisual synchrony perception for music, speech, and object actions. *Brain Res.* 1111, 134–142.
- Vroomen, J., Keetels, M., 2010. Perception of intersensory synchrony: a tutorial review. *Atten. Percept. Psychophys.* 72, 871–884.
- Welch, R.B., Warren, D.H., 1980. Immediate perceptual response to intersensory discrepancy. *Psychol. Bull.* 88, 638–667.
- Zampini, M., Shore, D.I., Spence, C., 2003. Audiovisual temporal order judgments. *Exp. Brain Res.* 152, 198–210.
- Zhang, D., Hong, B., Gao, X., Gao, S., Röder, B., 2010. Exploring steady-state visual evoked potentials as an index for intermodal and crossmodal spatial attention. *Psychophysiology*. doi:10.1111/j.1469-8986.2010.01132.x.