UNIVERSITE CATHOLIQUE DE LOUVAIN
BELGIUM

UNIVERSITE DE MONTREAL
CANADA

EXPLORING THE NEURAL ENTRAINMENT TO MUSICAL RHYTHMS AND METER:

A STEADY-STATE EVOKED POTENTIAL APPROACH

Sylvie NOZARADAN, MD

Promotors : Dr. André Mouraux (UCL)
Dr. Isabelle Peretz (UdeM)

Dissertation submitted for the degree of Doctor of Philosophy, Faculty of Medicine

March 2013

Ces années de thèse font sans doute partie de mes meilleurs moments de vie. Elles ont été pour moi l'occasion de découvrir une passion pour la recherche, et de combiner mes centres d'intérêts, a priori éloignés. Mais cela n'aurait pas été possible sans l'encadrement optimal dont j'ai bénéficié durant ces années

Tout d'abord, c'est toujours une rare chance de croiser sur sa route un mentors. Je mesure la chance que j'ai eue d'en recontrer dans le cadre de ma thèse. Mes remerciements vont donc en premier à André, mon superviseur à l'UCL. Etre sa première doctorante est un grand honneur, dont j'espère avoir été à la hauteur. Je n'exagère pas en affirmant qu'il a été, et restera sûrement, un modèle, tant au niveau de ses très nombreux talents dans chaque étape du travail de chercheur qu'au niveau de ses qualités humaines. Associée dans cette supervision en équipe, je dois beaucoup à Isabelle, mon superviseur à Montréal, de m'avoir permis de découvrir au Brams, déjà avant la fin de mes études de médecine, la richesse de cette thématique de recherche. Je la remercie chaleureusement pour son soutien, ses encouragements et sa confiance depuis le début.

Je remercie vivement les membres de mon comité d'accompagnement et du jury, Pascal Kienlen-Campard, Philippe Lefèvre, Marcus Missal, Alexandre Zénon, Charles Schroeder et Marc Schoenwiesner, pour leur lecture attentive de la thèse, ainsi que pour leurs commentaires et conseils tout au long du parcours. Je remercie aussi Etienne, Julie et Alex, ainsi que leur équipe, pour leur écoute et commentaires lors des journal clubs auxquels j'ai pu presenter mes travaux dès le début de la thèse, ainsi que leur bons conseils.

Je ne garderais sûrement pas un aussi bon souvenir de ces années si je n'avais eu la chance de réaliser cette thèse dans une si bonne ambiance. Je remercie donc les collègues de Nocions (dont Elisabeth, la plus agréable collègue de bureau qui soit!), et de COSY, dont le sourire et la bonne humeur sont autant de cadeaux que j'ai reçus chaque jour et qui vont me manquer. Que de bons souvenirs des conférences SfN, des verres du vendredi, de la campagne de vol… Merci également d'avoir été les meilleurs sujets d'expérience qui soient. Un grand merci aussi à l'équipe administrative et technique de COSY pour leur soutien et leur efficacité. J'ai eu la chance de trouver du côté Brams une ambiance tout aussi chaleureuse, qui aura rendu Montréal particulièrement difficile à quitter chaque fois. J'espère avoir encore de nombreuses occasions d'y séjourner et de collaborer avec le Brams.

Merci aussi à Vincent (aussi un de mes sujets d'expérience favoris!), Arnould, Leon van Noorden et le groupe Rhythm de Bruxelles, ainsi qu'aux chercheurs de la communauté Rhythm (RPPW et PoRT) pour ces échanges inspirants dans une ambiance amicale. La tonalité de ces rencontres est à cultiver, et a déjà débouché sur de belles amitiés. A propos d'amitié, je remercie bien sûr aussi Jessica Phillips-Silver et Coralie de Hemptinne, qui toutes deux ont fait preuve de patience pour m'accompagner dans mes premiers pas au labo, au Brams et à l'UCL. Je remercie aussi chaleureusement Marc Crommelinck, éternel créateur d'enthousiasme, qui m'a mis entre les mains, dès la 2ème candi, des lectures de neuroscience passionnantes (que je ne finis pas d'essayer de comprendre, alias Edelman!), et Bruno, qui après Marc a essayé de me les faire comprendre à son tour. Son enthousiasme ne cesse de m'inspirer depuis.

Ces réflexions sur la musique ne me sont pas tombées du ciel. Je remercie donc très chaleureusement ici mes mentors du côté musique, Françoise, Burkard, Jean-Claude, Santi et Caroline, qui continuent de m'inspirer un modèle de rigueur au travail, de curiosité et de créativité, qui s'applique si bien tant au métier d'artiste qu'à celui de chercheur.

Et bien sûr, un énorme merci à mon copain et ma famille de m'avoir supporté (au sens français et anglais du terme!) dans ce cheminement, et pour toute la force qu'ils me donnent chaque jour.

SUMMARY **9**

FOREWORD **11**

I. ENTRAINMENT IN NEURAL SYSTEMS **13**

I.1. ENDOGENOUS OSCILLATIONS 13

I.1.1. Spontaneous neural oscillations. *14*

I.1.2. Frequency tuning function. *16*

I.1.3. Synchronization. *17*

I.1.4. A link between synchronization of oscillatory activities and brain function. *17*

I.2. ENTRAINMENT TO OSCILLATORY INPUTS 20

I.2.1. Synchronous oscillation to sound envelope in the auditory system *20*

I.2.1.1. What is sound envelope? 20

I.2.1.2. Frequency decomposition of the sound. 21

I.2.1.3. Temporal coding versus rate coding. 23

I.2.2. The steady-state evoked potential approach *27*

I.2.2.1. Nature of the steady-state evoked potentials. 28

I.2.2.2. Frequency tagging. 29

I.2.2.3. Methodological considerations. 30

II. ENTRAINMENT TO MUSICAL RHYTHMS **35**

II.1. BEAT IN MUSIC: A UNIVERSAL HUMAN ABILITY? 35

II.1.1. Are beat and meter induced in all musical styles? *35*

II.1.2. Language and music rhythms. *37*

II.1.2.1. Rhythm in speech. 37

II.1.2.2. Music mimicking speech rhythms and *vice versa*. 38

II.1.3. Cultural differences *39*

II.1.3.1. The Groove as a cultural specificity. 40

II.1.3.2. Meter, binary bias and integer ratios across cultures. 40

II.1.4. Human development *48*

II.1.4.1. Production. 48

II.1.4.2. Perception. 49

II.1.4.3. Enculturation. 52

II.1.5. Evolutionary perspective *54*

II.1.5.1. Innateness. 55

II.1.5.2. Domain-specificity and inter-individual differences. 55

II.1.5.3. Human specificity. 58

II.2. SENSORIMOTOR COUPLING 64

II.2.1. Production of periodic signals 64

II.2.2. Perception of periodic signals 67

II.2.2.1. Models and empirical evidence on time processing. 67

II.2.2.2. Neural correlates of time processing. 69

II.2.2.3. Rhythm processing models. 74

II.2.2.4. Beat induction. 81

II.2.2.5. Neural correlates of rhythm processing. 86

II.2.2.6. Tagging the neural entrainment to beat and meter: our novel approach 103

II.2.2.6.1. Study 1: TAGGING THE NEURONAL ENTRAINMENT TO BEAT AND METER (Sylvie Nozaradan, Isabelle Peretz, Marcus Missal, André Mouraux) 103

II.2.2.6.2. Study 2: SELECTIVE NEURONAL ENTRAINMENT TO THE BEAT AND METER EMBEDDED IN A MUSICAL RHYTHM (Sylvie Nozaradan, Isabelle Peretz, André Mouraux) 124

II.2.3. Coupling of sensorimotor periodic signals 156

II.2.3.1. Characteristics of sensorimotor synchronization to the beat. 156

II.2.3.2. A frequency tuning function for beat perception and synchronization. 160

II.2.3.3. Neural correlates of sensorimotor synchronization. 170

II.2.3.4. Study 3: CAPTURING WITH EEG THE NEURONAL ENTRAINMENT AND COUPLING UNDERLYING SENSORIMOTOR SYNCHRONIZATION TO THE BEAT (Sylvie Nozaradan, Younes Zerouali, Isabelle Peretz, André Mouraux) 179

II.2.3.5. Perception leading to movement. 212

II.2.3.6. Movement influencing perception. 214

II.2.3.7. Auditory prominence for sensorimotor coupling. 216

II.2.3.8. Study 4: STEADY-STATE EVOKED POTENTIALS AS AN INDEX OF MULTISENSORY TEMPORAL BINDING (Sylvie Nozaradan, Isabelle Peretz, André Mouraux) 220

III. DISCUSSION AND PERSPECTIVES 245

III.1. STUDY 1: TAGGING THE NEURONAL ENTRAINMENT TO BEAT AND METER 245

III.1.1. Making a bridge between beat and meter-related SS-EPs, transient ERPs and ongoing oscillatory activities. 246

III.1.2. Some remarks on the frequency domain analysis of SS-EPs. *248*
III.1.3. Musicians versus nonmusicians. *248*
III.1.4. Beat and meter-related SS-EPs and head movement artifacts. *249*
III.2. STUDY 2: SELECTIVE NEURONAL ENTRAINMENT TO THE BEAT AND METER EMBEDDED IN A MUSICAL RHYTHM 250
III.2.1. The neural sources of beat and meter-related SS-EPs. *251*
III.2.2. Similarities between pitch and meter processing. *253*
III.2.3. Retrieving time resolution and phase from SS-EPs. *254*
III.3. STUDY 3: CAPTURING WITH EEG THE NEURONAL ENTRAINMENT AND COUPLING UNDERLYING SENSORIMOTOR SYNCHRONIZATION TO THE BEAT 256
III.4. STUDY 4: STEADY-STATE EVOKED POTENTIALS AS AN INDEX OF MULTISENSORY TEMPORAL BINDING .. 257
III.4.1. Cross-frequency modulation studied using frequency tagging. *258*
III.4.2. Is periodicity special? *258*
IV. GLOSSARY 261
IV.1. ENTRAINMENT IN PHYSICS, BIOLOGICAL SYSTEMS AND NEUROSCIENCE 261
IV.2. MUSICAL TERMS 265
V. REFERENCES 269

SUMMARY

The ability to perceive a regular beat in music and synchronize to it is a widespread human skill. Fundamental to musical behavior, beat and meter refer to the perception of periodicities while listening to musical rhythms, and usually involve spontaneous entrainment to move on these periodicities. However, the neural mechanisms underlying entrainment to beat and meter in Humans remain unclear. The present work tests a novel experimental approach, inspired by the steady-state evoked potential method, to explore the neural dynamics supporting the perception of rhythmic inputs. Using human electroencephalography (EEG), neural responses to beat and meter were recorded in various contexts: (1) mental imagery of meter, (2) spontaneous induction of a beat from rhythmic patterns, (3) multisensory integration, and (4) sensorimotor synchronization. Our results support the view that entrainment and resonance phenomena subtend the processing of musical rhythms in the human brain. Furthermore, our results suggest that this novel approach could help investigating the link between the phenomenology of musical beat and meter and neurophysiological evidence of a bias towards periodicities arising under certain circumstances in the nervous system. Hence, entrainment to music provides an original framework to explore general entrainment phenomena occurring at various levels, from the inter-neural to the inter-individual level.

FOREWORD

One of the richest features of music is its temporal structure. In particular, the *beat*, which usually refers to the perception of periodicities while listening to music, can be considered as a cornerstone of music and dance behaviors. Even when the music is not strictly periodic, humans perceive periodicities and are spontaneously entrained to move on these periodicities. Moreover, the beat can be grouped or subdivided in *meters*, which correspond to harmonics or subharmonics of the beat frequency (as in a waltz, which is a three-beats meter) (see also *Glossary*).

Getting entrained to music is an extremely common human activity, shared by humans of all cultures. It is a highly complex activity, which involves auditory (and also visual, proprioceptive and vestibular) perception, attentional capacities, as well as motor synchronization, performance and coordination. Hence, it is not surprising that a large network of brain structures is involved in entrainment to music, and that there is a growing interest in understanding the functional and neural mechanisms of the entrainment to music.

One of the major goals of this dissertation was to narrow the gap between scientific studies on neural entrainment on the one hand and entrainment to musical rhythms on the other hand. In both, entrainment processes and biases towards periodicity have been described as fundamental functional characteristics. Considering this, we tested whether periodicities induced by musical rhythms could entrain neural activities at frequencies corresponding to these periodicities.

In the present work, we have used the electroencephalogram (EEG), a technique particularly well suited to study a system that changes dynamically over short periods of times. We developed an original EEG approach to capture the processing of beat and meter periodicities. This approach is based on the long-standing observation that when the brain is stimulated periodically, it synchronizes its activity to the inputs and produces periodic output (Lunel & Van der Tweel, 1965 ; Regan, 1966). This neural activity can be captured objectively in the form of a steady-state evoked potential (SS-EP) identified by analyzing the EEG in the frequency domain.

Our experiments show the interest of this approach to study various aspects of beat perception in normal individuals: elicited by mental imagery paced onto periodic sounds (Nozaradan et al., 2011), emerging spontaneously when listening to rhythmic patterns (Nozaradan et al., 2012), elicited by sensorimotor synchronization to the beat (Nozaradan et al., in revision), and finally, elicited by simultaneous auditory and visual beats which were temporally congruent or not (Nozaradan et al., 2012).

Several terms, either from the neural oscillation or the musical rhythm literatures, are recurrent in the present work. For a definition of these terms, as well as a description of some important concepts related to these terms, the reader is referred to the section *Glossary*.

The present work is attached with several media files. These are the stimuli of Studies 1 to 4, and also audio tracks which illustrate some of the musical aspects addressed in the theoretical parts.

I. ENTRAINMENT IN NEURAL SYSTEMS

Part I of the present thesis reviews the evidence supporting the view that our neural system is biased towards periodicity, under certain circumstances at least, and can act in some contexts as multiple coupled oscillators. This question is of particular interest in regard to research on rhythm, pulse and meter perception. Indeed, in order to explain the underlying mechanisms that lead to this ubiquitous human ability, a theoretical model of resonance for pulse and meter (see Section II.2.2.3.3.) has proposed to link the phenomenology of pulse and meter with the concepts of neural oscillation (Large and Kolen, 1994; Large, 2008). The basic idea of this model is that some neural oscillations, possibly dispersed across cortical and subcortical areas and spanning a range of natural frequencies for beat and meter induction in music, entrain to the rhythm of the auditory sequence.

Following a review of the possible tendencies towards periodicity and entrainment in the activity of neurons (Section I.1), we will review the evidence of neural entrainment in the particular case of synchronization to oscillatory inputs (Section I.2). We will then discuss the neurophysiological evidence for entrainment in the auditory system (Section I.2.1), whose stimulation forms can be seen as oscillatory in nature. Finally, we will focus on SS-EPs (Section I.2.2), an electrophysiological method making a specific use of periodic repeated stimulation to tag brain activity and which inspired the present experimental work.

I.1. ENDOGENOUS OSCILLATIONS

There is a large amount of evidence for rhythmic neural activities. To characterize these dynamic patterns, the term “oscillation” was first mentioned by Hans Berger in 1929, to describe cyclical fluctuations of the electrical currents of the human scalp at approximately

10 Hz. This electrical activity was enhanced when participants closed their eyes and constituted the first description of the alpha band, an ongoing neural activity between 8 and 12 Hz typically enhanced when the eyes are closed (see e.g., Klimesch, 1999, for a review). Since this seminal observation, numerous studies have explored the relationship between dynamic patterns recorded with EEG or other techniques, and behavioral states.

The human brain, with its numerous connections between areas, displays low-frequency and fast rhythmic patterns grouped within complex wave-sequences (Steriade, 2006). Some of these oscillations are due to intrinsic neuronal properties, while others arise from the large interconnections of neurons across distant brain areas. From this perspective, the mechanisms underlying oscillatory activities, synchronization across neurons and the emergence of a frequency tuning function within one neuronal population may be interpreted as different aspects of a common phenomenon.

I.1.1. Spontaneous neural oscillations. Two mechanisms at least have been proposed to explain the oscillatory behavior of neuronal discharge: (1) the mutual interconnection between an excitatory neuron and an inhibitory interneuron, or between two inhibitory interneurons, and (2) the pacemaker neuron.

The hypothesis of an oscillatory activity emerging from a mutual interconnection of at least two cells including an inhibitory neuron was proposed for the first time to account for the fast oscillations (between 6 and 10 Hz) observed in the rat hippocampus (Wang and Buzsaki, 1996). Such network models of oscillatory activity were further studied in the context of central pattern generators. In many animal species, functional units of a few cells, located in the spinal cord, have been shown to generate continuous periodic activity responsible for automatic movements such as locomotion in many animal species (Marder and Bucher, 2001). In these network models, two neurons reciprocally inhibit each other. When isolated, these neurons do not fire in repetitive bursts. However, when they are coupled, they produce alternating patterns of activity (Fig. I.1.1.). The transition between activated and inhibited states occurs via various mechanisms. For instance, if the neuron shows spike-frequency adaptation, the active neuron may slow down or stop firing, thus releasing the other neuron from inhibition. Alternatively, the inhibited neuron may escape from inhibition due to its intrinsic membrane properties and, in turn, activate or inhibit the first inhibiting neuron. This postinhibitory rebound has been shown to be crucial for the timing of firing of the central pattern generator unit (Marder and Bucher, 2001; Calabrese, 1998). By extension, similar mechanisms have been described to explain the oscillatory activity of thalamic neurons as well as neurons in the globus pallidus for instance (Bevan et al., 2002), based on the interplay between low-threshold excitatory calcium current and burst of GABA-mediated inhibition.

According to the second kind of mechanism, some neurons are intrinsically rhythmic, and fire either endogenously or in response, for instance, to neuromodulatory substances such as neurotransmitters. Examples of such neurons have been observed in the inferior olive. Neurons of this structure exhibit sustained oscillatory activity that are generally observed between 4 and 10 Hz and are explained by the interplay between various ionic currents and their particular dynamics across the membrane (Bal and McCormick, 1997). Hence, when they receive stimulation, the dynamic of their responses lie within a narrow frequency range that coincides with their natural frequency of resonance, such that the transient response to the transient input takes the form of a transient oscillation. Neurons that are strongly oscillatory can provide important timing inputs for neuronal networks, by driving neurons that are not themselves intrinsically rhythmic (Fig. I.1.1.). However, they are more difficult to entrain or reset, except within a small frequency range.

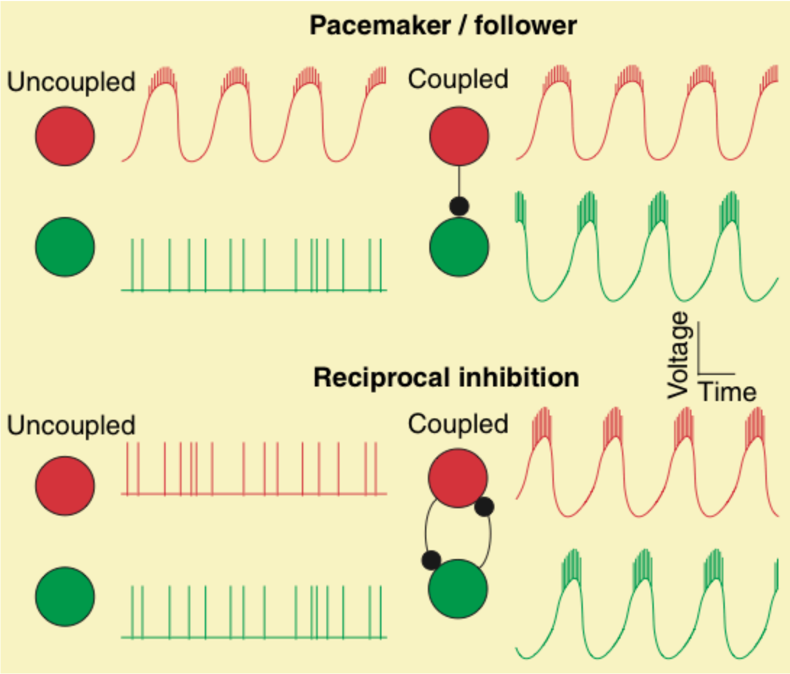


Figure I.1.1. From Marder and Bucher (2001). Upper panel. Rhythmic network based on the coupling between a pacemaker neuron (in red) and a non pacemaker neuron. Bottom panel. Rhythmic network based on the reciprocal inhibition between two non rhythmic neurons.

I.1.2. Frequency tuning function. The frequency range in which sustained oscillatory activities are observed is determined by structural aspects, acting as bandpass filters, at the level of the single neuron and the network.

Low-pass and high-pass filtering is mainly constituted by timing constraints due to the conductance (which can be defined as the ease at which an electric current crosses the membrane) and capacitance (which can be defined as the ability of the neuron to store an electric charge) of the neuronal membrane (Hutcheon and Yarom, 2000). The combination between low-pass and high-pass filtering properties determines the frequency tuning function of individual neurons. If the dynamic activity is limited to a narrow frequency range, it results in an almost periodic activity, as found in pacemaker neurons for instance.

Moreover, this leads to cases where neurons would show a quasi-periodic discharge in response to non-periodic input such as white noise (Joris et al., 2004).

In addition, the frequency tuning function of neural oscillatory behaviors can also result from the physical architecture of neuronal networks and the limited speed of neuronal communication due to axon conduction and synaptic transmission (Buzsaki and Draguhn, 2004). That is, the size of the synchronous group also influences the period of oscillation. Higher frequency oscillations can involve a small neuronal space, whereas very large networks are only able to synchronize to slow oscillations. Hence, the frequency tuning function of the network is determined by both the properties of the individual neuron, and the properties of its interconnections.

I.1.3. Synchronization. It is generally assumed that integration of information requires synchrony, or coincidence, of convergent inputs (Buzsaki and Draguhn, 2004). Synchrony is defined as the simultaneous occurrence of activity in two or more cells (see the Glossary for more details on the concept of synchrony). Oscillation-based synchrony is thought to be the most efficient physical mechanism for temporal coordination (Pikowski et al., 2001). Oscillatory synchronization, as a synonym with entrainment, can be achieved through networks that include pacemaker or inhibitory processes (Fig. I.1.1.), thus emerging from the synaptic connections and their intrinsic properties.

I.1.4. A link between synchronization of oscillatory activities and brain function. Given the diversity of the voltage-dependent channels and the intrinsic properties of the cellular membrane within the whole brain, it is likely that every neuron has a resonance curve and the potential to exhibit oscillatory activity under certain circumstances. Whether resonance and oscillatory synchrony in neurons are simply epiphenomena or whether they are used to integrate and communicate information is still debated (Hutcheon and Yarom, 2000).

However, several mechanisms can be proposed to explain the advantages for neural systems to act as coupled oscillators.

First, oscillatory behaviors in neuronal groups may influence the response chronometry of the oscillating neurons, because their excitability becomes phase dependent (Llinas, 1988, Hutcheon and Yarom, 2000; Engel et al., 2001). The oscillatory fluctuation of the membrane potential of a given neuron creates predictable time windows during which the neuron is more likely to respond to external input. If the input occurs at an inappropriate time according to the excitatory phase, the neuronal response is dampened and/or delayed.

Second, oscillatory synchronization may act as a filter and amplificator of the inputs. The amplification can be explained by the beacon effect. At equal input strength of each upstream neuron, the impact of the inputs ensemble is greater on the target cell when the inputs are synchronous. The filtering of the inputs is achieved based on the excitatory phase of the target neuron, but also based on the frequency of the stimulation. The inputs are thus selected when they fall within the frequency preference of the neuron, according to its intrinsic resonant oscillatory features or the resonant properties of the network to which the neuron is interconnected. This would thus determine the “sampling rate” of the neural network for a given input (see also Section I.2.2.1.).

Third, oscillatory synchronization could serve as a mean to bind cell assemblies. This is based on the assumption that information in the brain is processed, transferred and stored by flexible cell assemblies. These assemblies are defined as neuronal groups that are transiently synchronized (Edelman, 1978; 1989). Indeed, the binding of the neurons may depend on the coupling strength, itself influenced by the distribution of the resonant frequencies of the individual neurons of the group. As long as the frequencies of the coupled oscillators remain similar, synchrony can be sustained even with very weak synaptic links (Buzsaki and

Draguhn, 2004; Engel et al., 2001; Varela et al., 2001). This flexibility based on oscillatory synchronization is hypothesized to play a role in learning. In the rat hippocampus, a structure thought to play a crucial role in memory, brief pulse trains delivered at the peak of the neuronal oscillations induce long-term potentiation, whereas the same train applied out-of-phase weaken the previously strengthened inputs (Csicsvari et al., 2003).

I.2. ENTRAINMENT TO OSCILLATORY INPUTS

In the previous section, we briefly reviewed the neurophysiological bases of the natural propensity of neurons to generate oscillatory activity, and the possible role of this oscillatory behavior in brain function. The present section addresses the question of how neurons behave in contact to external inputs that are themselves oscillatory (auditory stimuli). Finally, we will examine how such repeated stimuli can be used to “tag” neural processes using electrophysiology (steady-state evoked potential approach).

I.2.1. Synchronous oscillation to sound envelope in the auditory system

I.2.1.1. What is sound envelope? Acoustic stimuli contain multiple temporal dimensions. They can be summarized in at least two components, the “fine structure” and the “envelope”, which are usual terms to describe waveforms in physics. In acoustic, the fine structure is determined by the fast pressure variations corresponding to the spectral content of the sound. The processing of fine structure is involved in pitch perception, which can be defined as the perceptual phenomenon of sounds organized within a scale from low to high tones (Schnupp et al., 2010). The fine structure is itself modulated in amplitude, and the dynamic of this amplitude modulation constitutes the sound envelope. In humans, amplitude modulations produce various hearing sensations depending on the modulation frequency. Rhythms and fluttering, as well as most amplitude modulation frequencies found in ordinary speech for instance, correspond to envelope frequencies up to 20 Hz whereas roughness and pitch correspond to amplitude modulation frequencies above 20 Hz. As we will see, it has been proposed that this perceptual boundary may be related to a change in the coding form of these sound inputs (Eggermont, 2001).

Whether envelope processing is embedded in the auditory system is an important question, regarding the ecological prominence of low frequency sound modulations. Indeed, low-frequency amplitude modulations (beneath 20 Hz) are prominent in acoustic natural environments (Nelken, 1999), and contain essential information, particularly for vocalization. For example, they have been shown to be necessary and almost sufficient for speech intelligibility (Shannon, 1995). This was evidenced by comparing speech intelligibility when manipulating the speech signal either by blurring the frequency content of the signal corresponding to the fine structure while keeping intact the frequency content up to 20 Hz, or by doing the inverse manipulation. Regarding the topic of the present thesis, the dynamics of amplitude modulation of the sound specifically beneath 5 Hz, constitutes a crucial, although not unique, cue for beat and meter perception. Hence, in the four experiments reported in the present work, the auditory stimuli have been designed such as to induce the beat and meter exclusively based on the dynamics of amplitude modulation of a pure tone.

I.2.1.2. Frequency decomposition of the sound. It is generally assumed that sounds are processed in the nervous system by cells responding to a specific frequency band. Already within the cochlea, these groups of cells are functional units defined by the strong correlation in the firing of individual cells in response to a preferred frequency band of stimulation. For this reason, these neuronal groups are often compared to mechanical filters, or a filterbank organized in a tonotopic map. This array of band-pass filters is assumed to decompose the sound input into several frequency bands according to the banwidth of frequency range to which each neuronal group responds preferentially (i.e., according to their frequency tuning function bandwidth). However, while a periodotopic organization of the cell groups encoding sound envelope has been widely observed in subcortical structures, whether such functional arrangement exists at the level of the cortex is still debated. For example, the bandwidth of preferred frequency range appears to vary considerably across cortical neurons in the auditory cortex (Joris et al., 2004).

Importantly, groups of cells manifest nonlinear behaviors in the decomposition of the sound input that are crucial for sound envelope processing. Particularly, the cochlea exhibits patterns of responses similar to a demodulation function, which can be observed by comparing the spectrum of responses recorded in the cochlea to the spectrum of the eliciting stimulus, as represented in Figure I.2.1.2.

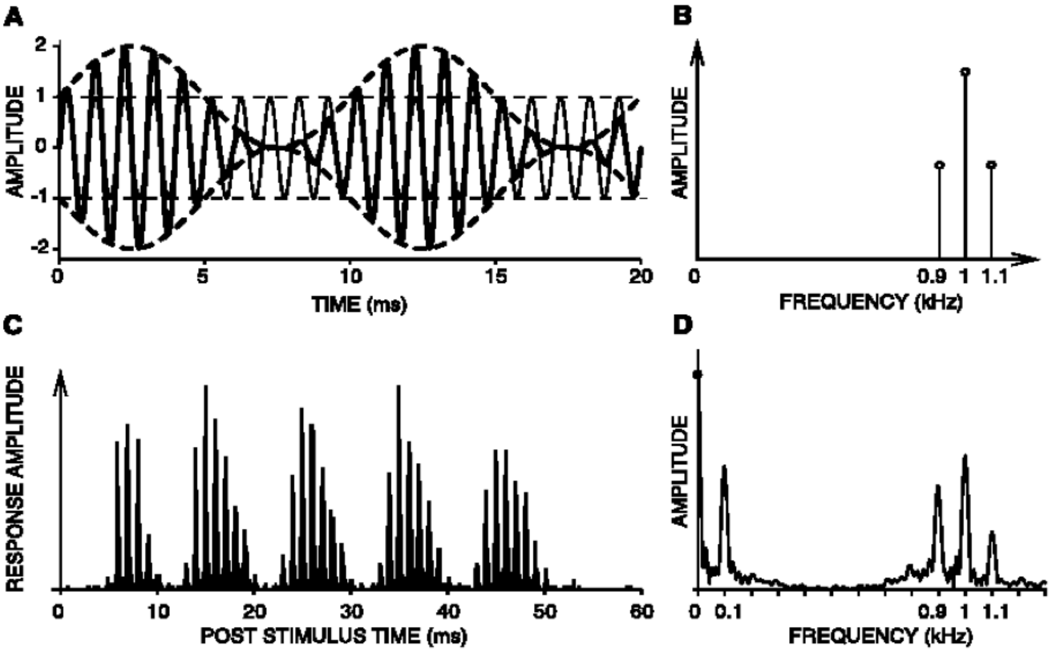


Figure I.2.1.2. From Joris et al. (2004). A. Superimposed waveforms of an unmodulated 1000 Hz tone (thin line) and the same tone sinusoidally amplitude modulated (thick line) at 100% depth with a frequency of 100 Hz. The dashed lines correspond to the sound envelope. B. Spectrum of the sound in A. C. Average nerve fiber response (poststimulus time histogram). D. Spectrum of the response. The activity at the carrier frequency and at sidebands indicates that there is a phase-locking to the fine structure of the sound. However, an additional component at the modulation frequency (0.1 kHz) emerges, compared to the sound spectrum.

I.2.1.3. Temporal coding versus rate coding. The first neural representation of the sound, and the sound envelope, is in the cochlea, as the first relay of the ascending auditory pathway (Fig. I.2.1.3.). Subsequently, this representation is conveyed by multiple, parallel pathways constituted by the auditory nerve to the cochlear nucleus. These parallel pathways, hypothesized to convey a detailed representation of the stimulus, diverge or converge further toward the inferior colliculus and finally the cortex (Fig. I.2.1.3.).

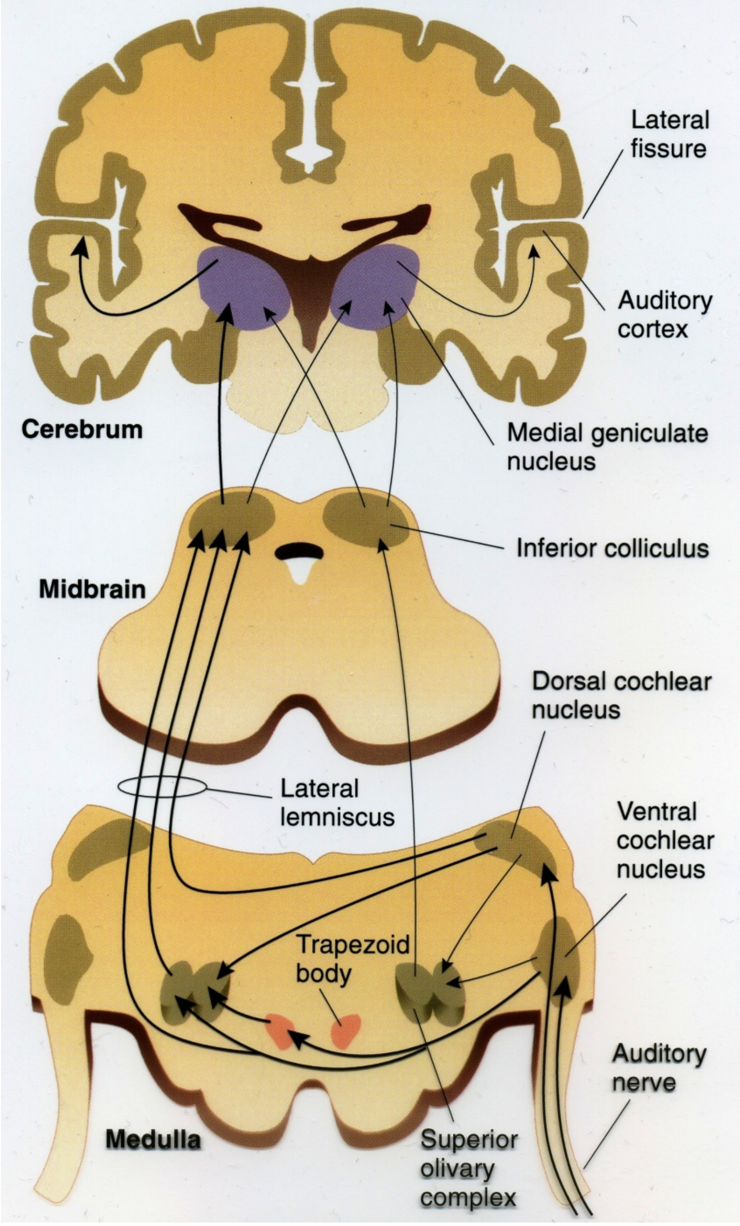


Figure I.2.1.3. Auditory pathway, from cochlea to cortex (copyright Allyn and Bacon, 2005).

Already at the level of the cochlea, auditory cells respond to sound envelope, as well as to the fine structure of sounds, in the form of firing patterns phase-locked to the temporal structure of the input. This temporal coding of information, referred to as *envelope-locking* in the context of sound envelope processing, contrasts with another form of encoding found extensively in the nervous system, based on the *average firing rate*. Referred to as the average number of spikes per unit time, the rate coding assumes that the firing rate increases with increasing stimulus intensity or change in one stimulus feature.

Along the ascending auditory pathway, the sound envelope information is transmitted from the cochlea to the brainstem, through a principle of synchronization, or phase-locking, via the auditory nerve (Fig. I.2.1.3.). The temporal coding of sound envelope information is widely maintained across the subcortical structures, while expressed differently within each structure regarding the bandwidth of frequency tuning function of the cells groups, the possible gain modulations, etc. The temporal coding is then transformed, at least in part, in an average rate coding in structures as the superior olivary complex of the brainstem and in higher relays of the auditory pathway. On the whole, responses to amplitude modulation become more phasic than sustained at these stages. Sustained responses are transformed into on/off patterns of response that are hypothesized to serve several functions such as improving the signal-to-noise ratio under naturalistic listening or detecting changes in the stimulus content (Eggermont, 2001).

The transformation from a temporal coding to an average rate coding is assumed to occur due to the intrinsic and network properties of neurons that are no longer able to produce sustained frequency following. From this loss of synchrony with the stimulus dynamic, the transition to an average rate coding is performed for instance by a process of convergence of inputs from neurons differing in their response dynamics. The target neuron produces an output only if the timing and nature of the diverse combined inputs are such that the target neuron is depolarized strongly and quickly enough to reach threshold (Eggermont, 2001). This process thus requires coincident input from a relatively large number of input neurons. The diversity of the dynamic of the inputs makes this coincidence occurring transiently.

Importantly, numerous studies have brought evidence for a loss of sound envelope locking for modulation frequencies above 100 Hz in cortical areas (Eggermont, 2001). Indeed, as auditory information reaches progressively higher levels of processing, higher frequencies are progressively more represented using rate coding. However, frequencies below 100 Hz are still represented by means of temporal coding. The substantial lowpass filtering of the temporal coding of sound envelope is due, at least in part, to the fact that from the first to the last relays of the ascending auditory pathway, the neurons respond by temporal coding with a certain amount of temporal jitter. This jitter might be transmitted, and possibly amplified along the auditory pathway, making the transduction of high frequencies with sufficient temporal accuracy very limited.

The fact that, unlike the processing of tone, envelope processing at low frequencies remains largely represented as temporal coding at the level of the cortex constitutes an important argument in favor of the SS-EP approach to study rhythm processing. Indeed, the SS-EP approach might have been less adequate if the frequency range for rhythm perception was rate-encoded at the level of the cortex.

Nevertheless, as we will see, the perception of musical rhythm and meter do not only rely on the information conveyed by amplitude modulation, but also exploits harmonic structure, timbre modulations or even endogenous imagery of a temporal structure that can be imposed onto the sound (see Section II.2.2.4.1. for more details on the generation of metrical accents). In theory, one could hypothesize that these numerous features, processed by independent neurons, would be integrated within a unified representation corresponding to the percept of beat and meter. Such across-feature interactions may be hypothesized to emerge when the sound envelope and the other features set up widespread synchrony at low frequencies across cortical neurons, thus adjusting to each other by synchrony of the periodic modulation of their responsiveness (Eggermont, 2001).

I.2.2. The steady-state evoked potential approach

The previous sections have briefly reviewed the evidence corroborating the view that neurons could show a natural propensity to generate oscillatory activity, and to synchronize to external inputs that are themselves oscillatory (auditory stimuli). Here, we examine how stimuli made oscillatory for experimental purpose can be used to “tag” neural processes using electrophysiology (steady-state evoked potential approach).

A large number of investigators have used non-invasive EEG techniques to study how the human brain processes external or endogenous inputs. The EEG signals recorded on the scalp lack spatial resolution. Indeed, the potentials measured at a given scalp position are not systematically determined by the activity of the cortical region located immediately underneath the electrode (Nunez and Srinivasan, 2005). However, it offers the advantage of measuring neural activity at the millisecond time scale.

The majority of studies have relied on the recording of event-related brain potentials (ERPs), i.e., changes in the ongoing electrical brain activity time-locked to a transient event, like the sudden onset of a sensory stimulus. In 1966, Regan introduced the approach of “steady-state visual evoked potentials” (SSVEP) as an alternative to characterize stimulus-evoked activity in the ongoing EEG. Unlike conventional transient ERPs, which Regan described as “*the response to a kick in the system*”, SSVEPs reflect a sustained cortical response induced by the long-lasting periodic repetition of a feature in the input stimulation, described by Regan as “*the response to a gentle shake of the system at a fixed repetition rate*“ (Regan, 1989). Regan named this response “steady-state" because it remains virtually constant in phase over time. Such responses have been described not only in the visual domain, but also in the auditory (Galambos et al., 1981), and somatosensory domains (Galambos, 1982). They will be referred in the present work as SS-EPs.

One of the advantages of the SS-EP approach is the prior knowledge of the frequency at which the neural response to the repeated stimulation should appear in the EEG, thus making the technique more objective. Moreover, since the response is expected to be concentrated within a very narrow frequency band, the technique has a very high signal-to-noise ratio (Regan, 1989).

I.2.2.1. Nature of the steady-state evoked potentials. SS-EPs are thought to result from an entrainment or resonance of a population of neurons responding to the stimulus at the frequency of stimulation (Vialatte et al., 2010). An alternative to this view is that SS-EPs result from the linear superposition of independent transient responses elicited by the fast repetition of the stimulus (Regan, 1989; Capilla et al., 2011). Thus, how these activities emerge within the human EEG and their relationship with transient ERP and ongoing oscillatory activities remains a matter of debate.

Irrespective of the outcome of this debate, the SS-EP phenomenon appears to be caused by an increase in the neuronal response synchronization as a result of the presentation of a repetitive external input whose temporal presentation rate is close to the temporal activation cycle of a neuronal group. When a given input is repeated at fixed time intervals, it may force the group to respond at a certain rate, biasing the propagation of excitatory and inhibitory postsynaptic potentials. If the inter-stimulus interval coincides with the neuronal activation cycle, then a higher amount of neurons are available to respond to the input and can synchronize their response properly. This would result in an amplification of the output along the targeted neural network, causing a noticeable increase in the amplitude of the signal registered on the scalp (Buszaki, 2006).

I.2.2.2. Frequency tagging. Another advantage of the SS-EP approach allows tagging responses to multiple inputs, based on their respective frequencies, and these inputs can even overlap each other spatiotemporally. Indeed, following Regan and Heron (1969), several studies have shown that different stimulation frequencies can be used to “tag” different inputs presented concurrently and, thereby, isolate the neural activity related specifically to the processing of each stream of stimulation (e.g., Morgan et al., 1996; Toffanin et al., 2009; Giani et al., 2012). For example, simultaneously presenting one stimulus modulated at frequency F1 and another stimulus modulated at frequency F2 elicits two distinct peaks in the EEG spectrum, at frequencies F1 and F2 (e.g., Chen et al., 2003). Based on this principle, the frequency-tagging method offers great advantage in studying for instance the processing of multisensory inputs, by disentangling the processing of each sensory input based on their respective frequencies of response (see Study 4 of the present thesis). Moreover, as movement-related SS-EPs can also be elicited by periodic hand movements (Gerloff et al., 1998), the frequency-tagging method may be suitable to study sensorimotor synchronization to a periodic input (see Study 3).

Importantly, when stimulating with two or more frequencies, peaks of activities can also emerge in the EEG spectrum at frequencies different from the frequencies of stimulation and their harmonics. These additional responses can appear at frequencies corresponding to the sum or difference of the stimulation frequencies (and/or their harmonics) (see also the *Glossary*). These *crossmodulation products* are hypothesized to result from the non-linear convergence or integration of the two input signals (Giani et al., 2012; Regan, 1989; Sutoyo and Srinivasan, 2009). The reader is referred to the Glossary for a more detailed description of these concepts. The concept of crossmodulation is also addressed in Study 3 (Section II.2.3.4).

I.2.2.3. Methodological considerations. Several parameters of the periodic stimulation can be varied to test the response of the neural system experimentally.

I.2.2.3.1. Modulation waveforms. A large range of modulation waveforms, from pure sinusoids to square waves, can be used to elicit SS-EPs. The consistency of the neural response may be hypothesized to depend directly on the stimulation waveform. Indeed, if the neural system responds to the stimulation train based on the detection of periodic contrast changes in the stimulation, the neural assemblies would not respond in the case where the stimulus modulation is too slow to be perceived as a change in the input. In contrast, the abrupt contrast generated by a square waveform of stimulation should elicit more consistent neural responses at every stimulus with less jitter along time, compared to a sinusoidal waveform of stimulation.

However, a pure sinusoid waveform of stimulation could offer several advantages. First, in line with the view that SS-EPs constitute neural responses whose nature is distinct from transient ERPs, it can be hypothesized that a square wave train of stimulation elicits a larger amount of transient responses than a sinusoidal stimulation. This has been shown by comparing for instance the neural responses to a train of auditory clicks with a train of sinusoidal amplitude modulations of a sound (Draganova et al., 2002). Second, when stimulating with a pure sinusoid, the response is expected to result theoretically in a unique frequency in the EEG spectrum, corresponding to the frequency of the stimulation (Victor and Zemon, 1984). In reality, it is not systematically the case, as distinct frequency components at higher harmonics appear concomitantly in the EEG spectrum. Therefore, the appearance of harmonics in the EEG spectrum can be considered as the product of non-linear brain processes (Regan, 1989). When obtained from scalp recordings, these non-linear responses could reveal either that the neural response was not sinusoidal, or that neural populations actually responded to the stimulation at harmonic frequencies.

I.2.2.3.2. Frequency of the modulation. It has been shown that for a given stimulus modulation within a given input modality, there is a preferred frequency range of stimulation, which can be assimilated to a frequency tuning function, or resonance curve. At these preferred frequencies, the neural network is hypothesized to respond with maximum amplitude to a given stimulus. The evidence of such preferred frequency allows understanding how the SS-EP approach may be useful to obtain the frequency tuning function in response to a given stimulus. The frequency tuning function would then give indications on the *sampling rate* of the responding neural network, and not only on the latency to process a single input, as does the majority of the ERP paradigms.

However, capturing the frequency tuning function in response to a given stimulation using EEG presents some limitations. First, the amplitude of the background noise in the EEG spectrum is not equivalent across different frequency ranges (specifically, it presents a 1/f distribution; e.g., Klimesch, 1999). This may thus bias the comparison of the amplitude of SS-EPs elicited at different frequency ranges. Second, when recorded at the level of the scalp, the obtained potentials are assumed to reflect a sum of brain processes, originating for example from “low-level” cortical areas and areas of higher level of processing, both being entrained by the stimulation train at a similar frequency. Thereby, it is possible that if a given neural population is unable to respond at a too fast stimulation rate, another population takes over from it, giving the illusion of a broader frequency tuning function when registered at the surface.

In sum, the exact physiological mechanisms that explain the human brain preference for certain stimulation frequencies to process a given stimulus feature are not yet fully understood. Nevertheless, this aspect is crucial in the SS-EP approach, as it may condition directly the emergence of a response.

I.2.2.3.3. Frequency domain analysis. Based on the prior knowledge of the expected frequency of the neural response, SS-EPs are usually measured in the frequency domain, using discrete Fourier transform for instance. Mathematically, the Fourier analysis consists in the decomposition of a possibly complex waveform into a set of sinusoidal basis functions. For example, the discrete Fourier transform calculates amplitude and phase spectra by projecting input signals onto pure sine waves, which are spaced linearly along the frequency axis. Thus, the discrete Fourier transform calculates exactly one Fourier component for each harmonic of some chosen lowest fundamental frequency.

Analyzing the EEG signal in the frequency domain using a Fourier transform is based on the assumption that the signal remains steady in phase and amplitude along the entire epoch on which the transform is computed. It also requires a sufficient length of the epoch, to obtain a sufficiently high frequency resolution. Logically, the length of the epoch is determined as a function of the frequency of the expected SS-EP. In other words, when stimulating with slow frequencies, longer epochs are required to obtain a sufficient number of stimulation cycles.

Part I briefly discussed the evidence that the nervous system presents properties that could explain a bias towards periodic behaviors (either detection of periodic signals or production of periodic activity), and that these behaviors are used for brain function under certain circumstances. Furthermore, we introduced the SS-EP approach, an electrophysiological method that takes advantage of this property of the nervous system to entrain to periodic inputs, to “tag” neural responses based on the prior knowledge of this frequency. Building on these concepts, we hypothesize that musical beat and meter, as periodic percepts in nature, would be a nice ecological framework to explore these neural properties. Interestingly, as we will see in following sections, several characteristics of periodic entrainment observed at the neural level (frequency tuning function, nonlinearities, frequency locking) can also be observed at the interindividual level, through an artistic expression: music. The following sections focus on the evidence regarding the perception of musical rhythms and entrainment to the beat and meter in music, followed by the empirical contributions of the present thesis. Hence, we aimed at making a bridge between these two literatures, dealing with neural entrainment on the one hand and with entrainment to musical rhythms on the other hand, by testing in the present thesis whether the neural correlate of beat and meter could be captured in the form of SS-EPs.

II. ENTRAINMENT TO MUSICAL RHYTHMS

II.1. BEAT IN MUSIC: A UNIVERSAL HUMAN ABILITY?

The perception of musical features has already been explored extensively in music theory, anthropology and psychology. In contrast, it is a relatively new field in neuroscience. Yet, studying the biological foundations underlying the perception and production of music could provide insights on numerous fundamental brain mechanisms (McAdams, 1989). This assumption is strengthened by the fact that entrainment to the beat seems ubiquitous in human musical cultures. However, whether the ubiquity of musical features is a consequence of history and geography, i.e., a consequence of the migration of populations having spread out these features, or whether they have an intrinsic biological foundation remains an open question. Therefore, rather than searching for universals, an approach that could overlook informative differences, researchers have preferred analyzing diversity (Stevens, 2012). In the following paragraphs, we will review (1) whether entrainment to musical rhythms is found in all musical styles, (2) the relationship between musical rhythms and speech, (3) whether musical rhythms differ across cultures, (4) whether the perception of musical rhythms is present in early developmental stages, and (5) whether the perception of musical rhythms is specific to the human species.

II.1.1. Are beat and meter induced in all musical styles?

Music always escapes definitions, probably because there are as many musical forms as musicians and listeners. It can be considered as a communication and signaling process such as language, but remains above all an artistic form of expression. This implies that humans possess the capacity to “decontextualize” the form of this expression and generate it independently of all contexts (Arom, 2001). Music may thus be viewed as self-referential, in contrast to language which is shaped by signifier-signified constraints (Arom, 2001).

In line with these considerations, it is perfectly conceivable to find music that does not contain beat and meter, either because the composers did not write the music by means of a periodic reference frame, and/or, because we do not perceive any beat when listening to these musical pieces. Music can thus be categorized as either measured (or metric) or unmeasured (or non-metric) forms (cf. Audio tracks II.1.1a and b[1], for examples of non-metric and metric forms respectively) (Arom, 1991). The first forms contain a beat and meter structure, in contrast to the second ones. As a proof of concept, one can ask individuals to move on such musical pieces. In unmeasured music, the observed movements are not periodic, and often, these musical pieces do not entrain individuals to move spontaneously. Examples of music that do not contain an isometric structure of time are found in the cantus planus from the medieval Gregorian tradition (Arom, 1991) (Audio track II.1.1c[2]), or in the melodic recitation of poems from the classic Persian tradition, accompanied by instrumental parts following the recitative phrases of the vocal parts (Nelson, 1985).

Hence, beat and meter do not constitute an obligatory ingredient of music, although this periodic reference frame is widely induced across musical genres and cultures. Actually, its use is likely to be related to the goal of musical expression. When music aims at conveying coordination across individuals, beat and meter is a powerful mean to improve it, as we will see in further sections.

---

[1] Audio track II.1.1a and b. These instrumental pieces from the musical classic Persian tradition illustrates nicely the contrast between measured and unmeasured music. In Audio Track II.1.1a, the musical phrases contains a rich rhythmic structure, but does not contain meter. In contrast, Audio Track II.1.1b, which follows the first one in the composition, is metric (with an important role of drum and of metric modes to induce it).

[2] Audio track II.1.1c. Pater Noster, from the monodic Gregorian tradition, typically driven by verbal rhythm.

II.1.2. Language and music rhythms

Music is not the unique form of expression involving auditory-motor coordination and sound production: speech is another. Given the similarities that music perception and production appear to share with language, researchers have explored the similarity between speech and musical rhythms. Do speech and music rhythms influence each other? Why are some temporal features such as isochrony particularly developed in music but not in speech?

II.1.2.1. Rhythm in speech. There is increasing evidence suggesting that the production and perception of language and music share overlapping mechanisms (Patel, 2008). Regarding rhythmic aspects, speech is constituted, such as music, by a large variety of rhythmic modulations of sound envelope. The temporal contour of speech is a necessary, and almost a sufficient, cue for speech intelligibility (see also Section I.2.1.1) (Shannon et al., 1995). These rhythms are even used as a basis to classify languages. English, for instance, is classified as "stress-timed", because it contains highly varied syllable durations due to flexible durations of the vowels. In contrast, other languages such as French are "syllable-timed", because they exhibit less variability in the syllable duration (Grabe and Low, 2002).

Interestingly, it has been demonstrated that these varying patterns of vowel duration in speech are closely paralleled by patterns of note duration found in music within one culture (Patel and Daniele, 2003). According to Patel and Daniele (2003), the music repertoire that has been composed by musicians in a stress-timed language context (e.g., English repertoire) exhibits greater sequences of highly varied note durations, in contrast to music that has been written by composers from a syllable-timed language, which present less duration variability in the notes succession. This corroborates the idea that speech and music, even in purely instrumental genres, strongly influence each other (Patel and Daniele, 2003; Hannon, 2009).

There are other examples of the influence of speech rhythm on the perception of rhythmic input in non-speech contexts. For instance, the tendency to perceive inputs as grouped along time is not identical across cultures. Even in non-speech contexts, these grouping processes seem to be influenced by the syntactic or stress differences found in speech across languages. Indeed, it has been shown that the preference for grouping in a succession of long and short notes (Fig. II.1.1.1.) differs between individuals that speak a language placing the functors before the content word (such as most European languages), and individuals that speak a language placing the functor after the content word (such as Japanese or Korean) (Iversen et al., 2008). The preference for grouping non-speech auditory stimuli in "short-long" groups is prevalent in English speakers: they tend to perceive the long event within a sequence as an ending. In contrast, Japanese speakers prefer "long-short" grouping, thus mirroring the grouping processes found in speech due to syntactic rules (Fig. II.1.1.1.). This observation suggests that grouping acoustic inputs (and by extension music) could be influenced by the exposure to speech along life (Iversen et al., 2008).

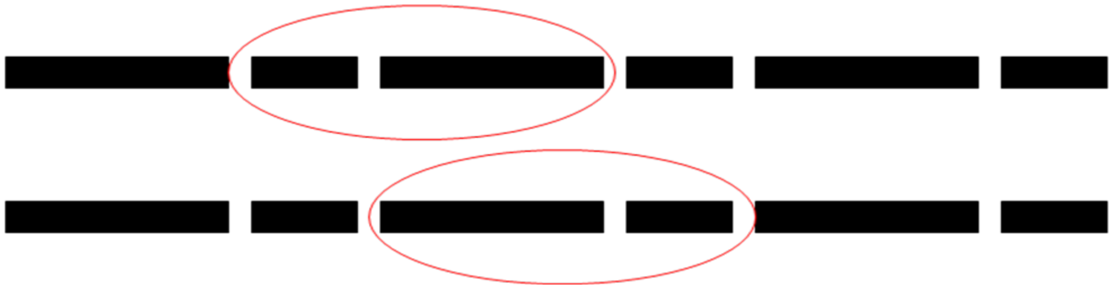


Figure II.1.1.1. From Iversen et al. (2008). Short-long sequences of sounds. First line: grouping found in English speakers. Second line: grouping found in Japanese speakers.

II.1.2.2. Music mimicking speech rhythms and *vice versa*. Contrasting with most musical forms, ordinary speech is not driven by periodicities, neither in its acoustic content nor in the perception of this content (Patel, 2006; 2008). Generalized across cultures, this fact is usually considered as a fundamental dissimilarity between music and language.

One explanation of this difference between speech and music may originate from their differing function (Knight and Cross, 2012). This is illustrated by the fact that speech exhibits various tendencies to isochrony, depending on the goal of the speaker. For instance, oratory speech, typically occurring in an asymmetric conformation between one orator and a crowd, manifests a high degree of isochronous regularity compared to other speech forms such as teaching or conversation (Knight and Cross, 2012). This increased periodicity has been hypothesized to serve persuasion and entrainment of the crowd to a common way of thinking, beyond the simple information transmission necessity (Knight and Cross, 2012).

In a similar vein, music also induces isochrony at various degrees depending on the goal of the producer. Indeed, when musicians aim at making people move spontaneously, as in popular dance music, isochronous pulses are induced using unambiguous acoustic cues presented at particular rates (e.g., amplitude modulation at 120 beat per minute). In contrast, music is often perceived as unmeasured when the goal is narrative. In this case, the rhythms are clearly inspired by the temporal contour of speech, as in the *recitativo* of the Italian opera tradition from the XVIth century (cf. Audio track II.1.2.2[3]).

This observation highlights the fact that, depending on the goals, rhythms in speech tend to mimic those of musical forms, and *vice versa*.

II.1.3. Cultural differences

To explore the biological foundations of beat and meter properly, it is important to be aware of the diversity encountered across cultures regarding the rhythmic material and the

---

[3] Audio track II.1.2.2. Recitativo from Orfeo, Monteverdi. The verbal rhythm predominates on meter, and the instrumental part contributes to underline the narrative content.

metrical forms. Since most of the empirical research on musical rhythm has been performed on Western individuals, the literature concerning beat and meter is probably biased. This section briefly reviews the cross-cultural studies having addressed this issue.

II.1.3.1. The Groove as a cultural specificity. As one could expect, rhythm has not been similarly developed across musical cultures (Pressing, 2002). Some traditions, such as black Atlantic music (i.e., the musical traditions originating from West Africa and their evolution across West African diasporas), have given to rhythmic aspects a prominent importance in their musical behaviors (Pressing, 2002). Particularly, the black Atlantic music has developed a strong culture of *groove* in music, which refers to the urge to move in contact with music (Pressing, 2002; Iyer, 2002). The various musical features leading to groove (as described in Section II.2.2.4.2.) are often found in funk, soul, hiphop, triphop, drum'n bass, house or jazz, i.e., music genres predominantly originating from the black Atlantic tradition (Witek et al., 2009).

Rather than directing tensions and expectations towards the musical form at a large temporal scale, as does most classical music for instance (which could be more pertinently referred to Western European written music), a groove mode immerses the listener into circular attention, generated by the repetition and familiarity of rhythmic patterns. However, repetitive groove-based music is not experienced as "static", but rather as moving in a circular manner. Repetition in this context should be viewed as promoting reinforcement and entrainment to move, rather than habituation (Witek et al., 2009).

II.1.3.2. Meter, binary bias and integer ratios across cultures. When listening to music, we frequently perceive more than one periodicity. These various periodicities are usually derived from several musical aspects (e.g., sounds duration, loudness, pitch, or harmonic changes; see also Section II.2.2.4.1.). Moreover, these periodicities are often hierarchically organized. In this hierarchy, the beat, also called *tactus* in music, corresponds to the most salient periodicity. The interactions between the distinct periodicities can be described in term of relative frequency and relative phase of one periodicity compared to the others (Lerdahl and Jackendoff, 1983). In music theory, the interactions between the different periodicities result in a percept of periodically alternating strong and weak beats, corresponding to the general definition of meter (Lerdahl and Jackendoff, 1983) (Fig. II.1.3.2a).

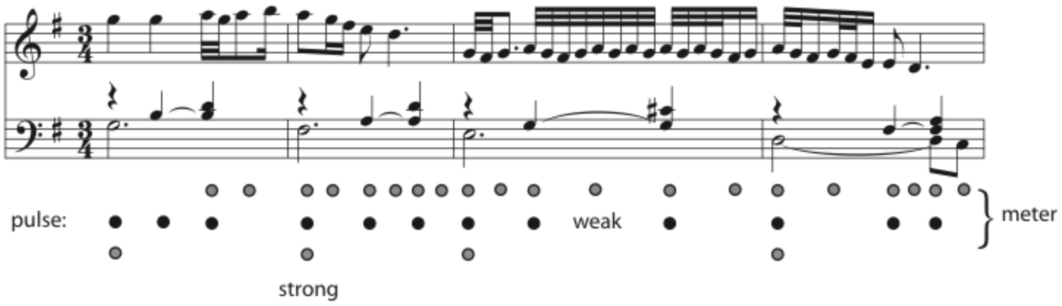


Figure II.1.3.2a. From Large (2008). Notation and music theoretic metrical structure for the first four bars of the Goldberg Variations *Aria (JS Bach).*

Most of the Western tonal music have either a duple or triple meter, i.e., the periodicities perceived at frequencies faster than the beat are usually either the second or third harmonic of the beat frequency (a 2:1 or 3:1 frequency relationship), and the periodicities perceived at frequencies slower than the beat usually correspond to either the second or third subharmonic of the beat frequency (a 1:2 or 1:3 frequency relationship). The phases of the different metric levels are aligned to each other. Hence, in Western tonal music, the different metric levels have integer ratios across each other (Fig. II.1.3.2a).

Simple integer ratios, based on a binary structure (i.e., 1:2, 1:4, 2:1 etc.), have been hypothesized to bias sensory perception at an early stage of processing. This hypothesis is based on the dominance of binary metric structure in Western music, mostly in popular styles (Pressing, 2002). Accordingly, it has been shown that listeners are better at discriminating, categorizing and reproducing rhythms containing 1:2 ratios as compared to 1:3 ratios (Drake, 1993).

The biological bases of this bias in favor of a binary structure remain unclear. It has been proposed that this bias for binary structure was the result of motor production constraints (e.g., bipedalism and left-right symmetry in locomotion) having shaped in turn the perceptual system. Alternatively, some authors hypothesized that it could be related to a bias in the way the auditory system encodes temporal structure at an early level of processing (Pablos Partin et al., 2007). Pablos Martin et al. (2007) explored this question by studying this bias for binary compared to non-binary ratios using EEG and mismatch negativity responses. Recorded on the scalp, the mismatch negativity potential is a negative deflection in the EEG signal appearing at a latency of around 150 ms when a stimulus deviates unexpectedly from a preceding train of auditory stimuli (May and Tiitinen, 2010). A mismatch negativity potential can be elicited by a deviant stimulus even when participants focus their attention on another modality, stimulus or task (Näätänen et al., 2007), thus suggesting that the mismatch negativity reflects neural processes that are independent of the focus of attention. Moreover, it is often considered that the mismatch negativity is related to the fact that the auditory system continuously extracts regularities from the environment and, thereby, creates expectations for the incoming sensory inputs. Finally, the amplitude and latency of the mismatch negativity depends on the magnitude of the violation (Näätänen et al., 2007), thus suggesting that it is related to the intrinsic saliency of the deviant event and, therefore, to the strength of the expectations induced by the preceding events. Because it is related to regularity and expectation, the mismatch negativity potential had been proposed as a tool to examine metrical expectation and its pre-attentive nature (Honing, 2012). In Pablos Martin et al. (2007) study, the train of stimuli consisted in a train of periodic tones, in which the deviant stimuli were obtained by reducing the inter-tone interval according to a percentage either in agreement with a binary interval ratio (i.e., corresponding to the half of the standard inter-tone interval), or out of this binary interval ratio (Fig. II.1.3.2b). When comparing the mismatch negativity potentials elicited by these deviants, the authors observed a significantly reduced latency of the mismatch negativity potentials in the binary interval ratios condition, thus suggesting a privileged perceptual status for binary rhythmical intervals already at a preattentive level (Fig. II.1.3.2c).

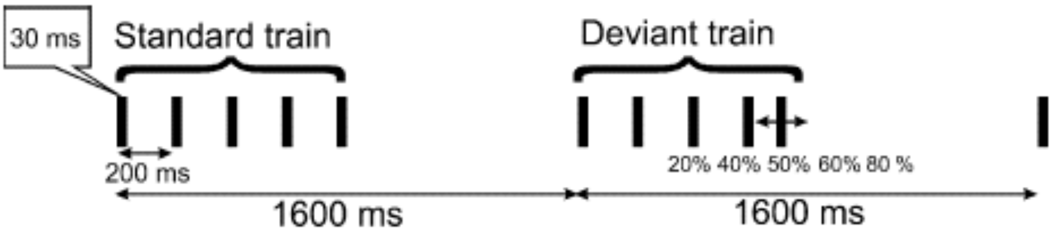


Fig. II.1.3.2b. From Pablos Martin et al. (2007). Scaled representation of the standard and deviant trains. Each tone lasted 30 ms, whereas the inter-tone intervals lasted 200 ms. Deviant intervals were obtained by reducing the inter-tone interval according to the percentages indicated in the Figure.

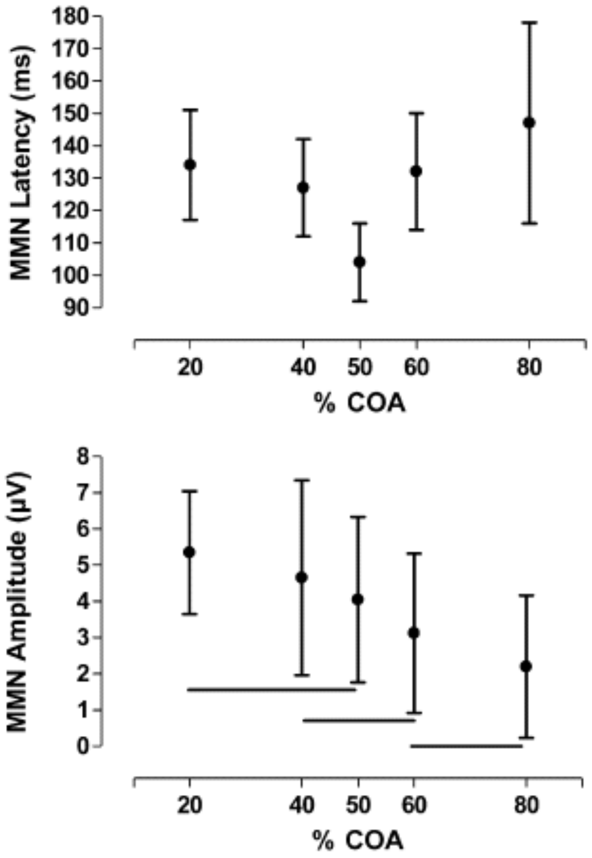


Fig. II.1.3.2c. From Pablos Martin et al. (2007). Mismatch negativity potential latency and amplitude values as a function of rhythmical contrast values (%COA). While the amplitude of the response decreased as a function of the reduction of the contrast value, the latency-contrast value function was marked by a clear reduction at the 50% contrast in an otherwise flat function.

The bias for binary structure is also illustrated by the *subjective rhythmization* phenomenon (Bolton, 1894; Brochard et al., 2003; Potter et al., 2009). Indeed, when listening to an equitone isochronous sequence, i.e., a succession of identical tones occurring at a regular pace, some tones are perceived as more salient (louder, longer, or both) than others (Bolton, 1894). This might explain for instance why people usually perceive a "ticktock" rather than a "tick tick" when they hear a clock. However, no physical characteristic of the sound accounts for the difference perceived. Several researchers have studied this phenomenon, thought to reveal an intrinsic bias of the auditory system for binary meter. For instance, Povel and

Okkerman (1981) studied the strength of perceived accents in equitone sequences of two alternating time intervals. They found that to make the tones perceptually identical the physical intensity of every other tone had to be increased.

Subjective rhythmization was also investigated using EEG, to avoid possible bias due to explicit responses given by the subjects. Brochard and colleagues (2003) presented participants with sequences of equitone sounds. They recorded with EEG the electrophysiological responses elicited while disrupting listeners’ expectancies in different positions of auditory equitone sequences. The disruption elicited a P300, i.e., a positive deflection of the EEG signal elicited about 300 ms after the occurrence of a deviant and thought to reflect attention and cognitive processing of the disruption in contrast with the preceding repeated stimulus. Significant differences in amplitude were observed between the P300 elicited on odd-numbered compared to even-numbered positions, suggesting that a default binary metric structure of cognitive, attention-dependent origin was perceived (Fig. II.1.3.2d).

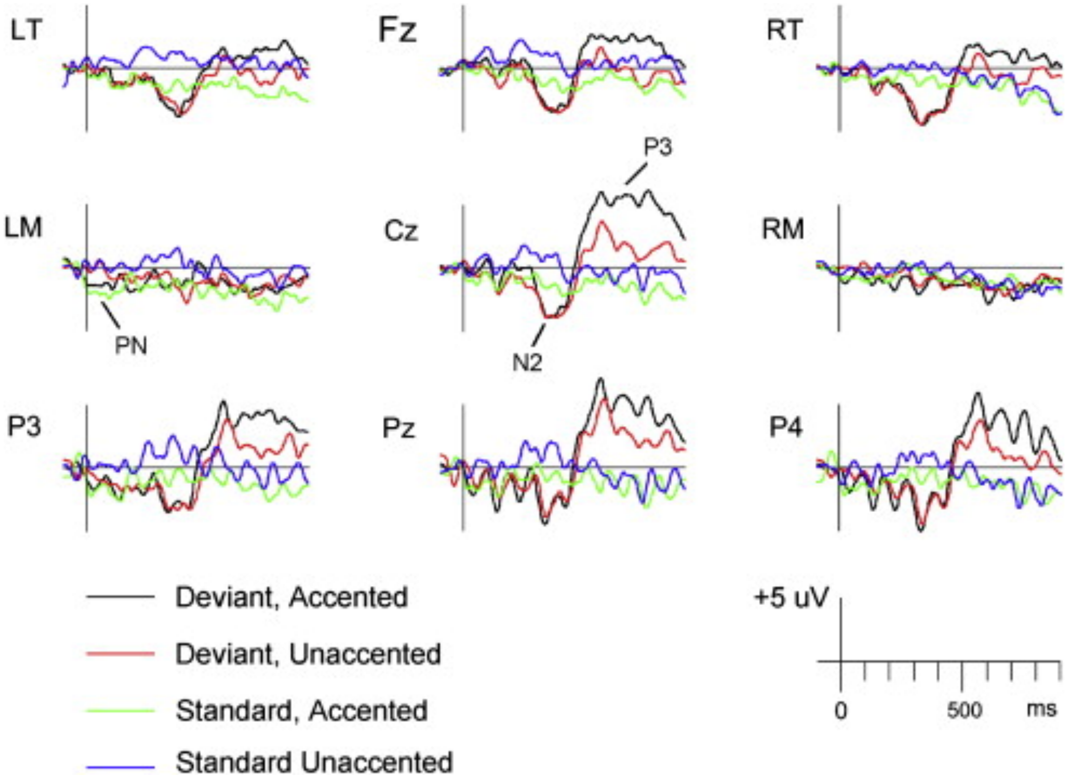


Figure II.1.3.2d. From Potter et al. (2009), who replicated Brochard et al. (2003) experiment. ERPs elicited when listening to isochronous tones (standard accented, unaccented) or deviant to these tones (deviant accented, unaccented). The accented tones are considered as the tones occurring at the odd number in the tones sequence, according to the "ticktock" hypothesis. As shown in the ERPs, deviant to accented tones elicited increased P300 compared to unaccented positions in the sequence.

Nevertheless, these behavioral, electrophysiological and musicological observations of the bias for binary meter have been tested only on individuals of similar lifespan exposure to Western musical traditions. Importantly, whether the "ticktock" effect is a cultural bias or not remains an open question, and it would be worth testing participants with long-time exposure to other musical habits for instance, to assess the impact of enculturation on this binary bias.

Contrasting with Western music, Macedonian dance music is an example of musical tradition using unequal meter lengths widely, in the form of additive meters (e.g., the grouping of meters by 3 and 2 beats, resulting in higher order cycles by 5 or 7 beats) (Pressing, 2002;

London, 2004) (cf. Audio track II.1.3.2[4]). A usual Balkan meter such as 7/8 (i.e., 7 note units per measure) contains isochronous temporal units at the slowest (the measure) and fastest level. Yet, at the intermediate level it contains alternation of non-isochronous groups of two or three notes (Fig. II.1.3.2e).

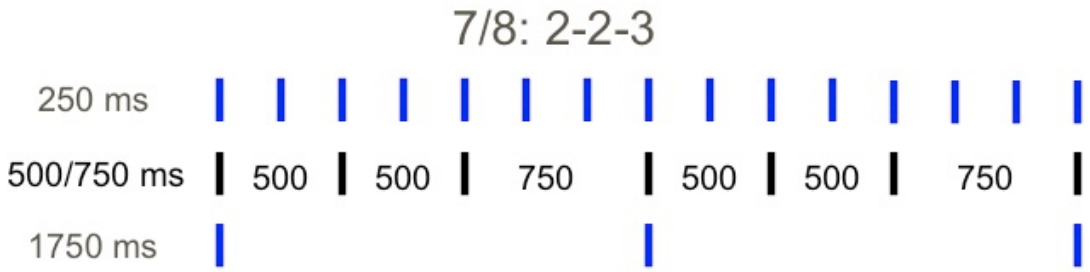


Figure II.1.3.2e. Non-isochronous meter (7/8). The slowest (1750 ms) and the fastest metric levels (250 ms) are isochronous, whereas the intermediate level alternates between binary and ternary groups.

Cross-cultural investigations have found that enculturation shaped our perceptual capacities since early in childhood. Western adults, possibly due to a prolonged exposure to Western rhythms, have been shown to present difficulties in perceiving, producing and synchronizing their movements to rhythmic patterns with non-isochronous meters (Fraisse, 1982; Essens, 1986; Hannon an Trehub, 2005a). When asked to reproduce rhythms containing complex interval ratios such as 3:2, Western adults tend to distort and assimilate the ratio to a simpler 2:1 ratio, in agreement with Western meter (Hannon et al., 2011). In contrast with Western listeners, non-Western adult listeners (e.g., from Bulgaria and Macedonia) accurately detect disruptions in rhythms having either 3:2 or 2:1 ratios, presumably because music in their culture contain both isochronous and non-isochronous metric levels (Hannon

---

[4] Audio track II.1.3.2. Goran Bregovic, serbian musician. This elaborated dance piece illustrates the complexity of metric structures found in most of this folk musical tradition.

and Trehub, 2005a). Thus, listening experiences and familiarity with rhythmic patterns play an important role in perception.

II.1.4. Human development

The study of infants and their interaction with musical rhythms is relevant to understand the biological basis of beat and meter perception/production. However, observations on children are particularly difficult to interpret because perception and production develop asymmetrically in childhood, due to the distinct maturation speeds of the systems responsible for motor output, for processing sensory input, etc.

II.1.4.1. Production. A vast literature exists on the rhythmic sensorimotor synchronization abilities in children. Using similar paradigms as those used for adults (e.g., finger or hand tapping to auditory stimuli), sensorimotor synchronization cannot be achieved before the age of 4 (McAuley et al., 2006; Repp, 2005; Patel, 2006). At that age, the tempo range accessible for sensorimotor synchronization is smaller than that of adults, and coincides with the range of the spontaneous motor tempo of children of this age (interonset interval of 400 ms) (Drake et al., 2000; Kirschner and Tomasello, 2009; McAuley et al., 2006; Provasi and Bobin-Bègue, 2003).

Before 4-years old, music elicits rhythmic behaviors, although not precisely synchronized with the beat of music (Zentner and Eerola, 2010). Already at 9 months, toddlers engage in significantly more rhythmic movement to music and other rhythmic sounds than to speech for instance, and exhibit tempo flexibility to some extent (Zentner and Eerola, 2010). From 2.5 years of age, children are sometimes able to tap in synchrony with an isochronous beat, after training and nonverbal reinforcement (Provasi and Bobin-Bègue, 2003). However, they remain unable, at least until the age of 3, to clap their hands with a metronome (Fitzpatrick et al., 1996). When focusing on spontaneous whole-body dancing movements, 2- to 4-year olds sometimes exhibit synchronized motions to music, but do not adjust the period of their hopping to a music played at tempo slower than their spontaneous motor tempo (Kirscher and Tomasello, 2009). Interestingly, children of this age are able to adjust their drumming tempo to a beat outside their spontaneous motor tempo, only when asked to drum with a social partner compared to a drumming machine or a drum sound (Kirscher and Tomasello, 2009). This observation supports the view that moving with a social partner improves joint action and increases motivation, which is necessary to achieve such coordinated movements.

II.1.4.2. Perception. Although rhythmic coordination requires several years to reach its maturity, several observations suggest that children are highly sensitive to rhythm and meter at an early age. To test the ability of young children to process rhythms despite limitations in movement production, investigators have used perceptual tasks, habituation procedures and the measure of fixation times, known to reflect the processing and categorization of external inputs in children. It has been shown that 2- to 7-month children discriminate simple auditory sequences based on their rhythms (Chang and Trehub, 1977; Demany et al., 1977; Trehub and Thorpe, 1989) and metric structures (Hannon and Johnson, 2005; Phillips-Silver and Trainor, 2005). For example, children as young as 7 months infer different meters from the same auditory rhythm when they get bounced at different periodic accents, illustrating the multisensory interactions between rhythm perception and body movement (Phillips-Silver and Trainor, 2005).

Nevertheless, infant looking time procedure is not ideal because it reveals whether or not children discriminate two stimuli, but not how well these stimuli are actually processed (Hannon et al., 2011). Moreover, the procedure requires behavioral responses which are often difficult to compare across different ages. EEG and evoked potentials could possibly overcome these limitations, by recording electrophysiological responses without the need of explicit behavioral responses. In line with this view, Winkler et al. (2009) used EEG to test whether newborns could exhibit discriminative brain responses to violations of metrically regular rhythmic patterns (Fig. II.1.4.2a), as measured with the mismatch negativity potential (Fig. II.1.4.2b) (see also Sections II.2.2.5.2). Sleeping newborns were presented with a varying rhythm, consisting of five different patterns, based on a standard rock rhythm of eight sounds. In different positions, sounds were omitted to create variety in the rhythm. Regularity was established by four different standard patterns (S1-S4; Fig. II.1.4.2a) that were all strictly metrical. That is, omissions only occurred on metrically weak positions, leaving the metrical structure undisturbed. But the regularity was violated in one deviant pattern, in which the omission occurred on the first beat of the measure. The authors found that a mismatch negativity potential was significantly elicited only in response to the omission on the beat, thus concluding that newborns could differentiate between omissions in weak and strong positions without attending to the rhythm. However, a re-exploration of this procedure (Bouwer and Honing, 2012) recently cast doubts on the conclusions of the original study. In fact, in the condition in which the omission occurred on the strong beat, the deviant consisted in the omission of a hit-hat sound plus a drum bass sound, whereas in the conditions in which the omission occurred in weak beats (S1-S4), the deviants consisted only in the omission of a hit-hat sound, thus occurring much more frequently (Fig. II.1.4.2a). Therefore, whether the mismatch negativity potential observed in Winkler et al. (2009) study reflects the detection of a mismatch regarding an internal representation of metric structure or an internal representation of basic, non rhythm-related, acoustic features remains a debated question (Bouwer and Honing, 2012) (see also Section II.2.2.5.2).

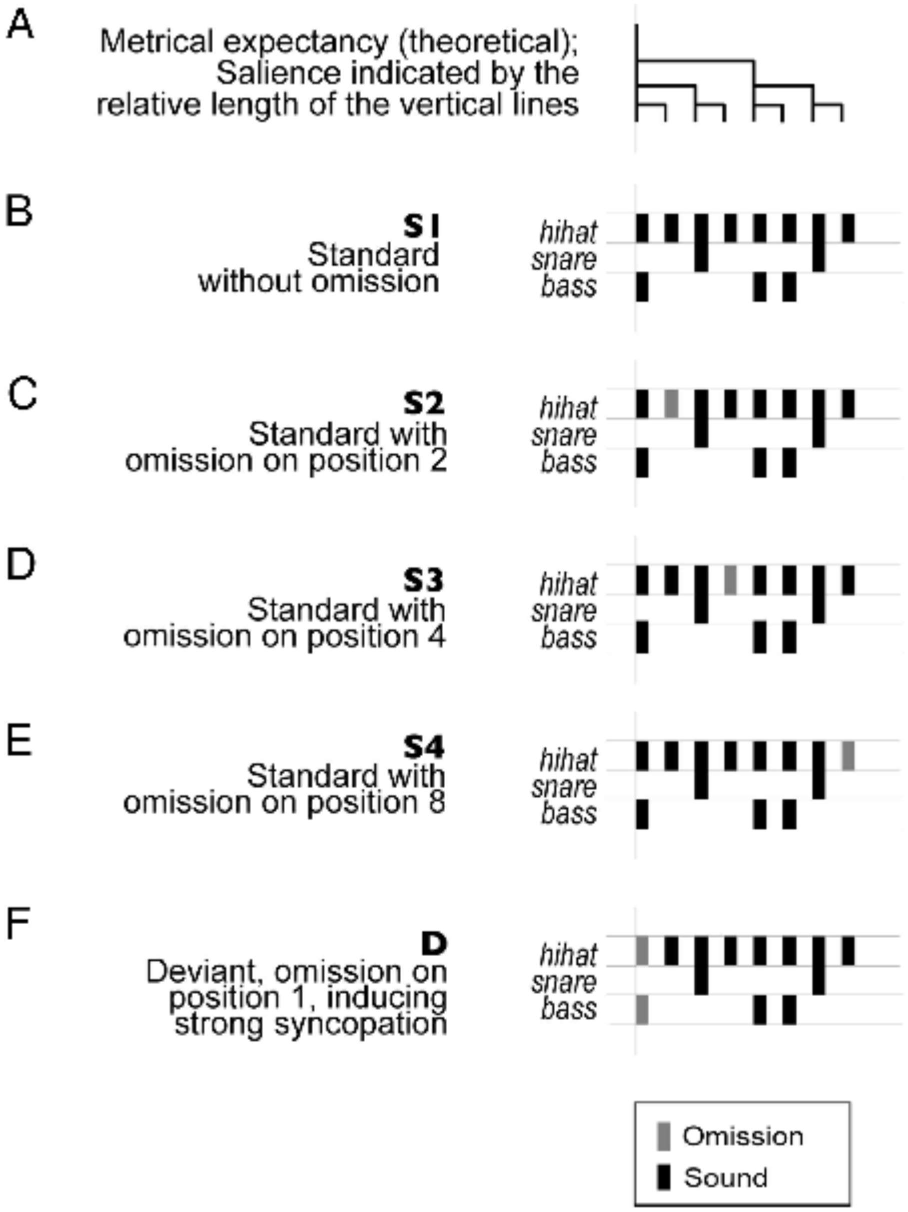


Figure II.1.4.2a. Schematic diagram of the rhythmic patterns used in Winkler et al. (2009) study.

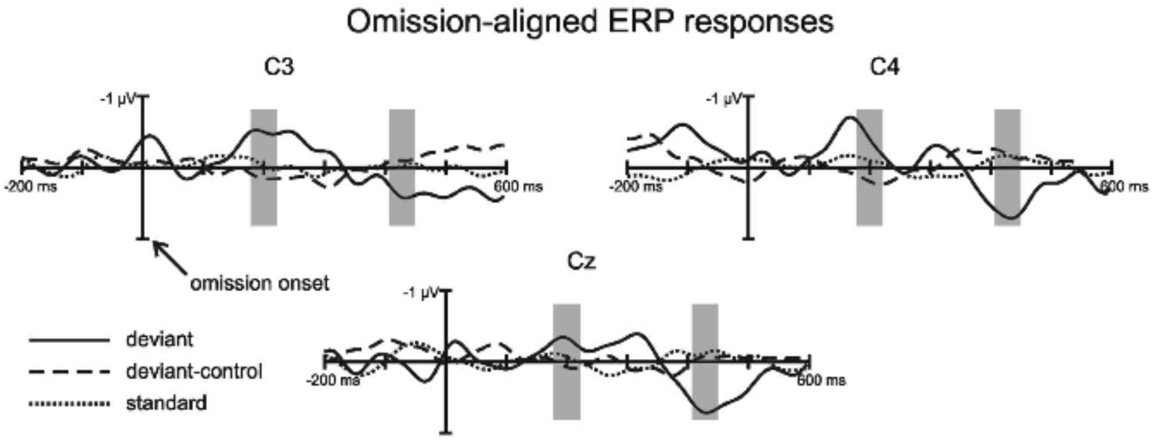


Figure II.1.4.2b. From Winkler et al. (2009). Group averaged (n=14) electrical brain responses elicited by rhythmic patterns in neonates. Responses to standard (average of S2, S3 and S4; dotted line), deviant (solid line) and deviant-control patterns (i.e., deviants appearing in the repetitive control stimulus block; dashed line) are aligned at the onset of the omitted sound (compared with the full pattern S1). Grey-shaded areas mark the time ranges with significant differences between the deviant and the other ERPs.

II.1.4.3. Enculturation. Interestingly, while newborns and young infants may grasp basic aspect of rhythm and meter, their listening experience rapidly influences how they respond to such structures. Indeed, several studies have found that 6-months old Western children who have far less exposure to music than adults, are able to discriminate rhythmic disruption in rhythms containing non-isochronous meters such as 3:2 ratios meters, whereas 12-month old Western children only discriminate rhythms having isochronous meters (Hannon and Trehub, 2005a; 2005b). This suggests that culture-specific metric representations begin to emerge and affect behavior between 6 and 12 months (Hannon and Trehub, 2005a, 2005b) (Fig. II.1.4.3).

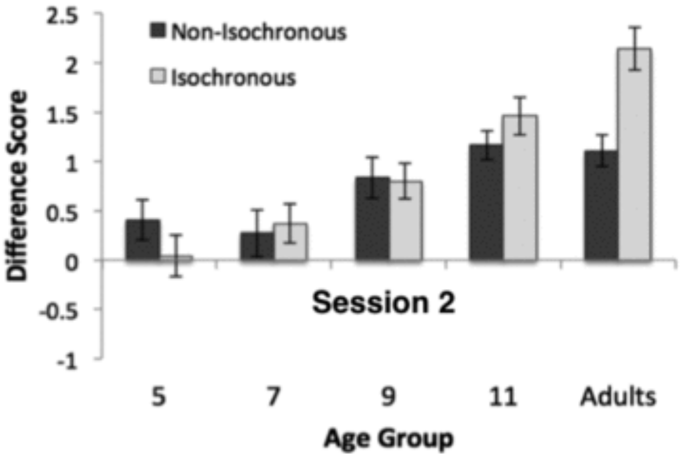


Figure II.1.4.3. From Hannon et al. (2011). Score calculated from the difference between discrimination of deviants in line with the measure vs. deviants disrupting the meter, in isochronous and non-isochronous meters, across the various age groups. As illustrated in this graph, the difference between isochronous and non-isochronous meters increases along lifespan. However, the discrepancy between methods measuring discrimination performance across age groups remains a problem for such studies.

In addition, it has been shown that North American 9-months old children are better at detecting disruptions to a duple-meter melody, far more common in Western music, than a triple-meter melody (Bergeson and Trehub, 2006). This tendency is reinforced in 7-months old children who received greater exposure to duple than triple meters through Kindermusik classes (Gerry et al., 2009). However, two weeks of at-home exposure to Balkan folk music containing non-isochronous meters have been shown to effectively reverse the culture-specific bias in Western young children (Hannon and Trehub, 2005a).

Nevertheless, rhythm processing is not infinitely flexible. Indeed, basic constraints limit the types of rhythmic patterns that can be grasped, as observed when testing young infants with meters of various degrees of complexity (Hannon et al., 2011; 2012). Indeed, it has been shown that young infants (5 to 7 months) who do not yet show culture-specific bias for simple (2:1) or complex (3:2) metric ratios and perform equally in discriminating both meters are already influenced by ratio complexity. When presented with highly complex ratios (7:4), they fail at discriminating targets within these meters, thus suggesting non culture-specific constraint (Hannon et al., 2011). Interestingly, highly complex rhythms such as 7:4 are also relatively rare in music throughout the world (Clayton, 2001; London, 2004). One of the explanations for the tendency to perceive simple integer ratios and the difficulty to process meters such as 7:4 is that adults and young infants alike tend to seek out isochronous, periodic, structures by subdividing intervals and using their common denominator as the primary temporal pulse (Hannon et al., 2011). For instance, when considering the 400 ms and 800 ms intervals of a given 2:1 meter, these intervals could be interpreted in terms of a single underlying 400 ms unit occurring 2 times per measure. In contrast, the common denominator in a 7:4 meter is inevitably found at fast rates that make difficult a robust entrainment and the perception of a salient beat. Likewise, increasing the duration of the common denominator such as placing it within a frequency range more comfortable to process would lead to an unusually long measure that would probably overwhelm the auditory working memory (Grondin, 2001).

Hence, general properties of the nervous system, such as those underlying expectation, working memory and movement, may bias humans towards some forms of regularity, in addition to experience with specific rhythm structures (Nakata and Mitani, 2005; Soley and Hannon, 2010).

II.1.5. Evolutionary perspective

There is a vigorous debate over the evolutionary status of music, and musical rhythms. Some argue that humans have been shaped by evolution to be musical (Mithen, 2005; Wallin, Merker and Brown, 2000). This was first proposed by Darwin in 1871, who referred to music and dance as courtship displays. In line with this view, a number of hypotheses have been proposed about the possible adaptive roles of music, and isochrony in music (Fitch, 2006). Others consider that musical abilities have not been a target of natural selection but, instead, reflect an alternative use of more adaptive cognitive skills, such as language, auditory scene analysis and sensorimotor coupling (Pinker, 1997). This alternative proposition had already been expressed by William James, who said that attraction toward music was "*a mere incidental peculiarity of the nervous system, with no teleological significance*" (cited in Langer, 1942). A way to solve this debate is to examine the innateness, the domain-specificity and the human-specificity of some rhythmic traits in the musical behaviors (Patel, 2006). As proposed by Patel (2006), this approach is useful because it links evolutionary studies of music to empirical research.

II.1.5.1. Innateness. The innateness of rhythmic abilities has been addressed in studies on human development, as reviewed above (see Section II.1.4). These studies have observed that young children are not able to synchronize to an external periodic stimulus, probably due to the immaturity of the motor system (Longhi, 2003), and that the procedures currently available are limited to examine the perception of beat and meter in infants. It is therefore difficult, from such observations, to conclude in favor of one particular hypothesis concerning the innateness and evolutionary status of musical rhythms.

II.1.5.2. Domain-specificity and inter-individual differences. The search for domain-specificity consists in determining whether one particular system processes only one kind of signal (Peretz and Coltheart, 2003). That is, here, whether musical rhythms are independent from speech rhythms for instance. Research on this issue has not particularly favored this modularity view. For instance, across distinct cultures, it has been shown that grouping processes of non-speech auditory stimuli paralleled speech grouping habits, probably due to familiarity to distinct syntactic rules (Iversen et al., 2008). Other work has shown that sequences of sounds of various durations in musical compositions paralleled speech rhythm within different traditions (Patel and Daniele, 2003). Building on these observations, there is no compelling reason to reject the hypothesis that human minds have not been specifically shaped by natural selection for musical rhythm, but that musical rhythm could be an offshoot of linguistic rhythm (Patel, 2006). However, the periodic isochrony spontaneously perceived in music cannot be fully explained by biological constraints common to language, as ordinary speech does not induce pulse perception (as discussed in Section II.1.2; see also Section II.1.5.3).

According to a modularist view, domain-specificity is supported if brain damage disrupts one process specifically, leaving another intact, and *vice versa* (i.e., if a double dissociation is found; Peretz and Coltheart, 2003). In this line, some authors aimed to explore using discrimination tasks whether brain damage disrupting one process, e.g. beat perception in music, also disrupted other cognitive abilities. The neuropsychological literature contains descriptions of individuals with musical rhythmic disturbance after brain damage, predominantly in the anterior part of the superior temporal gyrus, which was called "acquired arrhythmia" (e.g., Liégeois-Chauvel et al., 1998; Peretz, 1990, Wilson et al., 2002). Two notable findings from this literature are (1) that rhythmic abilities can be selectively disrupted, leaving pitch processing relatively intact, and (2) that there are dissociations in the performance of these patients between rhythmic tasks requiring simple discrimination of rhythmic patterns and those requiring the perception of periodicities from rhythmic patterns (see Peretz and Zatorre, 2005 for a review of the evidence).

In line with the neuropsychological approach and the search for dissociations between distinct cognitive skills, a first case of congenital disability of beat processing, labeled "beat deafness", was recently reported by Phillips-Silver et al. (2011). The studied individual was labeled as *beat deaf* on the basis of his apparent disability, compared to the normal distribution found in 33 age-matched individuals, to synchronize finger tapping and whole body movements with various musical pieces in contrast with moving on a metronome. The ability to perceive a beat was also examined in this study using discrimination tasks between various rhythmic patterns, and a task of detection of asynchrony between the auditory and visual parts of an audiovisual clip in which a dancer was presented as moving on music. Such as in the cases of acquired arrhythmia, Phillips-Silver et al. noted a dissociation between impaired processing of beat and preserved processing of melody in music, as well as preserved processing of rhythmic grouping.

Interestingly, similar dissociations have also been described between the processing of beat *versus* rhythmic grouping in Parkinson's patients suffering from damage of basal ganglia (Grahn and Brett, 2007; 2009) (see also Section II.2.2.5.1). This dissociated disruption appeared to be related to a loss of the ability to produce and perceive the periodic beat in rhythmic patterns, while the patients were not worse than controls in the context of unmeasured sequences.

However, regarding the issue of domain-specificity of beat processing, the fact that the relationship between deficits in beat processing and in other, nonmusical, cognitive skills was not examined, weakens the conclusion that these brain lesions or congenital/developmental disorders truly reveal domain-specificity of beat processing. Moreover, another aspect weakening the conclusions of these studies is that we still lack data concerning the ability to perceive and synchronize to the beat in music in the general population. This absence of data results probably from a lack of consensus on the methods to measure this ability across healthy individuals and patients, and across various cultures.

Furthermore, the choice of the stimuli remains a matter of debate, and addresses the question of the complexity rating and its impact in evaluating beat-processing abilities. Finally, the measures chosen to assess produced movements reflect important differences of concept in the exploration of sensorimotor synchronization in music (e.g., finger tapping to beeps vs. full-body movement to pop music). These various issues highlight the fact that searching for domain-specificity for beat perception and synchronization, and for beat-specific disabilities in individuals, while informative, is perhaps premature as long as clearer definitions of what is considered as a musical context, and how such context can be re-created in an experimental setup is not established.

II.1.5.3. Human specificity. General properties of the human nervous system, such as those underlying expectation, perceptual grouping or movement, may bias humans towards some form of temporal regularity (Nakata and Mitani, 2005; Soley and Hannon, 2010). These properties are sometimes expressed in other animal species as well, thus explaining why these species exhibit similar bias in perception. For example, anxiety-like behavior and sustained amygdala activity in mice and humans have been shown to be greater in the presence of temporally unpredictable than predictable sequences, suggesting that irregularity might be aversive (Herry et al., 2007). In addition to perceptual bias, animals exhibiting common motor constraints across species consequently show similar movement patterns, as the tendency for phrase-final notes to be relatively long, observed in the song production of both humans and certain bird species (Tierney et al., 2011).

Another temporal characteristic shared across species due to common motor constraints is the preferred frequency range found around 2 Hz for coordinated movements. The propensity to synchronize voice and locomotion across individuals of a group, a behavior highly represented in dance in humans, has been reported in apes too, in the form of repetitive hooting paced at 2 Hz and coordinated mobilization and displacement of the individuals (De Waal, 1988). This preferred frequency range common to the musical beat is probably due to constraints supporting these coordinated movements (see also Sections II.2.3.2).

Along the same line, a bias towards periodic behavior could possibly be observed in other animal species as well, because it relies on characteristics similarly found in these species. Hence, because isochrony improves prediction and thus interindividual coordination, given that optimal prediction allows movement synchronization to an external event, instead of reactive movement, it could be a mean to amplify signals through the summation and spread of the isochronous signal within a group of individuals (so-called “beacon effect”) (Merker et al., 2009). This would also explain why isochrony is present in many behaviors, for instance, the defense and alarm signals in some insects, which entrain the timing of their alarm signal to neighbors (Merker et al., 2009). It explains in contrast why isochrony is not driven in some human behaviors, such as in ordinary speech, which is not aimed at causing entrainment and movement coordination.

However, musical beat perception and synchronization in humans has several distinguishing features compared to rhythmic entrainment behaviors observed in other species (e.g., the rhythmic chorusing of certain frogs). For example, beat perception and synchronization involve (1) a periodic motor response to complex sound sequences, and not just pulse trains, (2) an adjustment to a broad tempo range, and (3) a crossmodal processing, with sounds inducing the synchronization of periodic movements that are not necessary to sound production (Patel et al., 2009). Surprisingly, animals as highly intelligent and close to humans as chimpanzees have never been shown to process musical beats, even in their most primitive forms and after training, whereas they can voluntarily produce rhythmic movements on a time scale appropriate for beat processing. Moreover, synchronization of movement to a musical beat is not commonly observed in domestic animals, such as dogs, that have lived with humans and their music for thousands of years (Fitch, 2009).

To explain this issue, Patel (2006) proposed the "vocal learning hypothesis". Vocal learning refers to the ability of animals to modify vocal signals as a result of experience with sounds usually produced by individuals of the same species. By extension, this definition has been restricted to cases where animals learn to mimic sounds that are not in their species repertoire. Vocal learning thus involves learning to produce vocal signals based on auditory experience and sensory feedback, also required for beat processing in music. Some animal species, such as songbirds, parrots or cetaceans, but not non-human primates, present this ability involving tight online auditory-motor coupling. The candidate brain structure that has been proposed as necessary for such vocal learning is basal ganglia, in charge of both interval timing and motor sequencing (Patel, 2006). Indeed, although basal ganglia are not particularly driven by one modality at first, some fundamental modifications of their structures, allowing tight audiomotor coordination, have been observed in species displaying vocal learning, thus corroborating the view that the basal ganglia of humans could have been modified by natural selection for vocal learning (Jarvis, 2004; Patel, 2006).

Recently, the evidence that specific species, such as parrots, presented abilities for beat processing has corroborated the vocal learning hypothesis: Patel and colleagues (2009) reported the case of one parrot exhibiting the ability to synchronize body movements with musical beat and to adjust the movements according to changes in the tempo. According to these authors, the fact that a non-human animal could acquire the ability to process beat from music through training, while unnatural, would suggest that this ability is not part of a selective adaptation for music (Patel, 2006).

However, while vocal learning might constitute a prerequisite for musical beat processing, one could ask the reason of the absence of isochrony in the perception and production of speech, as speech constitutes probably the most prominent locus of vocal learning in humans. This apparent paradox can be solved by considering the goal of ordinary speech, which is not to cause entrainment and movement coordination across individuals.

Importantly, when searching for evidence of beat processing as an ecologically natural behavior, *Homo sapiens* is the unique species manifesting spontaneous synchronization of periodic body movements to acoustic rhythms, engaging both sexes (Patel, 2006). Moreover, this skill develops relatively early in human ontogeny, long before sexual maturity (Fitch, 2006). Although, according to Patel's hypothesis, vocal learning would provide the neural circuitry required for beat processing in music, this is perhaps not sufficient for spontaneous entrainment to sounds (Fitch, 2009).

One possibility is that the propensity to engage in joint social action plays a crucial role in triggering and developing these rhythmic behaviors specific to music. This was suggested by the observation that young children improve their synchronization abilities when engaged in a joint action with an adult compared to a disembodied metronome, possibly through the building of a shared body representation and increased motivation (Kirschner and Tomasello, 2009) (see also Section II.1.4.1). Moreover, it has been shown that the groove in music is associated with positive affects in children and adults (Janata et al., 2012; Witek et al., 2009), and that interpersonal synchrony, even in non-music contexts, increases affiliation (Hove and Risen, 2009). Hence, music and dance can be considered as a powerful medium, alternative to speech, to inform on the physical ability and health, or to communicate recognizable emotions across individuals. The tight link between joint social action and musical behaviors would explain why such musical coordinated behaviors play an important role in collective work, rituals or war dance for example, widely across cultures (Hagen and Bryant, 2003) (cf. Audio track II.1.5.3[5]).

---

[5] Audio track II.1.5.3. Work song (James Carter and the Prisoners). Thought to give ardor at work, this song illustrates how the isochronous pulse, played by means of work tools, was used as synchronizer across workers. The frequency of the beat is slow due to biomechanical constraints from the gesture and tool manipulation, and also to synchronize across a maximum number of workers, even the weakests.

Taken together, the empirical work reviewed in Section II.1 show that the capacities to perceive and synchronize to temporal regularities are found in various forms throughout human cultures, and also across species that are genetically far from each other. These capacities are also found throughout the human ontogenic development, as observed in the developmental literature, through the importance of interpersonal synchrony in parent-infant interactions (Phillips-Silver et al., 2010).

However, isochrony, or periodic regularity, can be considered as a particular case in perception and action. Its prominent function is thought to be the improvement of prediction, given that optimal prediction allows movement synchronization to an external event, instead of reactive movement. Hence, beat and meter seem to have ubiquitously expanded in music traditions to promote inter-individual and intra-individual coordination (e.g., voice-limbs coordination). It is probably the reason why, very often, perception and movement production cannot be disentangled when a beat is induced in music.

II.2. SENSORIMOTOR COUPLING

Numerous studies have explored sensorimotor coupling within musical and non-musical contexts, to understand how the perception of isochrony seems to affect motor output and how movements appear to influence the perception of isochrony. Here, we review these studies on sensorimotor synchronization to the beat by analyzing separately, for the sake of clarity, three obligatory components of entrainment: production of periodicity, detection of periodicity, and the coupling of both (Todd et al., 2002). However, this separation may be artificial to some extent. Indeed, we will see that producing a periodic movement leads to the generation of periodic sensory inputs whereas perceiving periodic inputs can induce periodic movements.

II.2.1. Production of periodic signals

In humans as in non-human animals, periodic outputs are ecologically natural and found in a broad range of actions. Archetypical forms of periodic motion are often observed in cases where the action requires a sustained production of movements along time, such as for locomotion. However, periodic movements can also take a highly elaborated form such as dance.

Each effector producing a movement can be characterized by a resonance frequency. This resonance frequency refers to the frequency at which the production of periodic motion requires the minimum force to maintain the oscillations amplitude and, therefore, requires the minimum amount of muscle activity (White et al., 2008). The resonance frequency depends on external and internal constraints. An example of external constraint shaping body motion towards periodicity is Earth gravity in human locomotion. In this particular case, gravity is exploited in the generation of a sustained pendulum motion across the two legs. Moreover, gravity influences directly the resonance frequency, by conditioning the weight of body segments. Internal constraints are numerous as well, combining biomechanical characteristics such as the stiffness of the articular apparatus or the size of the limbs.

In the case of archetypical periodic motions, the nervous system controls the output in a reduced form, for instance at subcortical level (Rossignol and Frigon, 2011). Locomotion for instance has been shown to require a nervous unity located in the spinal cord: the central pattern generator. This nervous unity, which has been demonstrated to be necessary and even sufficient to walk (Rossignol and Frigon, 2011), can be reduced to two groups of neurons with mutual inhibitory connections (White et al., 2008) (see also Section I.1.1). The inhibitory connections generate antiphase oscillations in the activity of the two groups of neurons, alternately driving a pair of antagonist muscular groups (e.g., flexors-extensors) to induce periodic motion. In contrast with such archetypical periodic motions, periodic movements can also take a more elaborate form requiring cortical control. In either subcortical or cortical control of periodic movements repeated along time, incompressible temporal constraints exist, due, for example, to the times required for the propagation of action potentials and synaptic transmission. The constraints related to the control by the nervous system, combined with the resonance frequency due to biomechanical and dynamic constraints, determine the preferred frequency of an effector, sometimes referred to as the “spontaneous motor tempo” (Hogan and Sternad, 2007). Hence, the energy efficiency and robustness of periodic movements are due to the ability of the central nervous system to drive the system at a pace close to its resonance frequency (White et al., 2008).

Importantly, the production of periodic movements cannot be strictly isolated from the perception of periodic inputs, since the periodic movements are in turn responsible for the generation of periodic inputs (Todd et al., 2002). This is, once again, nicely illustrated in locomotion, which generates auditory, visual and vestibular rhythmic cues as a result of movement in space (Phillips-Silver et al., 2010). Moreover, these periodic inputs generated by the movement itself are necessary, to account for the control by the nervous system in the form of error corrections through feedback loops.

Finally, it is worth noting that a periodic movement becomes sustained, energy-efficient and replicable once its frequency reaches a pace close to the resonance frequency. In line with this assumption, studies using functional MRI have found that the amount of cortical activation during periodic movements is reduced as a function of the match between the frequency of the performed movement and the resonance frequency of the produced movement (Schaal et al., 2004), thus suggesting that a periodic movement executed at a pace far from the resonance frequency of the system requires more cortical control.

Taken together, these observations support the view that when continuous body movements are produced, the movements evolve towards a bias for periodicity, as a balance between sustainability of the produced movement and energy efficiency. Indeed, the production of periodic movements can be considered as a design feature of an adaptive system, to make a compromise between the energetic cost engendered by external or internal constraints and the benefit for instance from a “beacon effect” in the production of a sustained movement along time (i.e., a signal amplification; see also Section II.1.5.3).

II.2.2. Perception of periodic signals

Besides the production of periodic movements, another prerequisite for periodic sensorimotor coupling is the processing of periodic external inputs. Strikingly, periodicity is perceived from sounds, whether or not these sounds are actually periodic. Indeed, isochrony can be induced not only by isochronous pulses (as with a metronome) but also by complex rhythmic structures (Phillips-Silver et al., 2010).

The reasons why the auditory system seems biased towards periodicity and could act as a detector of periodicities within certain circumstances remain unclear. On the one hand, the perception of isochrony could be explained by intrinsic properties of time processing in the human brain. According to this view, the present section will discuss briefly the models and empirical work having studied time processing, as they could be relevant to understand beat perception. Second, the perception of isochrony could be explained by characteristics specific to rhythm processing. This has been suggested by models and empirical work having studied rhythm processing, which will be reviewed in this section as well. In particular, we will discuss the circumstances under which isochrony is induced from rhythmic patterns.

II.2.2.1. Models and empirical evidence on time processing. The world is filled with information ranging from the millisecond to several days. The temporal range of the events constituting musical rhythms lies between 200 ms and 2 s (London, 2004). This temporal range is also relevant for other contexts as speech, working memory or perceptual grouping for instance (Eagelman et al., 2005; Ivry and Spencer, 2004; Mauk and Buonomano, 2004). Two classes of models have been proposed to explain the processing of temporal information within this temporal range: the interval models and the entrainment models (McAuley, 2010).

Mostly based on evidence from non-human animals studies, interval models are generally constituted by three components, a clock component that estimates duration, a memory component that stores this duration, and a comparison component that compares the new duration to the duration stored in the memory component (Gibbon, 1977). Various formalizations of such models have been proposed. One of these is the scalar expectancy model, which postulates that the internal clock is constituted of a neural “pacemaker” pulsing as long as the timed stimuli are presented, coupled with an accumulator that codes for the time duration as a function of the number of pulses (Gibbon et al., 1984). Other formalizations consist for instance of process-decay models, which posit a tracking mechanism of the decay of neural activity following signal onset (Matell and Meck, 2000).

Alternatively to the interval models, it has been suggested that humans could show a natural tendency to perceive temporal information in term of relative rather than absolute durations and, thus, that successive time intervals are processed as ratios across each other (Grahn, 2012; Teki et al., 2012). This assumption is based on evidence that a context of periodic sound sequence improves temporal acuity to discriminate between pairs of single time intervals presented within or after the periodic sequence. This evidence suggests that time intervals of different duration can be processed when presented in parallel, and not only in series (Drake and Botte, 1993). In addition, it has been shown that jittered sound sequences were perceived as more periodic than they were in reality, as if an invariant representation of periodicity was extracted from the sequence as processed in a whole (Repp, 1992; Velasco and Large, 2012). To account for these observations, the entrainment models have been proposed as an alternative to the interval models. The basic assumption of entrainment models is that the timekeeper consists of one or more self-sustaining and entrainable oscillators that peak in amplitude at regular time intervals (Large and Jones, 1999). The adaptable period of each oscillator provides a reference for making judgments about timing.

II.2.2.2. Neural correlates of time processing. Functional neuroimaging studies, as well as studies of brain-damaged patients, have linked components of timing processes to several cortical and subcortical structures, including the cerebellum, basal ganglia, parietal cortex, prefrontal cortex, premotor cortex and supplementary motor area (Teki et al., 2012) (Fig. II.2.2.2a). These regions are also traditionally found to be involved in various aspects of movement, such as repetitive action and learning for the basal ganglia, fine-tuning of movement and sensorimotor integration for the cerebellum, or movement planification and execution for the premotor and supplementary motor areas.

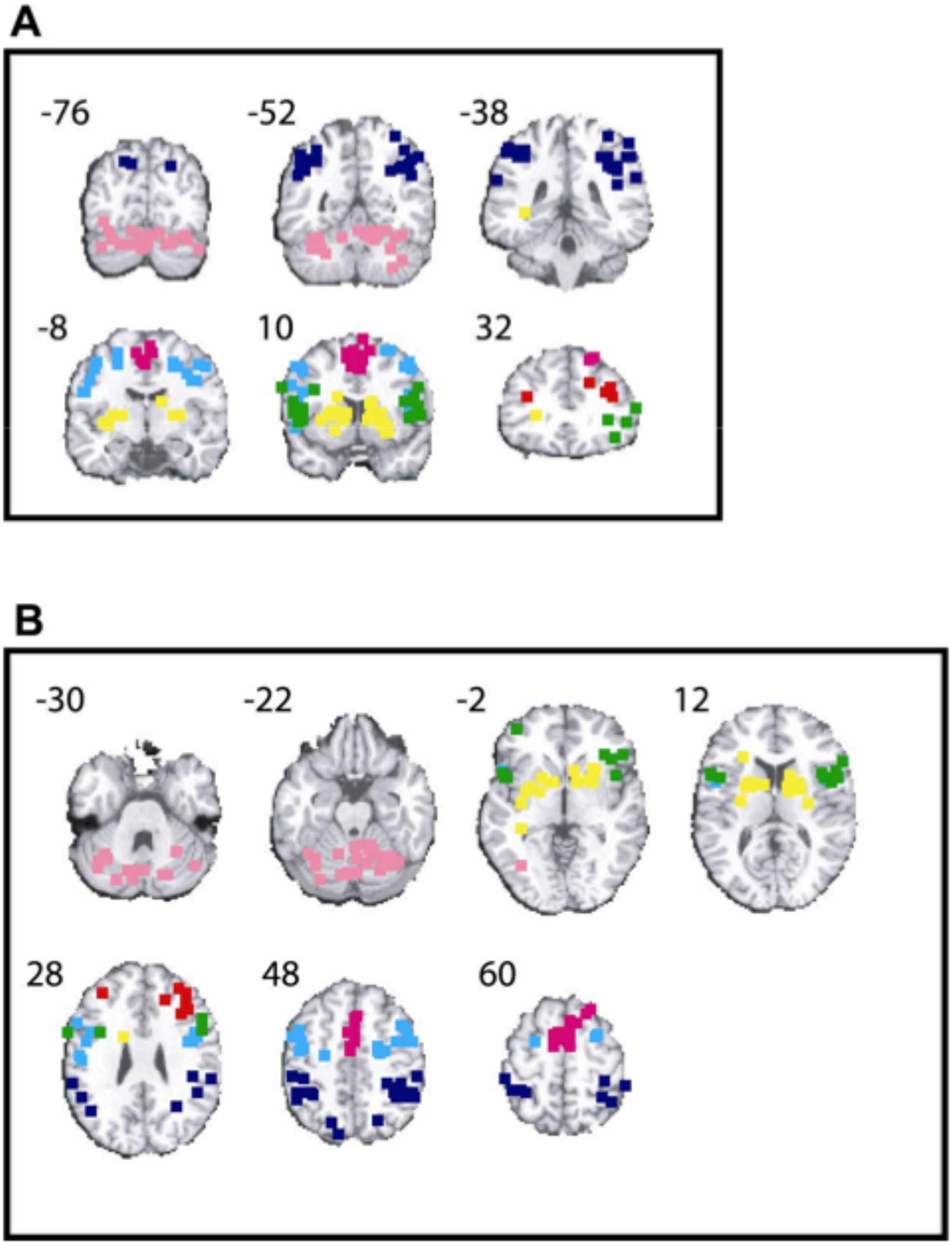


Figure II.2.2.2a. From Grahn (unpublished doctoral thesis). Areas of peak activation significantly reported across functional MRI studies based on timing tasks (coronal slices in panel A and transverse slices in panel B; the numbers indicate the position in the y and z plane respectively, in MNI coordinates): cerebellum (pink), basal ganglia (yellow), inferior frontal cortex (green), dorsolateral prefrontal cortex (red), premotor cortex (light blu), supplementary motor area (magenta) and inferior parietal cortex (dark blue).

Importantly, it has been shown that the network of activated brain areas changed depending on whether the task involved expectation, i.e., the allocation of greater attention to a particular point in time, compared to temporal discrimination, time estimation or working memory for instance (Grahn, 2012). In this line, two temporal aspects that are particularly relevant for the processing of musical rhythm and meter have been distinguished: the interval-based timing, also called “absolute” timing, and the beat-based timing, or “relative” timing (Teki et al., 2012; Grahn, 2012; Grube et al., 2010). To examine this distinction, various behavioral tasks have been used in the temporal range corresponding to musical rhythms. These tasks required either a beat-based timing, as in regularity detection, detection of deviation from isochrony or rhythm discrimination, or an interval-based timing, as in single interval duration discrimination (Fig. II.2.2.2b). Functional MRI data showed greater activity within the basal ganglia for beat-based timing and greater activity within the cerebellum for interval-based timing (Teki et al., 2011) (Fig. II.2.2.2c). Moreover, the fact that patients suffering from cerebellar damage, as well as healthy controls in which cerebellar function was disrupted by transcranial magnetic stimulation (TMS), exhibited reduced performance in interval-based as compared to beat-based timing tasks provides additional support to this view (Grube et al., 2010).

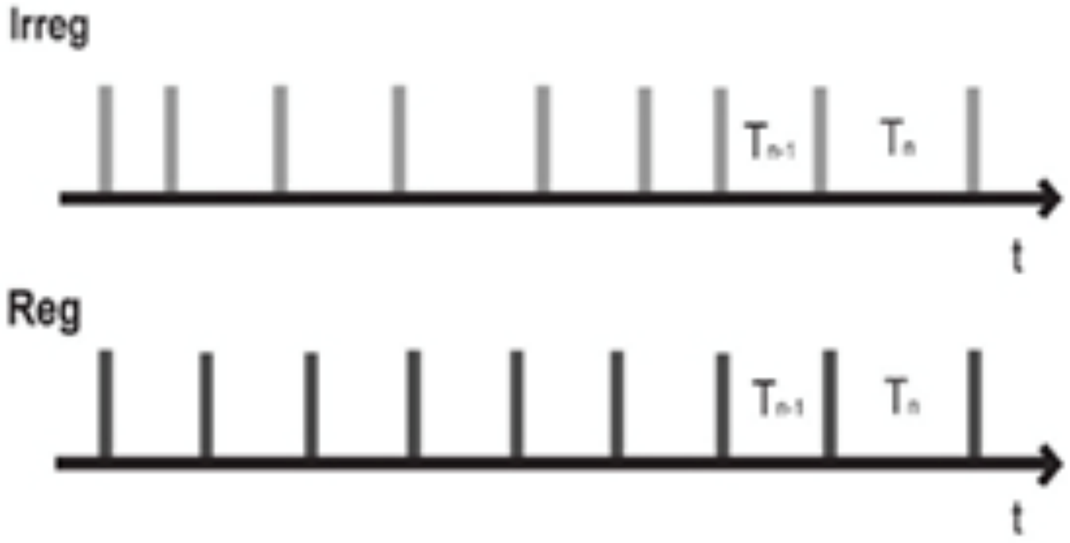


Figure II.2.2.2b. From Teki et al. (2012). Example of absolute vs. relative timing task. Irreg: Subjects have to detect a change of 30% of a time interval duration (T_n) compared to the penultimate time interval (T_{n-1}), within a sequence of jittered events. Reg: subjects have to detect a 15% change of a time interval duration (T_n) compared to time intervals from an isochronous series of events.

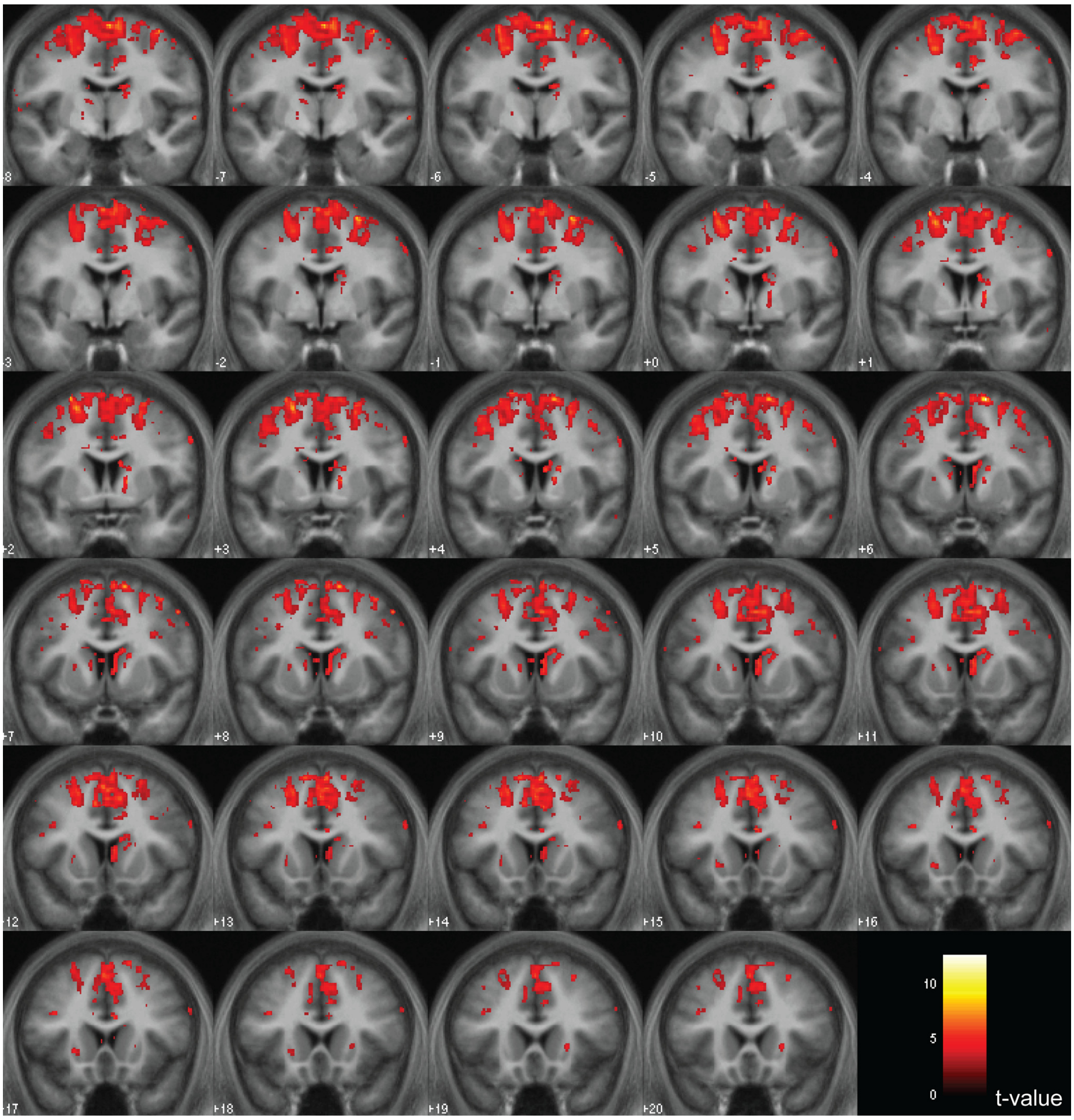


Figure II.2.2.2c. From Teki et al. (2011). BOLD activations for relative, beat-based timing (regular vs irregular) are shown in a series of coronal sections. Significant activations were found in the striatum, thalamus, premotor cortex, SMA, and prefrontal areas including the dorsolateral prefrontal cortex at a threshold of p=0.001 (uncorrected). The strength of activations (t-value) is graded according to the color scheme at the bottom right.

Nevertheless, besides the temporal aspect made prominent in a timing task, the temporal range at which temporal information occurs influences directly the processing of the information as well (Ivry and Spencer, 2004). Indeed, the processing of temporal ranges slower than two seconds approximately has been shown to differ from faster temporal ranges. These differences were observed on the basis of functional MRI studies and behavioral tests performed on patients having lesions in the cerebellum or suffering from basal ganglia damage. Using these methods, researchers established that the processing of events presented at slower temporal ranges involved preferentially the cerebellum, activated larger cortical networks and was less prone to induce synchronized movements (Ivry and Spencer, 2004). In contrast, the processing of faster temporal range preferentially involved the basal ganglia, activated the cortex to a lesser extent, and was more prone to induce movements.

Taken together, these observations suggest that the basal ganglia, i.e., a brain structure playing an important role in motor control, is involved in relative time processing even in the absence of explicit movement production. Moreover, the evidence of a network dedicated to beat-based compared to interval-based timing is fundamental to explain the spontaneous perception of beat in music.

II.2.2.3. Rhythm processing models. There is an obvious relationship between time and rhythm processing, as rhythms may be considered as sequences of time intervals. Thereby, in line with models proposed for time processing, various models have been proposed to account for rhythm perception. Specifically, as rhythms are by nature sequences of multiple intervals, these models focused on the processing of multiple intertwined time intervals, in contrast with timing models often based on the processing of single durations. These rhythm models distinguish between the processing of rhythms that lead to the non-isochronous grouping of events along time and those leading to a metric coding based on the isochronous grouping of the successive events (i.e., beat-based) (Povel and Essens, 1985).

Povel and Essens (1985) referred to the former as *nometrical* patterns and to the latter as *metrical* patterns.

II.2.2.3.1. Grouping models. Grouping refers to how a series of events along time is perceived as clustered or grouped together. Research on principles of grouping and their role in the figural coding of rhythms has a long history, sharing similarities with Gestalt principles of perceptual organization (Bolton, 1894; Wallin, 1911). Indeed, Gestaltist models, as formalized from visual perception, are based on the principle that perceptual experience tends to be ordered in a manner that is regular and invariant through several laws as laws of grouping by proximity, similarity or symmetry effect for instance (Wagemans et al., 2012). In regard to grouping processes, the various features constituting a series of temporal events and the patterning of these features along time lead to a figural coding of a rhythm, conveying to the perceiver a sense of inherent sequence organization based on clustering and accenting (Fraisse, 1956) (see also Section II.2.2.4.1). An example of such clustering in the auditory system is the grouping in a short-long sequence, the longest tone tending to be perceived as accented and as being the end of a group (see Fig. II.1.2.2). Importantly, several studies have shown that performers implicitly use their knowledge based on culture or lifelong exposure to emphasize some groupings (Iversen et al., 2008), thus suggesting that grouping principles are in fact not universally fixed (see Section II.2.2.4.1).

II.2.2.3.2. Metric coding models. Beyond the grouping phenomenon, much work has focused on the metric coding, i.e., the induction of a beat and the perception of a metric structure. In contrast with grouping, whose perception involves the segmentation of a rhythm into clusters of elements, metric coding involves the hierarchical representation of at least two levels of relative duration (McAuley, 2010). Various hypotheses have been proposed to account for beat induction, also called “beat finding”, or “beat extraction”. One category of models for metric coding of rhythm consists in rule-based models. These models may be viewed as an extension of the principles of grouping (Desain and Honing, 1999; Longuet-Higgins and Lee, 1982; Povel and Essens, 1985). These rule-based models share some similarities with interval timing models found in the timing literature (see Section II.2.2.1). According to these models, humans process rhythms by structuring their mental representation on the basis of an internal clock. This internal clock is hypothesized to involve a pacemaker-accumulator mechanism that ticks at regular intervals aligned with particular stimulus onsets that correspond to beats. A formalization of rule-based model has been proposed by Povel and Essens (1985), in a three-stage clock model. In the first stage, accents are assigned to certain events of a rhythm according to a set of perceptual rules. In the second stage, all the intervals that “fit” with the accents in the rhythm (clock intervals) are generated. Finally, in the “matching” stage, the amount of counter-evidence is calculated for each of the possible clock intervals, and the clock with the least negative evidence is determined to be the one most likely to correspond to the induced beat. Although these models provide an elegant explanation for a number of timing and rhythm phenomena, they present some shortcomings. For example, according to these models, the induction of an unambiguous beat should require systematic coincidence between the beats and actual accented tones. However, there are numerous examples showing that rhythmic patterns in which the strong beat does not systematically coincide with an actual sound in the sequence, i.e. syncopated rhythms, can induce a robust beat percept and convey a strong groove (Witek et al., 2012). In addition, these models are based on the principle of universal rules for accent generation and, therefore, cannot account for the evidence that these rules are highly dependent on familiarity and culture, and can also be modified by voluntary interpretation of the rhythms. Finally, another shortcoming of the internal clock approach is that it does not operate in real time, and is therefore unable to make adaptive, online predictions during the unfolding of the rhythm. Instead, the models must consider the fit of all possible clocks across the entire rhythm before settling on a solution (Grahn, 2012).

II.2.2.3.3. Entrainment models. Another class of models of rhythm processing and metric coding has been developed, relying on multiple self-sustaining internal oscillators entrained by external inputs, at periodicities corresponding to the different hierarchical metric levels of the rhythm (Eck, 2002; Large and Kolen, 1994; Toiviainen and Snyder, 2003). Importantly, in these models, the beat periods are based on subjective beats that may or may not correspond precisely with accented sounds present in the signal. That is, musical events for instance may be either temporally aligned (in phase) or misaligned (out of phase) with the periodic timing of peaks in oscillating amplitude of the subjective beats. Critically, two types of adaptive processes are assumed to facilitate entrainment in these models: phase correction and period correction. In such models, there is a range of rates, also called *tempi* in musical contexts, in which a stable entrainment can occur, referred to as entrainment region or resonance curve. Moreover, non-linearities can be introduced in these entrainment models, and appear to predict quite well the human perception of beat and meter induced, for example, by syncopated rhythms (Fig. II.2.2.3.3).

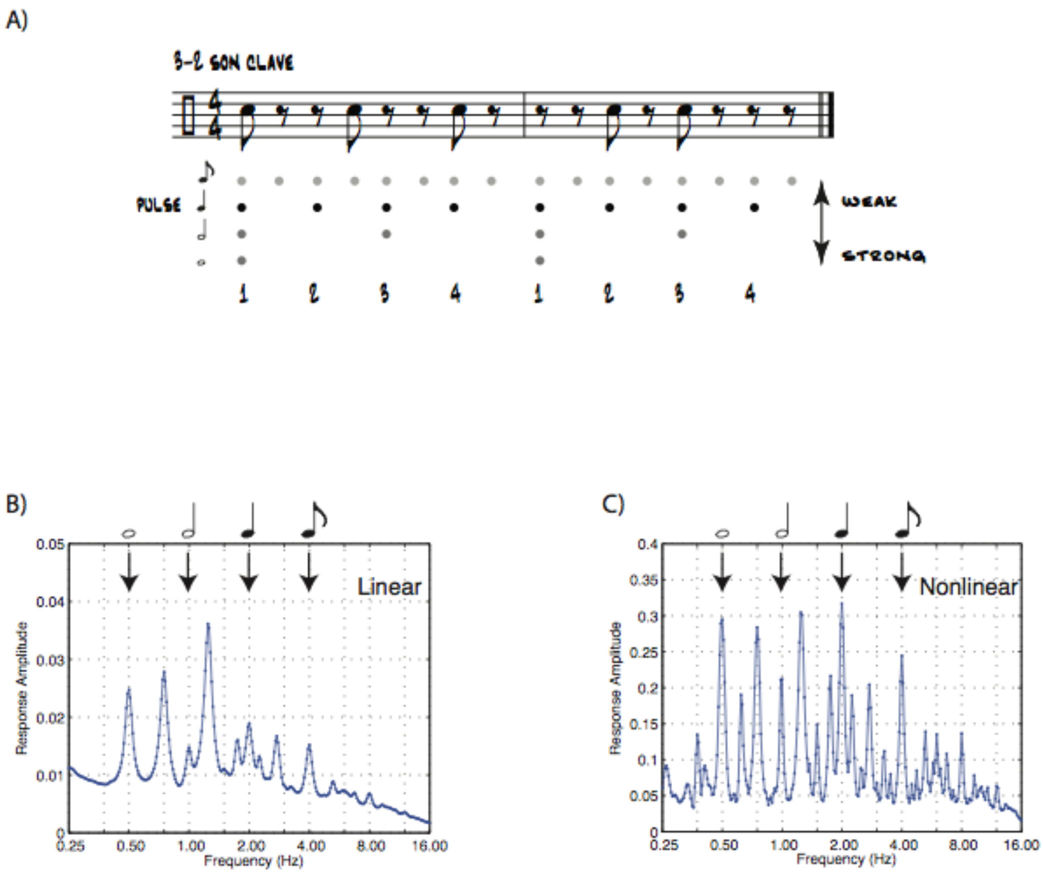


Figure II.2.2.3.3. From Large (2010). A) Syncopated rhythmic pattern (son clave, from the Cuban musical tradition), in musical notation and with symbols for the various metric levels contained in the syncopated rhythm. B) Prediction from a linear model of oscillators. The syncopation is not solved by the model, and the greatest peaks of energy are found at non metric frequencies. C) Prediction from a nonlinear model. The syncopation is close to be solved by the model, and peaks of energy are enhanced at metric frequencies.

Some of these entrainment models attempt to directly formalize the oscillators as neural oscillators in the brain (Large and Snyder, 2009). This formalization comes from neuroscience models of neural oscillators, in which the oscillations result from the interaction of excitatory and inhibitory neural populations (Buzsaki, 2004; Lakatos et al., 2005) (see Section I.1.1). The activity of different populations of neurons could thus oscillate at different periods, corresponding to different levels of the metrical hierarchy. This "neural resonance” approach, in which neural oscillators resonate to rhythmic stimuli, could naturally give rise to properties such as pulse, meter, and resonance curve, which are aspects of rhythm that are difficult to account for in other models. However, empirical evidence of such neuronal entrainment to beat and meter is lacking. Thereby, providing such evidence constitutes one of the goals of the present thesis.

II.2.2.3.4. Dynamic attending model. The entrainment models, in contrast to rule-based models, suggest that the tempo and rhythm of a series of events along time engage humans through realtime attentional synchrony (Jones and Boltz, 1989; Large and Jones, 1999; McAuley, 2010). This concept is formalized in the Dynamic Attending Theory, in which the internally driven rhythm is conceptualized as an attentional rhythm, i.e. a rhythmic modulation of selective attention (Jones, 1976). In this theory, dynamic attending would “*reflect the attender’s tacit use of an event’s dynamic structure*” (Jones and Boltz, 1989). This entrainable attending, referred to as “attunement”, would involve entrainment to one or multiple referent time periods and a selective attending to these time reference intervals.

Importantly, the phenomenon of dynamic attending can also be observed in the absence of periodicity. Indeed, the dynamic modulation of attention can be modeled as a hazard rate, in which dynamic attending increases as a function of the probability of an event to occur (de Hemptinne et al., 2007). Therefore, it would be intimately related to anticipation. Nevertheless, although dynamic attending is conceivable in a non periodic context, it would be optimally induced by periodic and hierarchical temporal structures, which allow the building of strong anticipation (Large, 2008).

Behavioral evidence supporting the Dynamic Attending Theory comes from a range of studies. For example, these studies have shown that simple rhythms enhance the detection and discrimination of rhythmically expected targets in perceptual monitoring tasks (e.g., a change in pitch, timbre or duration), thus suggesting a rhythm-induced dynamic modulation of selective attention (McAuley ad Jones, 2003). In addition, the recall of pitch sequences is enhanced when various accents (e.g., pitch skips, contour changes) are timed regularly as compared to when they occur irregularly (Boltz and Jones, 1986), thus suggesting a better memory encoding of regular *versus* irregular sequences.

To many authors, the Dynamic Attending Theory could account for the modulation of the ERPs elicited at different time points relative to the beat or meter (Brochard et al., 2003; Snyder and Large, 2005; Grube and Griffiths, 2009; Iversen et al., 2009, Schaefer et al., 2011) (see Section II.2.2.5.2 for detailed description of these studies). This interpretation is based on the fundamental assumption, first proposed by Bishop (1933), that cyclical fluctuations of the electrical currents produced by the nervous system under certain circumstances would actually reflect cyclical variations in neural excitability. This assumption was corroborated by numerous electrophysiological studies in primates, including humans (Lakatos et al., 2008; Schroeder and Lakatos, 2009; Busch et al., 2009; Varela et al., 1981; Buzsaki and Draghun, 2004). Taken together, these studies appear to converge to the view that when the activity of a neuronal population synchronizes at a given frequency, the phase of the induced oscillations could result in a cyclic fluctuation of the excitability of the responding neuronal population, which could explain the amplitude modulation of the event-related brain potentials generated by these neuronal populations, as well as the modulation of behavioral performance (Busch et al., 2009; Varela et al., 1981; Buzsaki and Draghun, 2004). This phenomenon has been observed as a consequence of endogenous entrainment of neuronal populations in the form of ongoing neural oscillations (Lakatos et al., 2005; Busch et al., 2009), but also due to the entrainment to an external periodic sensory input (Schroeder and Lakatos, 2008; see Henry and Obleser, 2012 for a recent observation in the auditory domain). These observations are crucial for the interpretation of the results of the present work, since they contributed to justify our experimental approach. These concepts are also discussed in Section III.1.

II.2.2.4. Beat induction. We have seen that beat, as described by music theorists, is not itself a stimulus property, although it is usually induced by a rhythmic stimulus (London, 2006; Large, 2008; Lerdahl and Jackendoff, 1983). Strikingly, although many frequency and phase combinations are available in a musical piece and could be chosen by individuals as their own perceived beat, even nonmusicians seem entrained at the similar frequency and phase, within a given culture. For the sake of clarity, the present section describes separately the perceptual cues that promote beat induction through the selection of a set of frequencies and phases (i.e., the generation of accents) and those inducing spontaneous movements on the perceived beat (as in groove-based music). However, these aspects are often tightly related in music, since the induction of a stable meter would logically facilitate moving body segments on this meter. Indeed, Janata et al. (2011) recently showed a strong positive correlation between the subjective groove feeling experienced by participants while tapping along different music pieces and the facility to produce consistent periodic tapping across the pieces, as the tapping is thought to reflect the stability of beat induction (see also Section II.2.3).

II.2.2.4.1. Generation of metrical accents. In the rhythm literature, an accent usually refers to an increase in salience when an event differs from surrounding events within a sequence (Parncutt, 1994). When the accents lead to meter induction, they are referred to *metrical accents* (Repp, 2010). The generation of metrical accents has been extensively studied by the authors having formalized rule-based models of rhythm processing. Indeed, to explain the mechanisms leading to the selection of one frequency and phase corresponding to the perceived beat in a rhythmic pattern, rule-based models suggest that some accents are assigned to certain notes of the rhythm according to a set of rules (Povel and Okkerman, 1981).

According to rule-based models, the *temporal* accents, i.e., accents which are generated by the sound intensity envelope and the inter-event intervals within a sequence, are particularly important for beat induction, as they are assumed to be sufficient to induce a beat (Povel and Essens, 1985). In such models, temporal accents would occur on (1) tones presenting a peak of intensity, (2) temporally isolated tones, (3) on the second in a group of two tones, and (4) on the first and last tone in a sequence of three or more notes (Povel and Essens, 1985). In addition, temporal accents are assumed to be more precisely determined temporally when they occur on notes having a sharp slope of intensity. Indeed, the perceptual moment of occurrence of an acoustic event does not systematically correspond to the onset of acoustic energy, but rather is influenced by other aspects such as the slope of rise in intensity of a given note.

Although there is consistent support that temporal accents are crucial for meter induction, other kinds of accents may also influence meter induction in music. For instance, *melodic* accents correspond to tones perceived as more salient because they are much higher or lower in pitch than their surrounding events (Lerdahl and Jackendoff, 1983). Moreover, *phrasal* accents are points of change in a musical contour. *Harmonic* accents arise from shifts in tonal stability within a particular musical context. *Timbral* accents are generated by changes in the sound texture, as extensively used in symphonic compositions for instance. All these accents are referred to *phenomenal metrical accents*, as they are conveyed by aspects of the physical sound structure (Repp, 2010).

In contrast with phenomenal accents, *endogenous determinants* can also influence the selection of a meter, as suggested by the fact that an induced metric structure can persist in the context of conflicting input such as syncopation or offbeat accents (Repp, 2010). Hence, the perception of metrical accents may be viewed as "*a mental construct, inferred from but not identical to the patterns of accentuation at the musical surface*" (Lerdahl and Jackendoff, 1983). According to the Dynamic Attending Theory, metrical accents would represent moments of increased attention, resulting from the interplay between phenomenal and endogenous metrical accents. In any case, it is worth noting that processing phenomenal accents still requires some interplay between the sound features and internal representation of these features. Thereby, a distinction between phenomenal and endogenous accents, referred to bottom up and top down processes by some authors (Repp, 2010) may be artificial to some extent.

Endogenous metrical accents can take various forms and are characterized by their flexibility, in contrast with the rule-based view which assumes fixed, bottom-up, meter induction. If the pattern of phenomenal accents is impoverished or ambiguous in a given musical exert, it can still induce a meter. A classic example is the so-called "tick-tock effect" (see Section II.1.3.2), also referred to "subjective rhythmization", which is based on an entirely endogenous induction of metrical accents. Moreover, the metric interpretation of an unaccented metronome can be primed by the preceding presentation of a sequence that unambiguously induces a given meter (Desain and Honing, 2003). Another way to prime a given metric interpretation is to present the music in a notation that has a time signature or bar lines (Repp, 2007). The notation indicates to the performer or to the listener, the alleged music meter, then influencing the mental construct of the music. Also, musicians can simply decide or be instructed verbally to think the music as embedded in one meter (see Study 1).

This is facilitated as long as the musical exert is sufficiently ambiguous to switch from one meter to another (Repp, 2005). Long-term exposure has also been shown to shape our perception of rhythms, as demonstrated in studies using culturally specific metric structures (Hannon and Trehub, 2005) (see also Sections II.1.3.2 and II.1.4.3). Finally, active or passive movements can also be used to impose a given meter when listening to a pattern (Phillips-Silver and Trainor, 2007) (see Section II.2.3.6).

As a pattern of metrical accent emerges in response to a rhythmic pattern, it stabilizes, becoming resistant to change (Large, 2008; London, 2004). Once stabilized, the presence of a single event with increased salience at the musical surface cannot change an unaccented pulse into an accented one. The pattern is de-stabilized only with the occurrence of strongly contradictory evidence (Cooper and Meyer, 1960). This concept of stability of metrical accents patterns allows explaining some musical structure such as syncopation. Inversely, multi-stable metric structures, in which more than one accent pattern can be perceived, offer opportunities to study how metrical accents are induced (Large, 2008; London, 2004).

Importantly, whether or not the generation of metrical accents for beat induction depends on the focus of attention is still controversial. According to the evidence reviewed in the preceding paragraphs, beat perception could be seen as the result of a wide range of processes involved in decoding sensory information, regardless of whether this leads to the emergence of a conscious experience (Bouwer and Honing, 2012). As noted by Bouwer and Honing (2012), it is most probable that both pre-attentive and attentive processes co-exist, and contribute differently to the perception of beat depending on stimulus complexity. When the beat is highly salient in the physical structure of the stimulus, automatic grouping that pertains to early auditory processes and is probably less dependent on selective attention would be sufficient to yield a robust beat perception. In contrast, when no beat can be unambiguously extracted from the acoustic stimulus, maintaining a beat perception would require selective attention and involve working memory.

II.2.2.4.2. Generation of spontaneous metrical body movements. Importantly, the perceptual cues that promote beat induction through the generation of metrical accents are sometimes distinct to those inducing spontaneous movements on the perceived meter. For example, syncopation, when constrained within a certain degree of complexity, is thought to promote the urge to move on the beat, possibly due to an automatic tendency to compensate, by the generation of an overt movement, the lack of actual acoustic cues produced on the beat (Witek et al., 2012).

Regarding the induction of metrical body movements, groove-based music is particularly interesting to study, as it refers to music typically leading to spontaneous movement and showing several characteristics that compel to move (see also Section II.1.3.1). These characteristics are for instance the induction of clear isochronous pulses through the generation of periodic peaks of sound amplitude. The periodic amplitude modulations are played at low pitch frequencies, for instance by drums, and according to a metronomic approach of timing centered around 2 Hz without *rubato* (i.e., large tempo variations, often used for expressive purpose in Western classical music). Moreover, groove-based music typically contains a large number of repetitions of a basic unit, generally consisting in 1 or 2 measures. The numerous repetitions of the basic units are aimed to generate robust predictions. However, subtle variations are introduced within and across the repeated basic units.

The *within* basic unit variations usually refer to microtiming in music (Witek, 2009; Iyer, 2002). More specifically, microtiming is a rhythmic structure in which the notes are jittered, at a temporally fine-grained degree, ahead or behind their metrical accent locations, thus playing rhythmic events “late” or “early” relative to expectations (Witek, 2009) (cf. Audio track II.2.2.4.2a[6]). These subtle time deviations are constantly repeated through the repetition of the basic units, and are integrated part of the stylistic expression of groove. Indeed, they tend to be reproduced even when participants are asked to create a perceptually regular performance on a computer by adjusting successive temporal intervals (Penel, 2000). These results suggest that these microscopic timing variations are obligatory components involved in the generation of temporal accents because they generate increased anticipation of the incoming, delayed, event (Repp, 2002).

In contrast, large-scale variations occur *acros*s the basic units, when the groove is temporally replaced with one or a few instruments performing a short and intensified rhythmic gesture, thus momentarily disrupting the repetition and taking the listener out of the groove state of listening (cf. Audio track II.2.2.4.2b[7]). These large-scale changes create structural tension towards the return of the original basic unit that has been removed (Butler, 2006).

II.2.2.5. Neural correlates of rhythm processing.

II.2.2.5.1. Functional MRI studies. Several studies have investigated the perception of musical rhythms using functional MRI. They found that rhythm perception recruits motor related areas, even in the absence of overt movement, showing activity in the premotor cortex, cerebellum, supplementary motor area and basal ganglia (Schubotz et al., 2000; Grahn and Brett, 2007; Chen et al., 2008a) (Fig. II.2.2.5.1a). The overlap with the brain activity elicited by tasks involving timing judgments (see Section II.2.2.1) suggests that

---

[6] Audio track II.2.2.4.2a. Hands up, by The Black Eyed Peas. Typically inspired by funk and soul styles, this song contains a repeated rhythmic pattern (drum, brass, etc.), the basic unit, in which microtiming variations in the accents and articulation of the tones contribute to the Groove.

[7] Audio track II.2.2.4.2b. Who Got the Rhythm, by Richard Grey vs Sugiurumn. Electro techno piece. The across-units variations contributes to re-entrain the dancefloor to the beat. Moreover, it makes proeminent the binary metric structure, within and across units.

rhythm processing shares some neural substrates with timing-related processes and motor-related processes, although it cannot be entirely excluded that these overlaps arise from activation of different neural subpopulations within the same brain region (Grahn, 2012).

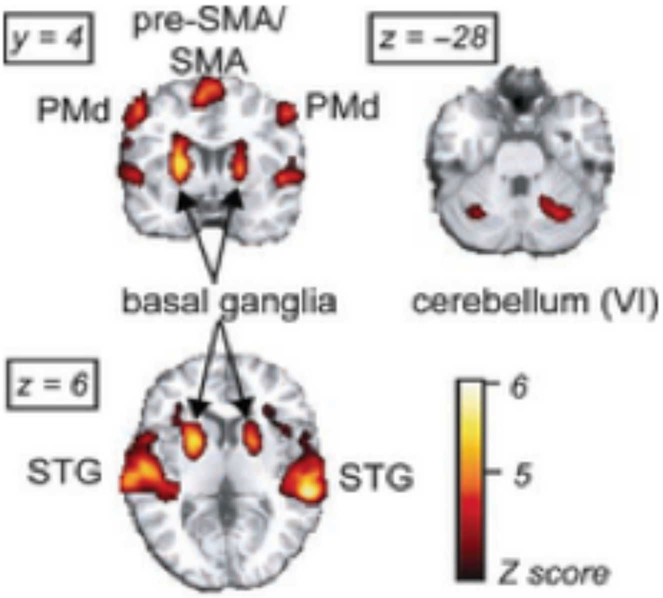


Figure II.2.2.5.1a. From Grahn (2009). Activation of the human brain recorded using functional MRI (expressed as z score values) during listening to a beat-inducing compared to a non-metric rhythm. The coronal and transversal views show activations in the supplementary motor areas (SMA) and pre-SMA, in the dorsal part of the premotor cortex (PMd), in the cerebellum, the basal ganglia and the superior temporal gyrus (STG).

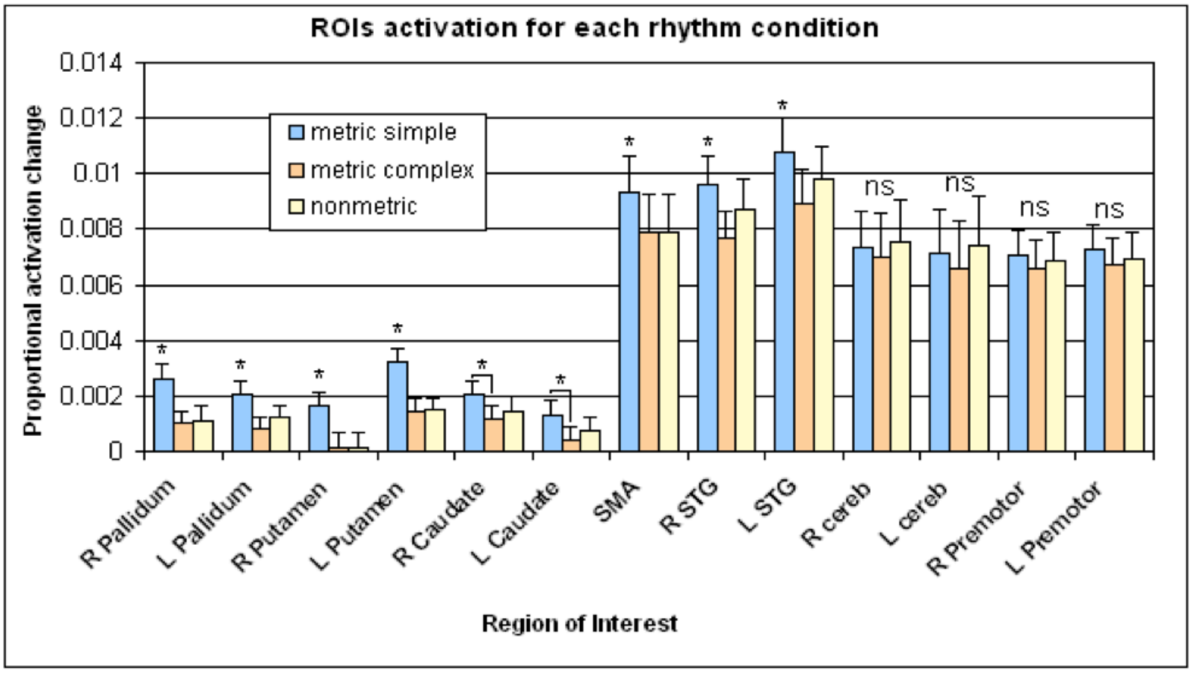


Figure II.2.2.5.1b. From Grahn (unplublished doctoral thesis). Levels of activation (proportion of blood-oxygenation-level-dependent signal change) in each region defined as region of interest for each rhythm condition (metric simple, metric complex and nonmetric). The basal ganglia, the supplementary motor area and the superior temporal plane show increased activation to metric simple rhythms, whereas the cerebellum and premotor areas do not respond significantly differently across rhythm types (ns = not significant).

A key part of rhythm perception that has been specifically addressed in functional MRI studies is beat perception. To address this question, researchers have first determined, based on performance in behavioral tasks as sensorimotor synchronization to the beat with finger tapping, the ability of some rhythms to induce a beat (Grahn and Brett, 2007; Grahn and Rowe, 2009). These authors have then classified these rhythms as metrically simple, metrically complex or non metric. Based on this classification, two functional MRI studies reported that activity in the basal ganglia was greater for regular rhythms that induce a beat perception as compared to rhythms that did not induce a beat percept (Grahn and Brett, 2007; Grahn and Rowe, 2009). Several experiments further replicated this finding by recording brain activations when participants listened without moving to rhythmic patterns classified as metric or not (Bengtsson et al., 2008; Chapin et al., 2011; Schubotz et al., 2000; Kornysheva et al., 2010).

Interestingly, the basal ganglia appear to be particularly activated during internal generation of a beat, as compared to simple external perception of a beat (Grahn, 2009). This was tested by asking participants to listen to rhythms containing strong syncopations and to selectively attend to the perceived beat, compared to rhythms in which the beat coincided systematically with actual sounds (Grahn, 2009). Again, this result was replicated further, and completed by the observation that maintaining the perception of a beat from a rhythm containing strong syncopations involves not only basal ganglia but also increased frontal lobe activity, a brain area usually active in tasks that involve working memory (Chapin et al., 2010; Vuust et al., 2006).

Regarding the role of basal ganglia in beat processing, a follow-up study performed in Parkinson's disease patients found a specific impairment in discriminating changes in the same beat-inducing rhythms. Because impairment of basal ganglia function is a key feature of Parkinson's disease, this corroborates the view that the basal ganglia may play a crucial role in beat feeling (Grahn and Brett, 2009). Further studies have confirmed the involvement of the basal ganglia in beat processing by showing impaired adaptation to tempo changes in Parkinson's disease patients (Schwartze et al., 2011). Interestingly, although Parkinson's disease patients show impaired beat processing when the beat is induced by complex rhythms, these patients seem to benefit from the pacing with a periodic sound provided either by a metronome or by a music containing clear periodic amplitude modulation of the envelope to induce the beat, to improve their ability at triggering spontaneous movements such as locomotion that are impaired by the disease (Thaut et al., 1996). Taken together, these observations on Parkinson's disease patients, while contradictory at first glance, suggest that basal ganglia is a prerequisite to process a beat when it has to be driven internally, but when impaired, this neural structure benefits from an external compensation, particularly from the auditory modality.

Recent functional neuroimaging techniques also allowed exploring, through the relative coincidence of the activation dynamics, the functional connectivity between distant brain areas. Concerning rhythm processing and beat induction, these approaches have suggested that listening to rhythms that have a perceived beat increases the functional connectivity between basal ganglia (putamen), cortical motor areas (premotor cortex and supplementary motor area) and the auditory cortex, as compared to rhythms that do not elicit a beat percept (Grahn, 2009). These results strengthen the view that beat perception involves integration and coupling between auditory and motor areas (Zatorre et al., 2007). In this line, it was shown that the planum temporale and the premotor cortex have correlated activations during the passive listening to rhythms (Chen et al., 2008) (see also Section II.2.3.5).

Importantly, while not producing movements during the fMRI acquisitions, participants are either asked to attend to the rhythms, in order to reproduce the rhythm by a movement after the scanning for instance (Chen et al., 2008), or to particularly attend to the beat when it is perceived (Grahn and Brett, 2007; Grahn and Rowe, 2009; Grahn, 2009; Chapin et al., 2011), or again to listen passively to the rhythms (Bengtsson et al., 2009) or attending to a distracting concomitant input (Chapin et al., 2011). The difference across tasks may probably explain the difference in the patterns of activation found across these studies. A striking example of discrepancy in the observed brain activations in contact with rhythms is given by the fact that some studies (Chen et al., 2008) did not find any significant activation in the basal ganglia, in contrast to others (Grahn and Brett, 2007; Grahn, 2009), which presented basal ganglia as critical for beat feeling.

II.2.2.5.2. EEG and MEG studies. Because of their fine temporal resolution, non-invasive electrophysiological techniques such as electroencephalography (EEG) and magnetoencephalography (MEG) are particularly well suited to study rhythm and beat perception in humans.

To study the electrophysiological correlates of beat perception, a number of researchers have based their approach on the fact that a key consequence of listening to rhythms is the setting up of expectancies. In order to generate these expectancies, participants are asked to listen to a continuously repeating rhythm, either an isochronous pulse as a metronome or a more complex rhythmic pattern. Occasionally, deviants are introduced in the repeating rhythm. These deviants can take the form of a change in the rhythm compared to the standard repeated rhythmic pattern, a change in the metric structure compared to the expected meter, or any other change related to the different features characterizing the sound (e.g., pitch or intensity) (Geiser et al., 2010).

By introducing such violations, researchers expect to observe ERPs related to the violation of expectancy. In the auditory modality, these components appear in the form of a mismatch negativity potential, thought to reflect brain processes involved in the detection of a mismatch between expectations and the incoming auditory input (May and Tiitinen, 2010) (see also Sections II.1.3.2 and II.1.4.2 on the concept of mismatch negativity). For example, Winkler et al. (2009) introduced sound omissions in rhythmic patterns thought to induce the perception of strong and weak beats (see Section II.1.4.2 for a more detailed description of this study). The omissions could occur either on the strong beat or on the weak beat. They found a larger mismatch negativity potential for the strong beat omission compared to the weak beat omission. A similar approach was applied to investigate metric coding in newborns (Winkler et al., 2009) (see Section II.1.4.2).

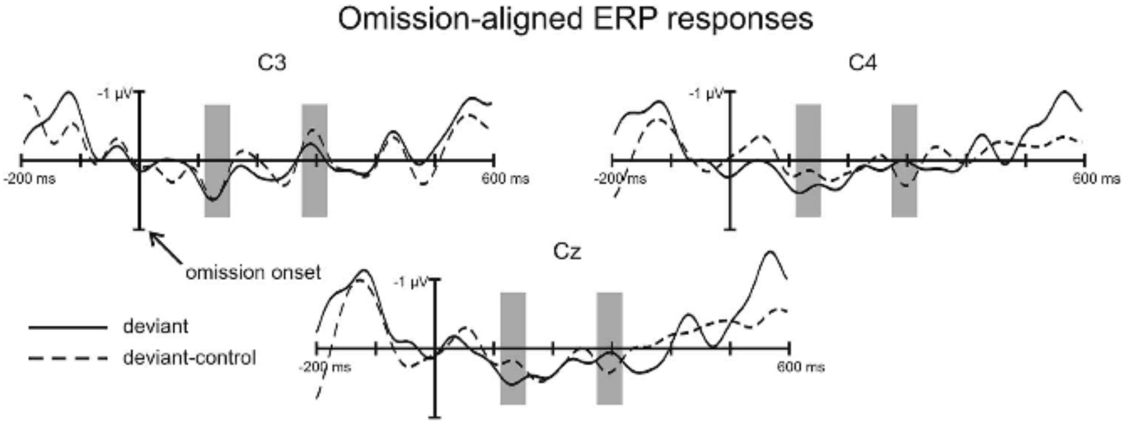


Figure II.2.2.5.2a. From Winkler et al. (2009). Mismatch-negativity potentials recorded with EEG on the scalp of adults and obtained when locked to the onset of the omission of a strong beat (deviant) compared to omission of a weak beat (deviant-control) (the grey bars correspond to the onset of following sounds in the rhythmic sequence) (see also Section II.1.4.2 for a detailed description of the paradigm).

This approach assumes that in order to be induced unambiguously a strong beat should coincide with an actual tone. However, numerous examples illustrate that highly syncopated rhythms can convey robust groove feeling and beat induction. Moreover, according to the authors having observed such mismatch negativity potential differences, the finding that rhythm-induced expectations may modulate the amplitude of the mismatch negativity argues in favor of the view that beat perception constitutes a pre-attentive process (Näätänen et al., 2007; May and Tiitinen, 2010). In fact, whether the mismatch negativity potential as evidenced in these studies reflects the detection of a mismatch regarding an internal representation of metric structure or an internal representation of basic, non rhythm-related, acoustic features remains a debated question (Bouwer and Honing, 2012) (see also Section II.1.4.2 for a more detailed description of the debate).

Another approach has been developed with the aim of identifying internal metric representations with EEG/MEG. Instead of introducing violations in a regular pattern, researchers asked the participants to listen to a rhythmic pattern and to imagine various metrical interpretations when listening to the pattern. The ERP elicited by a given sound of the pattern is then examined. If differences in ERP amplitude are observed as a function of the position of the sound relative to the imagined metric structure, this would constitute evidence that strong and weak beats are not processed equally. Fujioka et al. (2010) as well as Schaefer et al. (2011) showed that this was actually the case (Figs. II.2.2.5.2b to II.2.2.5.2e). However, although this approach allows capturing internal representations of meter, it is limited by the fact that such paradigm requires a voluntary interpretation imposed by an external instruction, and does not allow to study spontaneously induced meters directly. Moreover, because the sounds eliciting a beat percept are often not cleanly separated from one another by long-lasting periods of silence, response overlap can make it difficult to assess significant changes in ERP amplitude and latency in the time domain. Therefore, the identification of an ERP significantly elicited in response to a given sound could leave the door open to a lot of subjectivity.

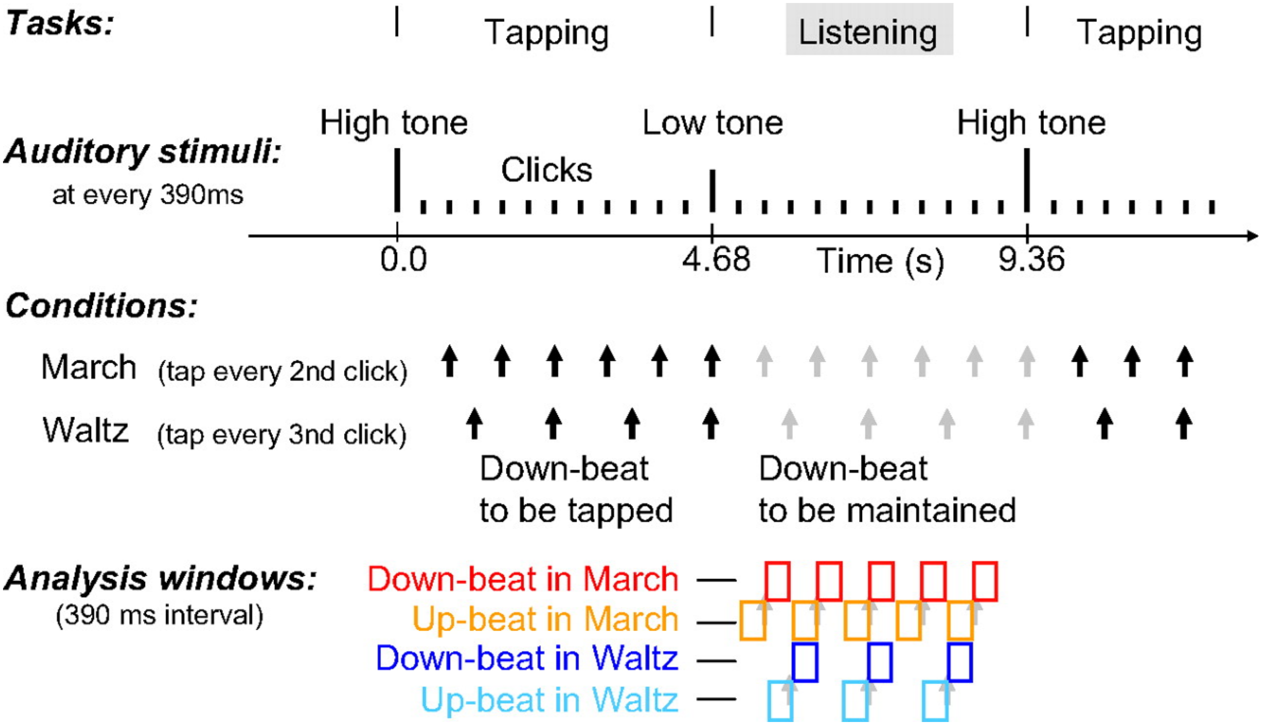


Figure II.2.2.5.2b. From Fujioka et al. (2010). Stimuli and task. Auditory stimuli were short tones presented every 390 ms. Changes in pitch cued the beginning and end of the tapping interval. After a high-pitched tone, the subjects tapped at every second click in the march (binary) condition or at every third click in the waltz (ternary) condition, in separate experimental blocks. Subjects stopped tapping at the low pitch tone and listened to the stimuli. The black arrows indicate the downbeats at which the subjects were tapping in each condition. The gray arrows indicate the subjectively maintained downbeat positions during the listening interval. Upbeats were the clicks immediately preceding the downbeat stimuli. The color-coded boxes indicate the time interval (0-390 ms from stimulus onset) of analyzed MEG data for the four conditions: downbeat in march (red), upbeat in march (orange), downbeat in waltz (blue), and upbeat in waltz (light blue), respectively.

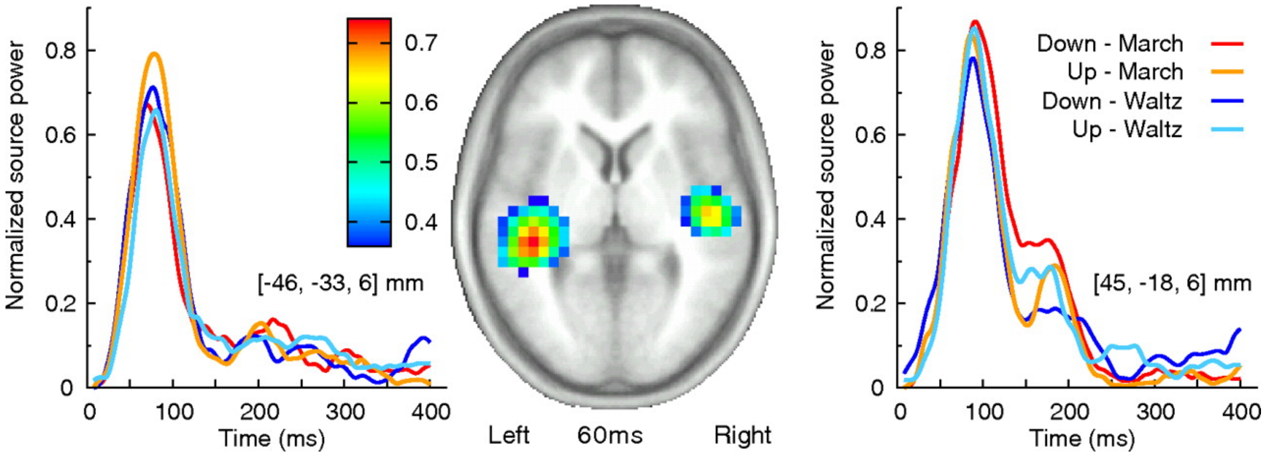


Figure II.2.2.5.2c. From Fujioka et al. (2010). Bilateral auditory evoked activities. The map (middle) shows the mean response of the four conditions in an axial slice at 60 ms latency. The left and right panels show the time course of the evoked response in each condition in the left and right auditory cortices.

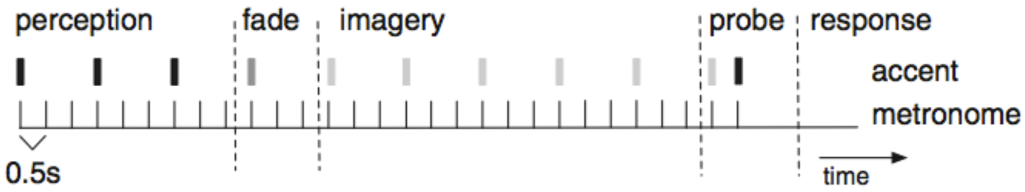


Figure II.2.2.5.2d. From Schaefer et al. (2011). A schematic overview of the stimulus sequence (here a ternary meter pattern), with a probe on an unaccented position. With an inter-onset interval of 500 ms between events, the sequence consists of 3 perceived meters, one transition meter (with fading of the accentuation), and 5 imagery meters. The first meter of perceived and imagery parts are not used for analysis. The meters were either binary, ternary, or by four beats. Comparisons were further performed between the downbeat and the first, second or third (for the 4 beats meter) unaccented tones.

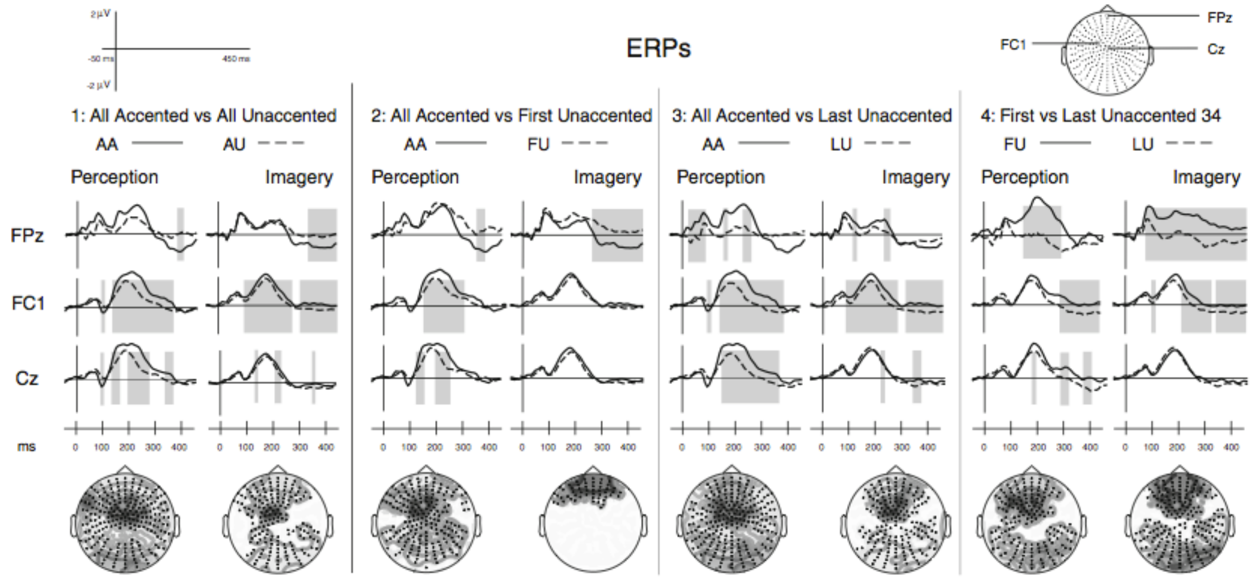


Figure II.2.2.5.2e. From Schaefer et al. (2011). ERPs of all the different comparisons, for perceived (i.e., first part of the stimuli) (left) and imagined (right) meters (x-axis: -50 to 450 ms after metronome click; y-axis: -2 to 2 µV). The distribution of channels with significant differences between these events is plotted below (significant cluster of electrodes highlighted), and the shading depicts the duration of the significant difference.

Finally, in addition to studying ERPs elicited by the different sounds constituting a rhythm, investigators have examined the dynamic increase or decrease in magnitude of beta-band (13-30 Hz) and gamma-band (> 30 Hz) ongoing oscillations. These changes were measured in studies using isochronous sequences of tones in which deviant tones were occasionally introduced, similarly as in mismatch negativity potentials paradigms (Snyder and Large, 2005; Fujioka et al., 2009) (Fig. II.2.2.5.2f). Researchers found that induced gamma-band oscillations (i.e., oscillation magnitudes obtained when averaging the trials in the time-frequency domain, thus preserving the activities non phase-locked across trials; see Fig. II.2.2.5.2g) were enhanced following occasionally omitted tones which should have occurred "on the beat" as compared to omitted tones which should have occurred "off the beat"

(Snyder and Large, 2005) (Figs. II.2.2.5.2f and II.2.2.5.2g). In this study, gamma-band oscillations were evaluated between 20 and 60 Hz.

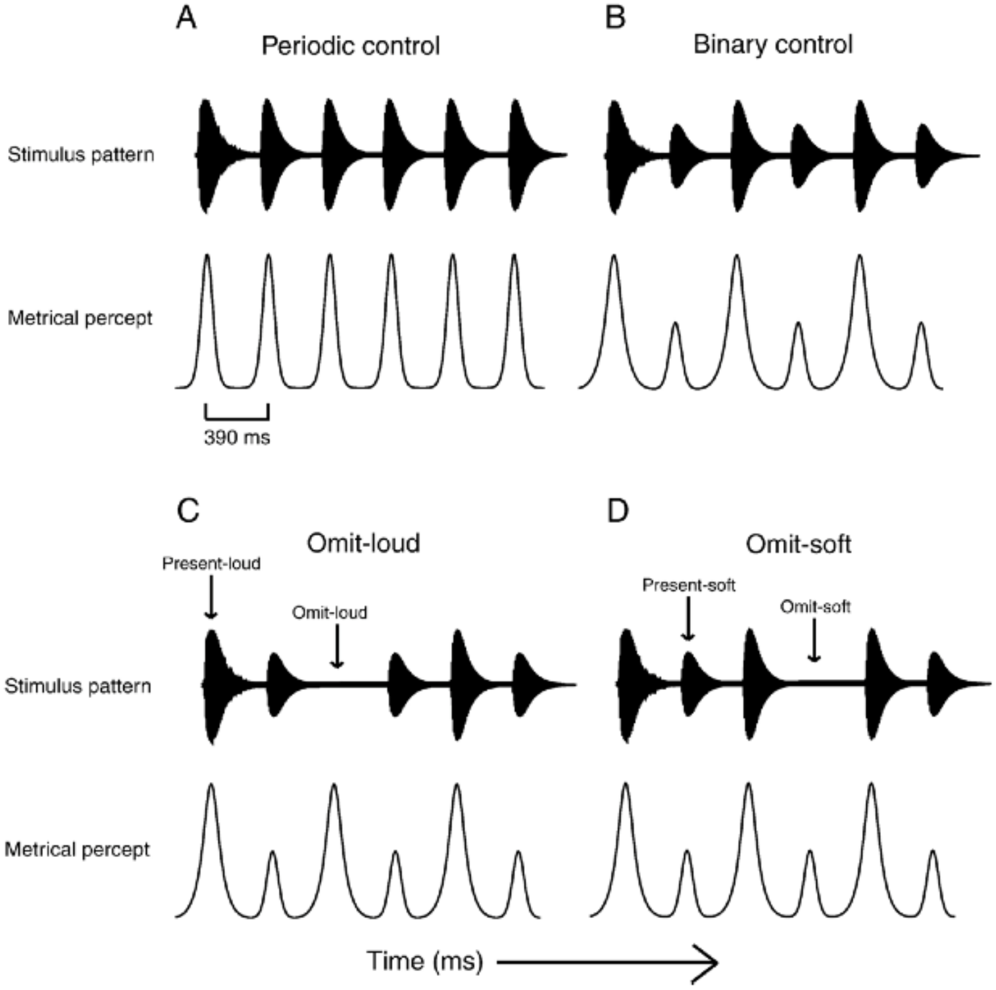


Figure II.2.2.5.2f. From Snyder and Large (2005). Pure-tone (262 Hz, 50 ms duration) stimulus patterns are shown with inter-onset intervals of 390 ms (above) and schematized metrical accent representations (below). The periodic control condition consisted of isochronous tones designed to elicit a simple pulse perception (A). The binary control condition consisted of alternating loud and soft tones, designed to elicit a duple meter perception (B). The omit-loud condition consisted of the binary control pattern with missing loud tones on 30% of two-tone cycles (C). The omit-soft condition consisted of the binary control pattern with missing soft tones on 30% of two-tone cycles.

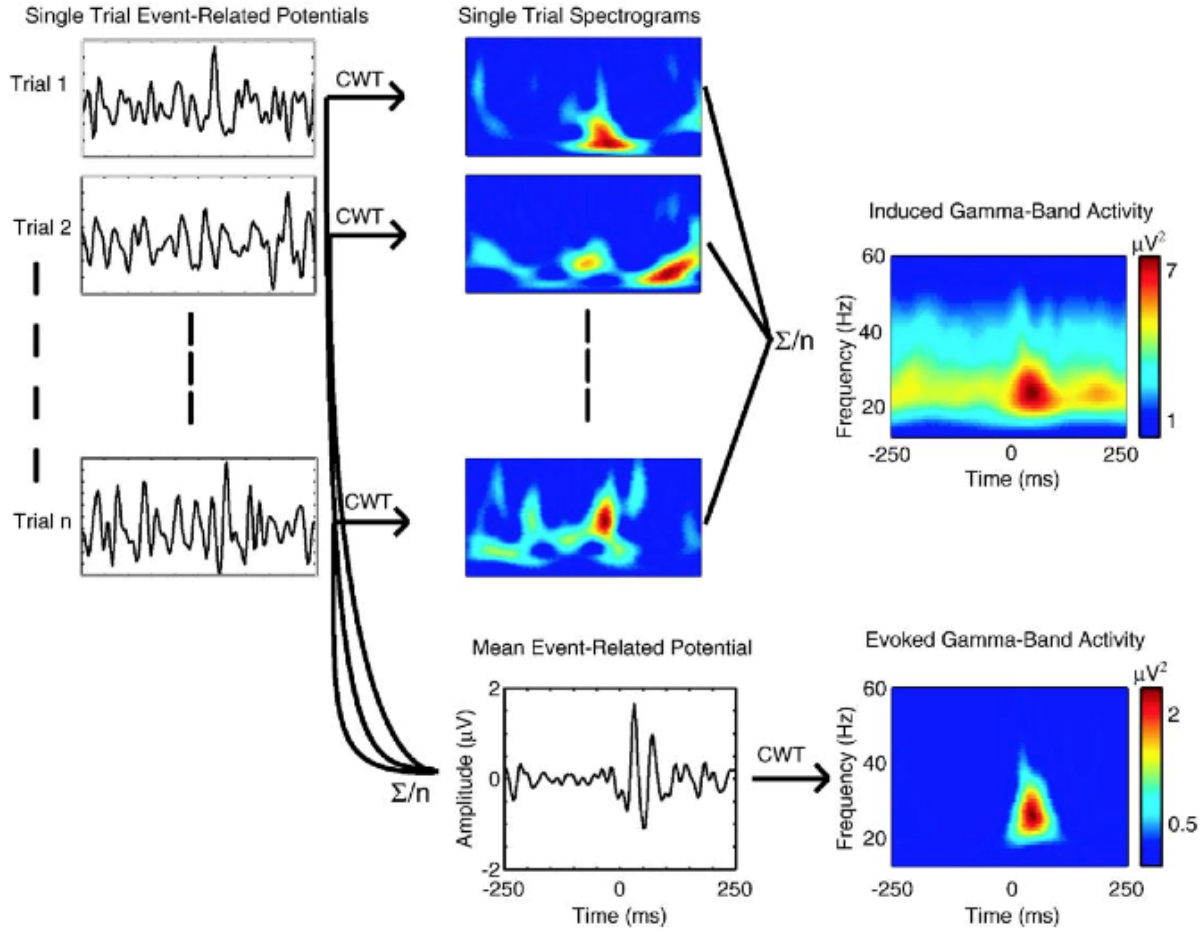


Figure II.2.2.5.2g. From Snyder and Large (2005). Calculating evoked and induced gamma-band activity.

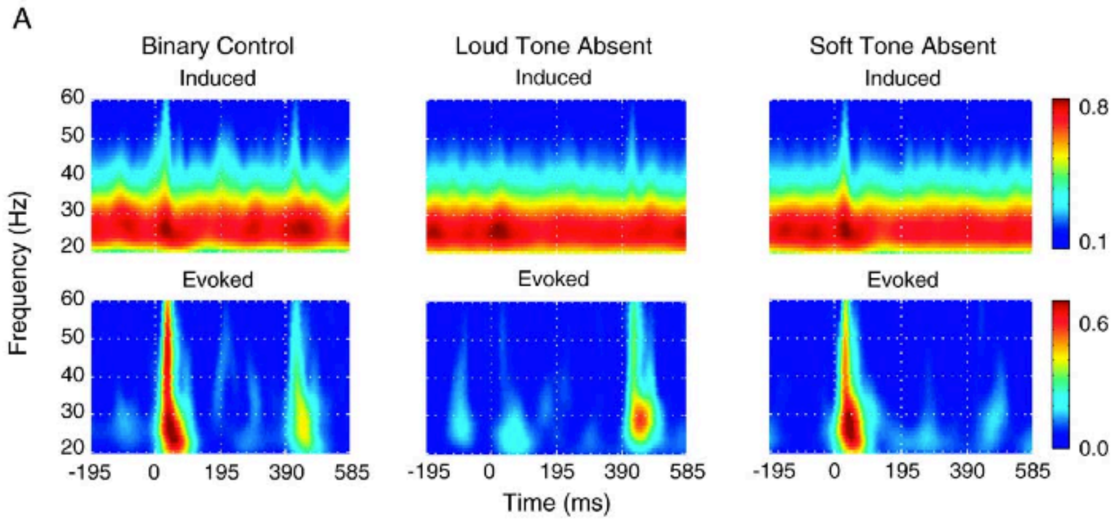


Figure II.2.2.5.2h. From Snyder and Large (2005). Grand-mean (n=8) normalized plot of time-frequency evoked and induced power for the binary control, omit-loud and omit-soft conditions.

Another study using this approach found that *evoked* gamma-band magnitude (i.e., oscillation magnitudes obtained when averaging the trials in the time domain, thus reducing or cancelling the activities non phase-locked across trials; see Fig. II.2.2.5.2g) increased immediately not only after each tone in a sequence of regularly presented tones but also after an unexpected omission, suggesting an association with anticipation (Fujioka et al., 2009) (Fig. II.2.2.5.2i). In this study, gamma-band oscillations were defined as the activity above 30 Hz. In complement, these authors also investigated beta band activity (between 15 and 30 Hz). They found that this beta activity decreased after each tone, followed by an increase, but that beta decrease was absent after omissions (Fig. II.2.2.5.2i).

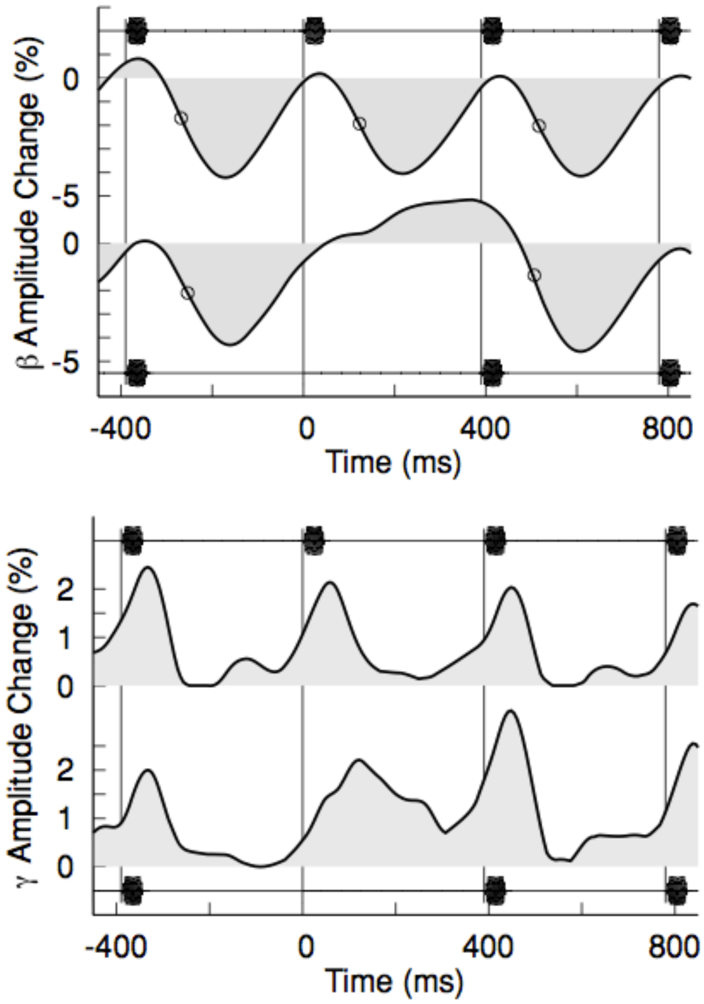


Figure II.2.2.5.2i. From Fujioka et al. (2009). Top panel. Event-related changes in beta band (15-20 Hz) activity for the stimuli (top line) and the omission (bottom line). Bottom panel. Event-related changes in gamma band (28-48 Hz) activity for the stimuli and omissions.

Using the mental imagery of meter approach as described in the preceding paragraphs, Iversen et al. (2009) found that metric interpretation changed oscillatory EEG activity in the upper beta-band range (20-30 Hz) (Fig. II.2.2.5.2h). The beta response to a tone was stronger when the tone was imagined to be the beat, as compared to when it was not. Moreover, the beta increase was very similar to that observed when the tones were physically accented.

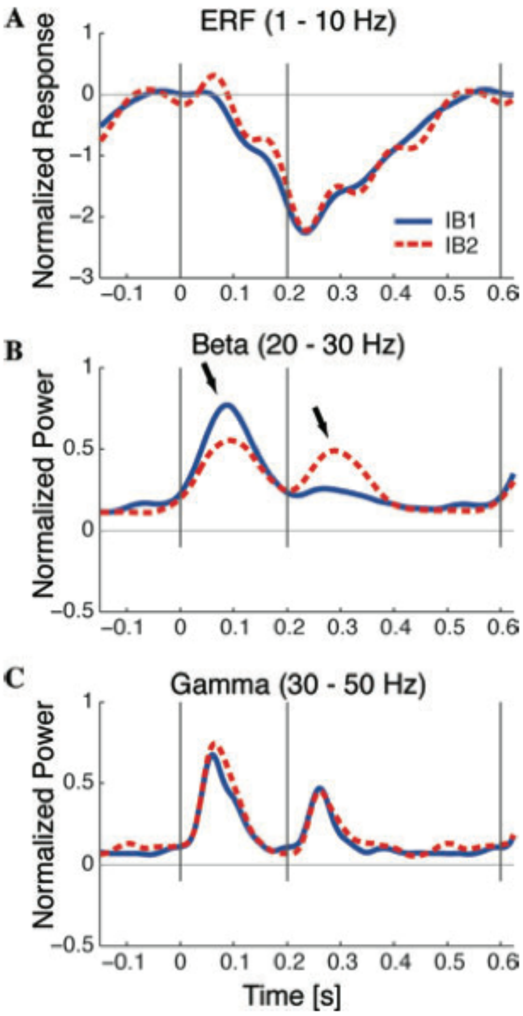


Figure II.2.2.5.2j. From Iversen et al. (2009). Across-participant (n=10) grand means of normalized evoked responses for the two imagined beat conditions (solid blue line: beat imagined on tone 1 indicated by the first black vertical line; dashed red line: beat imagined on tone 2 indicated by the second black vertical line). A. Evet-related field (1-10 Hz), B. beta (20-30 Hz), and C. gamma (30-50 Hz). For beta and gamma, the mean power envelopes (obtained after filtering and extracting the envelope using a Hilbert transform) were averaged across individuals after first normalizing each individual's peak power across both conditions to one. The arrows indicate the difference between the response to the imaging downbeat compared to imagined upbeat.

Taken together, in all these experiments investigating fast ongoing oscillations, the relevance of exploring high-frequency EEG oscillations in complement to event-related potentials was justified by the view that oscillations in the beta band may reflect processes occurring in motor cortices, basal ganglia and cerebellum (Pfurthscheller and Lopes da Silva, 1999), and that activities in the gamma band may be associated with attention, memory, anticipation and feature-binding (Pfurthscheller and Lopes da Silva, 1999). Moreover, because these ongoing oscillations have a short latency compared to classical ERPs, the authors argued in favor of studying them in response to rhythmic sequences of events to prevent from overlap between responses to successive events that are close in time (Snyder and Large, 2005). However, determining the region of interest in the time-frequency representation is not easy, as we have seen in the discrepancies across the studies in the definition of the frequency bands of interest.

II.2.2.6. Tagging the neural entrainment to beat and meter: our novel approach

Here, we present our novel approach developed in the present work. This approach is proposed as an alternative, complementary, method to tag, hopefully more directly, the neural entrainment to the beat and meter, based on their frequencies. Inspired by the SS-EP approach, which takes advantage of the experimentally determined periodicity of the input signal to identify the neural response in the frequency domain, the present approach was first tested using a paradigm of mental imagery of meter (Study 1), and then using the ability of rhythmic patterns to induce beat and meter spontaneously (Study 2).

II.2.2.6.1. Study 1: TAGGING THE NEURONAL ENTRAINMENT TO BEAT AND METER (Sylvie Nozaradan, Isabelle Peretz, Marcus Missal, André Mouraux)

This article has been published in the Journal of Neuroscience in 2011.

II.2.2.6.1.1. Abstract

Feeling the beat and meter is fundamental to the experience of music. However, how these periodicities are represented in the brain remains largely unknown. Here, we test whether this function emerges from the entrainment of neurons resonating to the beat and meter. We recorded the electroencephalogram while participants listened to a musical beat and imagined a binary or a ternary meter of this beat (i.e., a march or a waltz). We found that the beat elicits a sustained periodic EEG response tuned to the beat frequency. Most importantly, we found that meter *imagery* elicits an additional frequency tuned to the corresponding metric interpretation of this beat. These results provide compelling evidence that neural entrainment to beat and meter can be captured directly in the electroencephalogram. More generally, our results suggest that music constitutes a unique

context to explore entrainment phenomena in dynamic cognitive processing at the level of neural networks.

II.2.2.6.1.2. Introduction

Beat perception in music refers to the ability to perceive periodicities from sounds that are not necessarily periodic in reality (e.g. spontaneous head bouncing, foot tapping on the beat when listening to music) (Large, 2008). Beats can be organized in meters, corresponding to subharmonics - i.e., integer ratios - of the beat frequency.

How beat and meter are processed in the human brain remains largely unknown. The resonance theory for beat and meter perception (Large, 2008; Large and Kolen, 1994; van Noorden and Moelants, 1999), proposes that beat perception emerges from the *entrainment* of neuronal populations resonating at the beat frequency, itself giving rise to higher order resonance at subharmonics of beat frequency, corresponding to the meter. Several studies have explored the neural processes underlying beat and meter perception using original but indirect approaches, to examine how beat and meter structures create temporal expectancies and metrical representations. For example, investigators have examined how beat and meter may influence the transient evoked potentials (transient EPs) elicited by brief violations or accentuations inserted in a beat structure (Brochard et al., 2003; Snyder and Large, 2005; Grube and Griffiths, 2009), or how the transient EPs elicited by auditory stimuli may be modulated as a function of their position with respect to the beat structure (Iversen et al., 2009; Fujioka et al., 2010; Schaefer et al., 2010). All these findings are compatible with the theory of neural resonance. Indeed, neuronal oscillations have been shown to be instruments of entrainment in sensory cortices. In particular, Lakatos et al. (2008) showed that neuronal oscillations in primary sensory cortices may entrain to attended rhythmic streams. Hence, the observed modulations of transient EPs by beat and meter could be interpreted as the *consequence* of a periodic modulation of the responsiveness of the neuronal populations giving rise to these transient EPs. However, *direct* neurophysiological evidence supporting this resonance theory in the context of musical beat and meter is still lacking.

Here, we chose a novel approach: we hypothesized that neuronal entrainment to beat and meter can be captured directly as a *steady-state* evoked potential (steady-state EP), that is, the electrocortical activity generated by a population of neurons resonating by entrainment at the frequency of a periodic stimulus (Regan, 1989; Plourde, 2006; Galambos et al., 1981; Draganova et al., 2002). To this aim, we asked participants to listen to a sound pattern from which they could perceive a 2.4 Hz beat. They were asked to voluntarily imagine the meter of this beat as either binary or ternary (Fig. 1). We predicted that the beat extracted from the sound pattern would elicit a steady-state EP at the beat frequency ($f = 2.4$ Hz), and, most importantly, that the meter imagery would elicit a distinct steady-state EP at the metrically-related subharmonic of beat frequency ($f/2$ and $f/3$ for binary and ternary meters, respectively). Thereby, we aimed to tag directly the neuronal entrainment hypothesized to subtend beat perception, as well as the higher order resonance phenomenon hypothesized to underlie the representation of meter.

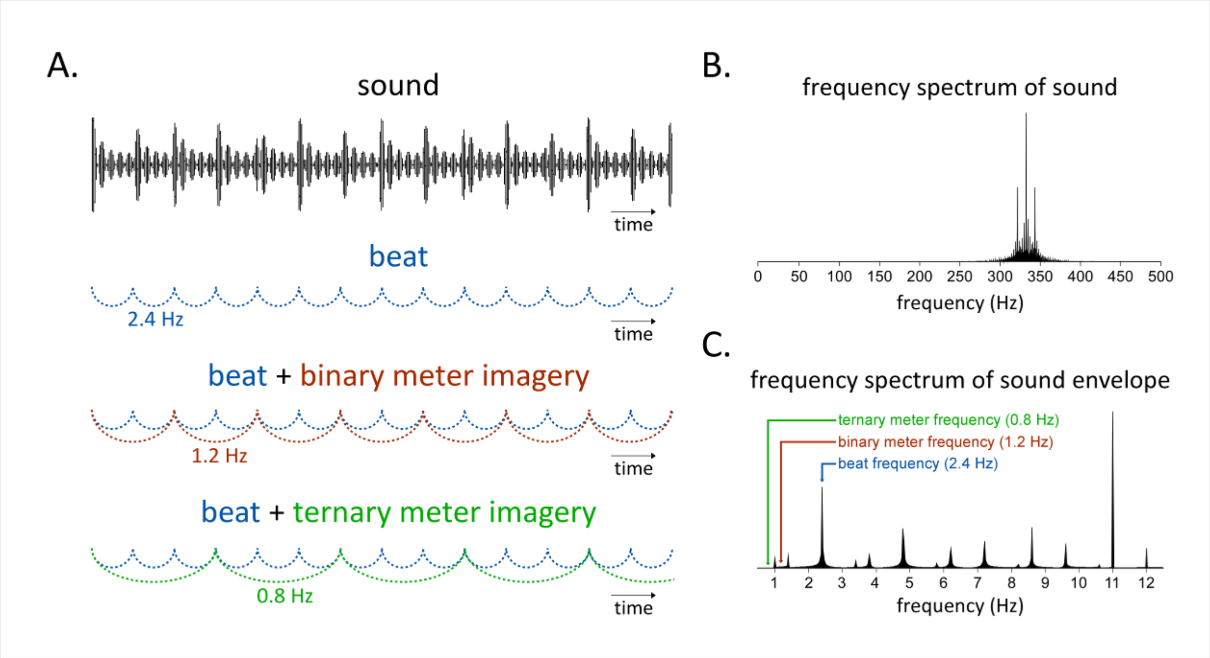


Figure 1. Experimental design. Panel A represents a 6 s excerpt of the 33-s auditory stimulus (x-axis: time; y-axis: sound amplitude). Note the pseudo-periodic beat structure, visible as a slightly irregular modulation of amplitude. From this pseudo-periodic stimulus, subjects perceived the 2.4 Hz beat represented in blue. In the binary and ternary meter conditions (shown in red and green, respectively), subjects were asked to imagine a binary (1.2 Hz) and ternary (0.8 Hz) metric structure onto this beat. Panel B represents the frequency spectrum of the sound stimulus. Note the peak corresponding to the frequency of the tone (333.33 Hz), as well as the sideband frequencies resulting from the convolution of that carrier frequency with the two different amplitude modulation frequencies (2.4 Hz and 11 Hz). Panel C represents the frequency content of the sound envelope obtained by convoluting the two different amplitude modulation frequencies (2.4 Hz and 11 Hz). Note that the sound envelope contains a peak at the frequency corresponding to the beat (2.4 Hz), but does not contain any sideband frequencies at the frequency of the binary (1.2 Hz) and ternary (0.8 Hz) meters.

II.2.2.6.1.3. Materials and methods

Participants

Eight healthy volunteers (3 females, 7 right-handed, mean age 30 ± 4 years, aged between 22 and 32) took part in the study after providing written informed consent. They all had musical experience, either in performance (3 participants with 15-25 years of practice) or as amateur listeners or dancers (5 participants). They had no history of hearing, neurological or psychiatric disorder, and were not taking any drug at the time of the experiment. The study was approved by the local Ethics Committee.

Auditory stimulation

Each auditory stimulus lasted 33 s. The stimulus consisted of a 333.33 Hz pure tone in which a 2.4 Hz auditory beat was introduced by modulating the amplitude of the tone with a 2.4 Hz periodicity (i.e., 144 beats per minute), using an asymmetrical Hanning envelope (22 ms rise time and 394 ms fall time, amplitude modulation between 0 and 1) (cf. Audio track Study 1). A 2.4 Hz periodicity was chosen because (1) pilot participants were comfortable in imagining binary (1.2 Hz) and ternary (0.8 Hz) rhythms using this 2.4 Hz tempo and (2) these tempi lie in the ecological range of tempo perception and production (Drake and Botte, 1993). The sound was then amplitude-modulated using an 11 Hz sinusoidal function oscillating between 0.3 and 1. Because the 2.4 Hz frequency was not an integer ratio of the 11 Hz frequency, the convolution of the two frequencies generated subtle irregularities in terms of amplitude and occurrence of the beats, thus resulting in a pseudo-periodic beat structure (Fig. 1A). Importantly, the frequency content of the sound envelope obtained by convoluting the two different amplitude modulation frequencies (2.4 Hz and 11 Hz) contained a peak at the frequency of the beat (2.4 Hz), but did not contain any sideband frequencies corresponding to the frequencies of the binary or ternary meters (i.e., 1.2 Hz and 0.8 Hz, respectively) (Fig. 1C). The subtle irregularities of the beat were perceived by all subjects and were purposely created to avoid induction of an involuntary binary subjective meter in the control condition (Bolton, 1894; Vos, 1973). Furthermore, the pseudo-periodicity of the beat structure, resulting from these irregularities, was closer to the more ecological situation where beat perception refers to the perception of periodicity from a non-strictly periodic framework (Large, 2008).

The auditory stimuli were generated using the PsychToolbox extensions (Brainard, 1997) running under Matlab 6.5 (The MathWork, USA), and presented binaurally through earphones at a comfortable hearing level (BeyerDynamic DT 990 PRO, Germany).

Meter mental imagery and control conditions

Participants were asked to perform three different tasks: a control task, a binary meter imagery task and a ternary meter imagery task, in separate conditions (Fig. 1). Each condition consisted of 10 trials during which the 33 s auditory stimulus was presented after a 3 s foreperiod. Stimulus presentation was self-paced. During the first condition, participants performed the control task. They were asked to detect a very short (4 ms) sound interruption which was inserted at a random position in two additional trials interspersed within the block. This control task required a sustained level of attention as the stimulus had a complex structure. The two trials containing a short interruption were excluded from further analyses. During the second condition, participants performed the binary meter imagery task. They were asked to imagine a binary metric structure onto the perceived beat ($f/2 = 1.2$ Hz). During the third condition, they performed the ternary meter imagery task, by imagining a ternary metric structure onto the beat ($f/3 = 0.8$ Hz). Before the binary and ternary meter conditions, to ensure that participants understood the task, they were asked to perform overt movements (hand tapping, aloud counting, etc.) paced to the imposed metric structure, first with the help of the experimenter, and then alone. Subjective evaluation by the experimenter of the synchrony of those movements with the meter indicated that all participants performed the task without difficulty. The participants were then asked to begin their meter imagery as soon as they heard the first auditory beat of the stimulus, and to maintain this imagery as consistently as possible throughout the entire trial. During debriefing, participants reported that they had performed the mental imagery task without difficulty, although it did require a relatively high level of attention.

EEG recording

Subjects were comfortably seated in a chair with their head resting on a support. They were instructed to relax, avoid any head or body movement during the recordings, and keep their eyes fixated on a point displayed on a computer screen in front of them. The experimenter remained in the recording room to monitor compliance to these instructions. 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 kΩ. The signals were amplified, low-pass filtered at 500 Hz and digitized using a sampling rate of 1000 Hz and referenced to an average reference (64-channel high-speed amplifier, Advanced Neuro Technology, The Netherlands).

EEG analysis

Continuous EEG recordings were filtered using a 0.1-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 to +33 s relative to the onset of the auditory stimulus at the beginning of each trial, thus yielding 10 epochs for each subject and condition. The EEG recorded during the first second of each epoch was removed (1) to discard the transient auditory evoked potential related to the onset of the stimulus (Saupe et al., 2009), (2) because previous studies have shown that steady-state EPs require several cycles of stimulation to be steadily entrained (Regan, 1989) and (3) because previous studies have shown that several repetitions of the beat are required to elicit a steady perception of beat and meter (Repp, 2005). These EEG processing steps were carried out using Analyzer 1.05 (Brain Products, Germany).

The following EEG processing steps were carried out using Letswave (Mouraux and Iannetti, 2008), Matlab (The MathWorks, USA) and EEGLAB (http://sccn.ucsd.edu).

Artifacts produced by eye blinks or eye movements were removed using a validated method based on an Independent Component Analysis (ICA) (Jung et al., 2000), using the *runica* algorithm (Makeig et al., 1996; Bell and Sejnowski, 1995), as implemented in EEGLAB. For each subject and condition, EEG epochs were averaged across trials to reduce the contribution of activities non phase-locked to the stimulation train. The time-domain averaging procedure was used to enhance the signal-to-noise ratio by attenuating the contribution of activities that were not strictly phase-locked across trials, i.e., activities that were not phase-locked to the sound stimulus.The obtained average waveforms were then transformed in the frequency domain using a discrete Fourier transform (FFTW) (Frigo and Johnson, 1998), yielding a frequency spectrum of signal amplitude (µV) ranging from 0 to 500 Hz with a frequency resolution of 0.031 Hz (Bach and Meigen, 1999).

Within the obtained frequency spectra, signal amplitude may be expected to correspond to the sum of (1) EEG activity induced by the auditory beat and/or the meter imagery task, i.e., beat- and meter-related steady-state EPs and (2) unrelated residual background noise due, for example, to spontaneous EEG activity, muscle activity or eye movements. Therefore, to obtain valid estimates of the beat- and meter-related steady-state 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 (2 frequency bins ranging from -0.15 to -

0.09 Hz and from +0.09 to +0.15 Hz relative to each frequency bin). The validity of this subtraction procedure relies on the assumption that, in the absence of a steady-state EP, the signal amplitude at a given frequency bin should be similar to the signal amplitude of the mean of the surrounding frequency bins.

Finally, the magnitude of beat- and meter-related steady-state EPs was estimated by averaging the signal amplitude measured at the three frequency bins centered on the target frequency of each steady-state EP (i.e., 2.4 Hz: bins ranging from 2.356-2.418 Hz; 1.2 Hz: bins ranging from 1.178-1.240 Hz; 0.8 Hz: bins ranging from 0.775-0.837 Hz; 1.6 Hz: bins ranging from 1.581-1.643 Hz), thereby accounting for a possible spectral leakage due to the fact that the discrete Fourier transform did not estimate signal amplitude at the exact frequency of each steady-state EP.

Statistical analyses

For each participant, condition and target frequency, the magnitude of steady-state EPs was averaged across all scalp electrodes, thus excluding any electrode selection bias (Fig. 2 and 3). This approach was used because there was no a priori assumption on the scalp topography of the beat- and meter-induced responses. Group-level results were expressed using the median and interquartile range (Fig. 4). To examine whether the beat and meter induced a significant steady-state response, one-sample t-tests were used to determine whether the noise-subtracted amplitudes measured at the target frequencies were significantly different from zero. Indeed, in the absence of a steady-state response, the average of the *subtracted* signal amplitude may be expected to tend towards zero.

To compare the beat- and meter-induced steady state responses across experimental conditions, for each target frequency, a one-way repeated-measures ANOVA was used to compare the noise-subtracted amplitudes obtained in the control, binary meter and ternary meter conditions. Degrees of freedom were corrected using the Greenhouse-Geisser correction for violations of sphericity. Size effects were expressed using the partial Eta-squared. When significant, post-hoc pairwise comparisons were performed using paired-sampled t-tests. Significance level was set at p <0.05.

Transient auditory event-related potentials

In order to examine whether the beat and meter elicited transient auditory event-related potentials that could be identifiable in the time-domain, average waveforms were computed after band-pass filtering (0.3 Hz to 30 Hz) and epoch segmentation from -1 s to +33 s relative to the onset of the sound stimulus (Figure 5).

II.2.2.6.1.4. Results

As shown at the individual level (Fig. 2) as well as in the group-level average of the global field amplitude spectra (Fig. 3), the auditory beat elicited, in all three conditions, a clear increase of EEG signal amplitude at 2.4 Hz, corresponding to the frequency of the beat, and referred to as *beat-related* steady-state EP. Furthermore, binary meter imagery elicited an additional response at 1.2 Hz (corresponding to the frequency of the binary meter) whereas ternary meter imagery elicited an additional response at 0.8 Hz (corresponding to the frequency of the ternary meter) and 1.6 Hz (corresponding to the first upper harmonic of the frequency of the ternary meter), referred to as binary and ternary *meter-related* steady-state EPs, respectively.

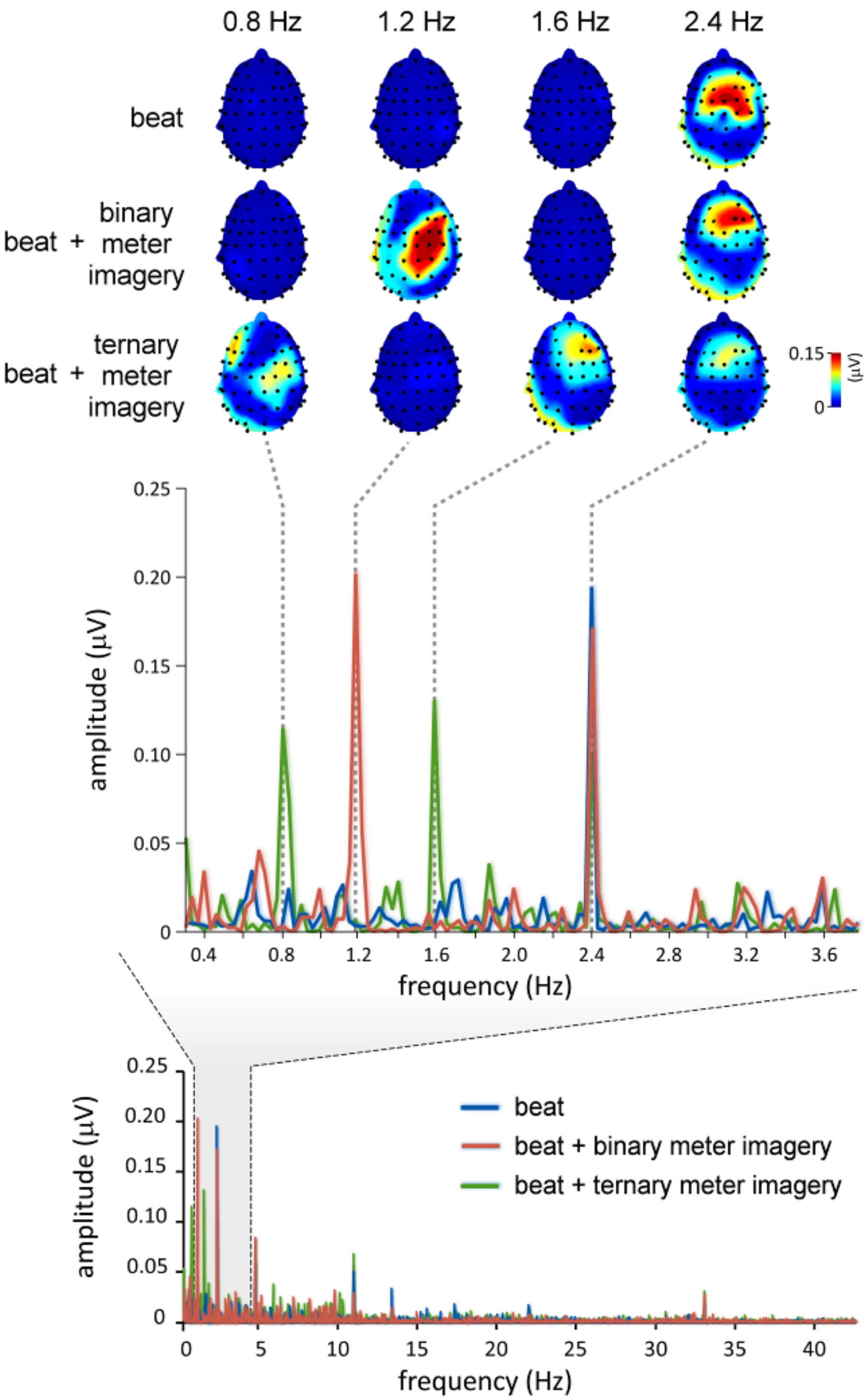


Figure 2. Beat- and meter-related steady-state EPs recorded in a single representative subject. The bottom panel represents the EEG amplitude spectrum (in microvolts, µV) from 0 to 45 Hz, averaged across all scalp electrodes, after applying the noise subtraction procedure (see Methods). The EEG spectrum obtained in the control condition is shown in blue, whereas the EEG spectra obtained in the binary and ternary meter imagery conditions are shown in red and green, respectively. The middle panel represents the EEG amplitude spectrum (µV) within a frequency range comprising the frequency of the beat (2.4 Hz) and the frequency of the imagined binary and ternary meters (1.2 Hz and 0.8 Hz, respectively). Note that in all three conditions, the auditory stimulus elicited, at f=2.4 Hz, a clear beat-

related steady-state EP. Also note the emergence of a meter-related steady-state EP at 1.2 Hz in the binary meter imagery condition, and at 0.8 Hz and 1.6 Hz in the ternary meter imagery condition. The upper panel represents the topographical maps of EEG signal amplitude at 0.8 Hz, 1.2 Hz, 1.6 Hz and 2.4 Hz, obtained in each of the three conditions.

Beat-related steady-state EP

The noise-subtracted amplitude of the 2.4 Hz beat-related steady-state EP, averaged across all scalp electrodes, was 0.23 µV (0.12 – 0.27 µV; median and interquartile range) in the control condition, 0.20 µV (0.12 – 0.28 µV) in the binary meter condition and 0.22 µV (0.14 – 0.35 µV) in the ternary meter condition (Figs. 3 and 4). This increase in signal amplitude was significant in all three conditions (control condition: t =-8.1, p <0.0001; binary meter condition: t =-4.3, p =0.003; ternary meter condition: t =-4.4, p =0.03).

The scalp topography of the beat-related steady-state EP was widely distributed over both hemispheres, and most often maximal over fronto-central and temporal regions (as shown in Fig. 2 in one subject).

The magnitude of the beat-related steady-state EP was not significantly different across conditions ($F_{1.6,11.5}$ =0.7, p =.494, η^2 =0.09) (Figs. 2, 3 and 4).

Meter-related steady-state EPs

In the binary meter condition, the noise-subtracted amplitude of the 1.2 Hz meter-related steady-state EP was 0.12 µV (0.05 – 0.24 µV). This increase in signal amplitude was significant (t =-3.1, p =0.01). In contrast, the amplitude of the noise-subtracted signal at 1.2 Hz obtained in the control condition (0.004 µV [-0.04 – 0.03 µV]) and in the ternary meter condition (0.004 µV [-0.01 – 0.03 µV]) was not significantly greater than zero (control condition: t =-0.2, p =0.81; ternary meter condition: t =-0.10, p =0.92) (Figs. 2 and 3).

The magnitude of the 1.2 Hz meter-related steady-state EP differed significantly across conditions ($F_{1.2,13.4}$ =11.5, p =0.008, η^2 =0.62). Post-hoc comparisons revealed that the magnitude of the EEG signal at 1.2 Hz was significantly greater in the binary meter condition than in the control condition (t =3.4; p =0.012) and the ternary meter condition (t =3.6; p =0.009) (see also Fig. 4).

In the ternary meter condition, the noise-subtracted amplitude of the 0.8 Hz meter-related steady-state EP was 0.18 µV (0.10 – 0.21 µV) and that of the 1.6 Hz meter-related steady-state EP was 0.08 µV (0.05 – 0.14 µV). Both increases in signal amplitude were significant (t =5.7, p =0.001 and t =26.8, p <0.0001, respectively). In contrast, the amplitude of the noise-subtracted signals obtained at 0.8 Hz and 1.6 Hz in the control condition (*0.8 Hz:* -0.02 µV [-0.05 – -0.01 µV]; *1.6 Hz:* 0.02 µV [-0.02 – 0.023 µV]) and in the binary meter condition (*0.8 Hz:* -0.002 µV [-0.07 – 0.02 µV]; *1.6 Hz:* -0.003 µV [-0.01 – 0.02 µV]) were not significantly greater than zero (control condition at *0.8 Hz:* t =0.1, p =0.91 and *1.6 Hz:* t =-0.3, p =0.74; binary meter condition at *0.8 Hz:* t =1.4, p =0.21 and *1.6 Hz : t* =-0.4, p =0.71) (Figs. 2 and 3).

The magnitude of the 0.8 Hz and 1.6 Hz meter-related steady-state EP differed significantly across conditions (*0.8 Hz:* $F_{1.8,12.4}$ =12.5, p =0.001, η^2 =0.64; *1.6 Hz:* $F_{1.2,7.9}$ =22.1, p =0.001, η^2 =0.76). Post-hoc comparisons revealed that both at 0.8 Hz and 1.6 Hz, the magnitude of the EEG signal was significantly greater in the ternary meter condition than in both the control condition (*0.8 Hz:* t =4.1, p =0.005; *1.6 Hz:* t =6.1, p <0.001) and the binary meter condition (*0.8 Hz*: t =4.0, p =0.005; *1.6 Hz*: t =10.8, p <0.001) (see also Fig. 4).

The topographical distribution of the 1.2 Hz, 0.8 Hz and 1.6 Hz meter-related steady-state EPs varied largely across subjects (see Fig. 2 for the scalp topographies obtained in one subject).

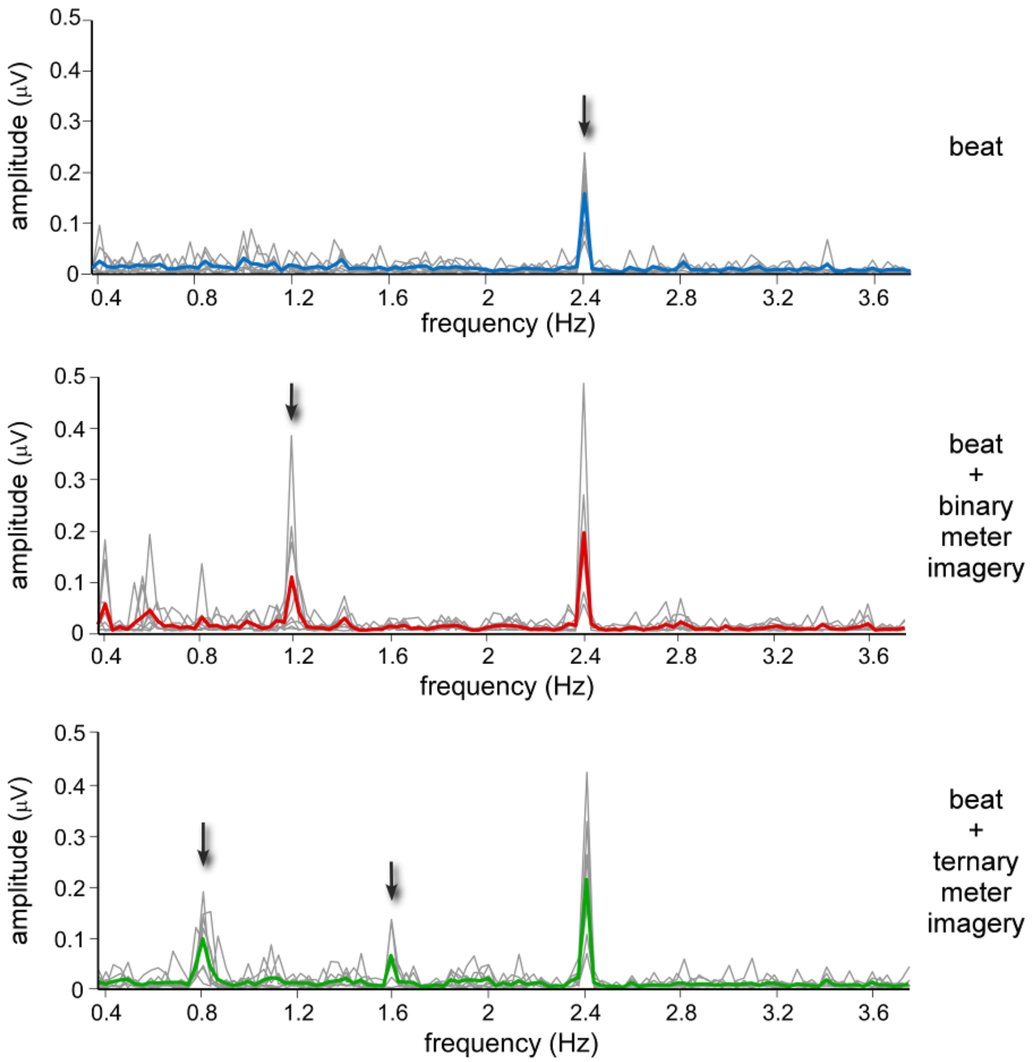


Figure 3. Group-level average of the beat- and meter-related steady-state EPs elicited by the 2.4 Hz auditory beat in the control condition (upper panel), the binary meter imagery condition (middle panel) and the ternary meter imagery condition (lower panel). The frequency spectra represent the amplitude of the EEG signal (µV) as a function of frequency, averaged across all scalp electrodes, after applying the noise subtraction procedure (see Methods). The group-level average frequency spectra are shown using a thick colored line; while single-subject spectra are shown in grey lines. Note that in all three conditions, the auditory stimulus elicited a clear beat-related steady-state EP at f=2.4 Hz. Also note the emergence of a meter-related steady-state EP at 1.2 Hz in the binary meter imagery condition, and at 0.8 Hz and 1.6 Hz in the ternary meter imagery condition.

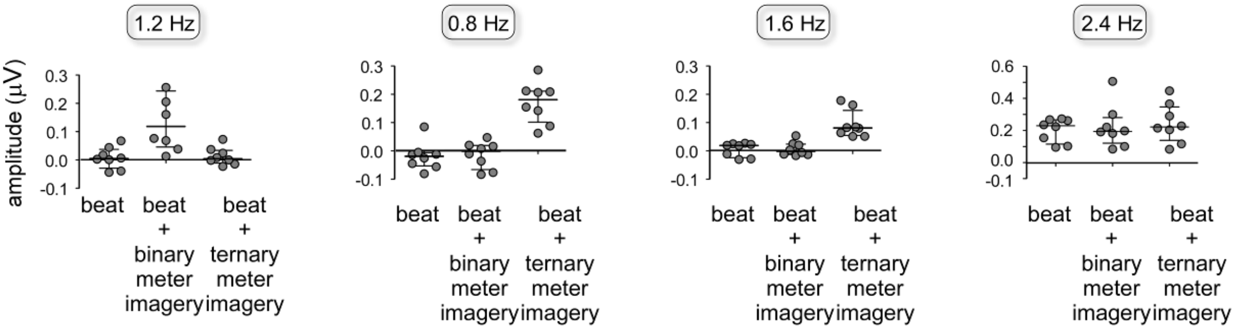


Figure 4. Amplitude of the beat- and meter-related steady-state EPs elicited in the control condition, the binary meter imagery condition and the ternary meter imagery condition. Dots represent individual amplitude values of the EEG signal for each experimental condition at each target frequency (1.2, 0.8, 1.6 and 2.4 Hz), averaged across all scalp electrodes after applying the noise subtraction procedure. The whisker plots represent the group-level median and interquartile range.

Transient auditory event-related potentials

As shown in Fig. 5, the onset of the auditory stimulus elicited a clear auditory evoked potential (N1 and P2 peaks), maximal at frontocentral electrodes. However, probably because the auditory beats were not induced by abrupt changes in the sound stream such as auditory “clicks”, the beat onsets did not a measurable transient event-related potential.

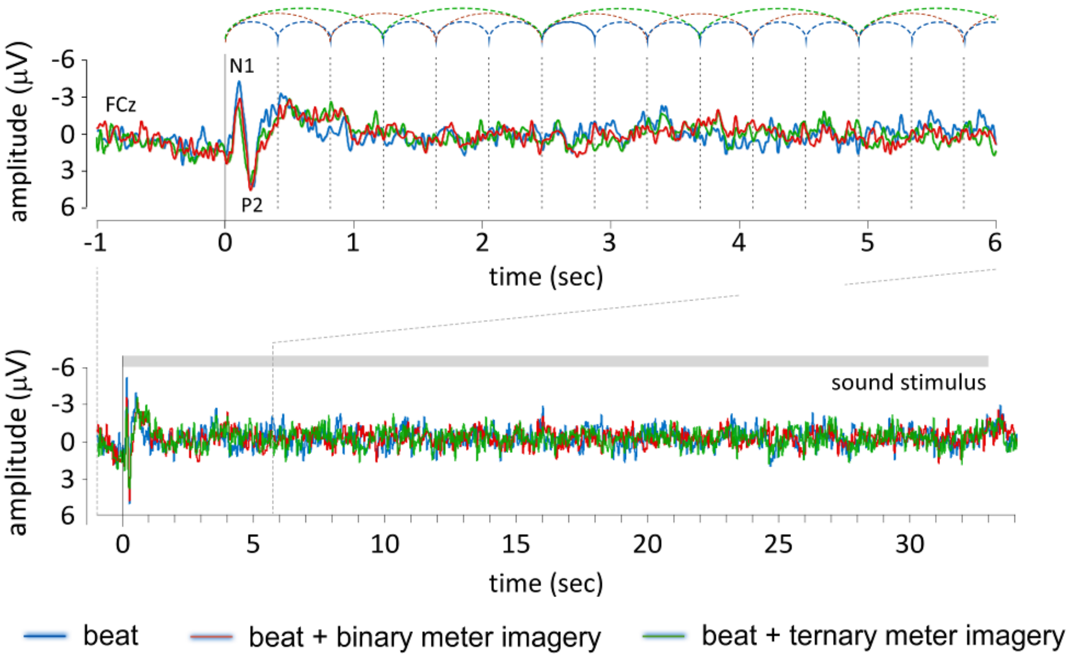


Figure 5. Transient auditory event-related potentials elicited by the 33-s sound stimulus (group-level average waveforms recorded at FCz) in the control condition (blue), the binary meter condition (red) and the ternary meter condition (green). As shown in the upper-panel, the onset of the auditory stimulus elicited a clear auditory evoked potential consisting of a negative peak (N1) followed by a positive peak (P2). In contrast, beat onsets (represented by the dashed vertical lines) did not elicit a measurable transient event-related potential.

II.2.2.6.1.5. Discussion

Our results show that neural entrainment to beat and meter can be captured directly in the human EEG as a periodic response entrained at the frequency of the beat and meter, respectively. Indeed, we found that the perception of a beat in a complex auditory signal was related to the emergence of a sustained periodic response in the EEG spectrum, appearing as a steady-state EP at the beat frequency ($f = 2.4$ Hz). More importantly, although the auditory stimulus was identical in all experimental conditions, we found that the voluntary metric interpretation of the beat as binary or ternary induced the emergence of an additional periodic signal in the EEG, at the corresponding subharmonic of beat frequency ($f/2$ =1.2 Hz and $f/3$ = 0.8 Hz for binary and ternary interpretations, respectively).

These findings provide strong and direct support to the resonance theory for beat and meter processing, which proposes that beat perception is subtended by the entrainment of neurons resonating at the beat frequency, and, that endogenous metric representations of this beat are subtended by higher order resonance products at frequencies corresponding to specific subharmonics of the beat frequency (Large, 2008; Large and Kolen, 1994; van Noorden and Moelants, 1999; Large and Snyder, 2009).

Neuronal entrainment to the beat

In the present study, the beat was induced by a continuous sound pattern whose amplitude envelope was not strictly periodic, thus requiring the endogenous reconstruction of beat periodicity. Musical beats are usually induced by acoustic features, such as a periodic modulation of loudness, melodic or harmonic accents or timbre variations (Drake et al., 2000; London, 2004). Musical beats can also be generated by mental representations shaped by prior musical experience, by the expectation of a periodicity and by a natural human tendency to generate periodic motions at rates corresponding to musical tempo range (London, 2004; Palmer and Krumhansl, 1990; van Noorden and Moelants, 1999).

How the brain performs these processes remains unclear. In the dynamic attending model proposed by Jones and collaborators (Jones and Boltz, 1989; Large and Jones, 1999), beat perception is considered to be the result of a dynamic process in which the periodic structure of the beat synchronizes or entrains the listener’s attention, leading to a periodic modulation of expectancy as a function of time. Building on this notion, the resonance theory for beat and meter perception (Large, 2008; Large and Kolen, 1994; van Noorden and Moelants, 1999) proposes that beat perception is subtended by neuronal entrainment at the beat frequency. Here, we show that the periodic activity resulting from this neuronal entrainment can be captured directly in the human EEG, as a steady-state EP. Nevertheless, whether or not the beat- and meter-induced steady-state responses reported in the present study reflected the entrainment of the neuronal populations contributing to transient auditory-evoked potentials remains, at present, an open question (Navarro Cebrian and Janata, 2010).

Furthermore, we propose that the beat-induced periodic EEG response identified in the present study may constitute a direct correlate of the actual mechanism through which attentional and perceptual processes are dynamically modulated as a function of time. Indeed, the responsiveness of the neuronal population that is entrained to the beat may be expected to vary according to the phase of the beat-induced cycle. Most importantly, this hypothesis would account for the previous observations that event-related potentials elicited at different time points relative to the beat or meter cycle exhibit differences in amplitude (Brochard et al., 2003; Snyder and Large, 2005; Grube and Griffiths, 2009; Iversen et al., 2009; Fujioka et al., 2010; Schaefer et al., 2010; Pablos Martin et al., 2007). Indeed, several electrophysiological studies in primates including humans have suggested that when the activity of a neuronal population synchronizes at a given frequency, the phase of the induced oscillations can induce a cyclic fluctuation of the excitability of the responding neuronal population, leading to an amplitude modulation of the event-related brain potentials that can be generated by these populations (Lakatos et al., 2008; Schroeder et al., 2008; Buzsáki and Draguhn, 2004; Sirota et al., 2008; Varela et al., 1981; Haig and Gordon, 1998; Makeig, 2002), but also a modulation of behavioral performance (Busch et al., 2009).

Neuronal entrainment to the meter

Musical beats can be organized in meters. As for the beat, meters in music are usually induced by accents, defined as periodic physical changes in the beat sequence such as changes in duration, loudness, timbre or pitch (Lerdahl and Jackendoff, 1983). When these accents are impoverished, ambiguous or even absent, the perception of a meter can still emerge, based on mental representations of meter (Lerdahl and Jackendoff, 1983; Repp, 2010). Perception of meter can emerge involuntarily as in the "tick tock" phenomenon (Brochard et al., 2003; Bolton, 1894; Vos, 1973) or, as in the present study, be induced voluntarily by imposing onto the beat the mental imagery of a given meter (Iversen et al., 2009; Fujioka et al., 2010; Schaefer et al., 2010).

Perceiving a given metric structure introduces additional periodicities, corresponding to integer ratios, or subharmonics, of the beat frequency (e.g. $f/2$ for a binary meter, as in a march; $f/3$ for a ternary meter, as in a waltz). A number of psychophysical and electrophysiological studies have shown that humans have a natural preference for such integer ratios in timing perception and production (Essens, 1986; Repp, 2005; Pablos Martin et al., 2007; Brochard et al., 2003), corroborating the resonance theory for beat and meter processing (Large, 2008). Our finding that an internally-driven metric structure applied onto the beat induces a periodic response in the EEG at the frequency of the applied meter suggests that the metric interpretation of the beat could emerge from the *facilitation or enhancement of specific subharmonics* within the neuronal network entrained by the beat. Importantly, because, in the present study, this metric interpretation was entirely driven by mental imagery, one must hypothesize that the emergence of these subharmonics of beat frequency can be generated through a dynamic top-down biasing of auditory beat processing.

Interestingly, when participants performed the ternary meter imagery task, this led not only to the emergence of a steady-state EP at the frequency of the corresponding meter ($f/3$ =0.8 Hz), but also at twice this frequency (1.6 Hz). The frequency of this additional steady-state EP corresponds to the frequency of the first upper harmonic of the meter frequency (i.e., $2*(f/3) = 1.6$ Hz), and could thus reflect the involuntary emergence of an additional metric level, here corresponding to a binary metric level spontaneously emerging alongside the voluntary ternary metric representation. This interpretation would suggest that the representation of multiple metric levels could be represented by such higher order resonance products, and would agree with the natural human bias for binary structures in timing perception and production (Essens, 1986; Repp, 2005; Pablos Martin et al., 2007; Brochard et al., 2003). Conversely, the frequency of this additional steady-state EP also corresponds to the frequency of the cross-modulation product between beat and meter frequencies (i.e., $f - (f/3) = 1.6$ Hz). Several studies have shown that when two or more steady-state EPs are elicited simultaneously, cross-modulation products can appear due to the non-linear convergence of the two oscillators (Regan, 1989; Sutoyo and Srinivasan, 2009). Hence, this additional steady-state EP could reflect an interaction between beat and meter processing in the human brain, in the form of a specific higher order resonance product, or “integration frequency” (Regan, 1989; Sutoyo and Srinivasan, 2009).

Conclusion

The results of the present study constitute direct experimental evidence for the entrainment of neuronal populations at the frequency of the beat, and at the subharmonics corresponding to the metric interpretation of this beat. These findings thus bring strong support for the resonance theory for beat and meter perception in humans (Large, 2008; Large and Kolen, 1994; van Noorden and Moelants, 1999). More generally, the finding that a mental representation of a given metric structure can induce a marked neuronal entrainment at the frequency of the meter provides a compelling empirical evidence of the theory of neuronal oscillations for dynamic cognitive processing (Buzsáki and Draguhn, 2004) and suggests that, due to their inherent periodic temporal structures, music and dance constitute a unique context to explore the phenomenon of entrainment at the level of neural networks.

Acknowledgements : S.N. and M.M. are supported be 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 and the Canadian Institutes of Health Research and by a Canada Research Chair. Authors would like to thank Giandomenico Iannetti (University College London), Valery Legrain (Ghent University) and Bruno Rossion (Université catholique de Louvain) for their highlighting comments and suggestions.

II.2.2.6.2. Study 2: SELECTIVE NEURONAL ENTRAINMENT TO THE BEAT AND METER EMBEDDED IN A MUSICAL RHYTHM (Sylvie Nozaradan, Isabelle Peretz, André Mouraux)

This article has been published in the Journal of Neuroscience in 2012.

The stimuli used in Study 2 are available in the Audio track file Study 2.

II.2.2.6.2.1. Abstract

Fundamental to the experience of music, beat and meter perception refers to the perception of periodicities while listening to music, occurring within the frequency range of musical tempo. Here, we explored the spontaneous building of beat and meter, hypothesized to emerge from the selective entrainment of neuronal populations at beat and meter frequencies. The electroencephalogram (EEG) was recorded while human participants listened to rhythms, consisting in short sounds alternating with silences, to induce a spontaneous perception of beat and meter. We found that the rhythmic stimuli elicited multiple steady-state evoked potentials (SS-EPs), observed in the EEG spectrum at frequencies corresponding to the rhythmic patterns envelope. Most importantly, the amplitude of the SS-EPs obtained at beat and meter frequencies were selectively enhanced, even though the acoustic energy was not necessarily predominant at these frequencies. Furthermore, accelerating the tempo of the rhythmic stimuli, so as to move away from the range of frequencies at which beats are usually perceived, impaired the selective enhancement of SS-EPs at these frequencies. The observation that beat- and meter-related SS-EPs are selectively enhanced at frequencies compatible with beat and meter perception indicates that these responses do not merely reflect the physical structure of the sound envelope but, instead, reflect the spontaneous emergence of an internal representation of

beat, possibly through a mechanism of selective neuronal entrainment within a resonance frequency range. Taken together, these results suggest that musical rhythms constitute a unique context to gain insight on general mechanisms of entrainment, from the neural to individual level.

II.2.2.6.2.2. Introduction

Feeling the beat is fundamental to the experience of music (London, 2004). It refers to the spontaneous ability to perceive periodicities (as expressed for instance through periodic head nodding or foot tapping), in musical stimuli that are not strictly periodic in reality (London, 2004; Phillips-Silver et al., 2010). This phenomenon is well illustrated by syncopated rhythms, that is, rhythmic patterns in which the perceived beat does not systematically coincide with an actual sound (Velasco and Large, 2011; Fitch and Rosenfeld, 2007). Moreover, beats are usually perceived within meters (e.g. a waltz, which is a three-beat meter), corresponding to (sub)harmonics – i.e., integer ratios – of beat frequency. These multiple periodic levels are nested hierarchically. Among these, the beat may be considered as the most prominent periodicity (London, 2004). Finally, perception and movement on beat and meter has been shown to occur within a specific range of tempo, corresponding to frequencies around 2 Hz (van Noorden and Moelants, 1999; London, 2004; Repp, 2005; 2006), and assimilated to a resonance frequency range within which internal representations of beat and meter would be optimally induced by external inputs (van Noorden and Moelants, 1999; Large, 2008).

How the brain spontaneously builds internal beat representations from music remains unknown. Recently, we showed that listening to simple periodic sounds elicits periodic neuronal activities frequency-tuned to the sound envelope periodicity. This neuronal entrainment was captured in the human electroencephalogram (EEG) as a beat-related steady-state evoked potential (SS-EP), appearing at the exact frequency of the beat in the EEG spectrum (Nozaradan et al., 2011).

Here, using this novel approach, we explored the EEG activity induced by listening to complex rhythmic patterns, which can be assimilated to syncopated rhythms and are commonly found in Western music. The patterns consisted in sequences of events, that is, sounds alternating with silences (Fig. 1), such as to induce a spontaneous perception of a beat and meter, based on the preferential grouping of four events (Povel and Essens, 1985). This was confirmed by a task performed at the end of the EEG recording, in which participants were asked to tap along the beat spontaneously perceived in each pattern. Therefore, building on prior assumptions (Povel and Essens, 1985), which were confirmed by the tapping task, the multiple frequencies constituting the envelope spectrum of the five sound patterns were categorized as either (a) related to the beat and metric levels (integer ratio subdivisions and groupings of the beat period) or (b) unrelated to beat and meter frequencies.

We expected that these patterns would elicit multiple SS-EPs at frequencies corresponding to the patterns envelope spectrum in the EEG spectrum. Most importantly, we aimed to capture the spontaneous building of internal beat and meter representations, hypothesized to emerge from a non-linear transformation of the acoustic inputs. Specifically, we examined whether the SS-EPs elicited at frequencies corresponding to the expected perception of beat and meter were selectively enhanced, as such an observation would constitute evidence for a selective neuronal entrainment underlying beat and meter neuronal representations.

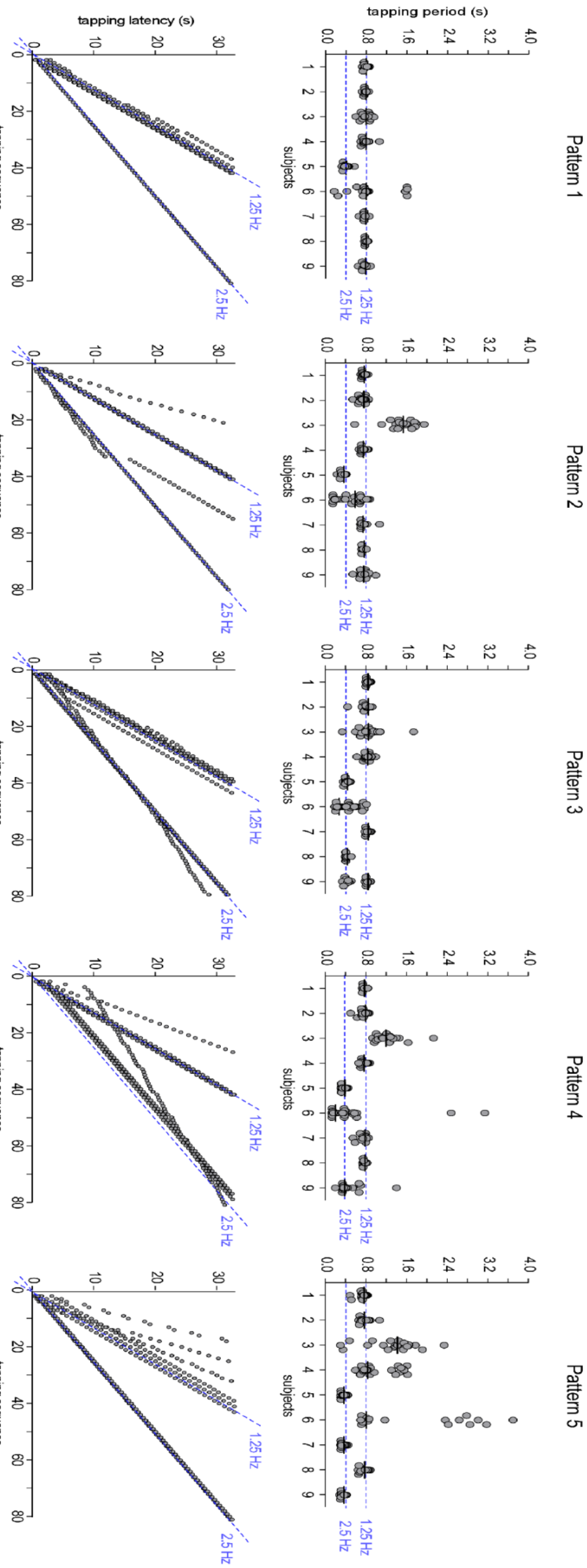

Pattern 1
Pattern 2
Pattern 3
Pattern 4
Pattern 5
tapping period (s)
0.0
0.8
1.6
2.4
3.2
4.0
subjects
1
2
3
4
5
6
7
8
9
1.25 Hz
2.5 Hz
tapping latency (s)
0
10
20
30
tapping sequence
0
20
40
60
80

Figure 1. In the last trial of each block, participants were asked to perform a rhythmic hand tapping movement synchronized to the perceived beat. Upper graphs. Tapping periods produced by each participant while listening to each of the five rhythmic patterns. Each dot corresponds to an individual tapping period. The median tapping period is represented by the horizontal black line. Lower graphs. Tapping latencies along the entire trial for each participant and each sound pattern. Note that most participants tapped at a frequency corresponding to a grouping by 4 events (1.25 Hz), except in pattern 5 where tapping was much less consistent.

II.2.2.6.2.3. Materials and methods

Participants

Nine healthy volunteers (4 females, all right-handed, mean age 29 ± 4 years) took part in the study after providing written informed consent. They all had musical experience in Western music, either in performance (3 participants with 15-25 years of practice) or as amateur listeners or dancers. None had prior experience with the tapping task used in the present study. They had no history of hearing, neurological or psychiatric disorder, and were not taking any drug at the time of the experiment. The study was approved by the local Ethics Committee.

Experiment 1: SS-EPs elicited by five different rhythmic patterns

Auditory stimuli

The stimulus consisted of 5 distinct rhythmic patterns lasting either 2.4 s (patterns 1, 3 and 4) or 3.2 s (patterns 2 and 5), looped continuously during 33 s. The structure of the patterns was based on the alternation of events, i.e., sounds and silence intervals of 200 ms duration. The sounds consisted of 990 Hz pure tones lasting 200 ms (10 ms rise and fall time). The patterns, inspired by the work of Povel and Essens (1985), were designed to induce the perception of a beat based on the preferential grouping of 4 events (i.e., a period of 0.8 s, corresponding to a 1.25 Hz beat), and at related metric levels. The related metric levels were constituted (a) by the subdivision of the beat periods by 2 (2.5 Hz) and by 4 (5 Hz, thus corresponding to the unitary event period at 0.2 s), and (b) by the integer ratio grouping of beat period by 2 (0.625 Hz) and 4 (0.312 Hz) in patterns 2 and 5 (because these patterns contained 16 events, thus allowing groupings by 2 x 4 and 4 x 4 events, respectively), and by 3 (0.416 Hz) in patterns 1, 3 and 4 (because these patterns contained 12 events, thus allowing groupings by 3 x 4 events) (Fig. 2).

The auditory stimuli were created in Audacity 1.2.6 (http://audacity.sourceforge.net/) and presented binaurally through earphones at a comfortable hearing level (BeyerDynamic DT 990 PRO, Germany), using the PsychToolbox extensions (Brainard, 1997) running under Matlab 6.5 (The MathWork, USA).

Experimental conditions

The five rhythmic patterns were presented in separate blocks. In each block, the 33-s auditory pattern was repeated 11 times. The onset of each pattern was self-paced, and preceded by a 3-s foreperiod. The order of the blocks was counter-balanced across participants.

During the first 10 trials of each block, participants were asked to listen carefully to the stimulus in order to detect the occurrence of a very short acceleration (duration of two successive events reduced by 10 ms, i.e., 190 ms) or deceleration (duration of two successive events increased by 10 ms, i.e., 210 ms) of tempo, inserted at a random position in two of the trials interspersed within the block. The participants were instructed to report the detection of the change in tempo at the end of each trial. This task ensured that participants focused their attention on the temporal aspects of the presented sound. The two trials containing a short tempo change were excluded from further analyses.

During the 11th trial of each block, participants were asked to perform a tapping task, in order to assess their perception of a periodic beat in each of the five patterns. They were instructed to tap to the regular periodic strong beat of the patterns, similarly as what they would in a concert when spontaneously entrained to clap their hands on the beat of music. Moreover, participants were asked to start tapping as soon as they heard the first beat of the trial, and to maintain their movement accurately paced on the beat that they spontaneously perceived from the patterns. The tapping was performed using their right hand, with small up and down movements of the hand starting from the wrist joint, maintaining the forearm and elbow fixed on an armrest cushion. When performing the tapping movement, the fingers of the tapping hand came transiently in contact with the armrest cushion. All participants naturally synchronized their movement such that the occurrence of this contact coincided with the occurrence of the beat. The experimenter remained in the recording room at all times, to monitor compliance to these instructions. The tapping movements were recorded using an accelerometer placed on the tapping hand (as explained below).

EEG recording

Participants were comfortably seated in a chair with the head resting on a support. They were instructed to relax, avoid any unnecessary head or body movement and keep their eyes fixated on a point displayed on a computer screen in front of them. 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 on the inferior and superior areas of the left orbit. Electrode impedances were kept below 10 kΩ. The signals were recorded using an average reference, amplified, low-pass filtered at 500 Hz and digitized using a sampling rate of 1000 Hz (64-channel high-speed amplifier, Advanced Neuro Technology, The Netherlands).

Hand movement recordings

Movements of the hand were measured using a 3-axis accelerometer attached to the hand dorsum (MMA7341L, Pololu Robotics & Electronics, USA). The signals generated by the accelerometer were digitized using three additional bipolar channels of the EEG system. Only the vertical axis of the accelerometer signal was analyzed, as it sampled the greatest part of the accelerations related to the tapping movement.

Hand movement analysis

The accelerometer signals recorded when participants performed the hand tapping movements in the last trial of each block were analyzed by extracting, for each rhythmic pattern, the latencies at which the fingers hit the armrest cushion, corresponding to the time points of maximum deceleration (Fig. 1). Tapping period estimates were then obtained by subtracting, from each tapping latency, the latency of the preceding tapping (Fig. 1).

Sound pattern analysis

To determine the frequencies at which steady-state evoked potentials (SS-EPs) were expected to be elicited in the recorded EEG signals, the temporal envelope of the 33-s sound patterns was extracted using a Hilbert function, yielding a time-varying estimate of the instantaneous amplitude of the sound envelope, as implemented in the MIRToolbox (Lartillot and Toiviainen, 2007). The obtained waveforms were then transformed in the frequency domain using a discrete Fourier transform (Frigo and Johnson, 1998), yielding a frequency spectrum of envelope magnitude (Bach and Meigen, 1999). The frequencies of interest were determined as the set of frequencies ≤ 5 Hz, i.e., the frequency corresponding to the 200 ms period of the unitary event of the patterns. As shown in Figure 2, the envelopes of patterns 1, 3 and 4 consisted of 12 distinct frequencies ranging from 0.416 Hz to 5 Hz with an interval of 0.416 Hz, whereas the envelopes of patterns 2 and 5 consisted of 16 distinct frequencies ranging from 0.312 Hz to 5 Hz with an interval of 0.312 Hz.

Within each pattern, z score values were then computed across the magnitude obtained at each of these frequencies in the spectra of the patterns envelope, as follows: $z = (x - \mu)/\sigma$, where μ and σ corresponded to the mean and standard deviation of the magnitudes obtained across the different peaks. This procedure allowed assessing the magnitude of each frequency relative to the others and, thereby, to determine which frequencies stood out relative to the entire set of frequencies.

EEG analysis in the frequency domain

The continuous EEG recordings were filtered using a 0.1-Hz high-pass Butterworth zero-phase filter to remove very slow drifts in the recorded signals. Epochs lasting 32 s were obtained by segmenting the recordings from +1 to +33 s relative to the onset of the auditory stimulus. The EEG recorded during the first second of each epoch was removed (a) to discard the transient auditory evoked potentials related to the onset of the stimulus (Saupe et al., 2009; Nozaradan et al., 2011; Nozaradan et al., 2012), (b) because steady-state evoked potentials require several cycles of stimulation to be steadily entrained (Regan, 1989) and (c) because several repetitions of the beat are required to elicit a steady perception of beat (Repp, 2005). These EEG processing steps were carried out using Analyzer 1.05 (Brain Products, Germany).

Artifacts produced by eye blinks or eye movements were removed from the EEG signal using a validated method based on an Independent Component Analysis (Jung et al., 2000), using the runica algorithm (Bell and Sejnowski, 1995; Makeig 2002). For each subject and condition, EEG epochs were averaged across trials. The time-domain averaging procedure was used to enhance the signal-to-noise ratio of EEG activities time-locked to the patterns. The obtained average waveforms were then transformed in the frequency domain using a discrete Fourier transform (Frigo and Johnson, 1998), yielding a frequency spectrum of signal amplitude (µV) ranging from 0 to 500 Hz with a frequency resolution of 0.031 Hz (Bach and Meigen, 1999). This procedure allowed assessing the neuronal entrainment to beat and meter, i.e., the appearance of frequency components in the EEG elicited by the frequency components of the sound patterns and induced beat percept (Pikovsky et al., 2001). Importantly, the deliberate choice of computing Fourier transforms of long-lasting epochs was justified in the present experiment by the fact that (1) beat and meter perception is assumed to be stationary enough along the trials, as suggested by the results of the tapping task, and (2) it improves the resolution of the obtained EEG frequency spectrum. Indeed, this allows concentrating the magnitude of the SS-EP into a very narrow band, necessary to enhance their signal-to-noise ratio as well as to disentangle between nearby SS-EP frequencies in the EEG spectrum (Regan, 1989). These EEG processing steps were carried out using Letswave4 (Mouraux and Iannetti, 2008), Matlab (The MathWorks, USA) and EEGLAB (http://sccn.ucsd.edu).

Within the obtained frequency spectra, signal amplitude may be expected to correspond to the sum of (a) stimulus-induced SS-EPs and (b) unrelated residual background noise due, for example, to spontaneous EEG activity, muscle activity or eye movements. Therefore, to obtain valid estimates of the 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 (2 frequency bins ranging from -0.15 to -0.09 Hz and from +0.09 to +0.15 Hz relative to each frequency bin). The validity of this subtraction procedure relies on the assumption that, in the absence of an SS-EP, the signal amplitude at a given frequency bin should be similar to the signal amplitude of the mean of the surrounding frequency bins (Mouraux et al., 2011; Nozaradan et al., 2011; Nozaradan et al., 2012). This subtraction procedure is important (a) to assess the scalp topographies of the elicited SS-EPs, as the magnitude of the background noise is not equally distributed across scalp channels, and (b) to compare the amplitude of SS-EPs elicited at distinct frequencies, as the background noise magnitude is not equally distributed across the frequency spectrum.

The magnitude of the SS-EPs was then estimated by taking the maximum noise-subtracted amplitude measured in a range of 3 frequency bins centered over the expected SS-EP frequency, based on the spectrum of the sound envelope. This range of frequencies allowed accounting for possible spectral leakage due to the fact that the discrete Fourier transform did not estimate signal amplitude at the exact frequency of any of the expected SS-EPs (Nozaradan et al., 2011; Nozaradan et al., 2012).

To exclude any electrode selection bias, SS-EP magnitudes were averaged across all scalp electrodes, for each rhythmic pattern and participant (Fig. 2). A one-sample t-test was then used to determine whether the average SS-EP amplitudes were significantly different from zero (Fig. 2). Indeed, in the absence of an SS-EP, the average of the noise-subtracted signal amplitude may be expected to tend towards zero. Finally, for each frequency, topographical maps were computed by spherical interpolation (Fig. 3).

Like for the sound pattern analysis, the amplitude of the SS-EPs obtained at the expected frequencies were expressed as z-scores, using the mean and standard deviation of the magnitudes obtained across the different peaks, to assess how each of the different SS-EPs stood out relative to the entire set of SS-EPs (Figs. 3 and 4).

To assess specifically whether SS-EPs elicited at frequencies related to beat and meter perception (0.416 Hz, 1.25 Hz, 2.5 Hz and 5 Hz in patterns 1, 3 and 4; 0.312 Hz, 0.625 Hz, 1.25 Hz, 2.5 Hz and 5 Hz in patterns 2 and 5) were selectively enhanced, the average of the z score values representing SS-EP amplitude at beat- and meter-related frequencies was compared to the average of the z-score values representing these same frequencies in the sound pattern envelope, using a one-sample t-test (Fig. 3). A similar procedure was used to compare the magnitude of SS-EPs and the magnitude of the sound envelope at frequencies unrelated to the beat and meter. Significance level was set at $p < 0.05$.

Finally, to compare the magnitude of each of the different SS-EPs obtained in each of the five sound patterns, relative to the magnitude of the sound envelope, a one-sample t-test was used to compare the standardized SS-EP amplitudes to the standardized sound envelope magnitudes, at each frequency constituting the envelope spectrum of the rhythmic patterns (Fig. 4).

EEG analysis in the time domain

The time course of the EEG signals recorded during presentation of the sound patterns of Experiment 1 was examined, to give a better sense of the stimulus-response relationship. These signals were obtained after band-pass filtering the signals between 0.1 and 30 Hz, segmenting the EEG epochs according to the length of each pattern (2.4 s length in patterns 1, 3 and 4; 3.2 s length in patterns 2 and 5), and averaging these epochs (Fig. 5).

Experiment 2: SS-EPs elicited by pattern 1 presented at upper musical tempi

All participants took part in a second experiment, performed on a different day. In this experiment, participants listened to pattern 1 of the first experiment, presented at either two or four times the original tempo. The two accelerated sound patterns were presented in separate blocks. The order of the blocks was counterbalanced across participants. The faster tempi were obtained by reducing the duration of the pattern events from 200 ms to 100 ms (tempo x 2) and 50 ms (tempo x 4). The task was similar to the task performed in Experiment 1. The sound pattern, EEG and movement signals were analyzed using the same procedures (Figs. 6-8). Pattern 1 was chosen for this second experiment because it appeared to elicit the most consistent beat percept, as compared to the other patterns (Fig. 1). Importantly, increasing the tempo did not substantially distort the envelope spectra of pattern 1 (Fig. 7). Indeed, it kept relatively intact the balance of frequencies relative to each other in the envelope spectrum. For example, in both the original and the accelerated versions of pattern 1, the most salient frequency remained the third frequency present in the envelope spectrum. These accelerated patterns could thus be considered as suitable to examine the effect of tempo on the magnitude of the elicited SS-EPs.

Experiment 3: SS-EPs elicited by pattern 2 presented at upper musical tempi

Six participants (three of which took part in Experiments 1 and 2) took part in a third experiment, performed on a different day. In this experiment, participants listened to pattern 2 of the first experiment, presented at the original tempo, and at two and three times the original tempo. The faster tempi were obtained by reducing the duration of the pattern events from 200 ms to 100 ms (tempo x 2) and 66 ms (tempo x 3). The task was similar to the task performed in Experiment 1. The three tempi were presented in separate blocks. The order of the blocks was counterbalanced across participants. The sound pattern, EEG and movement signals were analyzed using the same procedures as described in Experiment 1 (Figs. 6-8). As in Experiment 2, the spectrum of the sound envelope of the accelerated patterns was not substantially distorted compared to that of the original pattern (Fig. 7). Pattern 2 was chosen here because, in the 1st experiment, an apparent discrepancy was observed between the frequency of the SS-EP selectively enhanced in the EEG spectrum

(0.625 Hz, corresponding to the subharmonic of beat frequency, thus to a grouping by 8 events) and the frequency of the tapping on this pattern (1.25 Hz, corresponding to a grouping by 4 events). Therefore, Experiment 3 aimed at clarifying the adequacy between the frequency selectively enhanced in the EEG spectrum and the frequency selected in the tapping task by using different tempi.

II.2.2.6.2.4. Results

Experiment 1

Hand tapping movement. As shown in Figure 1, the participants tapped periodically at a frequency corresponding, in most cases, to a grouping by 4 events (1.25 Hz) and, in some cases, at related metric levels (corresponding to a grouping by 2 or 8 events, thus at 2.5 Hz or 0.625 Hz respectively). Importantly, in patterns 3 to 5, the acoustic energy was not predominant at the frequency of the perceived beat selected on average for the tapping (Fig. 2) but at distinct, non meter-related, frequencies, thus confirming that the frequency at which a beat is perceived does not necessarily correspond to the frequency showing maximum acoustic energy in the physical structure of the pattern envelope.

As compared to the tapping performed on the other sound patterns, the distribution of tapping periods in pattern 5 showed a much greater variability, both within and across participants, indicating that this sound pattern did not elicit a stable and unequivocal beat percept (Fig. 1).

Detection task. During the recording, participants performed the detection task with a median score of 8.5/10 (interquartile range: 8 - 10), with no apparent difference in difficulty reported between the patterns.

Steady-state evoked potentials. For each of the five sound patterns, most of the frequencies constituting their envelope elicited clear SS-EPs in the EEG frequency spectrum (Fig. 2). SS-EP amplitudes were significantly different from zero at most of the frequencies corresponding to the perceived beat and related metric levels (Fig. 2). The scalp topography of the elicited SS-EPs was generally maximal over fronto-central regions, and symmetrically distributed over the two hemispheres (Fig. 3). Moreover, as illustrated in Figure 3, the scalp topography of the SS-EPs elicited at beat and meter related frequencies did not substantially differ from the SS-EPs elicited at unrelated frequencies, thus suggesting that they originate from similar neuronal populations, or that the EEG did not allow disentangling the different spatial location of the underlying neural activity in our experiment.

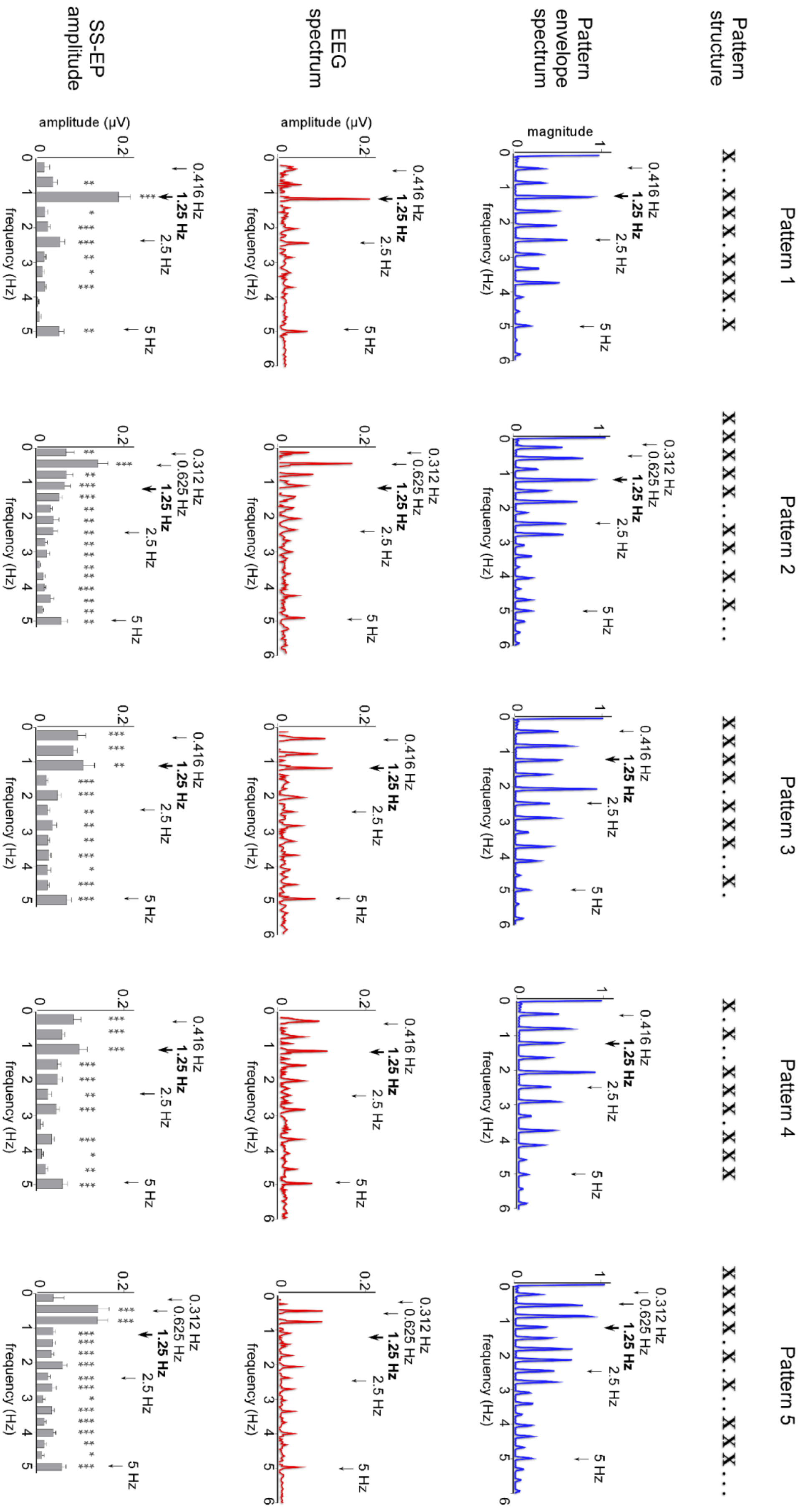

Pattern 1
Pattern 2
Pattern 3
Pattern 4
Pattern 5
Pattern structure
X..XXX.XXX.X
XXXXX..XX.X.X...
XXXX.XXX..X.
X.X..XXX.XXX
XXXX.X.X..XXX...
Pattern envelope spectrum
magnitude
EEG spectrum
SS-EP amplitude
amplitude (µV)
frequency (Hz)
0.312 Hz
0.416 Hz
0.625 Hz
1.25 Hz
2.5 Hz
5 Hz
0
0.2
1
2
3
4
5
6
*
**
***

*Figure 2. First line. Structure of the five rhythmic patterns, which consisted of a sequence of 200 ms pure tones (represented by a cross) and 200 ms silences (represented by a dot). Patterns 1, 3 and 4 contained a succession of 12 events, whereas patterns 2 and 5 contained a succession of 16 events. Second line. Frequency spectrum of the patterns sound envelope. The thick vertical arrow marks the expected beat frequency. The thin vertical arrows mark the related meter frequencies. Note that, in patterns 3 and 5, the acoustic energy did not predominate at the frequency of the expected beat frequency. Third line. Frequency spectrum of the EEG recorded while listening to each of the five sound patterns (noise-subtracted amplitude, averaged across all scalp channels, in microvolts). Fourth line. Mean magnitude (± standard error of the mean) of the SS-EPs elicited by each of the five sound patterns (noise-subtracted amplitude, averaged across all scalp channels, in microvolts). SS-EP amplitudes significantly different from zero are marked by * ($p<0.05$), ** ($p<0.01$) or *** ($p<0.001$).*

At the frequencies expected to relate to beat and meter, the standardized estimates of the SS-EP amplitudes, averaged across the five sound patterns, were significantly enhanced as compared to the standardized estimates of the sound envelope at these frequencies ($t=15.85$, $df=8$, $p<0.0001$) (Fig. 3). Conversely, at the frequencies unrelated to beat and meter, the standardized estimates of SS-EP amplitudes were significantly reduced ($t=16.62$, $df=8$, $p<0.0001$) (Fig. 3).

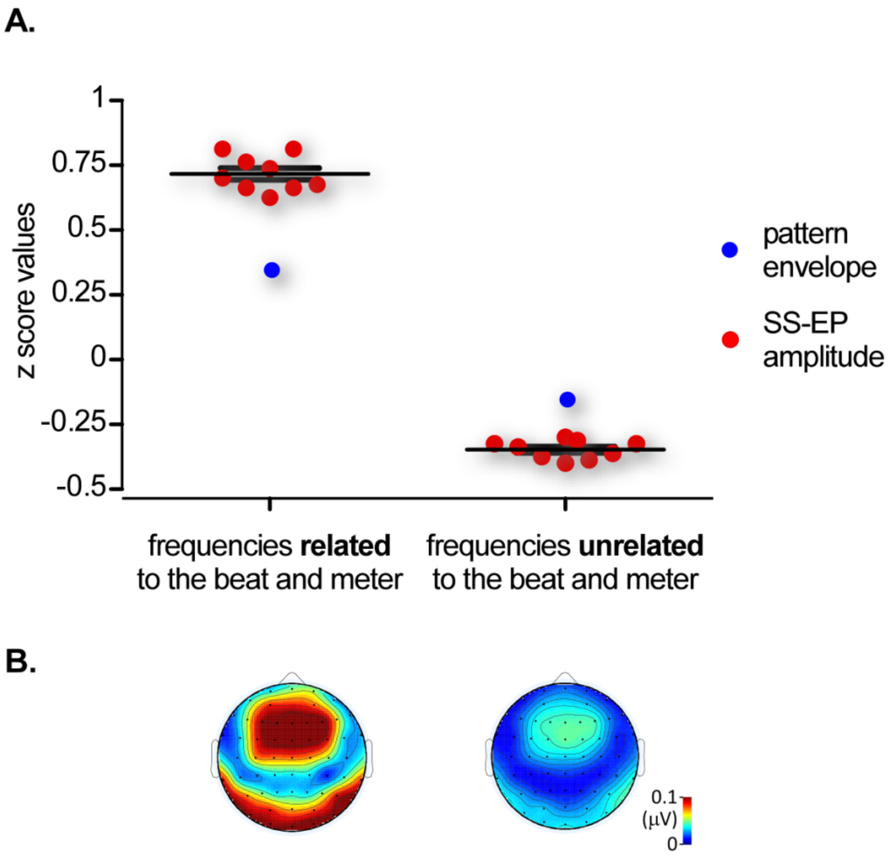


Figure 3. A. The red dots represent the z score values of the magnitude of the SS-EPs elicited at frequencies related to the beat and meter (left) and at frequencies unrelated to the beat and meter (right), averaged across all five sound patterns, in each participant. Note that, as compared to the average z score of the sound pattern envelope at corresponding frequencies (blue dots), the magnitude of the beat- and meter-related SS-EPs was markedly enhanced, whereas the magnitude of SS-EPs unrelated to the beat and meter was markedly dampened. B. Average topographical map of the SS-EPs elicited at frequencies related to the beat and meter (left) and at frequencies unrelated to the beat and meter (right), averaged across all patterns and participants (noise-subtracted amplitude, in microvolts).

When examining the sound patterns separately, the SS-EP elicited at the expected beat frequency (1.25 Hz) was significantly enhanced in patterns 1 (t=3.15, p=0.01, df=8), 3 (t=2.55, p=0.03, df=8) and 4 (t=3.31, p=0.01, df=8) but not in patterns 2 (reduced; t=4.04, p=0.003, df=8) and 5 (t=0.46, p=0.65, df=8) (Fig. 4). In pattern 2, the SS-EP corresponding to the beat frequency assessed in the hand tapping condition was not significantly enhanced.

However, the SS-EP corresponding to its subharmonic (grouping by 8 events instead of 4 events) was enhanced (t=2.48, p=0.03, df=8; Fig. 4). In pattern 5, except at the frequency of the unitary event (5 Hz), none of the SS-EPs elicited at beat and meter related frequencies appeared to be enhanced as compared to the sound envelope (Fig. 4). One possible explanation for the enhancement of the EEG signal observed at 5 Hz could be that, although none of the frequencies within this sound pattern was able to induce a stable perception of beat, the frequency corresponding to the rate at which the individual sounds were presented constituted a relatively salient feature in the pattern. The fact that this frequency did not elicit a perception of beat could be related to the fact that it lied outside the ecological range for beat and meter perception (Repp, 2006).

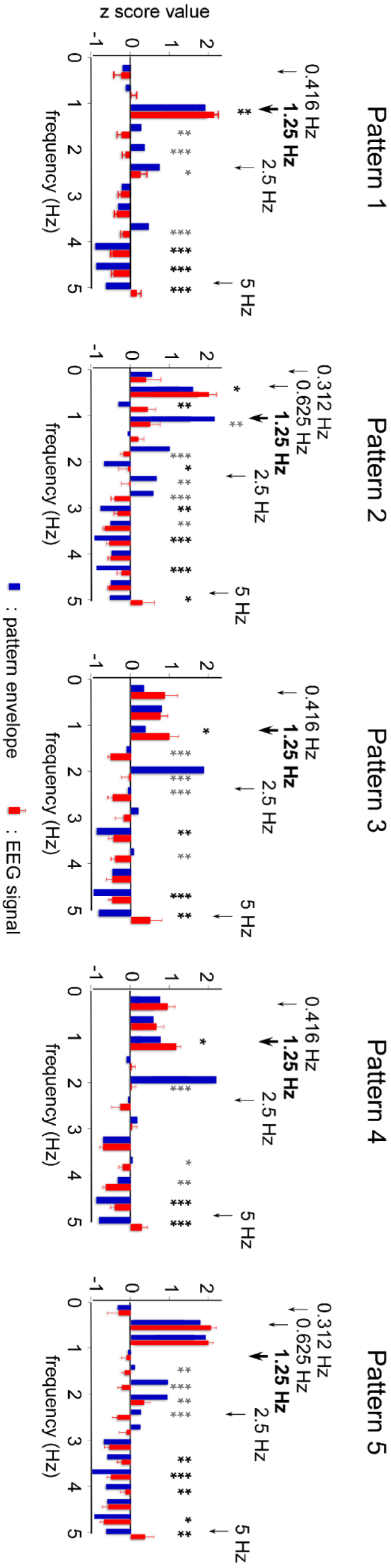
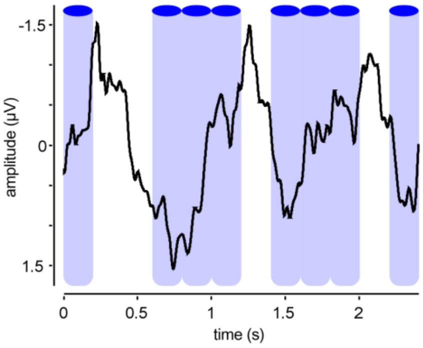

Pattern 1
Pattern 2
Pattern 3
Pattern 4
Pattern 5
0.416 Hz
0.312 Hz
0.625 Hz
1.25 Hz
2.5 Hz
5 Hz
z score value
-1
0
1
2
frequency (Hz)
0
1
2
3
4
5
*
**
***
🟦 : pattern envelope
🟥 : EEG signal

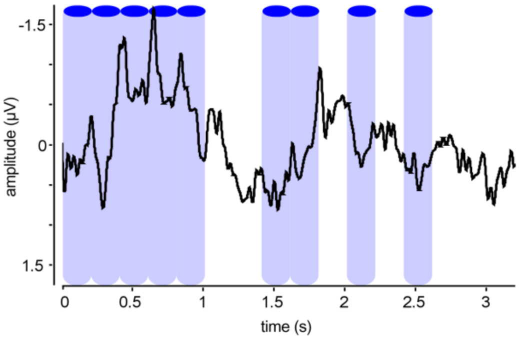
*Figure 4. The red bars represent the group-level mean z scores of the magnitude of the SS-EPs elicited between 0 and 5 Hz within each sound pattern (error bars: standard error of the mean). The blue bars represent the z scores of the magnitude of the sound envelope at corresponding frequencies. Significant differences between SS-EP and sound envelope amplitude are represented with a black (SS-EP > sound envelope) or grey (SS-EP < sound envelope) asterisk (one-sample t-test; * : p<0.05, ** : p<0.01, *** : p<0.001). The expected beat and meter frequencies are highlighted by vertical arrows (beat frequency shown in bold).*

Time domain analysis of the EEG signals. As displayed in Figure 5, some periodicity in the average signals could be observed in the time domain, mostly in pattern 1, in which the 1.25 Hz frequency was highly enhanced compared to other frequencies in the EEG spectrum. In the other patterns, these periodicities were more difficult to assess in the time domain, as several frequencies were present concurrently. The lack of clear time-locked evoked responses is most probably due to (1) the fact that the sounds used to elicit the beat percepts were not as transient as those used in previous studies and (2) the fact that in addition to the beat frequency, various metric levels were also enhanced and these various frequencies are mixed when observed in the time domain.

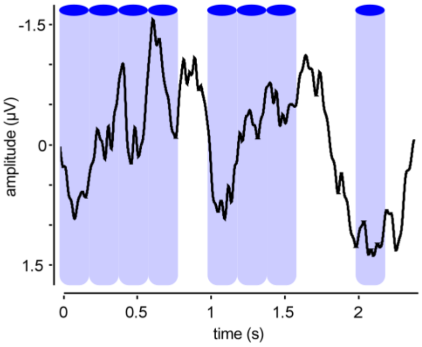
Pattern 1



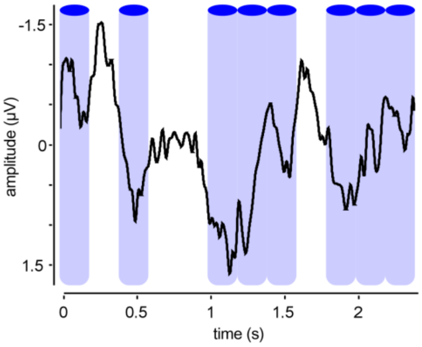
Pattern 2



Pattern 3



Pattern 4



Pattern 5

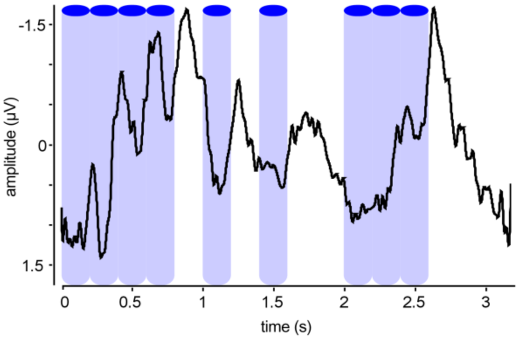


Figure 5. Time course of the EEG responses recorded in Experiment 1 (group-level average of the signal recorded at electrode FCz, segmented according to the length of each pattern: 2.4 s in patterns 1, 3 and 4; 3.2 s in patterns 2 and 5). The obtained signals were examined from electrode FCz, as this electrode displayed the maximum amplitude, independently of the elicited signals frequencies in the EEG spectra. Some periodicity in the average signals can be observed, mostly in pattern 1, in which the 1.25 Hz frequency was predominant in the EEG spectrum. In the other patterns, these periodicities are more difficult to assess in the time domain, as several frequencies were present concurrently. The sounds are represented by the blue bars.

Experiment 2

Hand tapping movement. As shown in Figure 6, the participants changed their tapping frequency across the different tempi of presentation of pattern 1. Specifically, they tapped periodically at a frequency corresponding, in most cases, to the grouping by 4 events (i.e., at 2.5 Hz) in tempo x 2, and to the grouping by 12 events (i.e., at 1.6 Hz) in tempo x 4 (Fig. 6). Moreover, the distribution of tapping periods showed a greater variability across participants when accelerating the tempo, thus confirming previous observations that a stable and unequivocal percept of beat is preferentially elicited by a pattern presented at tempi lying within a specific frequency range (Repp, 2005; 2006) (Fig. 6).

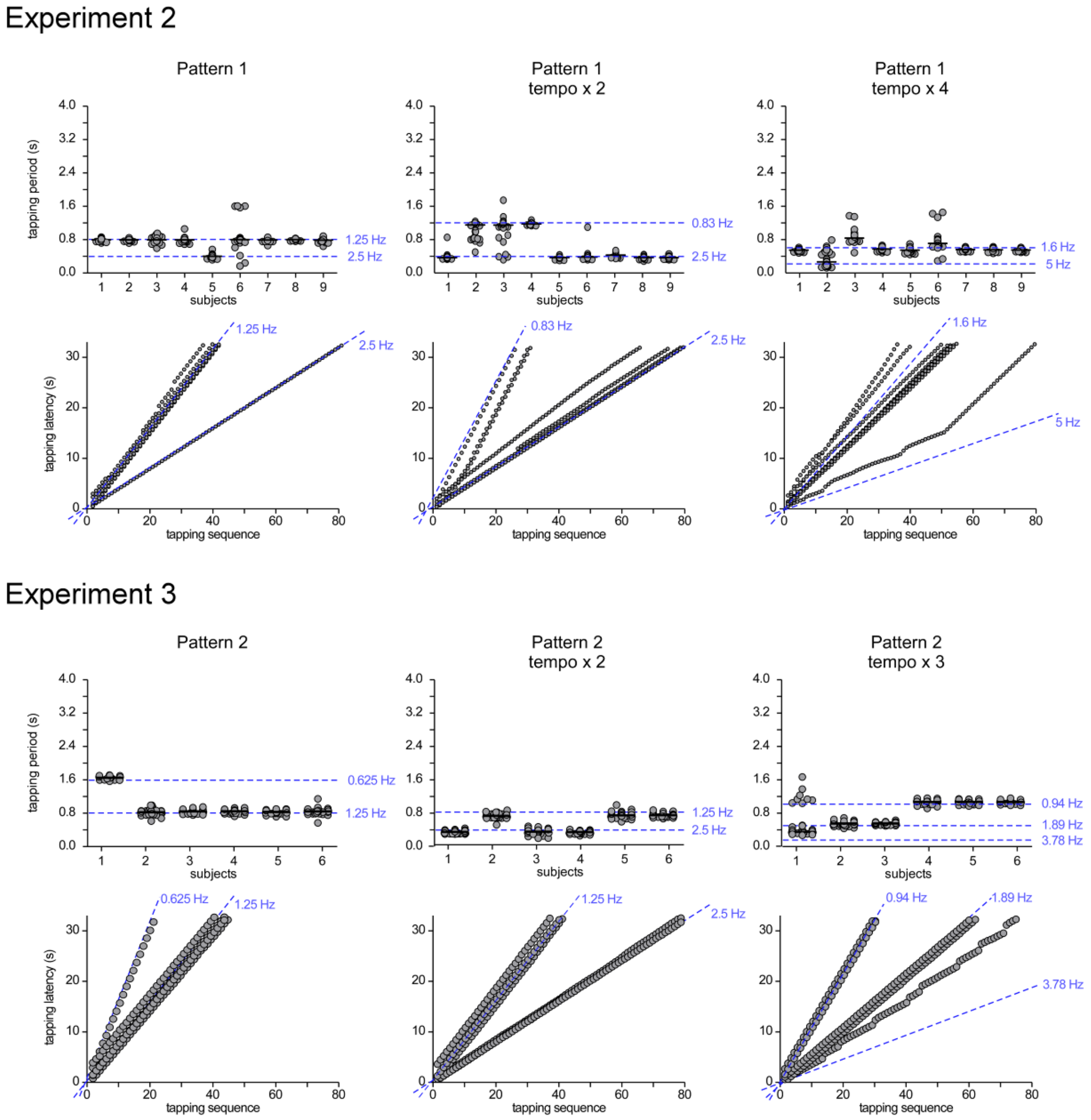


Figure 6. Experiment 2. Upper graphs. Tapping periods produced by each participant while listening pattern 1 presented at the original tempo, at tempo x 2 and at tempo x 4. Lower graphs. Tapping latencies along the entire trial for each participant and tempo. Experiment 3. Upper graphs. Tapping periods produced by each participant while listening to pattern 2 presented at the original tempo, at tempo x 2 and at tempo x 3. Bottom graphs. Tapping latencies along the entire trial for each participant and tempo.

Detection task. During the recording, participants performed the detection task with a median score of 3.5/4 (interquartile range: 3 - 4), with no apparent difference in difficulty reported between the different tempi.

Steady-state evoked potentials. Most of the frequencies constituting the envelope of the sound pattern elicited SS-EPs at corresponding frequencies (Fig. 7). Figure 8 shows the means and standard errors of the means of the z score values obtained for the EEG signals across participants, as well as the z score values obtained for the sound envelope of each tempo of presentation. Importantly, in contrast to the original tempo, the standardized amplitude of the SS-EP elicited at a frequency corresponding to the grouping by 4 events (at 2.5 Hz,) was not significantly greater than the standardized magnitude measured at that frequency in the sound envelope in tempo x 2 (t=0.01, p=0.98, df=8). Moreover, it was significantly smaller than the sound envelope in tempo x 4 (t=2.86, p=0.02, df=8) (Fig. 8). In contrast, the beat subharmonic (corresponding to a grouping by 12 events; 1.6 Hz) was significantly enhanced in tempo x 4 (t=4.03, p=0.003, df=8; Fig. 8).

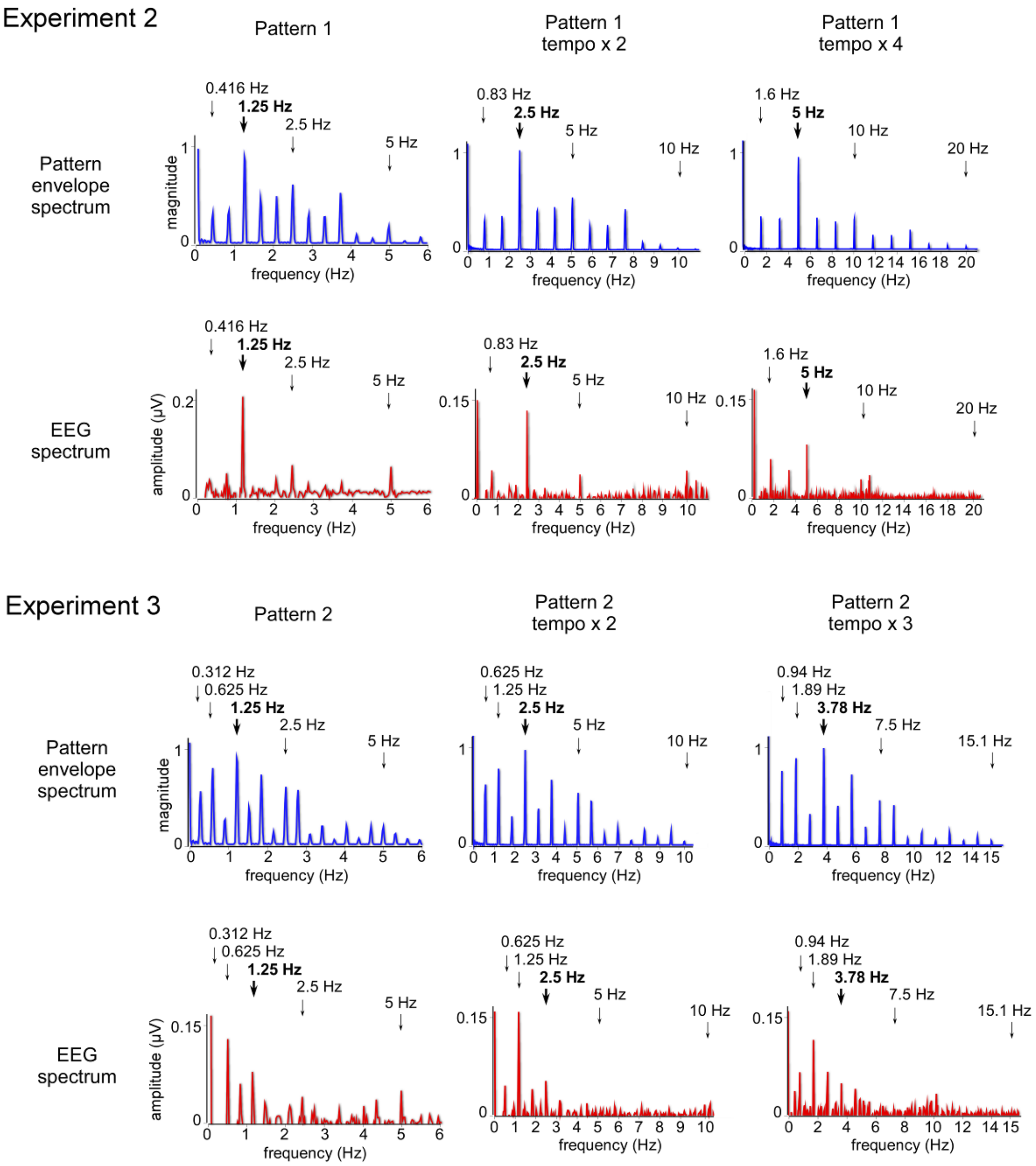


Figure 7. Upper graphs. Envelope spectrum of pattern 1 at the original tempo, tempo x 2 and tempo x 4 (experiment 2) and of pattern 2 at the original tempo, tempo x 2 and tempo x 3 (experiment 3). The expected beat and meter frequencies are highlighted by the arrows (the grouping by 4 events is shown in bold). Bottom graphs. Spectrum of the corresponding EEG (noise-subtracted amplitude averaged across all scalp channels, in microvolts).

Experiment 3

Hand tapping movement. As shown in Figure 6, the participants changed their tapping frequency across the different tempi of presentation of pattern 2. Specifically, they tapped periodically at a frequency corresponding, in most cases, to the grouping by 4 events (i.e., at 1.25 Hz) at the original tempo. At tempo x 2, half of the participants still tapped on a grouping by 4 events (at 2.5 Hz) and the other half tapped on a grouping by 8 events (at 1.25 Hz) (Fig. 8). At tempo x 3, some participants tapped to a grouping by 8 events (1.89 Hz), whereas others tapped to a grouping by 16 events (0.94 Hz).

Detection task. During the recording, participants performed the detection task with a median score of 3.5/4 (interquartile range: 3 - 4), with no apparent difference in difficulty reported between the different tempi.

Steady-state evoked potentials. Most of the frequencies constituting the envelope of the sound pattern elicited SS-EPs at corresponding frequencies (Fig. 7). Like in Experiment 1, presentation of the sound pattern using the original tempo elicited SS-EPs which were predominant at the frequency corresponding to the subharmonic of the beat (i.e., a grouping by 8 events) (Figs. 7 and 8). Figure 8 shows the means and standard errors of the means of the z score values obtained for the EEG signals across participants, as well as the z score values obtained for the sound envelope for each tempo of presentation. The z score value of the SS-EP corresponding to the grouping by 8 events (i.e., at 0.625 Hz, at 1.25 Hz and at 1.875 Hz in the three tempi respectively) was the highest in average. When compared to the sound, this z score value was significantly enhanced at tempo x 2 ($t=7.45$, $p=0.0007$, $df=5$), but not at the original tempo ($t=0.11$, $p=0.91$, $df=5$) and not at tempo x 3 ($t=0.67$, $p=0.52$, $df=5$) (Fig. 8).

II.2.2.6.2.5. Discussion

Participants listened to rhythmic sound patterns, expected to induce a spontaneous perception of beat and meter at 1.25 Hz and its (sub)harmonics (Povel and Essens, 1985). The multiple periodic features of these complex sound patterns elicited corresponding periodic signals in the EEG, identified in the frequency domain as steady-state evoked potentials (SS-EPs; Regan, 1989). This result is in line with previous evidence of envelope locking in neurons of the auditory cortex, that is, the ability of these neurons to synchronize their activity to the temporal envelope of acoustic streams (see Eggermont, 2001; Bendor and Wang, 2007 for some reviews).

Importantly, when comparing the frequency spectrum of the EEG to the frequency spectrum of the sound envelope, we found that the magnitude of the SS-EPs elicited at beat- and meter-related frequencies was significantly increased as compared to the magnitude of the SS-EPs elicited at frequencies unrelated to the beat and meter percept (Fig. 2). The fact that the magnitude of the elicited SS-EPs did not faithfully reflect the magnitude of the acoustic energy at corresponding frequencies (Fig. 3) suggests a mechanism of selective enhancement of neural activities related to beat and meter perception, possibly resulting from a process of dynamic attending (Jones and Boltz, 1989).

It is worth noting that this selective enhancement of beat- and meter-related SS-EPs was observed even in sound patterns in which the acoustic energy was not predominant at beat frequency (e.g. patterns 3 and 4; Fig. 4). Conversely, frequencies showing predominant acoustic energy in the sound envelope but unrelated to the beat and meter were markedly reduced in the EEG. Taken together, this suggest that beat and meter perception could involve spontaneous neural mechanisms of selection of beat-relevant frequencies in processing rhythmic patterns, and that these mechanisms can be captured with SS-EPs (Velasco and Large, 2011; Zion-Golumbic et al., 2012).

This interpretation may also account for previous observations that event-related potentials elicited at different time points relative to beat or meter cycle exhibit differences in amplitude when observed in the time domain (Fujioka et al., 2010; Iversen et al., 2009; Schaefer et al., 2011). Indeed, the observed SS-EP enhancement at beat-related frequencies could be interpreted as resulting from a stronger neural response to sounds, or silences, occurring “on beat”. Whether SS-EPs result from a resonance within neurons responding to the stimulus, or whether they are explained by the superimposition of transient event-related potentials remains a matter of debate (Galambos et al. 1981, Draganova et al. 2002).

As compared to analyses in the time domain, characterizing beat- and meter-related EEG responses using SS-EPs offers several advantages. First, it allows assessing a selective enhancement not merely at one frequency, corresponding to the beat, but also at other frequencies, corresponding to distinct metric levels associated to beat perception (London, 2004; Large, 2008). Second, because the sounds eliciting a beat percept are often not cleanly separated from one another by long-lasting periods of silence, response overlap can make it difficult to assess the enhancement of beat- and meter-related activities in the time domain (Fig. 5).

Interestingly, when examining SS-EPs elicited by each pattern separately, we found that pattern 5 did not elicit enhanced SS-EPs at the frequency of the expected beat (Figs. 3 and 4). Importantly, moving to the beat in this pattern showed variability both within and across participants (Fig. 1), suggesting that this pattern failed to elicit a stable and unequivocal beat percept. This was possibly due to the intrinsic complexity of its rhythmic structure (Thul et al., 2008) or, perhaps, to cultural or individual long-term exposure bias, yielding a blurred attraction for the frequency expected to correspond to the beat in this specific rhythmic pattern. Hence, the lack of beat-related SS-EP enhancement for this pattern actually corroborates the view that non-linear transformation of the sound by neural populations are involved in building beat and meter representations while listening to rhythms.

Most importantly, this interpretation is strengthened by the fact that this selective enhancement phenomenon appeared to be sensitive to the tempo at which the sound patterns were presented. This was tested in Experiment 2, in which the tempo was accelerated by 2 and by 4, thus reaching the upper limit for beat perception (grouping by four events at 2.5 Hz in tempo x 2, and 5 Hz in tempo x 4). Indeed, beat and meter perception is known to emerge within a specific tempo range, corresponding to frequencies around 2 Hz (Repp, 2005; 2006; van Noorden and Moelants, 1999). This is explained, at least in part, by the temporal limits for dynamic attending to discrete events along time (Jones and Boltz, 1989) and for perceptual grouping in the auditory system (Bregman, 1990; van Noorden, 1975). This tempo sensitivity for beat perception was confirmed in Experiment 2. Indeed, the tapping task showed that at twice the original tempo, participants still tapped, on average, to the grouping by four events (2.5 Hz). However, at four times the original tempo, participants no longer tapped to the grouping by four events (5 Hz) and, instead, tapped to a grouping by 12 events (1.6 Hz) (Fig. 6). Most importantly, the magnitude of the SS-EP appearing at the frequency corresponding to the grouping by four events was dampened at the fastest tempo. In contrast, the magnitude of the SS-EPs appearing at subharmonics lying within the frequency range for beat perception at this accelerated tempo was enhanced (Figs. 7 and 8). Hence, the shift in beat percept appeared to parallel the shift in the selective enhancement of SS-EPs, supporting the view that the enhancement of beat-related SS-EPs underlies the emergence of beat perception (Figs. 7 and 8). Moreover, as varying the frequency of stimulation can act as a probe of a resonance bandpass (Hutcheon and Yarom, 2000), these results suggest that beat- and meter-related SS-EPs could, at least in part, result from a resonance phenomenon (Large, 2008). Hence, it may corroborate hypothesis according to which beat perception in music emerges from the entrainment of neuronal populations resonating at beat frequency (Large, 2008; van Noorden and Moelants, 1999).

Finally, the relationship between beat perception as assessed through overt movement of a body segment and selective SS-EP enhancement observed in the absence of overt movement was explored further in pattern 2. Indeed, in this pattern, although participants consistently moved to a grouping by four events in the hand tapping task (1.25 Hz; Fig. 1), the SS-EPs recorded in the absence of hand tapping showed a relative reduction at that frequency in the EEG spectrum. Instead, a significant enhancement was observed for the SS-EP appearing at the frequency corresponding to a grouping by eight events. A possible explanation to this discrepancy is that the frequency of the beat perceived when listening without moving differed from the frequency of the beat perceived when performing the hand tapping task. Indeed, the hand tapping frequency could be biased by several constraints, such as biomechanical constraints. This interpretation is in line with evidence showing that humans are spontaneously entrained to move on musical rhythms using specific body segments depending on the tempo (e.g. slow metric levels preferentially lead to move axial body segments such as bouncing the head, whereas fast metric levels tend to entrain more distal body parts, such as foot tapping) (Toiviainen et al., 2010; van Noorden and Moelants, 1999; McDougall and Moore, 2005). To explore this further, we examined in Experiment 3 the effect of accelerating the tempo of pattern 2. The tapping showed that, at two times the original tempo, half of the participants still tapped to the grouping by four events (2.5 Hz, as in the original tempo), but the other half tapped to the grouping by eight events (1.25 Hz) (Fig. 6). At three times the original tempo, about the same tapping distribution was found between a tapping frequency corresponding to the grouping by eight events (1.89 Hz) and by sixteen events (0.94 Hz) (Fig. 6). In contrast, when listening without moving, a tendency to selective enhancement at frequencies corresponding to the same grouping by eight events was observed across the tempi in the EEG (Figs. 7 and 8). Hence, it could be that performing the hand tapping task engaged specific constraints, leading to differences between the frequency of the beat perceived in the hand tapping condition and the frequency of the beat perceived in the passive condition.

Taken together, the observation that SS-EPs are selectively enhanced when elicited at frequencies compatible with beat and meter perception indicates that these responses do not merely reflect the physical structure of the sound envelope but, instead, reflect the spontaneous emergence of an internal representation of beat, possibly through a mechanism of selective neuronal entrainment within a resonance frequency range.

Acknowledgements: S.N. is supported by the Fund for Scientific Research of the French-speaking community of Belgium (F.R.S.-FNRS). A.M. has received the support from the French-speaking community of Belgium (F.R.S.-FNRS) FRSM 3.4558.12 convention grant. 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.

II.2.3. Coupling of sensorimotor periodic signals

Following the review of the evidence on production and perception of periodic signals, the present section reviews the third component thought to be required to entrain movements on the beat: the coupling of both production and perception of periodic signals.

The coupling of perception of periodicities and periodic movement, referred to as sensorimotor synchronization, is fundamental to music performance and dance behaviors (Repp, 2005; 2006). For example, musical performance in a group requires the precise coordination of sound-producing movements with the sounds and movements produced by other musicians. Dance involves the coordination of movements with music. In some cases, the produced dance movements generate sounds, such as tap dancing and marching, and dancers coordinate each other movements based on the produced sounds. Practicing with a metronome is another case of sensorimotor synchronization in a musical context, as well as the spontaneous entrainment to move when listening to groove-based music.

Most studies from the sensorimotor synchronization literature concern simple movements such as finger tapping, and simple stimuli such as metronome-like sounds. Although they do not specifically focus on musical contexts, these studies are relevant to music because they reveal basic sensorimotor synchronization mechanisms and characteristics (Repp, 2006). These characteristics have been summarized in several reviews (Repp, 2005; 2006; Aschersleben, 2003). As some of these characteristics are important to know to better understand the motivation of Study 3 (Section II.2.3.4), we briefly describe them here, followed by a review of the evidence on the neural processes underlying sensorimotor synchronization to the beat.

II.2.3.1. Characteristics of sensorimotor synchronization to the beat. Sensorimotor synchronization usually refers to the *coordination of rhythmic movement with rhythmic* *sensory stimuli* (Repp, 2005). On the one hand, the term “rhythmic” in this literature frequently describes signals whose unfolding along time varies dynamically according to a certain temporal structure, periodic or not (Hogan and Sternad, 2007). Here, we will focus on periodic structure specifically, as it concerns entrainment to the musical beat. On the other hand, the term “coordination” refers to the time locking, ideally corresponding to a phase lag of zero, between the sensory signal and the produced movement (see also *Glossary*). In biological systems, the phase lag between sensory signal and produced movement is rarely equal to zero and, in the context of sensorimotor synchronization, the produced movement often tends to *precede* the sensory input. Therefore, the produced movement cannot be viewed as a simple reaction triggered by the sensory input but, instead, involves a mechanism of anticipation to compensate the incompressible delays induced by the conduction and processing of the sensory input as well as the generation of motor output.

A fundamental aspect of sensorimotor synchronization is *error correction*, since it prevents the inherent variability of the movement from having a cumulative effect on the times of occurrence of successive actions (Vorberg and Wing, 1996). Two distinct error correction processes have been hypothesized to occur, that can be modeled as *phase correction* and *period correction* in the particular case of periodic sensorimotor synchronization (Mates, 1994; Repp, 2006). In this model, phase correction would adjust the occurrence of each successive action but would not affect the period of the underlying timekeeper or oscillator which drives action’s rate. Phase correction would rely on the detection of asynchronies between the sound and the produced movement. In contrast, period correction would adjust the interval of the timekeeper or oscillator period. It would rely on the detection of a mismatch between the period of the internal timekeeper and the period of the pacing sequence.

The distinction between the two error corrections mechanisms has been modeled based on a wealth of evidence suggesting that phase correction is relatively automatic whereas period correction is consciously controlled (Repp, 2006). For example, when a phase perturbation is introduced in a metronome sequence on which participants have to synchronize, phase correction occurs regardless of whether the perturbation itself is consciously detected by the participants (Repp, 2000). Moreover, when participants are instructed not to react to a detectable phase perturbation, they can reduce their phase correction but not suppress it entirely (Repp and Keller, 2004). In contrast, period correction in response to tempo change in the pacing sequence is more effective when the tempo change is consciously detected (Repp, 2001b), is affected by the focus of attention and can be suppressed voluntarily (Repp and Keller, 2004). Therefore, it was suggested that phase correction is a relatively "low-level" process necessary to stabilize sensorimotor synchronization, whereas period correction is a "higher-level" process requiring modulation of an internal period (Repp, 2005).

However, even without any perturbation of the period or phase of the external input, an important characteristic of sensorimotor synchronization is its *variability*. As mentioned by Repp (2006), not only are movements inherently variable, but also the perception of a perfectly regular metronome is subject to variability and errors arising from the nervous system. One of the oldest findings concerning variability in sensorimotor synchronization is the *mean negative asynchrony*, i.e., the fact that during tapping in synchrony with an auditory metronome, tapping latencies tend to precede the tone onsets by a few tens of milliseconds, rather than being distributed symmetrically around the tone onset

(Aschersleben, 2002) (Fig. II.2.3.1). Although the size of this asynchrony can vary depending on the experimental conditions (e.g., amplitude and force of the performed movement), there are large interindividual differences as well, from almost no asynchrony to asynchrony as large as 100 ms without being aware of this anticipatory tendency. Indeed, participants are unaware of this tendency to anticipate the tone and, when they are trained with online verbal feedback to correct tapping latency and achieve a strict synchrony, they feel that they are being forced to delay their taps (Repp, 2006). Thus, the mean negative asynchrony seems to be necessary for persons to gain the subjective impression of being in synchrony.

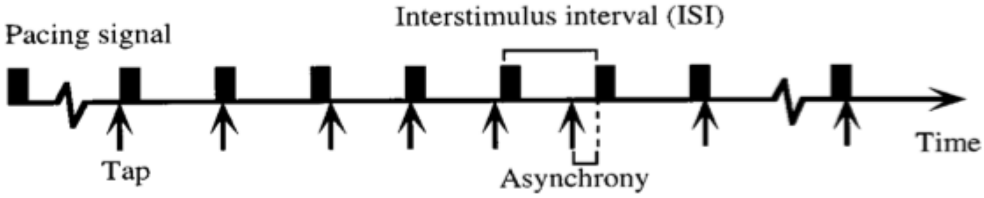


Figure II.2.3.1. From Aschersleben (2002). The sensorimotor synchronization tapping task seen along time. The tapping consistently anticipates the pacing signal.

Additional evidence concerning the mean negative asynchrony show that it vanishes when additional tones placed at a multiple of the metronome frequency are inserted between the metronome tones to tap on. This has been explained by a tendency to minimize tapping variance through shortening the intervals between tones (Vorberg and Wing, 1996). These observations could also account for particular behaviors in music, such as rushing when playing with others, or coming in too soon after a rest (Repp, 2006). As compared to untrained subjects, musically trained participants tend to show a smaller, and sometimes absent (Repp, 2005), mean negative asynchrony (Aschersleben, 2002), possibly because they automatically subdivide the inter-onset intervals internally.

Various explanations have been proposed to account for the mean negative asynchrony phenomenon. For example, it has been suggested that it is a consequence of the different neural transmission times of auditory and tactile/somatosensory inputs. This explanation was suggested based on the evidence that the negative mean asynchrony tends to disappear if, in a finger tapping task for instance, a sound is produced by the contact of the fingers on a surface (i.e., as the sound produced by a piano keyboard) (Fraisse, 1980; Aschersleben and Prinz, 1995). The sound produced by the fingers could thus be matched to the sound to which the fingers have to synchronize, generating sensory feedbacks within the same modality. However, what would explain the asymmetric distribution of the finger tapping *before* the sound input? It has been proposed that participants may be inclined to perceive the sounds on which they have to synchronize as consequences of their taps, at least in this kind of tapping paradigm (Repp, 2005).

II.2.3.2. A frequency tuning function for beat perception and synchronization. There is an apparent optimal tempo for repeating body movements elicited by music (van Noorden and Moelants, 1999). Moreover, even when we simply listen to music without performing any movement, we select an internal pulse from a relatively narrow range of frequencies. It is usually assumed that beat and meter perception in music occurs within a specific range of frequencies around 2 Hz. If music is performed too quickly, successive sounds become indistinguishable. Conversely, if music is performed too slowly, rhythmic organization tends to fall apart, leaving only a series of isolated sounds. Between the two extremes, music and other sound patterns have perceivable rhythm. This specific range of tempo is thought to relate (1) to the optimal rate at which successive events are perceived as segregated within a sensory system, (2) to the frequency at which repeated movements are performed according to constraints such as biomechanical limits, and (3) to the range of tempo required for the coupling of sensory and motor signals. Interestingly, this tempo range has often been assimilated to a resonance curve, or entrainment region, that is, a frequency range within which the human brain is more easily entrained by external sensory stimuli.

II.2.3.2.1. Lower rate limit. The lower rate limit (in Hz) is usually derived in the literature from the interonset interval beyond which, when asked to synchronize to isochronous auditory events, the width of the variability of asynchronies is no longer proportional to the interonset interval duration, but instead show a larger increase (Repp, 2006). Using this criterion, studies have shown that synchronization becomes difficult for interonset intervals greater than 1.8 s, corresponding to a lower rate limit of approximately 0.5 Hz (Woodrow, 1932; McDorman, 1962) (for the sake of clarity, the interonset intervals will be converted into their corresponding frequencies further in this section).

Another criterion for the lower rate limit is the emergence of a tendency to react to, rather than anticipate, sequences of tones. Using this criterion, Engström et al. (1996) found that reactions, instead of mean negative asynchrony indicating anticipation, begins to emerge for rates slower than 0.75 Hz, and becomes predominant at 0.25 Hz. However, the variability of asynchronies is reduced when the interonset intervals between target tones are subdivided explicitly by additional tones, thus turning the task to a 1:n tapping (Repp, 2003b; 2005). In trained musicians, this so-called subdivision benefit could be obtained by mental division of the interonset interval (London, 2004).

As described extensively by Repp (2006), at rates faster than 0.6 Hz, synchronizing with a sequence of external inputs seems effortless and automatic. In contrast, when the rate approaches 0.5 Hz, synchronization becomes laborious, and the task resembles more that of interval estimation. Each tap must be placed consciously at the remembered duration of the previous interonset interval, and error corrections are computed deliberately, rather than automatically as observed during sensorimotor synchronization to faster rates of stimuli. However, the lower rate limit can be circumvented by subdividing the inter-onset intervals internally, as reported by musically trained individuals in musical contexts (Aschersleben, 2002).

The lower rate limit is unlikely to constitute a sensorimotor limit, since the slow tempo allows ample time for error correction. But it could correspond to the lower rate limit for rhythm perception, below which successive tones are no longer perceived within a sequence, but rather, as unrelated events (Bolton, 1894; Fraisse, 1982). Along this line, Pöppel (1997) proposed a low frequency neural mechanism that would "*bind successive events of up to 3 s into perceptual units*". This time period was also referred to "subjective present" (Michon, 1978). Further, a link has been made between these concepts and working memory (Miyake et al., 2004). Indeed, for a tap to anticipate the occurrence of the next tone in a sequence, the previous tone must still be "present" in working memory. Using verbal memory tasks and reproduction paradigms, researchers have estimated the time span of auditory working memory at 2 s, which is consistent with a lower rate limit of 0.5 Hz found for sensorimotor synchronization and beat induction.

II.2.3.2.2. Upper rate limit. The upper rate limit tends to be determined by the maximum frequency at which the effector can move. For finger tapping, this upper rate limit has been observed at 5-7 Hz (thus corresponding to intertap intervals of 150-200 ms) (Repp, 2005). This biomechanical limit can be circumvented asking participants to perform a 1:n synchronization, so that a tap is made in only one out of every n tones (Bartlett and Bartlett, 1959). At higher rates, the synchronization between taps and tones usually drifts apart, and participants are unable to judge whether they are in synchrony or not.

In some cases, the upper rate limit is reached earlier, revealing that some aspects make the sensory or sensorimotor processing more difficult. For example, using a 1:n synchronization paradigm, it has been shown that the upper rate limit was significantly lower for the 1:6 and 1:9 tapping ratio, and much lower again for 1:5 and 1:7 tapping patterns, whereas it did not differ for 1:2, 1:3, 1:4 and 1:8 tapping-to-event ratio (Fig. II.2.3.2.2) (Repp, 2005d), suggesting that processing 5 or 7 units structures are more demanding than numbers that can be subdivided equally (see also Section II.1.3.2 for a review of the evidence on the integer ratio metric biases). Other aspects make the sensory or sensorimotor processing more difficult, thus leading to a sharpening of the frequency tuning function for meter perception and synchronization. These aspects are for instance the rhythm complexity, the grouping or metrical accents generated by the rhythmic pattern (Povel and Essens, 1985; Repp, 2005b), or off-beat compared to on-beat tapping (Keller and Repp, 2004).

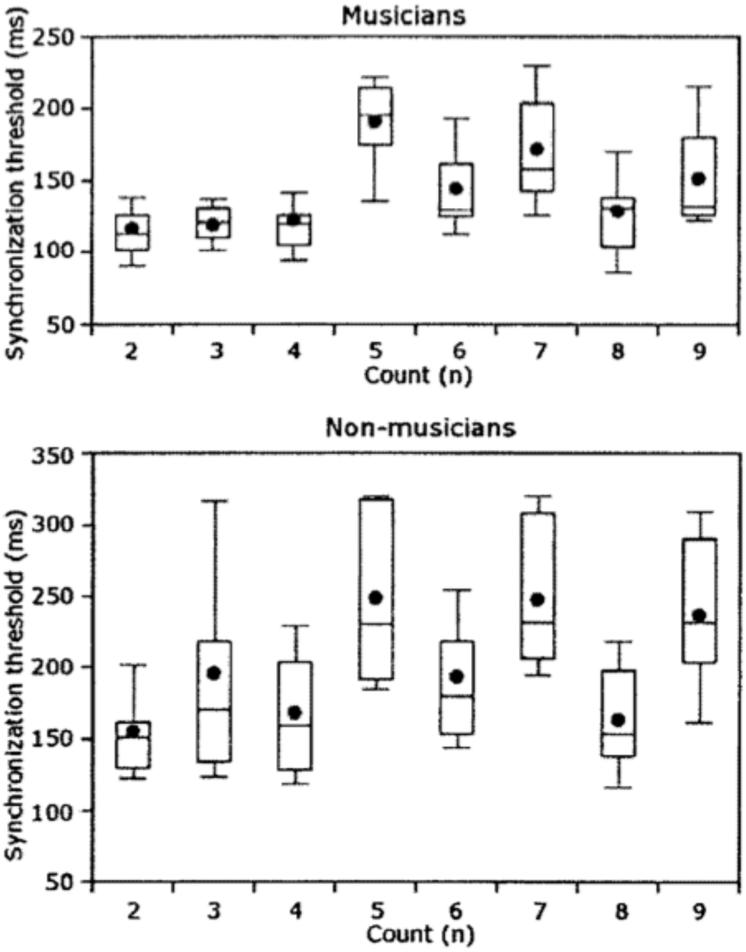


Figure II.2.3.2.2. From Repp, 2006. Synchronization task on 1:n ratio. In musician and nonmusicians, increased difficulty is observed for 5 and 7 ratio particularly, leading to a lower rate limit for synchronization threshold. However, this difference between tapping ratios tends to vanish in musicians.

The neurophysiological basis to account for the upper rate limit in sensorimotor synchronization remains unclear. Some explanations, belonging to attentional, perceptual and sensorimotor limits have been proposed (Repp, 2006). This upper limit could result from a limit of dynamic *attentional* processes, i.e., the upper limit at which these processes could be temporally modulated (1999) (see also Section II.2.2.3.4). This hypothesis is supported by several experimental observations showing temporal limits of attentional processes. For example, in the visual modality, detection of the second of two consecutive targets is impaired when the time-interval between the two targets is shorter than 500 ms, a phenomenon referred to as the “attentional blink” (Raymond et al., 1992). In the auditory modality, a similar “attentional blink” phenomenon has been found for inter-stimulus intervals of 100-200 ms (Arnell and Jolicoeur, 1998). This limit corresponds to the 5-7 Hz upper rate limit found in sensorimotor synchronization, suggesting that the upper rate limit corresponds to the upper limit rate at which attention can be temporally modulated.

Another hypothesis is that the upper limit for sensorimotor synchronization could reflect a limit for *perceptual* processing independently of attention. It has been proposed that the upper rate limit may reflect a temporal integration window within which successive events are not processed as segregated, even though their multiplicity is still perceived (Repp, 2006). This hypothesis is based on evidence of a significant increase of the threshold for tempo discrimination between auditory sequences whose rate is superior than 10 Hz (Michon, 1964). In addition, when investigating auditory sequences containing differences in pitch across tones, researchers have observed that segregation into separate streams occurs systematically when the rate of presentation is superior than 10 Hz, thus suggesting that, in absence of pitch difference, the successive events are perceptually merged (Bregman, 1990; van Noorden, 1975). Other researchers have used mismatch negativity potential paradigms (see also Section II.1.3.2) to investigate preattentional sensory integration and segregation processes (Tervaniemi et al., 1994). They provided evidence that two or more successive stimuli are treated as a single unit within the auditory system when they are separated by intervals of 150 ms or less.

The third hypothesis that has been proposed to explain the upper rate limit for sensorimotor synchronization pertains to sensorimotor limits, rather than attentional or purely perceptual constraints (Repp, 2006). According to this third hypothesis, the upper rate limit could result from incompressible times for sensorimotor neural loop in the human nervous system, required for online error correction processes. This could be the consequence of the fact that, at this upper rate, the perceptual information on which phase error correction is based becomes unreliable. If the event rate is high, there is not enough time to perceive asynchronies between sequence events and taps, and the error correction cannot be implemented in response to the most recent event.

As suggested by Repp (2006), considering the upper rate limit as a consequence of the inherent strength of the coupling between the processing of a rhythmic sensory stimulus within a given modality and motor processing may help explain for example the important differences observed across stimuli, as well as across individuals or sensory modalities (Repp and Penel, 2004; Grahn et al., 2011) (see Section II.2.3.7). Indeed, if the strength of the sensorimotor coupling decreases as the sequence rate is increased, the point at which decoupling is reached would be a measure of the degree to which a rhythmic stimulus in a given modality entrains the motor system.

II.2.3.2.3. Relevance in musical contexts. The rate limits observed in sensorimotor synchronization inevitably shape music performance and composition. For example, handbooks on music theory mention metronomic movements of 160-216 considered as fast tempi, 90-120 for the moderate tempi, and 40-60 for the tempi considered as slow (thus corresponding to 2.6-3.6 Hz, 1.5-2Hz and 0.6-1 Hz respectively; the metronomic movements traditionally refer to as the amount of click per minute), with range limits of 40 to 216 (i.e., 0.6 to 3.6 Hz). Another illustration of the rate limits is also found in music in the fact that tones separated by less than 100 ms do not serve as rhythmic elements but form ornaments attached to adjacent tones, arpeggi or glissandi. In addition, Van Noorden and Moelants (1999) tested the hypothesis that tempo distribution observed in music would follow the rate distribution obtained from the tapping literature when reaching minimum sensorimotor synchronization variability. After modeling a resonance curve of optimal beat perception and synchronization based on the results of tapping studies, they showed that the tempo distribution found across different music styles followed this resonance curve.

Nevertheless, musical performance uses specific techniques to overcome these rate limits. For example, fast music generally contains periodic regularities at slower time scales, whereas beat periods in slow music are subdivided by additional tones and are usually accompanied by visual cues, as after long rests, to stabilize movement synchronization. Conversely, when musicians aim to communicate a tempo that is relatively far away from the optimal range (i.e., very slow or very fast), the music has to provide strong cues to impose this “less natural” tempo, including the avoidance of regular divisions of a slow pulse and strong accentuations of every beat of a fast pulse (van Noorden and Moelants, 1999).

II.2.3.2.4. Tempo preferences. Within the tempo limits that define perceivable rhythms and allow sensorimotor synchronization, individuals demonstrate tempo preferences. The concept of preferred tempo has been widely studied, and various denominations have been used to describe it, as “mental tempo” or “internal tempo”, reflecting various assumptions about the origins of a preferred tempo (McAuley, 2010).

Stern (1900) was one of the first to suggest that the tempo of spontaneous motor activity may provide insight about the pace of mental activity. To investigate *spontaneous motor tempo*, he asked participants to tap the hand on a table at a rate that they considered “just right” (“*not too fast or too slow*”). Although this measure can reveal spontaneous motor tempo, it might be biased by the instructions given to the subjects, or the kind of movement asked to them (see Section II.2.1) (McAuley, 2010). A purer measure of the spontaneous motor tempo should preferably come from spontaneous rhythmic motor activities such as walking (Boltz, 1994), or instruction to chose a tempo allowing continuous regular movements sustained along time.

The most representative value of spontaneous motor tempo measured with tapping in adults reported in the literature is around 1.6 Hz (Fraisse, 1982, McAuley, 2010). However, there are also large differences across individuals, from 0.6 to 5 Hz, explained by biomechanical and dynamic constraints related to the type of movement (see Section II.2.1) (McAuley et al., 2006). Nevertheless, despite these large individual differences, measures of spontaneous motor tempo tend to be reliable within individuals, and stable along the sequence production (standard variation of about 5%).

Whereas spontaneous motor tempo refers to the natural rate for a rhythmic motor activity (for instance tapping), the *preferred perceptual tempo* refers to the rate at which a series of sounds is judged to be “just right”, i.e., neither too fast, nor too slow (Fraisse, 1982; McAuley et al., 2006; McAuley, 2010). Importantly, the most commonly reported value for the preferred perceptual tempo is around 1.6 Hz, and is thus comparable to the spontaneous motor tempo. This corroborates the view that motor and perceptual tempo preferences have common psychological bases (McAuley et al., 2006).

Some factors seem to affect the preferred perceptual and spontaneous motor tempi. For example, by investigating individuals from 4 to 95 years, two studies have observed that the preferred perceptual tempo slows as a function of age, thus paralleling a similar age-related motor slowing in spontaneous motor tempo (Drake et al., 2000; McAuley et al., 2006) (Fig. II.2.3.2.4): for children between 4 and 7 years, preferred tempo was typically of 2.5-3.3 Hz, whereas it was 1.6 Hz for adults, and close to 1.4 Hz for older adults (McAuley et al., 2006).

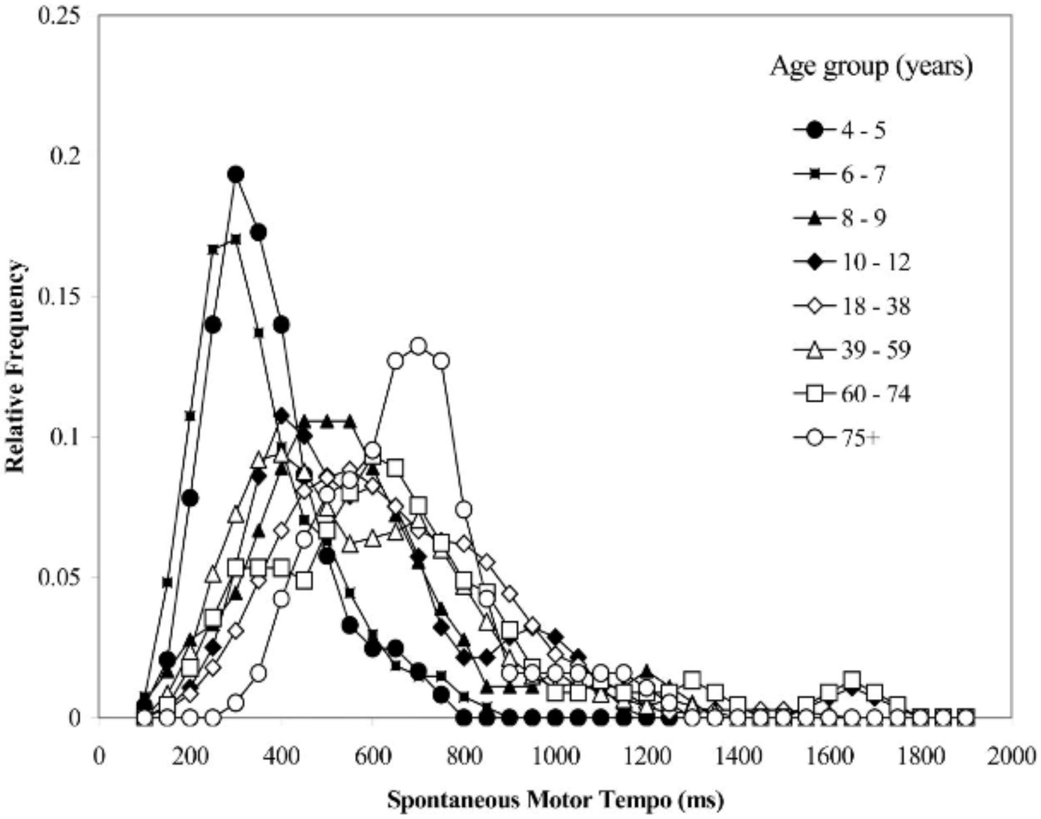


Figure II.2.3.2.4. From McAuley et al. (2006). Relative frequency distribution of spontaneous motor tempo across the various age groups.

This evolution of the preferred tempo across the lifespan seems itself flexible. For example, Drake et al. (2000) showed that preferred tempi of children with musical training tend to shift to adult rates sooner than children without musical training. This observation suggests that the developmental changes in preferred tempo may be a consequence of experience, rather than maturation, with more musical experience speeding this developmental change (McAuley, 2010). Interestingly, this learning perspective is consistent with other work having shown that even on the time scale of an experiment, participants seem to develop a global sense of the average rate of the events they are confronted to, and that this average rate influences judgement about the tempo or duration of events (Jones and McAuley, 2005).

Aside from effects of age and learning through exposure, researchers have investigated how individual differences for preferred tempo may be determined by body morphology. Indeed, due to biomechanical and dynamic properties of the body, each body segment should present its own resonance frequency (Toiviainen et al., 2010) (see also Section II.2.1). However, McAuley et al. (2011) failed to find a link between preferred perceptual tempo and physiological variables such as size of individuals (McAuley, 2010).

Similarly, one could expect that the preferred frequency for sensorimotor synchronization would depend on the body segment. For example, the preferred tempo for movements of the torso should be lower than the preferred tempo for movements of the limb extremities. This hypothesis is in line with evidence showing that humans listening to music are spontaneously entrained to move on distinct metric levels using specific body segments depending on the tempo (e.g., slow metric levels preferentially lead to move axial body segments such as bouncing the head, whereas fast metric levels tend to entrain more distal body parts, such as foot tapping) (Toiviainen et al., 2010; van Noorden and Moelants, 1999; McDougall and Moore, 2005).

II.2.3.3. Neural correlates of sensorimotor synchronization. *Functional MRI studies.* A large number of studies have explored sensorimotor synchronization using functional neuroimaging. They used finger tapping because (1) it is the most convenient movement to avoid concomitant head movements, and (2) it is flexible enough to accommodate various experimental conditions. These studies vary by the use or lack of pacing stimulus, and by the relative complexity of the tapping task (Witt et al., 2008).

Witt and colleagues (2008) performed an instructive meta-analysis of this literature to summarize the results found across these various experimental conditions. According to these authors, neuroimaging studies show concordant activation across the various pacing and task difficulty conditions in regions commonly associated with motor tasks, including the primary motor cortex (usually activated for the execution of simple or complex sequencial movements), supplementary motor area (also involved in the execution of simple voluntary movement, but also in cognitive functions as motor programming, motor learning, movement selection or initiation of the movement), basal ganglia (frequently observed as involved in the performance of simple repetitive movements as well as more complex sequential movements and internally generated movements; the latter case being observed with concomitant activation of the supplementary motor area) and cerebellum (also involved in the preparation, execution and timing of movements) (Fig. II.2.3.3a).

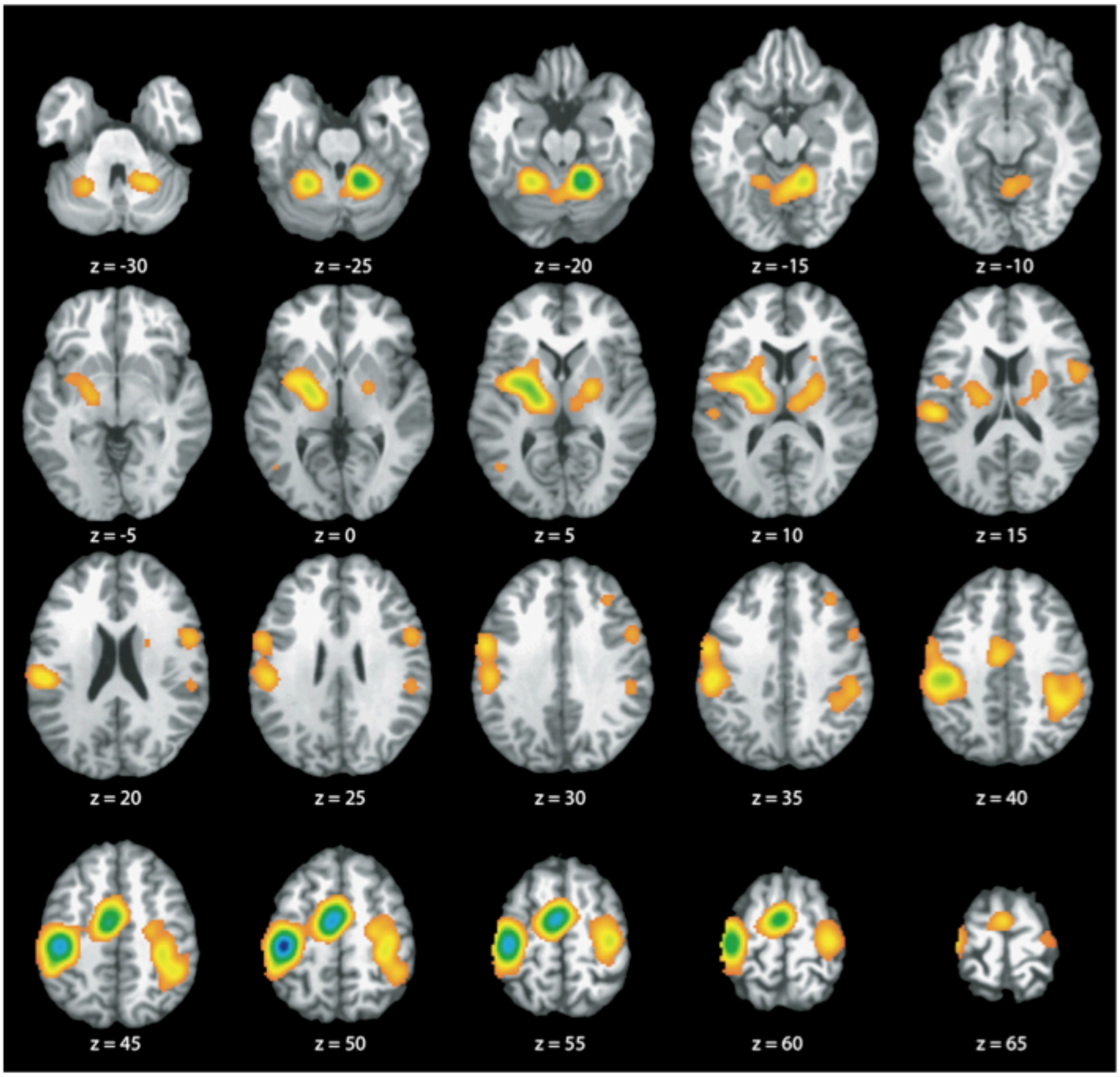


Figure II.2.3.3a. From Witt et al. (2008). Axial slices of the concordant activations found across all finger tapping task variations studied in this meta-anlysis (38 studies in total). Robust concordance is seen in bilateral sensorimotor cortices, supplementary motor area, left ventral premotor, bilateral inferior parietal cortices, bilateral basal ganglia and bilateral anterior cerebellum.

When comparing the neural activity elicited by a self-paced tapping and externally-paced tapping, three regions were found to be concomitantly activated in visually-paced, auditorially-paced and self-paced tapping: the primary sensorimotor cortex, supplementary motor area and anterior cerebellum (Witt et al., 2008). Interestingly, the basal ganglia, particularly the putamen, were activated only in the self-paced and auditorially-paced tapping but not with visually-paced tapping (the implications of these findings are discussed in Section II.2.3.7.). Moreover, the dorso-lateral prefrontal cortex, a region involved in sustained attention (Langner and Eickhoff, 2012), showed increased activation in self-paced tapping compared to externally-paced tapping. This result was interpreted as a consequence of the fact that self-paced tapping was more demanding than auditorially-paced tapping, which leads to more automatic repetitive movements (Rao et al., 2001). Other regions were found to be particularly activated during self-paced tapping compared to externally-paced tapping, including the cerebellum and inferior ventral premotor cortex, both regions whose activation is correlated with the complexity of motor tasks (Witt et al., 2008).

EEG/MEG studies. Self-paced and externally-paced finger tapping have been investigated to a lesser extent using EEG and MEG, probably because the lack of spatial resolution of these techniques does not allow disentangling easily movement-related potentials from potentials elicited by the processing of the external pace. In some of these studies, the electrophysiological activities elicited by the tapping movements were analyzed as single transient ERPs (Gerloff et al., 1997; Kopp et al., 2000; Gerloff et al., 1998; Müller et al., 2000; Pollock et al., 2003). By aligning the trials to the onset of the movement or to the tap, researchers recorded at the surface a response occurring around 100 ms before movement onset (Fig. II.2.3.3b). Source reconstructions of this activity indicated a generator within the primary motor cortex controlateral to the moving hand, suggesting that this potential reflected movement planning and execution. In addition, a potential was elicited around 100 ms after movement onset, whose source was located within the primary somatosensory cortex, suggesting that this potential reflected tactile and somatosensory feedback (Fig. II.2.3.3b).

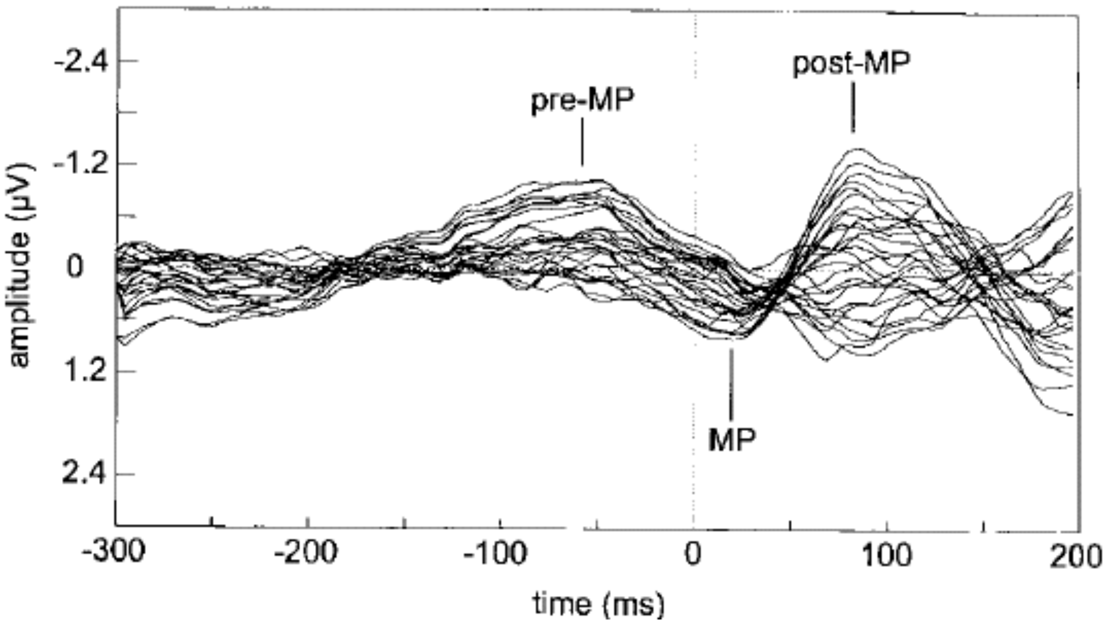
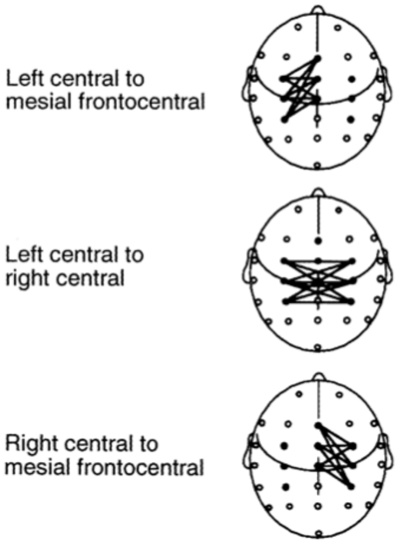
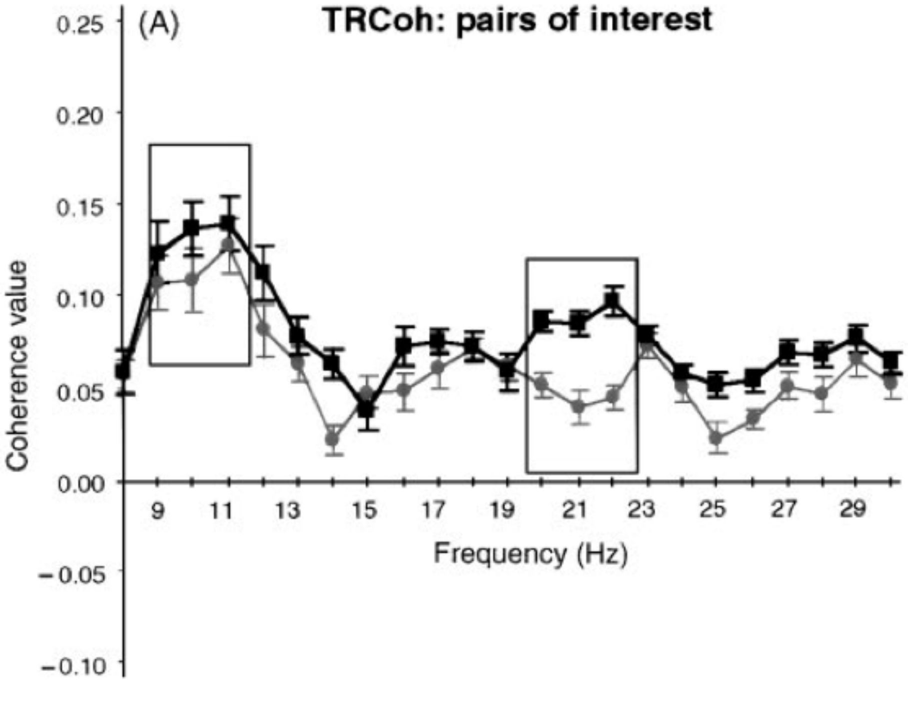


Figure II.2.3.3b. From Gerloff et al. (1998). Movement-related potentials elicited by repetitive finger movements of the right hand (overlay of 28 EEG channels, grand average of 10 subjects). The 0 ms time indicates movement onset.

By analyzing the data in the frequency domain, researchers observed that the repetitive movements elicited in the EEG spectrum steady-state evoked potentials at frequencies corresponding to the frequency of the periodic movement (Gerloff et al., 1997; 1998; Kopp et al., 2000; Osman et al., 2006; Bourguignon et al., 2012). This was presented as a powerful way to increase the signal-to-noise ratio with reduced testing duration. When trying to reconstruct the sources of this activity using the magnetic signals recorded with MEG at the surface, researchers found that most of this activity could be explained by dipoles located in the primary motor cortex contralateral to the moving segment, as well as dipoles in the contralateral primary sensory area (Bourguignon et al., 2012). In addition, dipoles located in the dorsolateral prefrontal cortex and posterior parietal cortex helped explaining the activity phase-locked to the periodic movement recorded at the surface, thus corroborating the results of fMRI studies (Bourguignon et al., 2012). However, results of source reconstructions vary across studies, and some authors could explain the main part of this activity by a single dipole located in the primary motor cortex contralateral to the moving hand, with no difference between self-paced and externally-paced conditions (Kopp et al., 2000).

Aside of studying single transient ERPs and SS-EPs elicited by repetitive movements, researchers also investigated modulations of the ongoing oscillations in various frequency bands (Gerloff et al., 1998; Manganotti et al., 1998; Pollock et al., 2005). The rationale of these investigations was that (1) several frequency bands, as beta band for instance, have been shown to be related to movement processing, either from EEG/MEG studies in humans or intracranial recording in non-human animals (Pfurtscheller and Lopes da Silva, 1999), and (2) the binding of distant brain areas necessary to synchronize a movement on an external stimulus was hypothesized to be supported by increased synchronization of the neural activity across distant brain regions (Pfurtscheller and Lopes da Silva, 1999; Varela et al., 2001; Singer et al., 1999). According to this view, researchers aimed to examine "*local activity*", represented by the local power, and "*synchronized oscillatory activity*", thought to reflect functional connectivity across spatially distributed neural networks (Pollock et al., 2006). This was assessed by calculating the degree of phase coherence across electrodes at the surface. Several MEG studies found a prevailing frequency of across-sensors coupling in two distinct frequency ranges, 8-12 Hz and 20-22 Hz (Gerloff et al., 1998; Manganotti et al., 1998; Pollock et al., 2005; 2006). This coupling was stronger in self-paced tapping conditions compared to externally-paced tapping. This result was interpreted as reflecting higher demands of the motor system during self-paced tapping. This interpretation was partly based on the fact that across-sensors coupling also increased with task complexity at these frequencies (Gerloff et al., 1998) (Fig. II.2.3.3c).


Left central to
mesial frontocentral
Left central to
right central
Right central to
mesial frontocentral


(A)
TRCoh: pairs of interest
Coherence value
0.25
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
9
11
13
15
17
19
21
23
25
27
29
Frequency (Hz)

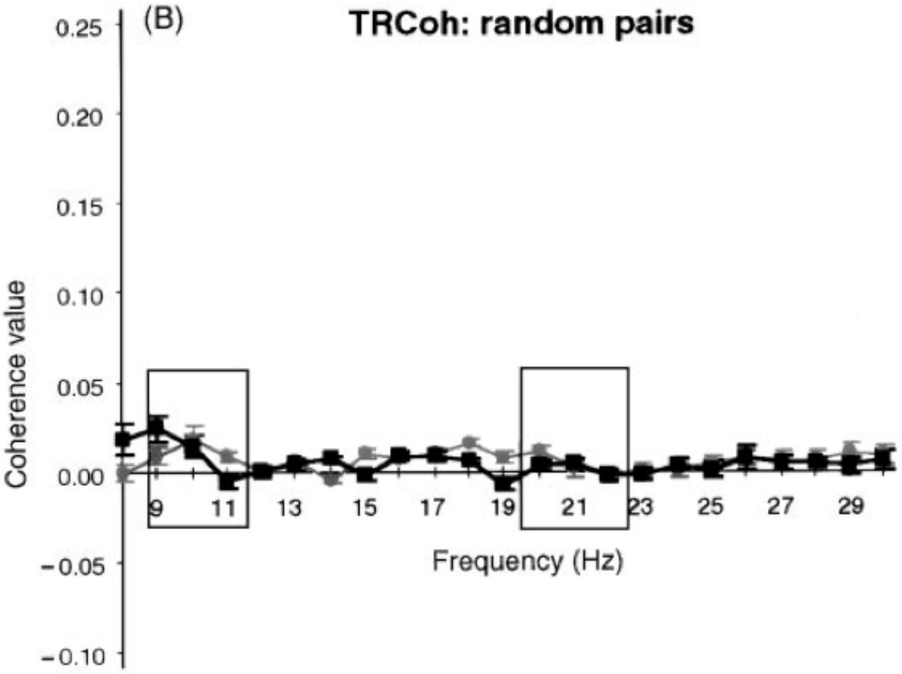


Figure II.2.3.3d. From Gerloff et al. (1998). Upper panel. In this study, the signal was recorded using a 28 electrodes EEG set-up. The Figure (upper panel) represents the electrode pairs of interest, determined based on the topography of the activities observed in the time domain (as represented in Figure II.2.3.3b). This leads to 27 electrode pairs of interest. Middle and bottom panels. Group data for chorence value during self-paced (black) and externally-paced (grey) movement (n=8) (error bars represent 2 SEM). Each data point represents an average of two 0.5 Hz bins of the fast Fourier spectrum and of 27 electrode pairs (A: electrode pairs of interest; b: random pairs). Windows indicate the frequency ranges in which the differences between the two movement conditions were most prominent (9-11 Hz, 20-22 Hz).

Nevertheless, indexing large-scale coupling using calculation of across-sensors phase coherence across a wide range of frequencies, while informative, could be biased by the neural responses to the external pacing stimulus, if present. Indeed, the frequency bands in which the highest coupling is found could contain both movement-related activity and activity related to the processing of an external pace. Moreover, the lack of spatial resolution across sensors remains an important limitation to interpret the results: an increased phase coupling within the same frequency band across various sensors could arise from a common source whose activity is projected across various sensor sites, or, it could arise from distant sources actually coupled. Finally, these studies did not examined, to our knowledge, the coupling occurring between the neural activity related to the movement and the activity elicited by the external stimulus whose movement has to be synchronized to. As explained in the next section, Study 3 of the present work aims to explore the neural correlates of sensorimotor synchronization to the beat, by overcoming these limitations using the SS-EP approach.

II.2.3.4. Study 3: CAPTURING WITH EEG THE NEURONAL ENTRAINMENT AND COUPLING UNDERLYING SENSORIMOTOR SYNCHRONIZATION TO THE BEAT (Sylvie Nozaradan, Younes Zerouali, Isabelle Peretz, André Mouraux)

This article is currently under revision.

Study 3 aims to explore the neural correlates of sensorimotor synchronization to the beat, by overcoming some limitations noticed in preceding studies. We hypothesized that movement-related and beat-related neural activities could be disentangled based on their respective frequencies, to further investigate their coupling thought to support sensorimotor synchronization. First, instead of searching for coupling across a wide range of frequency bands, we made a strong prediction of the frequencies at which the activities should be concentrated, based on the periodicity of the performed movement and the pacing sound. Second, the concentration of the movement-related and beat-related activities within very narrow frequency bands was hypothesized to help improving the signal-to-noise ratio of these activities, as this aspect is fundamental to further assess phase coherence and topographies. Third, instead of calculating coherence across electrodes within the same frequency band, the electrodes of interest were selected based on the topography of these activities, once observed in the frequency domain. Although the spatial resolution was still limited by the inherent constraints of the EEG, we hypothesized that this method could improve significantly the quality of the observed signals.

II.2.3.4.1. Abstract

Synchronizing movements with rhythmic inputs requires tight coupling of sensory and motor neural processes. Here, using a novel approach based on the recording of steady-state evoked potentials (SS-EPs), we examine how distant brain areas supporting these processes

coordinate their dynamics. The EEG was recorded while subjects listened to an auditory beat and tapped their hand on every second beat. We found that rhythmic sensorimotor synchronization was supported by a beat-related entrainment frequency-locked to the beat whose topography indicated activity originating from auditory areas, and a movement-related entrainment frequency-locked to the movement whose topography indicated activity originating from motor and/or somatosensory areas contralateral to the tapping hand. Most importantly, we found evidence for an interaction between sensory- and motor-related activities in the form of (1) a cross-modulation SS-EP appearing at the sum of beat- and movement-frequencies, indicating a non-linear process of sensorimotor integration, (2) a phase coupling of beat- and movement-related responses and (3) a selective enhancement of beat-related activities over the hemisphere contralateral to the tapping hand, suggesting a top-down effect of movement-related activities on auditory beat processing. Taken together, our results indicate that rhythmic sensorimotor synchronization is supported by a dynamic coupling of sensory and motor related activities.

II.2.3.4.2. Introduction

Synchronizing movements to external inputs is best observed with music (London 2004, Repp 2005, Janata 2011). Perception of beat in music refers to the spontaneous human ability to perceive periodicities from sequences of sounds (London 2004). The regular temporal structure of beats is thought to facilitate movement synchronization on musical rhythms. Indeed, one of the fascinating aspects of beat perception is its strong relationship with movement (Phillips-Silver and Trainor 2005, Madison et al. 2006, Grahn and Brett 2007, Chen et al. 2008, Phillips-Silver et al. 2010, Janata et al. 2011, Teki et al. 2011). First, music spontaneously entrains humans to move (van Noorden and Moelants 1999, Madison et al.

2006, Phillips-Silver et al. 2010, Janata et al. 2011). Second, it has been shown that movement influences the perception of musical rhythms (Phillips-Silver and Trainor 2005, 2007). Third, functional neuroimaging studies have shown that motor cortical areas are activated when listening to rhythmic sequences (Grahn and Brett 2007, Chen et al. 2008, Teki et al. 2011).

How distant brain areas involved in sensorimotor synchronization are able to coordinate their dynamics remains, at present, largely unknown. Two main theories have been proposed (Pressing 1998, Repp 2005). The information-processing theory considers responses to the beat as a succession of discrete events and proposes that sensorimotor synchronization is achieved through error correction mechanisms occurring from one event to the other (Vorberg and Wing 1995, Pressing 1998, Praamstra et al. 2003). Contrasting with this view, the dynamic systems theory postulates that sensorimotor synchronization results from a dynamic and continuous coupling of sensory and motor oscillators (Beek et al. 2002, Hogan and Sternad 2007). This second hypothesis is in line with the hypothesis that beat perception is represented in the human brain as a dynamic prediction or attending process, in which the periodic temporal structure of the auditory beat entrains the listener's attention, leading to a periodic modulation of expectancy as a function of time (Jones and Boltz 1999, Large and Jones 1999). Building on this view, the resonance theory for beat perception (Large and Kolen 1994, Large 2008) proposes that the perception of periodicities in music emerges from the entrainment of neuronal populations resonating at the frequency of the beat. Recently, we provided direct evidence of such neuronal entrainment to musical beats in humans using electroencephalography (EEG) (Nozaradan et al. 2011). Specifically, we showed that an auditory beat elicits a periodic neural response at the exact frequency of the beat. This oscillatory activity, frequency-locked to the beat and captured in the EEG signal in the form of a steady-state evoked potential (SS-EP) (Regan 1989), was hypothesized to reflect the neuronal entrainment underlying beat perception.

Here, we used this novel approach to explore the neural dynamics supporting sensorimotor synchronization to the beat, that is, the performance of overt movements paced on the beat. Specifically, we examined whether sensorimotor synchronization to the beat is supported in the human brain by two distinct neuronal entrainments: a neuronal entrainment at the frequency of the beat which would underlie beat processing (Large 2008, Nozaradan et al. 2011) and a distinct neuronal entrainment at the frequency of the movement which would underlie the production of synchronized movements (Gerloff et al. 1997, 1998, Kopp et al. 2000, Daffertshofer et al. 2005, Kourtis et al. 2008, Bourguignon et al. 2011). Most importantly, this frequency-tagging approach allowed disentangling, at least to some degree, beat- and movement-related EEG responses and, thereby, characterizing their coupling, hypothesized to support sensorimotor synchronization to the beat.

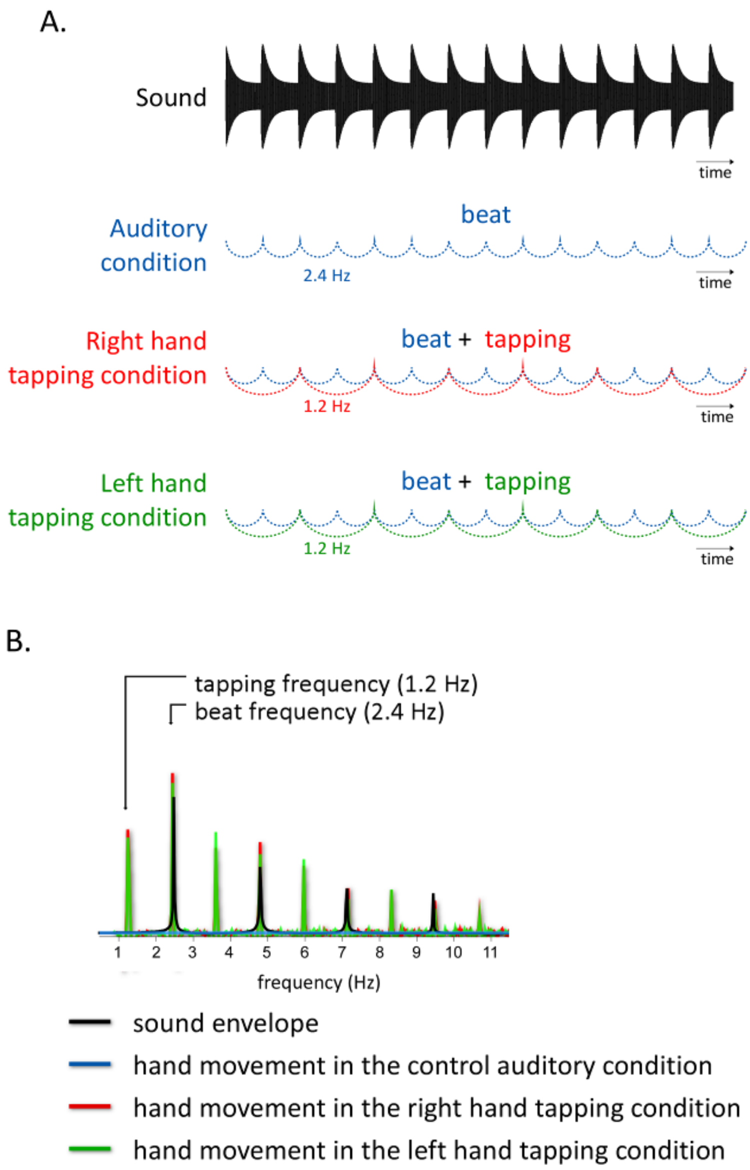


Figure 1. Experimental paradigm and hand movement signals. A. The auditory stimulus consisted of a 33 s 333.33 Hz pure tone in which a beat was introduced by modulating the amplitude of the tone with a 2.4 Hz periodicity. The upper graph shows a 6 s excerpt of its sound envelope. In the control auditory condition, the participants were asked to listen to the sound, in order to detect the occasional occurrence of a short discontinuity. In the left and right hand tapping conditions, the participants were asked to perform a hand tapping movement paced on every second beat in the sequence (i.e. at half the frequency of the beat). B. Frequency spectrum of the sound envelope (black) and the accelerometer signals recorded from the hand in the control auditory condition (blue), the right hand tapping condition (red) and the left hand tapping condition (green).

II.2.3.4.3. Materials and methods

Subjects

Eight healthy volunteers (3 females, all right-handed, mean age 27 ± 4 years, aged between 22 and 36) took part in the study after providing written informed consent. They all had musical experience, either in performance (3 subjects with 15-25 years of practice) or as amateur listeners or dancers. None had prior experience with the tapping task used in the present study. They had no history of hearing, neurological or psychiatric disorder, and were not taking any drug at the time of the experiment. The study was approved by the local Ethics Committee.

Auditory stimulation

Each auditory stimulus lasted 33 s. The stimulus consisted of a 333.33 Hz pure tone in which a beat was introduced by modulating the amplitude of the tone with a 2.4 Hz periodicity (i.e., 144 beats per minute), using an asymmetrical Hanning envelope (12 ms rise time and 404 ms fall time, amplitude modulation between 0.25 and 1). A 2.4 Hz periodicity was chosen because (1) this tempo lies within the ecological range of tempo perception and production (Drake and Botte 1993), (2) we previously showed that this beat frequency elicits a measurable beat-related steady-state evoked potential (SS-EP) in the human EEG (Nozaradan et al. 2011) and (3) pilot experiments showed that subjects are comfortable in tapping on every second beat using this beat frequency. The auditory stimuli were generated using the PsychToolbox extensions (Brainard 1997) running under Matlab 6.5 (The MathWork, USA), and presented binaurally through electromagnetically shielded insert earphones at a comfortable hearing level (ER 3A, Etymotic Research, Elk Grove Village, IL, USA).

Experimental conditions

Subjects were comfortably seated in a chair with their head resting on a support. They were instructed to relax, avoid any unnecessary head or body movement and keep their eyes fixated on a point displayed on a computer screen in front of them. Subjects were asked to perform three different tasks: a control auditory task, a right hand tapping task and a left hand tapping task, in three separate conditions (Fig. 1). The tapping task was performed using each hand to assess the hemispheric lateralization of EEG activities relative to the tapping hand. Each condition consisted of 6 trials during which the 33 s auditory stimulus was presented after a 3 s foreperiod. Stimulus presentation was self-paced. During the first condition, subjects performed the control auditory task. They were asked to listen carefully to the periodic sound in order to detect the occurrence of a very short (4 ms) sound interruption, inserted at a random position in two additional trials interspersed within the block. The subjects were instructed to report their detection at the end of each trial. This control task required a sustained level of attention, and thus ensured that attention was focused on the sound. The two trials containing a short interruption were excluded from further analyses. During the second condition, subjects performed the right hand tapping task. They were asked to perform tapping movements with their right hand, accurately paced on every second beat of the sequence, i.e., at half the beat frequency ($f/2 = 1.2$ Hz). During the third condition, they performed the same task, this time using their left hand. Before the right and left hand tapping conditions, a short training session ensured that subjects understood the task. The subjects were asked to start their tapping as soon as they heard the first auditory beat of the stimulus, and to maintain their movement as synchronized as possible throughout the entire trial. The tapping was performed with small up and down movements of the hand starting from the wrist joint, maintaining the forearm and elbow fixed on an armrest cushion. When performing the tapping movement, the fingers of the tapping hand came transiently in contact with the armrest cushion. All subjects naturally synchronized their movement such that the occurrence of this contact coincided with the occurrence of the beat. Importantly, the contact with the armrest produced tactile feedback, but did not produce any auditory feedback, as the subjects were fitted with earphone inserts. The experimenter remained in the recording room at all times, to monitor compliance to these instructions.

EEG recording

The 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 on the inferior and superior areas of the left orbit. Electrode impedances were kept below 10 kΩ. The signals were amplified, low-pass filtered at 500 Hz, digitized using a sampling rate of 1000 Hz and referenced to an average reference (64-channel high-speed amplifier, Advanced Neuro Technology, The Netherlands).

Hand movement recordings

Movements of the hand were measured using a 3-axis accelerometer attached to the hand dorsum (MMA7341L, Pololu Robotics & Electronics, USA). The signals generated by the accelerometer were digitized using three additional bipolar channels of the EEG system. Only the vertical axis of the accelerometer signal was analyzed, as it sampled the greatest part of the accelerations related to the tapping movement.

Frequency-domain analysis

SS-EP amplitude. Continuous EEG recordings were filtered using a 0.1-Hz high-pass Butterworth zero-phase filter to remove very slow drifts in the recorded signals. Epochs lasting 32 s were obtained by segmenting the recordings from +1 to +33 s relative to the onset of the auditory stimulus, thus yielding 6 epochs for each subject and condition. The EEG recorded during the first second of each epoch was removed (1) to discard the transient auditory evoked potentials related to the onset of the stimulus (Saupe et al. 2009, Nozaradan et al. 2011, Nozaradan et al. 2012), (2) because steady-state evoked potentials (SS-EPs) require several cycles of stimulation to be steadily entrained (Regan 1989) and (3) because several repetitions of the beat are required to elicit a steady perception of beat (Repp 2005). These EEG processing steps were carried out using Analyzer 1.05 (Brain Products, Germany).

Artifacts produced by eye blinks or eye movements were removed from the EEG signal using a validated method based on an Independent Component Analysis (Jung et al. 2000), using the runica algorithm (Bell and Sejnowski 1995, Makeig 2002). For each subject and condition, EEG epochs were averaged across trials. The time-domain averaging procedure was used to enhance the signal-to-noise ratio of beat- and movement-related EEG activities by attenuating the contribution of activities that were not strictly phase-locked across trials, i.e., activities that are not phase-locked to the sound stimulus. The obtained average waveforms were then transformed in the frequency domain using a discrete Fourier transform (Frigo and Johnson 1998), yielding a frequency spectrum of signal amplitude (µV) ranging from 0 to 500 Hz with a frequency resolution of 0.031 Hz (Bach and Meigen 1999). These EEG processing steps were carried out using Letswave (Mouraux and Iannetti 2008), Matlab (The MathWorks, USA) and EEGLAB (http://sccn.ucsd.edu).

Within the obtained frequency spectra, signal amplitude may be expected to correspond to the sum of (1) EEG activity induced by the auditory beat and/or the hand movement task, referred as to beat- and movement-related SS-EPs and (2) unrelated residual background noise due, for example, to spontaneous EEG activity, muscle activity or eye movements. Therefore, to obtain valid estimates of beat- and movement-related 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 (2 frequency bins ranging from - 0.15 to - 0.09 Hz and from + 0.09 to + 0.15 Hz relative to each frequency bin). The validity of this subtraction procedure relies on the assumption that, in the absence of an SS-EP, the signal amplitude at a given frequency bin should be similar to the signal amplitude of the mean of the surrounding frequency bins (Mouraux et al. 2011, Nozaradan et al. 2011, Nozaradan et al. 2012). This subtraction procedure is important because the magnitude of the background noise is not equally distributed across scalp channels (Supplementary Material). Indeed, without this subtraction procedure, the scalp topographies of the elicited SS-EPs would reflect a combination of the scalp distribution of the SS-EP response and the scalp distribution of the background noise present at that frequency.

The magnitude of beat- and movement-related SS-EPs was then estimated by averaging the noise-subtracted amplitudes measured at the frequency bins centered over the expected 1.2 Hz movement-related SS-EP (bins ranging from 1.178 - 1.240 Hz) and the expected 2.4 Hz beat-related SS-EP (bins ranging from 2.356 - 2.418 Hz). This range of frequencies allowed accounting for possible spectral leakage due to the fact that the discrete Fourier transform did not estimate signal amplitude at the exact frequency of each SS-EP (Nozaradan et al. 2011, Nozaradan et al. 2012). In the right and left hand tapping conditions, an additional SS-EP appeared at 3.6 Hz in the EEG frequency spectrum of all subjects (Figs. 2 and 3). The magnitude of this additional SS-EP was estimated by averaging the signal amplitude measured at the bins ranging from 3.562 - 3.624 Hz.

To exclude any electrode selection bias, the estimated magnitudes of the 1.2, 2.4 and 3.6 Hz SS-EPs were averaged across all scalp electrodes, for each condition and participant (Figs. 2 and 3). A one-sample t-test was then used to determine whether the average SS-EP amplitudes were significantly different from zero. Indeed, in the absence of an SS-EP, the average of the noise-subtracted signal amplitude may be expected to tend towards zero. Furthermore, a one-way repeated-measures ANOVA was used to compare the magnitude of the SS-EPs recorded in each of the three experimental conditions (control, left hand tapping, right hand tapping). Degrees of freedom were corrected using the Greenhouse-Geisser correction for violations of sphericity. Size effects were expressed using the partial Eta-squared. When significant, post-hoc pairwise comparisons were performed using paired-sampled t-tests. Significance level was set at $p < 0.05$. Finally, for each condition, topographical maps were computed by spherical interpolation for the 1.2, 2.4 and 3.6 Hz SS-EPs (Fig. 2).

Hand tapping movement. The accelerometer signals generated by the hand tapping movements were analyzed in the frequency domain using the same procedure used to analyze the EEG signals. A one-sample t-test was used to examine whether the noise-subtracted amplitude measured at 1.2 Hz in the tapping conditions was significantly different from zero. Indeed, in the absence of a periodic hand movement at 1.2 Hz, the average of the noise-subtracted signal amplitude may be expected to tend towards zero. Furthermore, a one-sample t-test was used to compare the noise-subtracted amplitude at 1.2 Hz obtained across the two tapping conditions (right and left hand tapping).

Time-domain analysis

Time domain analyses were carried out to assess the phase coupling between beat-related and movement-related EEG activities (Figs. 5, 6 and 7), as well as the phase coupling between these activities and the hand tapping movement (Figs. 6 and 7), as follows.

Hand tapping movement. The vertical acceleration signals recorded within each subject and tapping conditions were averaged across trials, to attenuate the contribution of unrelated signals not phase-locked to the sound stimulus. These average waveforms, lasting 33 s, were further segmented in epochs aligned to the occurrence of each beat onto which the subjects tapped, and extending from -0.416 to +0.416 s relative to the beat onset. Thus, these epochs expressed the across-trial average of the time course of the movement signal within each 1.2 Hz tapping-movement cycle (Fig. 6). The tapping latency relative to the beat occurrence was defined as the latency of maximum acceleration within each tapping-movement cycle (Figs. 6 and 7). To assess the synchronization lag of each subject, a one-sample t-test against zero was used to examine whether the mean relative tapping latencies were significantly different from zero in each of the two tapping conditions. Furthermore, a paired sample t-test was used to compare tapping latencies obtained in the left and right hand tapping conditions.

Beat-related SS-EP. For each subject and condition, EEG oscillations corresponding to the beat-related SS-EP were extracted using a narrow band-pass FFT filter centered at 2.4 ± 0.4 Hz as width of the FFT window (Hanning function). The filtered signals were then averaged and segmented using the same procedure as for the accelerometer signals (Fig. 6). The electrode displaying the maximum amplitude at 2.4 Hz within the noise-subtracted frequency spectra was chosen as electrode of interest (Fig. 5). This electrode selection criterion was used to maximize the signal-to-noise ratio of the analyzed signals and, thereby, reduce the contribution of residual noise when estimating phase coupling. Within each tapping-movement cycle, the relative latency of the beat-related SS-EP was arbitrarily defined as the latency of maximum amplitude of the oscillation following the occurrence of the beat (Figs. 6 and 7). A repeated-measures ANOVA was then used to compare the mean relative latencies obtained in each condition (control, right hand tapping and left hand tapping). When significant, post-hoc pairwise comparisons were performed using paired-sampled t-tests.

Movement-related SS-EP. The same procedure was used to extract EEG oscillations corresponding to the movement-related SS-EP, and to estimate their relative latencies in the right and left hand tapping conditions. The movement-related SS-EP was extracted using a band-pass filter centered at 1.2 ± 0.4 Hz. The electrode displaying the maximum amplitude at 2.4 Hz within the noise-subtracted frequency spectra was chosen as electrode of interest (Fig. 5). The relative latency of the movement-related SS-EP was arbitrarily defined as the latency of maximum value of amplitude of the oscillation preceding the occurrence of the beat (Figs. 6 and 7). A paired sample t-test was used to compare the mean relative latencies obtained in the left and right hand tapping conditions.

Phase coupling. The relative latencies of the beat- and movement-related SS-EPs as well as that of the hand tapping movements were averaged across the left and right hand tapping conditions, as the values obtained for each of the two conditions were not significantly different. A Pearson's correlation test was then used to examine the relationship between the latency of the beat-related SS-EP and the hand tapping movement, the movement-related SS-EP and the hand tapping movement, as well as the beat-related and movement-related SS-EPs.

Topographical distribution of the SS-EPs

Current source density (CSD) estimates of the band-pass filtered time-domain EEG signals averaged relative to beat onset were used to better assess the differences between the topographical distributions of the beat-related 2.4 Hz SS-EPs, the movement-related 1.2 Hz SS-EPs and the 3.6 Hz cross-modulation SS-EPs obtained in each condition (Fig. 4). Topographical maps were computed using the group-level average signals, after normalization of the values of each subject. The estimates were computed using spherical spline interpolation, as implemented in the CSD toolbox (Kayser and Tenke 2006a). For each frequency and condition, group-level average topographical maps were computed at the latency of maximum global field power (obtained after analyzing the signal as explained in the Time-domain analysis section above). A Pearson's correlation test was used to assess the relationship between the topographical distributions of the 1.2 and 2.4 Hz SS-EPs, the 1.2 and 3.6 Hz SS-EPs, as well as the 2.4 and 3.6 Hz SS-EPs.

Furthermore, we examined in these current source density estimates whether the topographical distributions of the movement-related 1.2 Hz SS-EP was lateralized relative to the tapping hand (Fig. 4). For each subject and hand tapping condition, the magnitude of the 1.2 Hz SS-EP recorded at the central electrode ipsilateral to the tapping hand (C4 and C3, for right and left hand tapping respectively) was compared to the magnitude recorded at the central electrode contralateral to the tapping hand (electrodes C3 and C4, for right and left hand tapping respectively), using a paired-sample t-test (Coles 1989). The same procedure was used to assess the lateralization of the beat-related 2.4 Hz SS-EP as well as the 3.6 Hz SS-EP, using electrodes FC2 and FC1, as these electrodes showed the highest magnitude in the group-level averaged maps (Fig. 4).

II.2.3.4.4. Results

Hand tapping movement

The noise-subtracted frequency amplitude spectra of the accelerometer signals was, at 1.2 Hz, significantly greater than zero in both the right (t = 4.21, p = 0.004, df = 7) and the left (t = 4.58, p = 0.002, df = 7) hand tapping conditions (Fig. 1). The magnitude of the 1.2 Hz hand movement signal was not significantly different across the two tapping conditions (t = 0.33, p = 0.74, df = 7) (Fig. 1).

Beat-related, movement-related and cross-modulation SS-EPs

As shown at the group-level average of the frequency spectra (Fig. 2) as well as the individual-level frequency spectra (Fig. 3), in all three conditions, the auditory beat elicited a clear increase of signal amplitude at 2.4 Hz, corresponding to the frequency of the beat, and referred to as beat-related steady-state evoked potential (SS-EP). The term ‘beat’ was used because the frequency of the periodic modulation of the auditory stimulus was within the frequency range of musical tempo, i.e., a frequency at which sensorimotor synchronization to periodic sensory input is frequently observed. Hence, the neuronal activity captured in this SS-EP could, at least in part, be functionally distinct from that captured by auditory SS-EPs elicited using higher stimulation frequencies (e.g. 20-40 Hz; Draganova et al. 2002).

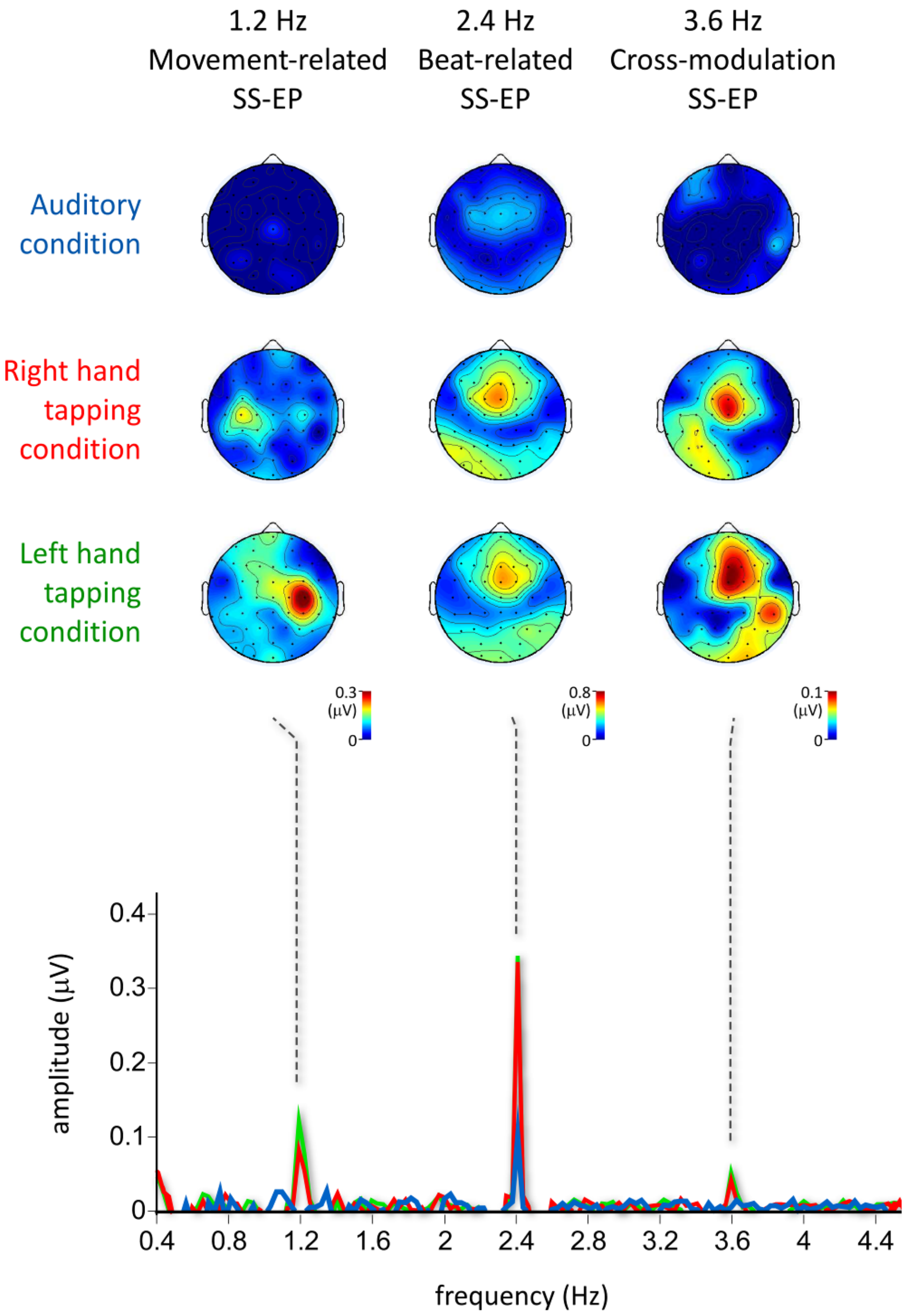


Figure 2. Group-level average frequency spectra (Hz) of the noise-subtracted EEG amplitude signals obtained in the control auditory condition (blue), the right hand tapping condition (red) and the left hand tapping condition (green), averaged across all scalp channels. In all conditions, the 2.4 Hz auditory beat elicited an SS-EP at 2.4 Hz. As shown in the corresponding topographical maps, this beat-related SS-EP was maximal over fronto-central electrodes. In the left and right hand-tapping conditions, the 1.2 Hz hand tapping movement was related to the appearance of an additional SS-EP at 1.2 Hz. As shown in the topographical maps, this movement-related SS-EP was maximal over the central electrodes contralateral to the moving hand. In these two conditions, an additional SS-EP

emerged at 3.6 Hz, referred to as cross-modulation SS-EP, whose scalp topography showed patterns similar to both beat-related and movement-related SS-EPs topographies.

In the left and right hand tapping conditions, an additional peak in the EEG frequency spectra was observed at 1.2 Hz, corresponding to the frequency of the tapping movement, and referred to as movement-related SS-EP. In these conditions, a third peak was also observed at 3.6 Hz, referred to as cross-modulation SS-EP.

Beat-related SS-EP. The noise-subtracted amplitude of the 2.4 Hz beat-related SS-EP, averaged across all scalp electrodes, was 0.09 ± 0.01 µV in the control auditory condition, 0.33 ± 0.03 µV in the right hand tapping condition and 0.34 ± 0.05 µV in the left hand tapping condition (mean ± standard error of the mean) (Figs. 2 and 3). The increase in signal amplitude at this frequency was significant in all three conditions (control auditory condition: $t = 6.05$, $p = 0.0005$, $df = 7$; right hand tapping condition: $t = 10.1$, $p < 0.0001$, $df = 7$; left hand tapping condition: $t = 6.6$, $p = 0.0003$, $df = 7$). The magnitude of the beat-related SS-EP was significantly different across conditions ($F = 31.7$, $p < 0.0001$, $\eta^2 = 0.82$, $df = 7$) (Figs. 2 and 3). Post-hoc comparisons showed that the magnitude of the EEG signal at 2.4 Hz was significantly greater in the tapping conditions than in the control auditory condition (respectively $t = 9.6$, $p < 0.0001$, $df = 7$ and $t = 5.3$, $p = 0.001$, $df = 7$) (Fig. 3).

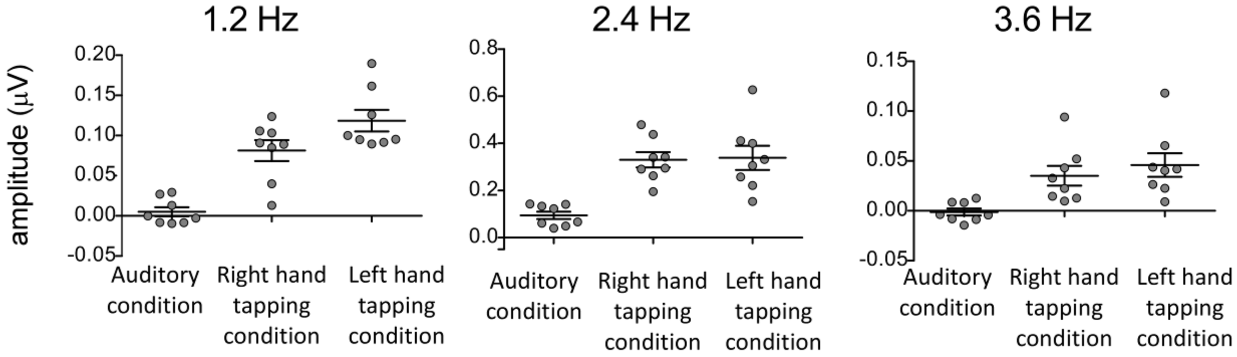


Figure 3. Amplitude of the 1.2, 2.4 and 3.6 Hz EEG activity in the control auditory condition, the right hand tapping condition and the left hand tapping condition. Dots represent individual noise-subtracted amplitude values obtained at each target frequency, averaged across all scalp electrodes. The whisker plots represent the group-level mean and standard error of the mean.

As shown in Figures 2 and 4, in the control auditory condition, the scalp topography of the beat-related SS-EP was maximal over frontal and temporal regions, and symmetrically distributed over both hemispheres. In contrast, in the left and right hand tapping conditions, the scalp topographies were clearly asymmetrical, and maximal over the hemisphere contralateral to the tapping hand. Comparison of the signals recorded at electrodes FC1 and FC2 confirmed this lateralization of the beat-related SS-EP in the hand tapping conditions (t = 5.38, p = 0.001, df = 7).

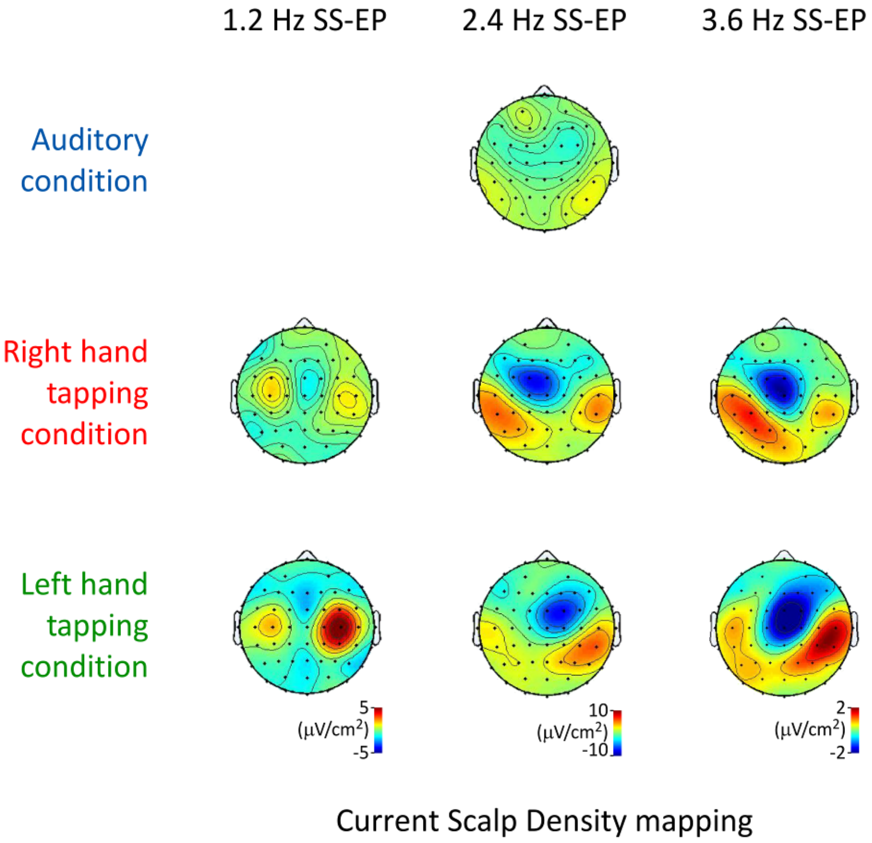


Figure 4. Topographic maps (group-level average current scalp density) of the 1.2 Hz, 2.4 Hz and 3.6 Hz SS-EPs obtained in the control auditory condition, the right hand tapping condition and the left hand tapping condition.

Movement-related SS-EP. The noise-subtracted amplitude of the 1.2 Hz movement-related SS-EP, averaged across all scalp electrodes, was 0.005 ± 0.006 µV in the control auditory condition, 0.08 ± 0.01 µV in the right hand tapping condition and 0.12 ± 0.01 µV in the left hand tapping condition (Figs. 2 and 3). The increase of signal amplitude at this frequency was significant in the right and left hand tapping conditions (right hand tapping condition: $t = 6.3$, $p = 0.0004$, $df = 7$; left hand tapping condition: $t = 8.9$, $p < 0.0001$, $df = 7$), but not in the control auditory condition ($t = 0.89$, $p = 0.4$, $df = 7$). The magnitude of the movement-related SS-EP was significantly different across conditions ($F = 19.1$, $p < 0.0001$, $\eta^2 = 0.73$, $df = 7$)

(Figs. 2 and 3). Post-hoc comparisons showed that the magnitude of the EEG signal at 1.2 Hz was significantly greater in the right and left hand tapping conditions than in the control auditory condition ($t = 4.8$, $p = 0.002$, $df = 7$ and $t = 8.4$, $p < 0.0001$, $df = 7$ respectively) (Figs. 2 and 3).

As shown in Figures 2 and 4, the scalp topography of the movement-related SS-EP was maximal over the central region contralateral to the hand movement. Comparison of the signals recorded at electrodes C3 and C4 confirmed a significant lateralization of the movement-related SS-EP ($t = 3.32$, $p = 0.01$, $df = 7$).

Cross-modulation SS-EP. The noise-subtracted amplitude of the 3.6 Hz SS-EP, averaged across all scalp electrodes, was -0.001 ± 0.003 µV in the control auditory condition, 0.035 ± 0.01 µV in the right hand tapping condition and 0.05 ± 0.02 µV in the left hand tapping condition (Figs. 2 and 3). The increase in signal amplitude at this frequency was significant in the right and left hand tapping conditions (right hand tapping condition: $t = 3.53$, $p = 0.01$, $df = 7$; left hand tapping condition: $t = 3.86$, $p = 0.006$, $df = 7$), but not in the control auditory condition ($t = 0.36$, $p = 0.73$, $df = 7$). The magnitude of this additional SS-EP was significantly different across conditions ($F = 10.5$, $p = 0.001$, $\eta^2 = 0.6$, $df = 7$) (Figs. 2 and 3). Post-hoc comparisons showed that the magnitude of the EEG signal at 3.6 Hz was significantly greater in the right and left hand tapping conditions than in the control auditory condition (respectively $t = 3.4$, $p = 0.012$, $df = 7$ and $t = 3.8$, $p = 0.007$, $df = 7$) (Fig. 3).

As shown in Figures 2 and 4, the scalp topography of the 3.6 Hz SS-EP observed in the left and right hand tapping conditions was very similar to that of the 2.4 Hz SS-EP obtained in these same conditions. Indeed, both displayed a maximum over frontal and temporal regions, were clearly asymmetrical, and maximal over the hemisphere contralateral to the tapping hand. Comparison of the signals recorded at electrodes FC1 and FC2 confirmed the lateralization of the 3.6 Hz SS-EP obtained in the left and right hand tapping conditions (t = 3.21, p = 0.01, df = 7).

A significant correlation was found between the scalp topographies of the 3.6 Hz SS-EP and the 2.4 Hz beat-related SS-EP (right hand tapping: $r^2 = 0.69$, $p < 0.0001$; left hand tapping: $r^2 = 0.64$, $p < 0.0001$). In contrast, there was no significant correlation between the scalp topographies of the 3.6 Hz SS-EP and the 1.2 Hz movement-related SS-EP (right hand tapping: $r^2 = 0.04$, p = 0.11; left hand tapping: $r^2 = 0.005$, p = 0.53). Furthermore, there was also no significant correlation between the 1.2 Hz and the 2.4 Hz SS-EPs in the left hand tapping condition ($r^2 = 0.01$, p = 0.29), and a weak negative correlation in the right hand tapping condition ($r^2 = 0.017$, p = 0.001).

Phase coupling

The mean latency of the tapping movement was -0.036 ± 0.01 s in the right hand tapping condition and -0.046 ± 0.01 s in the left hand tapping condition, relative to the actual occurrence of the beat (Fig. 7). These latencies were significantly different from zero (right hand tapping condition: t = 32.5, $p < 0.0001$, df = 7; left hand tapping condition: t = 39.69, $p < 0.0001$, df = 7). That is, subjects exhibited a mean negative asynchrony in their synchronization performance (Repp 2005; Aschersleben 2002), i.e., a systematic anticipation of the hand tapping movement relative to the beat onset, ranging from a hundred of ms before the beat for some subjects to no anticipation for other subjects (Fig. 7). The tapping latencies were not significantly different across the two tapping conditions (t = 1.17, p = 0.28, df = 7) (Fig. 8).

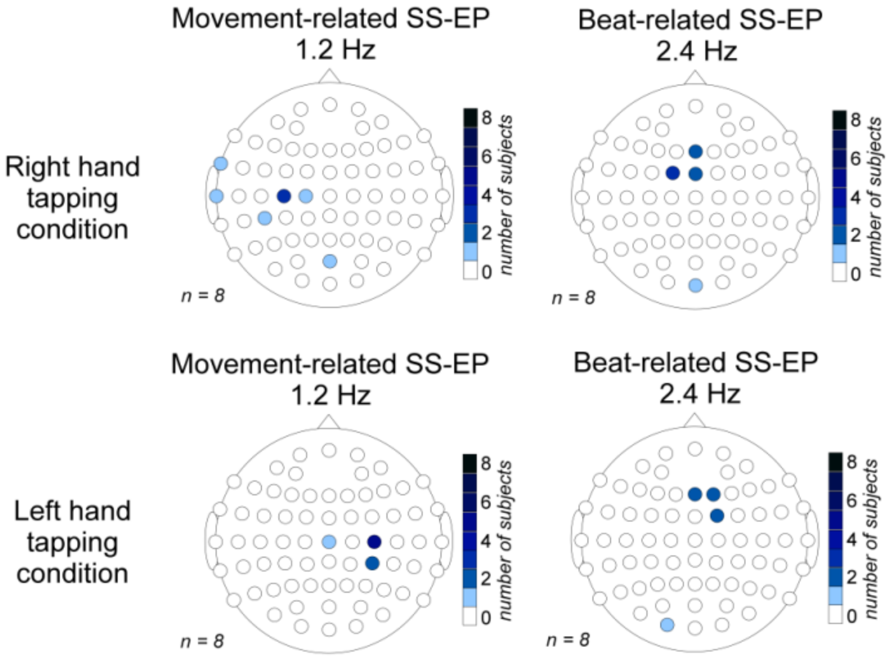


Figure 5. Topographical location of the electrodes at which the 1.2 Hz movement-related SS-EP and the 2.4 Hz beat-related SS-EP displayed maximal amplitude in each participant and hand tapping condition. Note that the beat-related SS-EP was predominant over fronto-central electrodes whereas the movement-related SS-EP was predominant over the central electrodes contralateral to the tapping hand.

The latency of the beat-related SS-EP, expressed relative to the actual beat occurrence, was 0.044 ± 0.02 s in the control auditory condition, 0.154 ± 0.01 s in the right hand tapping condition and 0.144 ± 0.01 s in the left hand tapping condition. These values were significantly different across conditions ($F = 12.37$, $p = 0.0008$, $\eta^2 = 0.63$, $df = 7$). Post-hoc comparisons showed that the relative latency of the beat-related SS-EP obtained in the control auditory condition was significantly different from the relative latency of the beat-related SS-EP obtained in both the right hand tapping condition ($t = 3.63$, $p = 0.008$, $df = 7$) as well as the left hand tapping condition ($t = 3.39$, $p = 0.01$, $df = 7$).

The relative latency of the movement-related SS-EP was of -0.336 ± 0.2 s in the right hand tapping condition and -0.306 ± 0.02 s in the left hand tapping condition. These values were not significantly different across the two tapping conditions (t = 1.77, p = 0.12, df = 7).

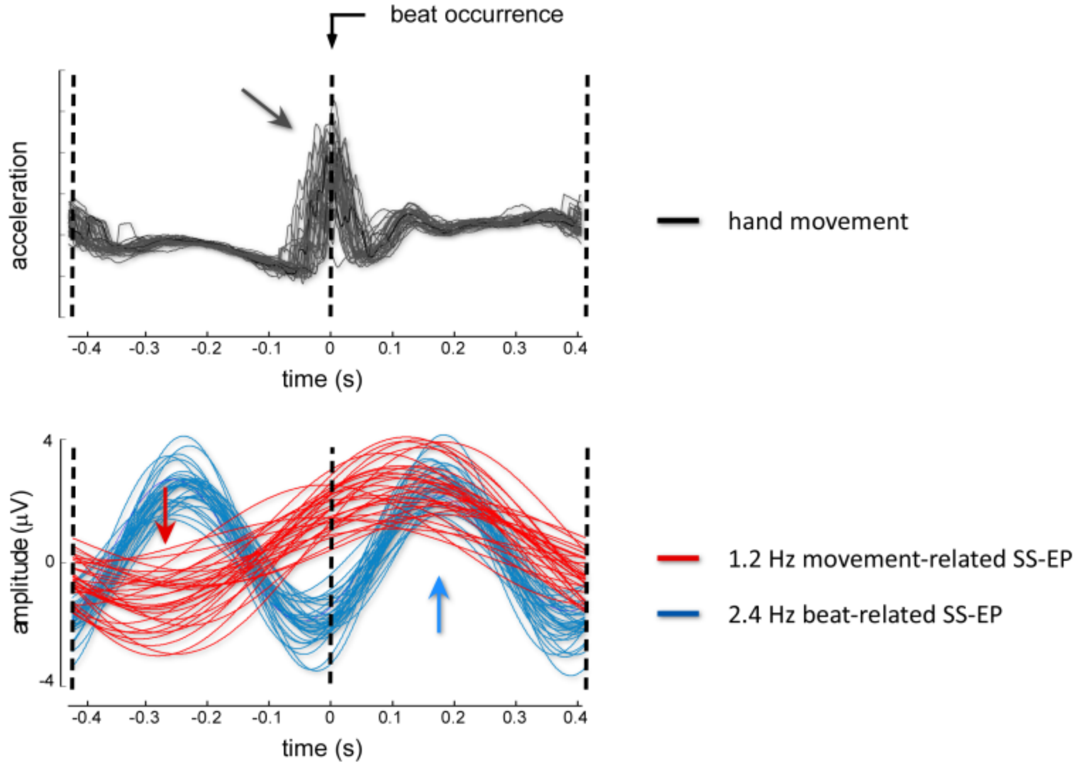
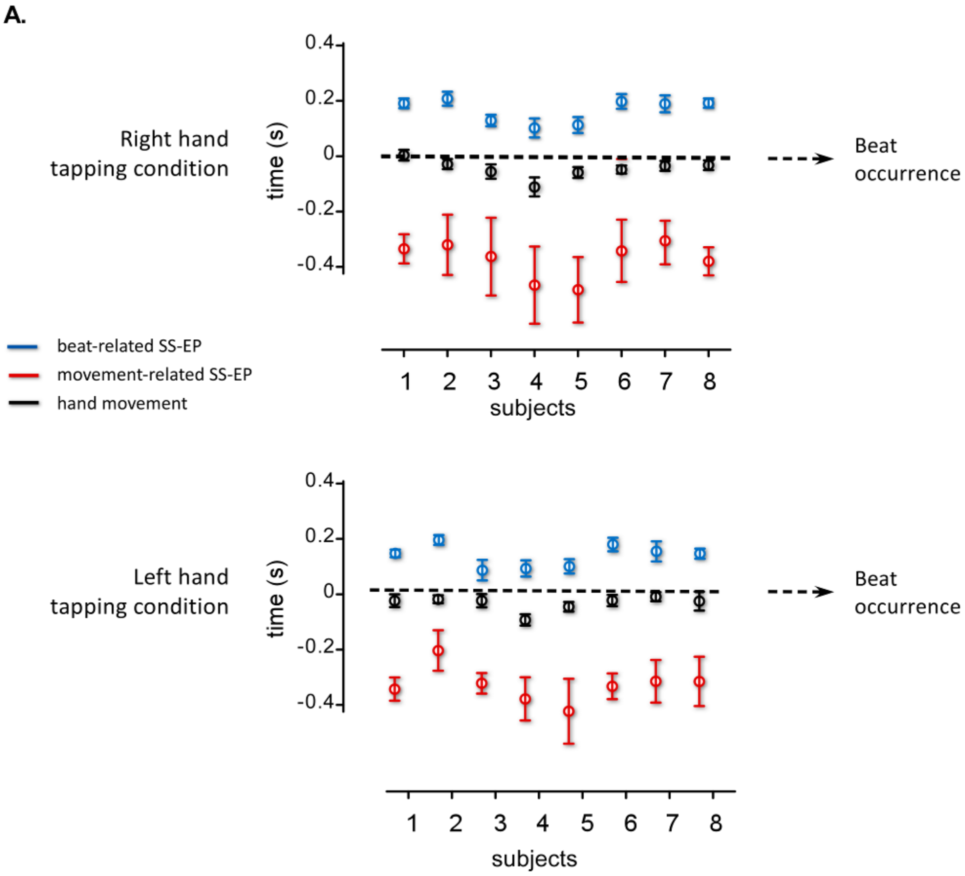


Figure 6. Temporal analysis of the hand movement (vertical acceleration), the beat-related SS-EP and the movement-related SS-EP signals in one representative participant (right hand tapping condition). The upper graph represents the superimposed hand movement signals obtained for each tapping-movement cycle. The lower graph represents the superimposed beat-related (blue) and movement-related (red) EEG signals obtained after narrow-band filtering around 2.4 and 1.2 Hz, respectively (see Methods for details). The dashed black lines represent the onset of the beat within these tapping-movement cycles. The trials are centered on the occurrence of the beats onto which the subjects synchronized their tapping. The colored arrows point to the relative latencies of the beat- and movement-related SS-EPs.

There was a significant correlation between the relative latencies of the beat-related SS-EP and the hand tapping movement ($r^2 = 0.53$, $p = 0.03$) (Fig. 7). There was also a significant correlation between the relative latencies of the movement-related SS-EP and the hand tapping movement ($r^2 = 0.48$, $p = 0.05$), and between the relative latencies of the movement- and beat-related SS-EPs ($r^2 = 0.66$, $p = 0.01$).


A.
Right hand tapping condition
time (s)
0.4
0.2
0
-0.2
-0.4
1 2 3 4 5 6 7 8
subjects
Beat occurrence
beat-related SS-EP
movement-related SS-EP
hand movement
Left hand tapping condition
time (s)
0.4
0.2
0
-0.2
-0.4
1 2 3 4 5 6 7 8
subjects
Beat occurrence

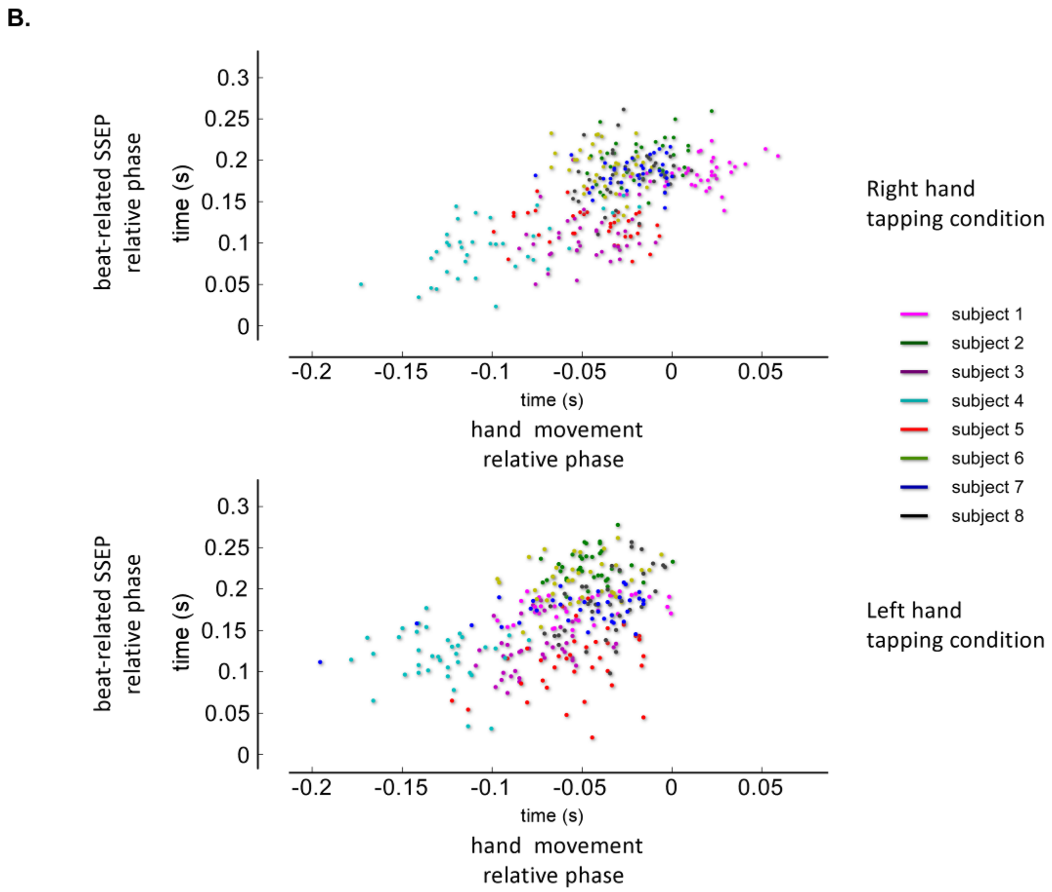


Figure 7. A. Relative latencies of the hand tapping movement (black), the beat-related SS-EP (blue) and the movement-related SS-EP (red) estimated for each individual subject (1-8). The upper graph shows the results obtained in the right hand tapping condition whereas the lower graph shows the results obtained in the left hand tapping condition. The whisker plots represent the mean and standard deviation of the latencies obtained in each hand tapping cycle. The horizontal dashed line represents the actual occurrence of the beat to which subjects synchronized their tapping. B. Scatter plot expressing the relative latencies of the beat-related SS-EP as a function of hand tapping movement in the left hand tapping condition (upper graph) and the right hand tapping condition (lower graph). Each dot represents a different hand tapping cycle. The different subjects are represented using different colors.

II.2.3.4.5. Discussion

The objective of the present study was to explore how distant brain areas supporting cooperative perception and action coordinate their dynamics during rhythmic sensorimotor synchronization. We found that moving to the beat is supported by distinct neuronal entrainments, i.e., periodic activities which can be captured in the human electroencephalogram (EEG) as steady-state evoked potentials (SS-EPs): a beat-related neuronal entrainment, locked to the frequency of the beat, and a movement-related neuronal entrainment, locked to the frequency of the movement (Fig. 2). Most importantly, we found possible evidence for a dynamic integration of these sensory- and motor-related periodic activities. First, moving to the beat led to the appearance of an additional SS-EP at a frequency corresponding to the sum of beat and movement frequencies, and thus indicating possibly the existence of a non-linear integration process, reflecting the activity of a population of neurons onto which sensory- and motor-related activities converge (Giani et al. 2012, Ding and Simon 2009, Wang et al. 2011). Second, we observed a tight phase coupling of the beat- and movement-related signals. Third, a selective enhancement of the beat-related SS-EP over the hemisphere contralateral to the moving hand was found, suggesting that the activity related to the rhythmic movement exerted a top-down effect on the processing of the auditory beat.

Beat- and movement-related SS-EP

Cortical responses were elicited in the present study by the long-lasting periodic repetition of the auditory stimulus, and were observed in the EEG spectrum in the form of SS-EPs at the exact frequency of the beat (Nozaradan et al. 2011; Regan 1989). Whether these activities result from the stimulus-driven entrainment of a network of neurons responding to the periodically-modulated feature of the stimulus, or whether they reflect the linear summation of successive transient ERPs elicited by the periodic repetition of the sensory stimulus remains an open question, and the two hypotheses may coexist (Galambos et al. 1981, Draganova et al. 2002).

In the control auditory condition, the topography of the beat-related SS-EP was maximal over frontal and temporal electrodes and symmetrically distributed over the two hemispheres (Figs. 2, 4, and 5). This topographical distribution was similar to that of mid- and late-latency auditory ERPs (Galambos et al. 1981, Draganova et al. 2002, Wang et al. 2012) as well as that of auditory SS-EPs elicited by higher stimulation frequencies (e.g. 20-40 Hz; Johnson et al. 1988). Hence, it seems likely that the bulk of the beat-related SS-EP, such as auditory ERPs and SS-EPs, mainly reflects activity originating bilaterally from auditory cortices.

In the hand tapping conditions, the rhythmic movements elicited repetitive cortical responses which could be observed in the EEG spectrum at a frequency corresponding to that of the rhythmic movement. Moreover, this activity exhibited at that frequency a scalp topography maximal over the central region contralateral to the moving hand (Figs. 2, 4 and 5). This could support the view that the production of periodic movements is related to periodic activities of neurons located within the hand representation of the contralateral primary motor and/or somatosensory cortex (Gerloff et al. 1997, 1998, Kopp et al. 2000, Daffertshofer et al. 2005, Kourtis et al. 2008, Bourguignon et al. 2011). However, as for the beat-related SS-EP, whether these activities result from the stimulus-driven entrainment of neurons, or whether they reflect the linear summation of successive transient ERPs elicited by the periodic repetition of the movement remains an open question.

Sensorimotor integration: evidence and alternative interpretations

(1) Cross-modulation SS-EP. In addition to the beat-related 2.4 Hz SS-EP and the movement-related 1.2 Hz SS-EP, moving to the beat elicited a third SS-EP appearing at 3.6 Hz. Such a response with the observed auditory topography cannot be explained by the simple linear summation of the 1.2 and 2.4 Hz oscillatory activities in the time domain, because this would not have resulted in an activity at 3.6 Hz in the frequency domain (Regan 1989, Giani et al. 2012). Hence, the 3.6 Hz activity could reflect a non-linear cross-modulation product of beat- and movement-related oscillations.

Indeed, when two distinct neuronal populations oscillate at two distinct frequencies, cross-modulation frequencies, corresponding to the sum or difference of the two main frequencies, may emerge if the signals conveyed by each of the two oscillating populations converge onto another population integrating these inputs (Zemon and Ratliff 1984, Regan 1989, Regan et al. 1995, Williams et al. 2004, Appelbaum et al. 2008, Sutoyo and Srinivasan 2009). Using scalp EEG recordings, non-linear cross-modulation SS-EPs have already been reported within visual (Zemon and Ratliff 1984, Regan 1989, Regan et al. 1995, Appelbaum et al. 2008, Sutoyo and Srinivasan 2009) and auditory (Wile and Balaban 2007, Purcell et al. 2007) sensory modalities; and the existence of neurons producing such responses has been confirmed using single-cell recordings (Williams et al. 2004). Hence, the finding that moving to the beat would not only be related to the presence of beat- and movement-related SS-EPs but would also lead to the emergence of a cross-modulation SS-EP appearing at 3.6 Hz, i.e. at the sum of the 2.4 Hz beat-related neuronal entrainment and the 1.2 Hz movement-related neuronal entrainment, could indicate the existence of a non-linear process of convergence of sensory- and motor-related periodic activities, possibly reflecting the activity of a population of neurons whose output corresponds to the product of the two input oscillations (Giani et al. 2012, Ding and Simon 2009, Wang et al. 2011).

Previous studies have already reported evidence for such sensorimotor interactions, in the form of a modulation of transient ERPs (Lutkenhoner et al. 2002, Praamstra et al. 2003), as well as modulation of oscillatory patterns of neuronal activity within the gamma (30 Hz and more) or beta frequency range (between 15 and 30 Hz) (see e.g. Roelfsema et al. 1997, or Donoghue et al. 1998). Importantly, our finding that moving to the beat does not only elicit beat- and movement-related SS-EPs at 1.2 and 2.4 Hz, but also elicits a non-linear cross-modulation product at the sum of these two frequencies, is not at all incompatible with these observations, nor with the hypothesis that synchronization between distant cortical areas is subtended by fast oscillatory activity that is not imposed by the periodicity of the stimulus. Indeed, the distinct form of sensorimotor interactions observed in the present study may occur specifically between oscillatory signals, thus allowing to identify these non-linear convergence products based on their predicted frequencies.

(2) Phase coupling of beat- and movement-related SS-EPs. On average, the hand tapping movement preceded the actual occurrence of the beat. This anticipation of the hand tapping movement is a well-described phenomenon, referred to as mean negative asynchrony (Repp 2005, Aschersleben 2002). Such as in other studies (Repp 2005, Aschersleben 2002), the negative asynchrony was reproducible within subjects and across the two hands, but varied greatly between subjects, from no anticipation to more than 100 ms (Fig. 7). The latency of the hand tapping movement was strongly correlated with the latency of the movement-related SS-EP, supporting the view that the production of rhythmic movements is related to neuronal activity time-locked to the produced movement and/or the somatosensory

feedback (Gerloff et al. 1997, 1998, Kopp et al. 2000, Daffertshofer et al. 2005, Kourtis et al. 2008, Bourguignon et al. 2011).

More surprising was the finding of a significant correlation between the latency of the hand tapping movement and the latency of the beat-related SS-EP (Fig. 7). Although a contamination of movement-related activities at 2.4 Hz cannot be entirely excluded, this observation suggests that the timing of the neuronal entrainment to the beat was not only dependent on the timing of the eliciting auditory beat, but was also modulated according to the timing of the produced synchronized movement. This observation suggests that the production of synchronized hand movements may modulate the neuronal representation of the auditory beat, possibly contributing to an accurate sensorimotor synchronization. Furthermore, it suggests that the amount of movement asynchrony is already encoded at an early sensory level, i.e. within the neuronal representation of the beat.

To summarize, our finding may indicate how, during rhythmic sensorimotor synchronization, coupling occurs between the distant brain areas, supporting beat- and movement-related neuronal entrainment.

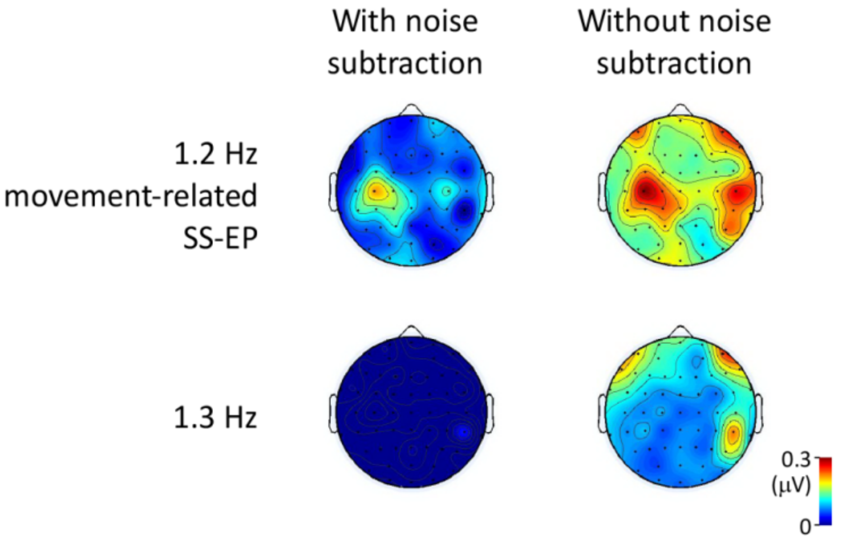
(3) Movement-induced enhancement of the beat-related SS-EP. Interestingly, the magnitude of the 2.4 Hz beat-related SS-EP was significantly enhanced in the hand tapping conditions as compared to the control auditory condition. Furthermore, this enhancement was much more pronounced over the hemisphere contralateral to the moving hand (Fig. 4). This observation suggests that movement-related activities exerted a top-down effect on the processing of the auditory beat, in particular, within the hemisphere involved in producing the hand tapping movement.

However, because of the low spatial resolution of EEG data, these findings have to be taken cautiously. Indeed, alternative interpretation of the complex signature of movement-related activities and its possible consequence when analyzed in the frequency domain (i.e. the generation of responses at harmonic frequencies which would not necessarily have the same scalp topography as the response obtained at 1.2 Hz) cannot be excluded. Future studies based on other methods to sample brain activity such as MEG, fMRI or the invasive recording of LFPs could address this crucial question more adequately.

Conclusion

Taken together, the results of the present study indicate that rhythmic sensorimotor synchronization could involve a dynamic coupling and interaction of sensory- and movement-related neuronal entrainments. More generally, our findings suggest that the recording of SS-EPs constitutes a promising approach to gain insight on the dynamics of neuronal activity underlying cooperative perception and action.

Acknowledgements: S.N. is supported by the Fund for Scientific Research of the French-speaking community of Belgium (F.R.S.-FNRS). A.M. has received the support from the Pain Research EFIC-Grünenthal Grant (EGG), as well as the IASP 2010 Early Career Research Grant. 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.



Supplementary Material. Scalp topographies of the signal amplitude at 1.2 Hz, i.e. expected to contain movement-related activity, as well as at 1.3 Hz, i.e. at neighboring frequency not expected to contain any SS-EP activity, with and without the noise subtraction procedure (group-level average of the right hand tapping condition). This illustrates that the magnitude of the background noise is not equally distributed across all scalp channels and that its contribution in the scalp topography of SS-EPs could be reduced using the noise subtraction procedure.

II.2.3.5. Perception leading to movement. As we have seen in the previous sections, a key concept that emerges from the results of the numerous studies on sensorimotor coupling is that perception of isochrony seems to affect motor output and in turn movements appear to influence the perception of isochrony. In particular, this concept is supported by the results of functional neuroimaging studies having shown that perception of musical rhythms involves integration and coupling across auditory and motor regions (Zatorre et al., 2007).

First, the activation of brain areas dedicated to movement when listening to beat-inducing rhythms suggests that both processes are not independent from each other (Chen et al., 2008; Grahn et al., 2009). For example, both the planum temporale and the premotor cortex have been shown to be recruited during passive listening to rhythms (Chen et al., 2008).

Second, it has been proposed that the coupling between auditory and motor areas when listening to rhythms may be mediated through the dorsal auditory pathway, i.e., a pathway connecting the posterior superior temporal gyrus (planum temporale) to prefrontal, premotor and motor cortices (Warren et al., 2005; Zatorre et al., 2007). The involvement of this pathway in auditory-motor interaction for rhythm processing is based on evidence of concomitant increased activity in these regions in the production of rhythmic sequences regardless of whether the rhythm is learned through auditory or visual modalities. Indeed, this observation suggests an auditory-motor representation of all rhythms learned for the purpose of production, regardless of the modality used to learn the rhythm (Karabanov et al., 2009). In the same line, Chen et al. (2006; 2008b) found that both the planum temporale and the premotor cortex were recruited when tapping to increasingly metrical rhythms, and were functionally correlated when tapping to increasingly complex rhythms.

Third, auditory-motor coupling may also be influenced by musical training. Chen et al. (2008) found that although left hemisphere auditory-motor coupling, assessed on the basis of concomitant increased activity in auditory and motor cortices, was present in musicians and nonmusicians when listening to music, only musicians showed significant coupling in the right hemisphere. In the same line, Grahn and Rowe (2009) found greater coupling in both hemispheres in musicians compared to nonmusician individuals in similar experimental conditions.

Taken together, these results suggest that passive listening to rhythms recruits a movement-related network of brain areas, which may explain why rhythmic auditory stimuli are likely to induce automatic movements as in groove-based music.

Investigating behaviors in contact to groove-based music could also be informative, in complement to neuroimaging studies, to understand how beat perception leads to movements (Janata et al., 2011) (see also Sections II.1.3.1 and II.2.2.4.2). Few studies have examined systematically the behaviors related to groove feeling in music (Witek et al., 2012; Janata et al., 2011). In one of these studies, Janata et al. (2011) let participants move their body freely on music and recorded their movements. The authors showed that isochrony was the major characteristic of the produced movement. Moreover, another movement pattern reoccurring across participants was head bobbing, which increased as a function of the subjective groove feeling rated by the participants. This head bobbing emerged as a spontaneous behavior, perhaps to reinforce entrainment to the beat.

The importance of head movement for beat perception, probably through the role of induced vestibular inputs, had already been highlighted in previous studies before Janata et al. (2011) study. Indeed, periodic vestibular inputs elicited by passive movement of the head have been shown to underlie the sense of meter in adults (Phillips-Silver and Trainor, 2007).

Also, when artificially induced using direct galvanic stimulations of the vestibular system, vestibular inputs similarly influences rhythm perception (Trainor et al., 2009). Vestibular stimulation also occurs when listening to loud (>90 dB) dance music and may thereby play a role in the entrainment to move spontaneously on music in such context (Todd and Cody, 2000).

Studying the neural correlates of groove feeling is still at its beginning. In fact, for most of the neuroimaging techniques, a fundamental issue remains the spontaneous movement. Indeed, as the methods used to sample brain activity usually require to refrain from moving (specifically the head), it is possible that the inhibition of movement may affect drastically these physiological responses.

II.2.3.6. Movement influencing perception. In parallel with ample evidence showing that hearing music can make individuals move, there is increasing evidence that body movements can shape the auditory perception of rhythms. This notion had already been expressed by Jacques-Dalcroze (1920) in his music teaching theory. According to his approach, rhythm is embodied and expressed through movements. Moreover, his theory aims at teaching that musical knowledge is the result of physical experience. Recently, Phillips-Silver and Trainor (2005) demonstrated that this cross-modal process is intrinsic to the human musical experience. They trained 7-month-old infants with a 2-minute repetition of a metrically ambiguous rhythm which can be perceived either in duple or triple meter (Fig. II.2.3.6). The infant subjects were bounced up and down while held in the arms of the experimenter, either on every second beat to induce perception of the rhythm in duple meter, or on every third beat, to induce the perception of the rhythm in triple meter. After training, the infants' preferences for an accented version of the auditory pattern in duple versus triple meter was tested using looking time procedure. Infants who were bounced in duple meter listened significantly longer to the duple version, whereas those who were bounced in triple meters listened significantly longer to the triple version of the rhythm. In complementary experiments, Phillips-Silver and Trainor (2007, 2008) showed that visual cues were not crucial to this effect. They replicated their experiment on normal adults by bouncing them using a rocking chair, and confirmed that neither visual cues nor active body movements were necessary to induce this effect, which appeared to depend only on head movement, even passive.

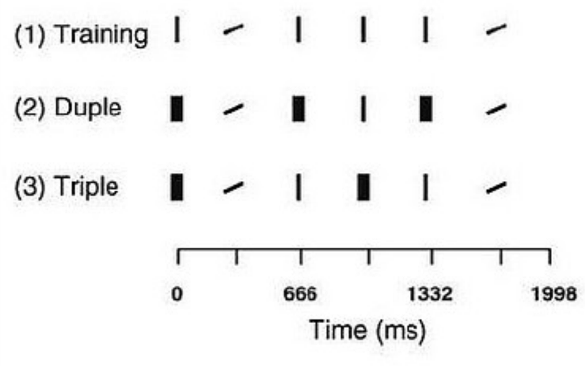


Figure II.2.3.6. From Phillips-Silver and Trainor (2005). Metrically ambiguous stimuli used in this experiment, and metrical priming induced by bouncing movement.

Another argument in favor of the view that musical beat induction derives from movement is that the tempo range of optimal musical pulse perception, around 300-900 ms according to most studies, is similar to that of several automatic body movement such as locomotion (Fraisse, 1982). Indeed, it is reasonable to hypothesize that biomechanical constraints and other bias shaping repetitive movements may shape perception (see also Section II.2.3.2.4 for a more detailed discussion on this view). Taken together, all these experiments support the view that music has to be studied as a sensory-motor and multisensory experience.

II.2.3.7. Auditory prominence for sensorimotor coupling. Although we have seen that entrainment to the musical beat is a multisensory experience, it is predominantly induced, even in its simplest form, by auditory inputs as compared to inputs conveyed through other sensory modalities. Indeed, one rarely moves in synchrony with rhythms defined by purely visual stimuli, certainly not spontaneously (Repp and Penel, 2004). For example, when observing people dancing on a muted television, individuals are not likely to tap their foot or bounce the head spontaneously. Sound, rather than light, is preferred as a medium for rhythmic, repeated, stimulation. However, the modality-specificity of beat perception is maybe not so trivial to explain, when considering that (1) time itself is assumed to be "amodal" and conveyed by multiple sensory modalities, (2) the neural substrates of auditory and visual timing tasks overlap substantially (Schubotz et al., 2000), and (3) any advantages that beat perception and the bias for periodicity may offer to auditory timing and sensorimotor synchronization accuracy should be equally beneficial to the visual modality for instance (Grahn, 2012).

Several studies have attempted to characterize and understand the prominence of the auditory system in conveying perception of isochrony and accurate movement synchronization. Psychophysical experiments have demonstrated that discrimination of short time intervals (50 to 2000 ms) is poorer in vision than in audition (Grondin et al., 1998). Moreover, discrimination and reproduction of temporal patterns are superior in the auditory modality (Repp and Penel, 2004). Also, when auditory and visual stimuli are in conflict within a timing task, judgments are typically more strongly influenced by the auditory than the visual temporal information (Shams et al., 2000).

This prominence of the auditory modality to extract timing information extends to the temporal control of action and sensorimotor synchronization. For example, there is evidence that the degree to which temporal information in vision is able to drive action is weaker than in audition (Patel et al., 2005). When participants are asked to tap in synchrony with isochronous auditory or visual sequences (typically tones *versus* flashes) at a given tempo, it has been shown that the variability of the asynchronies between taps and input sequences is much greater with visual input as compared to auditory input (Repp and Penel, 2002). Similarly, Repp and Penel (2002) found that the automatic phase correction process triggered by phase perturbations in simultaneous auditory and visual sequences is much stronger for auditory perturbations as compared to visual perturbations. Patel et al. (2005) also found that synchronization performance to visual rhythmic patterns is much less efficient than synchronization performance to auditory rhythms. In addition, using rhythmic patterns that varied from strongly metrical to weakly metrical, Patel and colleagues (2005) found that accuracy in the visual modality did not differ between strong and weak beat-based structure, in contrast with the synchronization performance to given rhythms conveyed by auditory inputs.

It has been proposed that visual inputs may not be as efficient to induce beat perception because visual stimuli may be unable to elicit perception of metrical accents (see Section II.2.2.4.1). For example, while Western European individuals tend to listen to the second sound of a pair of sounds as accented (Povel and Essens, 1985) (see Fig. II.1.2.2), there is no evidence that repeated pairs of light flashes convey a sense of perceptual accent on the second flash. Whereas spatial grouping processes have been extensively studied in the visual system through the Gestalt perspective (Wagemans et al., 2012b), the emergence of temporal grouping is unclear.

Functional neuroimaging studies have also brought insight on the specificity of the auditory system to convey beat perception and accurate sensorimotor synchronization. Although synchronization to auditory and visual stimuli elicit different patterns of brain activity (Jäncke et al., 2000), the perceptual monitoring of visual and auditory rhythms activates a relatively similar network, including motor regions (Shubotz et al., 2000), leading researchers to the conclusion that time processing is “amodal” by nature. However, Grahn and colleagues (2011) showed that the putamen, a part of the basal ganglia that responds to beat induction in the auditory modality (see Sections II.2.2.5.1 and II.2.3.3), is not activated when participants are presented with a given rhythm conveyed through the visual modality. Interestingly, a sense of beat could be induced by the visual rhythms, correlated with increased activity in the putamen, only when visual rhythms were preceded by auditory versions of the same rhythm but not *vice versa* (Grahn et al., 2011). This latter evidence suggests that the feeling of beat may be induced, under certain circumstances, by sensory modalities other than the auditory modality.

Interestingly, it has been proposed recently that the lack of evidence suggesting beat induction in the visual modality could result from the use of non optimal visual stimuli. Indeed, musicians synchronize their playing to the visually perceived movements of a conductor, a visual cue that differs greatly from the simple visual flashes used in most studies. To explore this possibility, Hove et al. (2010) compared tap synchronization performance at 1.6-3.3 Hz rates of presentation of flashes or different types of moving visual stimuli. Flashes elicited the worst synchronization performance of all tested stimuli, whereas the best performance was elicited by visual stimuli that moved in a way compatible with the tapping motion of the finger (e.g., synchronizing the bottom of the tap trajectory with the bottom of the visual stimulus trajectory). Thus, visual motion appears to improve tapping performance overall, although congruency of action motion and stimulus motion is crucial as well. Hove et al. (2011) concluded from their findings that visual spatiotemporal rhythm can entrain human movement nearly as well as an auditory metronome. This important finding opens an avenue of research to study the multisensory temporal binding related to entrainment to the beat in ecological contexts, particularly using visual representation of body movements. Although the influence of visually presented body movements compared to non biological motion has been demonstrated in entrainment to the beat in children (Kirschner and Tomasello, 2010), this research is still as its beginning.

II.2.3.8. Study 4: STEADY-STATE EVOKED POTENTIALS AS AN INDEX OF MULTISENSORY TEMPORAL BINDING (Sylvie Nozaradan, Isabelle Peretz, André Mouraux)

This article has been published in Neuroimage in 2012.

Here, we made a more "classic" use of the SS-EP approach. The stimuli were periodically modulated for experimental purpose, to "tag" the corresponding neural responses based on their frequencies. As suggested by the title of Study 4, we aimed to test whether this "classic" use of the frequency tagging method was suitable to study the dynamic multisensory interaction hypothesized to occur when a beat is conveyed through two distinct sensory modalities. To this aim, aside of periodically modulating the luminance of the visual stimuli in order to elicit a visual SS-EP, we also made the flickering visual object moving periodically, as visual movements convey more robust beat induction than static visual objects (Hove et al., 2011). However, in the present study, the visual periodic movements aimed at inducing the beat were non biological. Nevertheless, this study should be taken as a first step to further investigate beat induction through biological motion using the steady-state evoked potential approach.

II.2.3.8.1. 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.

II.2.3.8.2. Introduction

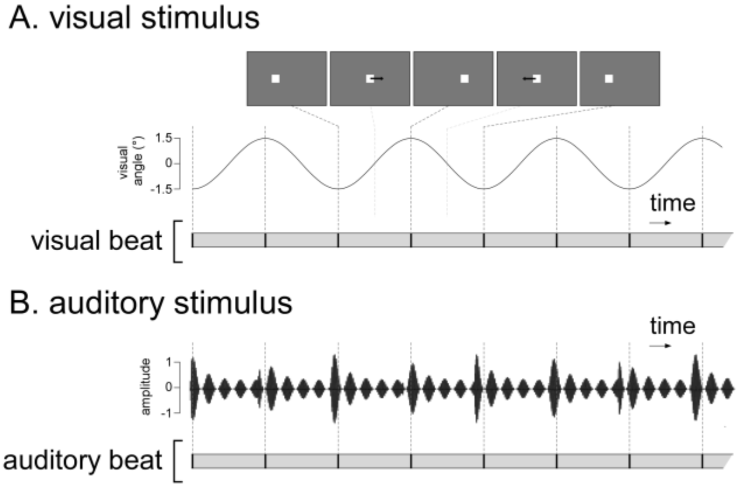
Building coherent representations of the external world requires integrating and merging information concurrently sampled through our different senses (Gibson, 1966; Spence and Driver, 2004). Most events occurring in the environment concomitantly activate afferents from different sensory modalities. 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; Stein and Meredith, 1993; Elliott et al., 2010).

Temporal congruency facilitates cross-modal integration (Vroomen and Keetels, 2010; Welch and Warren, 1980; Sekuler et al., 1997; Bertelson, 1999; Zampini et al., 2003; Fujisaki and Nishida, 2005; Vatakis and Spence, 2006; Petrini et al., 2009). 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; Senkowski et al., 2008; Schroeder 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 (Regan, 1989; Morgan et al., 1996; 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 focus of attention (Saupe et al., 2009; de Jong et al., 2010; Toffanin et al., 2009; Talsma et al., 2006).

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 (London, 2004; Large, 2008; van Noorden and Moelants, 1999). Although beats are preferentially conveyed by auditory input (Patel et al., 2005; Repp, 2005; Glenberg et al., 1989; Grahn et al., 2011), 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 and Jones, 1999; Large and Kolen, 1994; Large, 2008; van Noorden and Moelants, 1999; Nozaradan et al., 2011). 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-binded, 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.



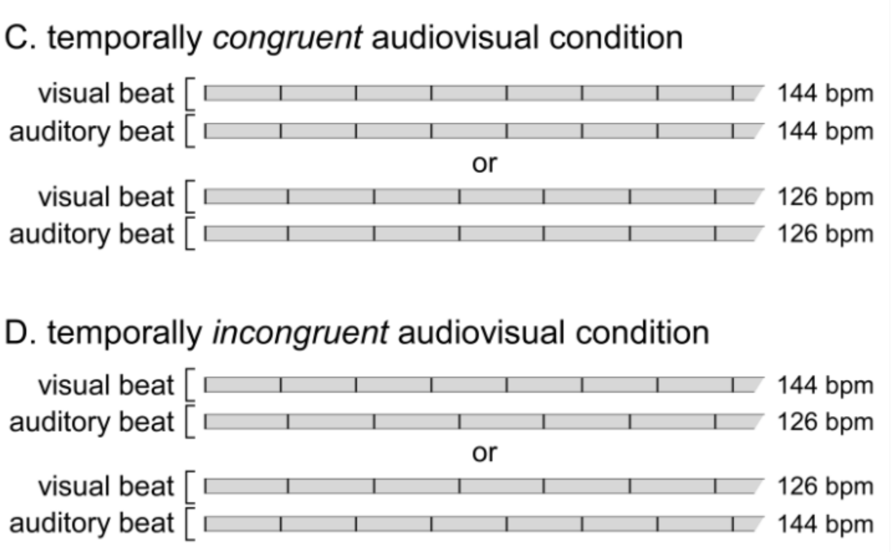


Figure 1. Audiovisual stimulation. A: Visual stimulus. A white square presented on a grey background moved periodically from left to right and from right to left. This sinusoidal periodic movement was used to induce a visual beat with a frequency of 2.1 or 2.4 Hz (126 or 144 beats per minute, bpm). B: 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.

II.2.3.8.3. 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 (Regan, 1989; Galambos et al., 1981).

Visual stimuli. The visual stimuli were presented using a CRT computer monitor (resolution: 1024 x 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 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 plesiochronous (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 Figure 1).

EEG recording

Subjects were comfortably seated in a chair with their head resting on a support. They were instructed to relax, 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 kΩ. 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 analyses 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 (Mouraux and Iannetti, 2008; Regan, 1989; Lachaux et al., 2000; Pfurtscheller and Lopes da Silva, 1999) (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 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 artefacts 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 9.968-10.031 Hz; 11 Hz auditory SS-EP: bins ranging from 10.968-11.031 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://sccn.ucsd.edu).

Statistical analyses

For each participant and analysis procedure (across-trial averaging in the time domain, 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-measured ANOVA was performed to compare the SS-EPs amplitudes 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-measured 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. 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 $p < 0.05$.

II.2.3.8.4. Results

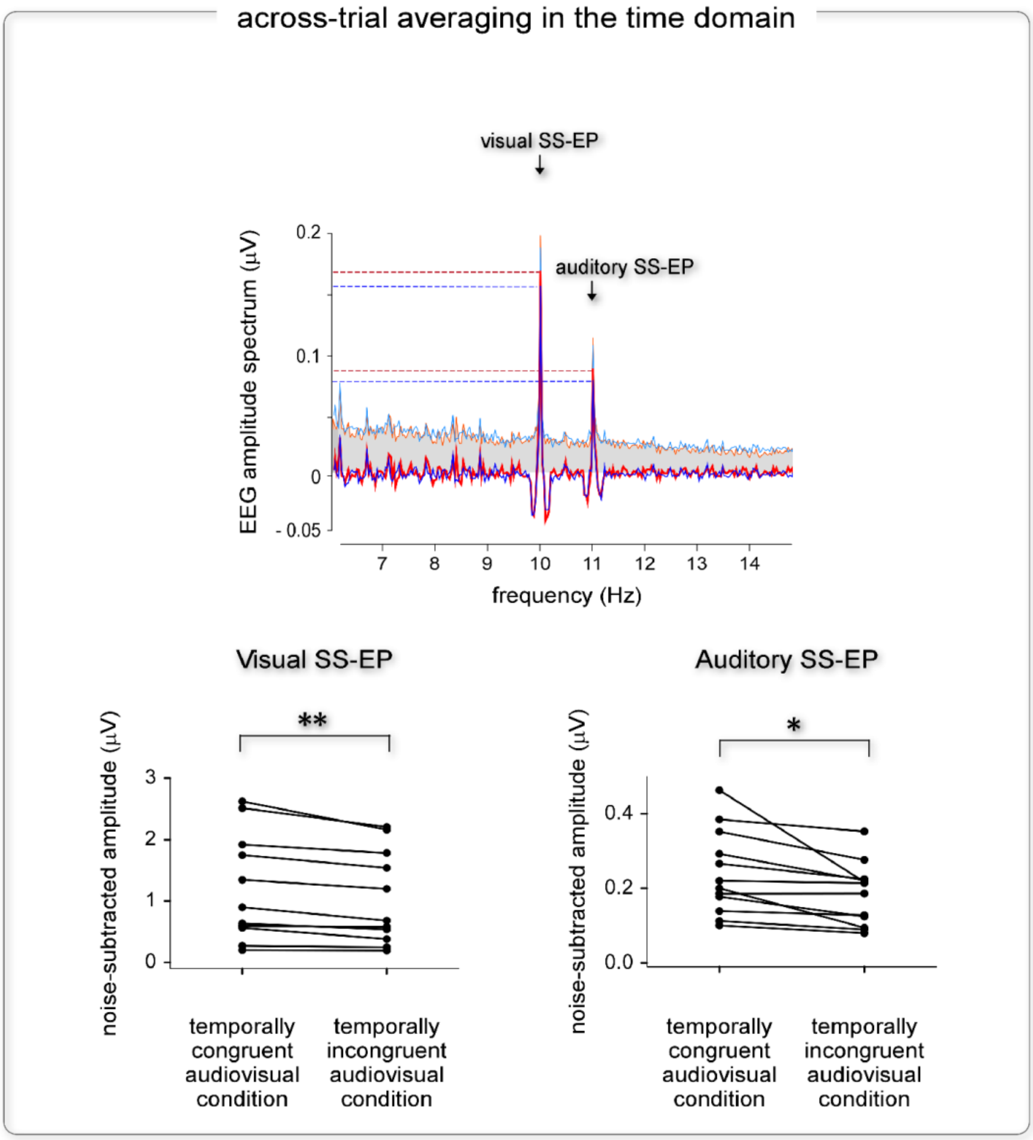
Behavioral results

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 ± 0.25 µV in the temporally congruent condition and 1.01 ± 0.21 µV in the temporally incongruent condition (mean ± 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 ± 0.26 µV in the temporally congruent condition and 1.10 ± 0.25 µV 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 ± 0.03 µV in the temporally congruent condition and 0.18 ± 0.05 µV 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 ± 0.03 µV in the temporally congruent condition and 0.16 ± 0.02 µV 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).


across-trial averaging in the time domain
visual SS-EP
auditory SS-EP
EEG amplitude spectrum (µV)
0.2
0.1
0
- 0.05
7
8
9
10
11
12
13
14
frequency (Hz)
Visual SS-EP
Auditory SS-EP
**
*
noise-subtracted amplitude (µV)
3
2
1
0
0.4
0.2
0.0
temporally congruent audiovisual condition
temporally incongruent audiovisual condition

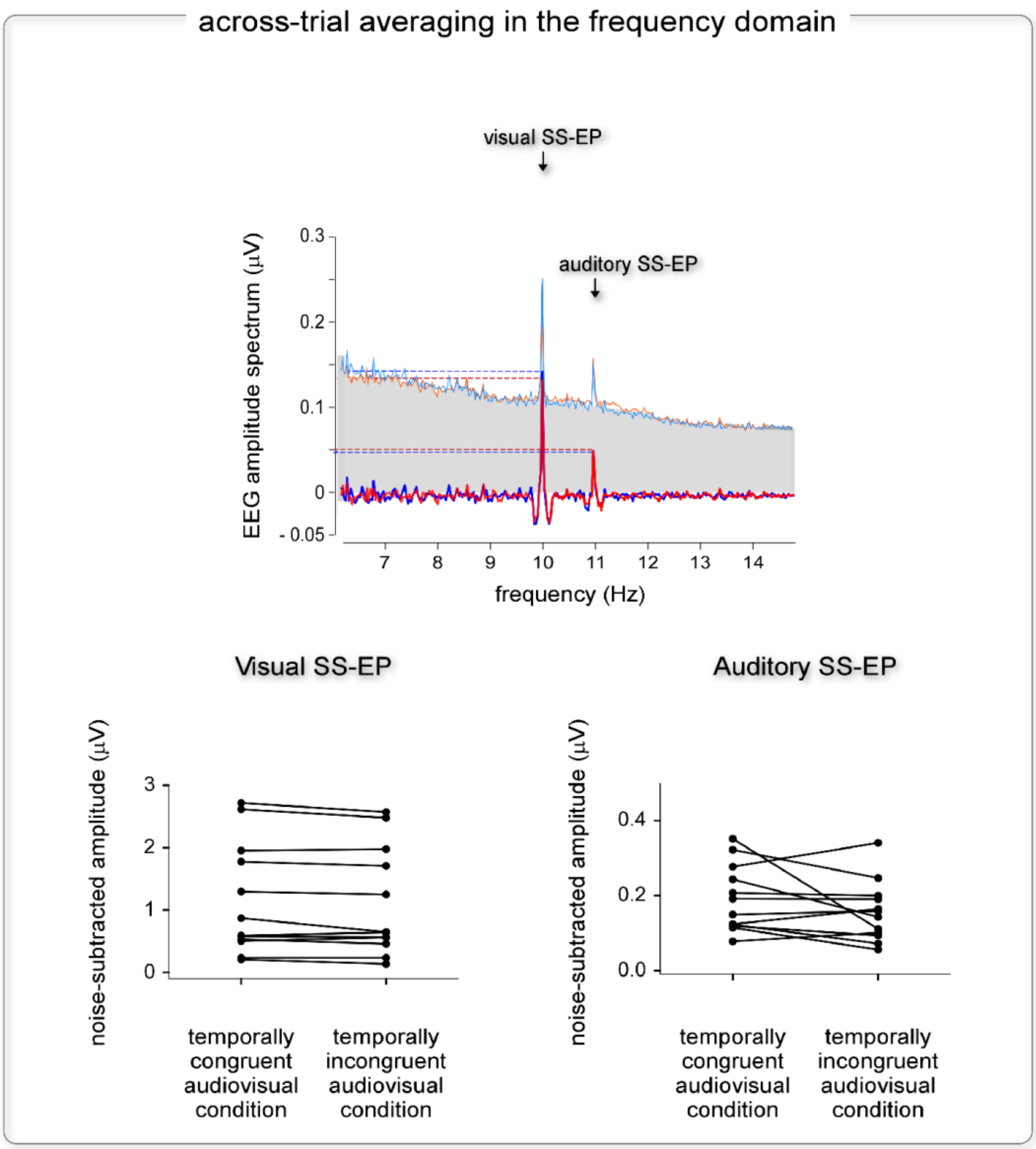


Figure 2. The upper parts of the panels represent the auditory and visual SS-EPs elicited by temporally congruent (red) and incongruent (blue) audiovisual stimuli. The up panel display the frequency spectra obtained after averaging trials in the time domain. The bottom panel display the frequency spectra 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.

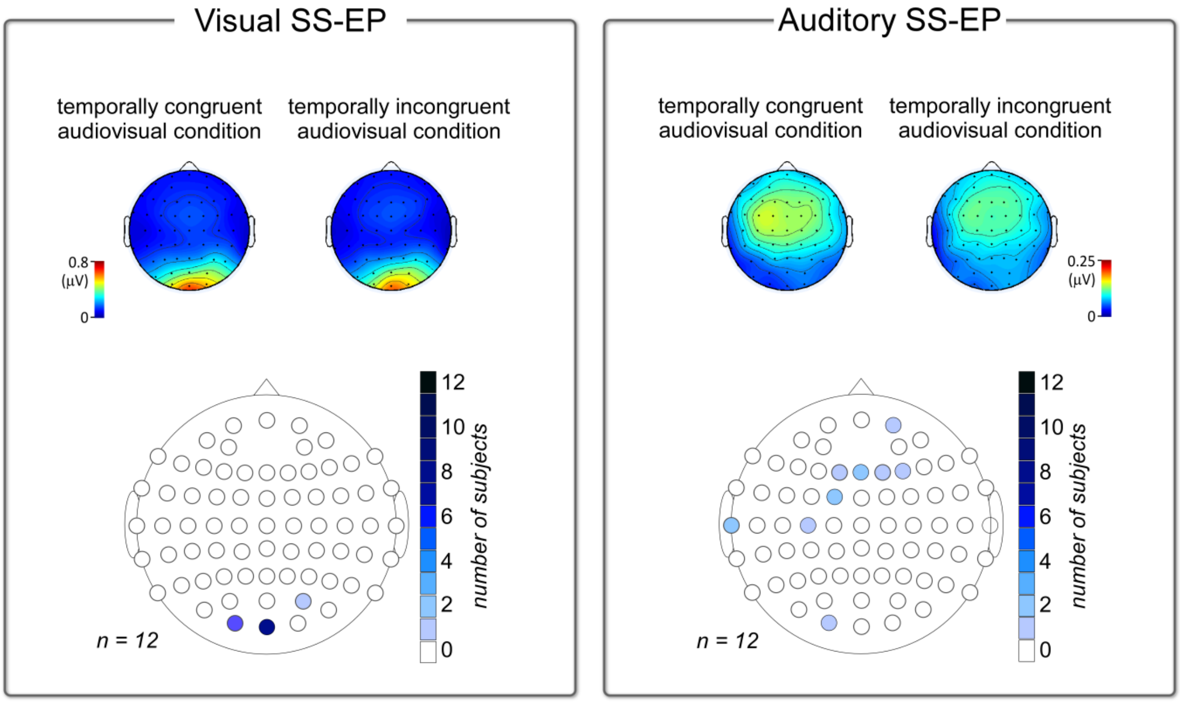


Figure 3. Topographical distribution of visual and auditory SS-EPs. The upper parts represent the group-level average topographical maps of the magnitude of the SS-EPs elicited by 10-Hz visual stimulation and 11-Hz auditory stimulation, obtained after averaging trials in the time domain. The bottom parts 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.

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, η^2 = 0.64, p = 0.001). There was a significant interaction between the factors 'temporal congruency' and 'modality' (F = 5.47, η^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, η^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, η^2 = 0.38, p = 0.025). There was no significant interaction between the factors 'temporal congruency' and 'modality' (F = 0.3, η^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 Figure 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 ($F = 22.21$, $\eta2 = 0.61$, $p = 0.001$). 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 ($F = 18.21$, $\eta2 = 0.62$, $p = 0.001$), but no interaction between the two factors ($F = 2.22$, $\eta2 = 0.17$, $p = 0.164$).

II.2.3.8.5. 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).

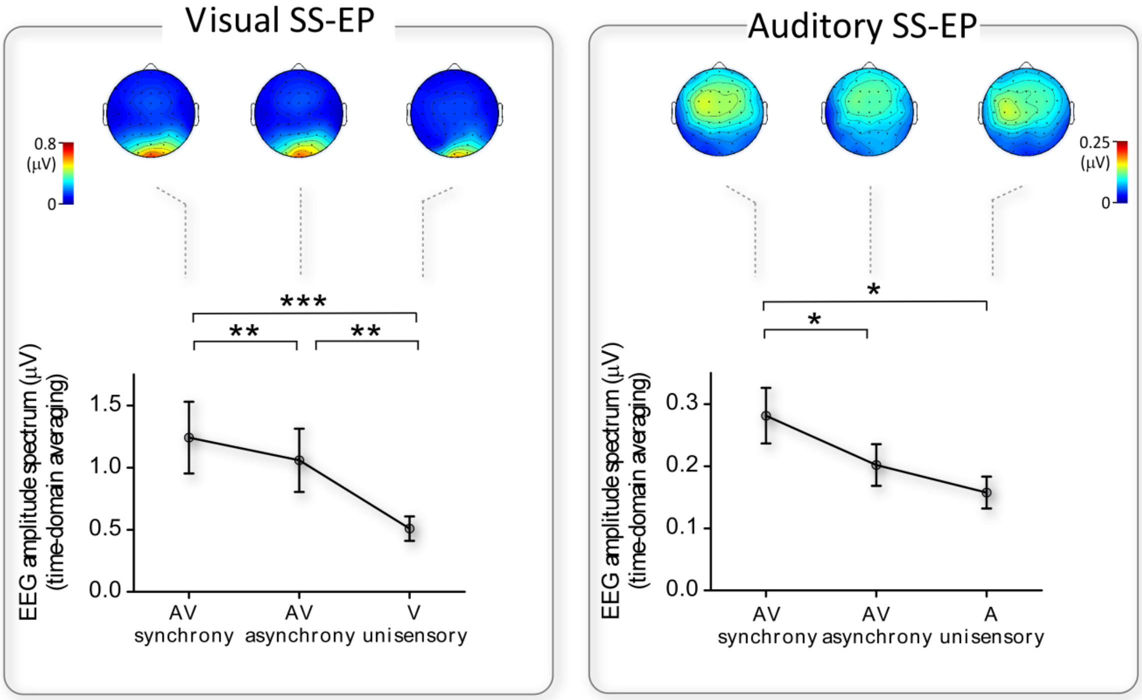
The finding that temporal congruency between the auditory and visual beats enhanced the phase coherence of the SS-EPs elicited by each of the two streams of sensory input indicates that, from trial to trial, there was less variability in the timing of the elicited SS-EPs when both sensory systems were 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., 2010; Kayser et al., 2008; Luo et al., 2010; Schall et al., 2009; Senkowski et al., 2008; Talsma et al., 2010; Schroeder et al., 2008). 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 (Senkowski et al., 2005; Kim et al., 2007; Zhang et al., 2010; Iversen et al., 2008; Lakatos et al., 2009; Kayser, 2009). 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., 2011). As previous studies have shown that the responsiveness to external stimuli of oscillating neurons is dependent on the phase of these oscillations (Fries, 2005; Bush et al., 2009), 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 remains 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; Krauzlis 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.

Acknowledgements : S.N. is supported be 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 this manuscript.



*Supplementary Figure. Magnitude of visual and auditory SS-EPs elicited by isolated auditory or visual stimulation (unisensory condition) and by combined auditory and visual stimulation (temporally-congruent and incongruent multisensory conditions). The magnitude of the visual and auditory SS-EPs obtained in the unisensory condition were significantly smaller than the magnitude of the visual and auditory SS-EPs obtained in the multisensory conditions, both when averaging the trials in the time domain, and when averaging trials in the frequency domain (paired-sample t-tests performed using the noise-subtracted amplitude measured at the electrode displaying maximal SS-EP amplitude in the spectrum averaged across the three conditions: unisensory, congruent multisensory, incongruent multisensory; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$). Note that the scalp topographies of the auditory and visual SS-EPs elicited in the unisensory and multisensory conditions were, however, very similar.*

III. DISCUSSION AND PERSPECTIVES

One of the major goals of this dissertation was to narrow the gap between scientific studies on neural entrainment on the one hand and entrainment to musical rhythms on the other hand. In both, entrainment processes and biases towards periodicity have been described as fundamental functional characteristics. Considering this, we tested whether periodicities induced by musical rhythms could entrain neural activities at frequencies corresponding to these periodicities. These neural activities were captured with the EEG in the form of steady-state evoked potentials (SS-EPs) identified by analyzing the EEG in the frequency domain.

This approach of using the SS-EP method to tag periodic percepts abstracted from sensory inputs is still in its infancy. Hence, new observations could be made in each of the four empirical studies. However, they certainly also present a number of shortcomings. These issues will be discussed in this last section.

III.1. Study 1: TAGGING THE NEURONAL ENTRAINMENT TO BEAT AND METER

In Study 1, we asked participants to listen to a periodic sound, and to voluntarily imagine the meter of this beat as either binary or ternary (i.e., as in a march or a waltz respectively). We observed that listening to a periodic sound and imagining a meter on this sound elicited neural activities at frequencies corresponding to the perceived and imagined beat and meter, that we called *beat*- and *meter-related steady-state evoked potentials* (SS-EPs).

Study 1 shows how our approach could provide advantages that are characteristic to the SS-EP method (Regan, 1989): (1) a more objective identification of the potentials evoked by the beat and meter based on the experimentally expected periodicities of these percepts; (2) a more straightforward measure of these potentials; and finally (3) an extremely high signal-to-noise ratio given the concentration of the response of interest within a narrow frequency band. Beyond these advantages provided by the SS-EP approach, the results of Study 1 could be interpreted as a direct evidence of internally-driven SS-EPs.

III.1.1. Making a bridge between beat and meter-related SS-EPs, transient ERPs and ongoing oscillatory activities. It could be hypothesized that the neural oscillatory activities observed in Study 1 are responsible for the emergence of our psychological sense of beat and meter. According to this interpretation, beat and meter-related SS-EPs would constitute the neural support of the entrainment models proposed to explain beat perception (Large, 2008) (see *Entrainment models,* Section II.2.2.3.3). However, evidence of neural synchronization to the beat and meter have previously been observed only in higher frequency bands (beta and gamma) (Fujioka et al., 2009; Iversen et al., 2009; Snyder and Large, 2005), thus contrasting with our finding of a robust synchronization of neural populations at frequencies corresponding *directly* to the beat and meter.

Such observations could be reconciled by considering that the amplitude dynamics of high frequency neural oscillations (beta and gamma) may be driven by slow frequency oscillatory activities in the delta band (1-4 Hz), these latter supporting the entrainment to the perceived beat and meter (Large, 2008). This view is in line with the concept of low frequency neuronal oscillations acting as an instrument of sensory selection (Schroeder and Lakatos, 2008; Lakatos et al., 2008). Several electrophysiological studies in animals including humans have corroborated this view (Varela et al., 1981; Haig and Gordon, 1998; Makeig, 2002; Buzsaki and Draguhn, 2004; Lakatos et al., 2008; Schroeder and Lakatos, 2008; Sirota et al., 2008). Collectively, these studies have shown that when the activity of a neuronal population synchronizes at a given low frequency, the phase of the induced oscillations can elicit a cyclic fluctuation of the excitability of the responding neuronal population, leading to an amplitude modulation of the ERPs or higher frequency ongoing oscillatory activities that can be generated by these populations, but also a modulation of behavioral performance (Busch et al., 2009; Henry and Obleser, 2012). This concept has been further extended from sensory to motor processing (Praamstra et al., 2006).

How can we make a bridge between evidence of responses to beat and meter in the form of modulation of the magnitude of high frequency ongoing activities and modulation of transient ERPs amplitude on the one side, and beat and meter-related SS-EPs on the other side? We propose that the beat-induced periodic EEG response identified in Study 1 may constitute a direct correlate of the actual mechanism through which attentional and perceptual processes are dynamically modulated as a function of time (see *Entrainment models* and *Dynamic attending models*, Sections II.2.2.3.3 and 4). That is, the responsiveness of the neuronal population that is entrained to the beat may be expected to vary according to the phase of the beat-induced cycle. If the beat-induced cycle, as observed in the form of beat and meter-related SS-EPs, reflects cyclic modulation of excitability in neural population, this would account for the previous observations that event-related potentials elicited at different time points relative to the beat or meter cycle exhibit differences in amplitude (Brochard et al., 2003; Snyder and Large, 2005; Pablos Martin et al., 2007; Grube and Griffiths, 2009; Iversen et al., 2009; Fujioka et al., 2010; Schaefer et al., 2011). In future research, it would be worth testing this hypothesis directly using rhythmic patterns that are known to induce a beat percept, and observing the transient responses to short gaps distributed uniformly with respect to the phase angle of the induced beat cycle (i.e., as in Henry and Obleser's study of 2012, with the difference that these authors used periodic frequency modulation inputs to entrain low frequency neural activities). In the same vein, a direct phase relation between beat-related neural entrainment and higher frequency ongoing neural activity remains to be tested.

III.1.2. Some remarks on the frequency domain analysis of SS-EPs. There is no compelling evidence that the transformation of a waveform into sinusoidal Fourier components is physiologically meaningful in itself. Indeed, although different frequency components may sometimes have quite different physiological properties resulting from the projection of different transient components of the response to different frequency components, this is not systematically the case in all SS-EP paradigms (Regan, 1972)[8]. That is, whether the different SS-EPs identified in the EEG spectrum following Fourier transform come from a single multi-phasic non sinusoidal neural response originating from a single source, or rather result from separate neural sources, remains an open question (see also Section III.2.1 for a development of the discussion on this point). In Study 1 for instance, it is impossible to disentangle whether the beat and meter-related SS-EPs originated from a unique or distinct neural sources (although the different topographies obtained for these activities suggest that their sources were distinct). Nevertheless, measuring the EEG signals in the frequency domain could be justified because the response of interest could be defined and measured more objectively by the prediction of the exact frequency of the expected neural responses in the EEG spectrum.

III.1.3. Musicians versus nonmusicians. In Study 1, the data obtained from musicians and nonmusicians participants were merged, because there were no significant differences between the two groups. Admittedly, we expected to find increased amplitude of the beat

---

[8] The situation is similar for the transient ERPs measured in the time domain: whether or not the different components whose latency and amplitude are measured in transient ERPs are actual distinct responses possibly originating from distinct neural sources has always been heavily debated (Luck, 2005).

and meter-related SS-EPs in musicians compared to nonmusicians, and one may wonder why there was no such difference in the results. One interpretation of this absence of difference could come from the opposite effects of two processes. On the one hand, musicians are certainly able to apply better than nonmusicians a mental imagery of meter with consistency and minimum temporal jitter along the trials. This would result in increased amplitude of the corresponding SS-EPs when observed in the frequency domain. On the other hand, the ease of musicians to perform such mental imagery could lead to a reduced active involvement and reduced attention to the task compared to nonmusicians. This would thus reduce the amplitude of the corresponding SS-EPs, given that SS-EPs at low frequencies have been shown to be influenced by attention (Ding et al., 2006). Taken together, these two processes could cancel out each other's effect.

III.1.4. Beat and meter-related SS-EPs and head movement artifacts. We must remain cautious regarding the possible contamination of the beat and meter-related SS-EPs by artifacts originating from periodic head movement produced at similar frequencies. Indeed, periodic head movements could be responsible for artifacts due to displacement of the electrode cap, thus mimicking the time course of the periodic beat and meter. Head movement related to beat perception constitutes a well-known limitation in the rhythm literature (see Section II.2.3.5).

However, there were several aspects of Study 1 suggesting that head movements did not contaminate, at least not significantly, the observed SS-EPs. First, participants' head rested on a chin rest, thus reducing the possibility of head bouncing as observed in ecological context with vertical position of the head. Second, participants were asked to fixate a cross displayed on a computer screen in front of them, to ensure minimum eye movements during the recordings. In addition, oculomotor signals were concomitantly recorded. Thus, head movements during visual fixation would have resulted in periodic modulation of the oculomotor signal, preferentially in the vertical axis. However, there was no significant periodicity emerging at frequencies corresponding to the beat and meter (or at (sub)harmonics) in the Fourier transform of the oculomotor signal, thus suggesting that the head remained fixed during the experiment. Third, periodic head movement such as head bouncing along the anterior-posterior axis should have resulted in a periodic artifact possibly overlapping the SS-EPs at corresponding frequencies. Yet, the scalp topography of this artifact would have likely been similar across participants, revealing activity over the most anterior and posterior sites. In contrast, the topographies of the beat and meter-related SS-EPs differed across participants, and were not found specifically at these sites.

Given these points we can be confident that the SS-EPs were not produced by significant artifacts due to head movements, even though one should always remain cautious about this issue. In the perspective of further research, it would be worth testing this issue specifically, by placing on the head a position tracker or an accelerometer, to record head movements directly and with better precision. Moreover, this method would allow studying more systematically the impact of head movement on the quality of the EEG signal in similar SS-EP and rhythm paradigms.

III.2. Study 2: SELECTIVE NEURONAL ENTRAINMENT TO THE BEAT AND METER EMBEDDED IN A MUSICAL RHYTHM

Based on the results of Study 1, Study 2 was conducted using rhythmic patterns that are known to induce spontaneous perception of beat and meter and are commonly used in Western music. The aim of this study was to induce beat and meter *spontaneously* instead of in response to an external instruction imposing a specific frequency and phase for the beat and meter, as it was the case in Study 1. Moreover, the sensory inputs were not strictly periodic within the frequency range for beat perception in Study 2, in contrast with Study 1. In fact, the sensory inputs contained periodicities, generated by the repetition of the rhythmic pattern along the trials, but the frequency of this repetition (0.32 and 0.416 Hz) was assumed to be too slow to correspond to a beat. Therefore, in contrast to Study 1, participants were free to select any frequency and phase available from the rhythmic patterns as their own perceived beat and meter.

Importantly, we observed a selective enhancement of the beat and meter-related SS-EPs in the EEG spectrum, compared to the frequencies contained in the rhythmic patterns that had no relevance for beat and meter. Moreover, this spontaneous selective neural entrainment at frequencies of the perceived beat and meter was disrupted when playing the rhythmic patterns faster or slower, such as to move the tempo away from the musical tempo. These results were interpreted as a direct evidence of the *perceptual selection process* underlying the perception of meter from complex rhythms. Moreover, these results suggest that our approach could allow measuring the quality of the beat percept induced by sounds, without the need of explicit overt movement. Finally, they provide a first evidence of the frequency tuning function, or resonance curve, shaping our perception of beat and meter in a similar manner than the enhancement of beat and meter neural entrainment.

The issues discussed above regarding the interpretation of the results, the frequency domain analysis or the possible artifacts are all relevant also for Study 2. However, the results of this latter study give the opportunity to raise additional points.

III.2.1. The neural sources of beat and meter-related SS-EPs. When analyzing neural responses in the frequency domain, the different SS-EPs identified in the EEG spectrum after Fourier transform (e.g., the beat- and meter-related and unrelated SS-EPs of Study 2) could

(1) result from the frequency decomposition of a unique multi-phasic non sinusoidal neural response from a unique source; (2) constitute separate neural responses from separate neural sources; or (3) reflect an overlap of activities originating from both (1) and (2). According to the first account, the beat and meter-related SS-EPs observed in the EEG Fourier transform could reflect the temporal dynamic of a signal whose amplitude is modulated according to the rhythm envelope, with additional periodic enhancement of the modulation at beat and meter frequency. According to the second account, the distinct neural sources could give rise to the beat and to the meter-related and unrelated SS-EPs (e.g., a network specifically engaged in the processing of sound amplitude modulations, periodic or not, *versus* a network specifically engaged in the processing of beat and meter periodicities). Alternatively, the beat and meter-related and unrelated SS-EPs could also originate from a common neural population but generated by distinct neural subpopulations characterized by distinct frequency tuning functions. According to this latter interpretation, it could be hypothesized that the processing of periodicities (detection and reconstruction) within the frequency range specific to beat and meter could be supported by brain areas specifically devoted to this processing and functionally organized as an array of bandpass filters (i.e., a model similar to models proposed for periodicity pitch) (Eggermont, 2001). Once selected, the periodicities of each beat and metric level would be followed by distinct subgroups of neurons characterized by distinct preferred frequencies of response (see also Section III.2.2 below).

These accounts remain largely speculative at this stage. For instance, in future research, it would be worth investigating the question of the neural sources specifically, by using other electrophysiological methods. Indeed, while EEG lacks spatial resolution to investigate the question of the neural sources, depth-electrode recordings in humans for instance may allow overcoming these limitations. Benefiting from high temporal and spatial resolution, this technique would make it possible to investigate the cerebral sources of beat-related SS-EPs, and relate their temporal course to concomittent signals registered at the surface (Kumar and Schoenwiesner, 2012).

III.2.2. Similarities between pitch and meter processing. Periodicity is the most important determinant of pitch, similarly to meter. Indeed, the auditory system is apparently highly sensitive to the similarity between the successive periods of an acoustic waveform (Bendor and Wang, 2010; Eggermont, 2001). As only a small number of repetitions of the period is necessary to perceive pitch, similarly only a small number of repetitions of a meter is sufficient to induce a meter percept, thus revealing the stability of this percept. Also, the nervous system is tolerant to perturbation or deterioration of this periodicity, as periodicities can be perceived from stimuli that are not strictly periodic in reality, suggesting that percepts of periodicity are supported by invariants abstracted from non periodic inputs. This property of the auditory system has been hypothesized to emerge from the fact that most natural sounds are not strictly periodic, neither within the frequency range of meter or within the frequency range of pitch (Large, 2008; Bendor and Wang, 2010; Schnupp et al. 2010).

Stability, tolerance and invariance in periodicity perception might result from non-linear transformations of the sound spectral content at various levels of the auditory pathway (Large and Almonte, 2012). This is illustrated for example by the missing fundamental phenomenon, in which a pitch can be induced at a given frequency although this frequency is not conveyed in the sound input in reality. Similarly, a beat percept can be induced by a rhythmic pattern at a frequency that is not present in the sound envelope, as illustrated in highly syncopated rhythms (Velasco and Large, 2011).

In fact, one may speculate that meter and pitch emerge from similar physiological properties of the auditory neurons, but occurring at different frequency ranges (between 30 and 20 000 Hz for the pitch and between 0.5 to 5 Hz for the meter). Therefore, investigating the parallel between pitch and meter periodicity using similar neurophysiological paradigms (e.g., jittered *versus* periodic inputs) may help understanding their respective phenomenology and underlying neural mechanisms.

III.2.3. Retrieving time resolution and phase from SS-EPs. The SS-EP approach may appear as an EEG method that misses time resolution, since the elicited activities are identified in the frequency domain rather than in the time domain as ERPs. However, the SS-EP approach may offer the possibility to study the *frequency tuning function* corresponding a given stimulation (see Section I.2.2.3.2). The frequency tuning function is thought to give indication on the “sampling rate” of the neural population, i.e., not only the latency to process a single input but also the timing necessary between successive inputs to be processed. This concept was used in Study 2 to recover the resonance curve thought to shape beat and meter perception. By showing that the selective neural entrainment corresponding to beat and meter perception occurs within a specific frequency range, the results of Study 2 suggest that beat and meter perception are supported by a neural entrainment and by a resonance phenomenon within the responding neural network.

In addition to this aspect of frequency tuning function, another temporal aspect crucial to beat perception is the *phase* selected for the beat within a given rhythmic pattern. In Study 2, participants were free to select a beat percept among the various frequencies and phases available from the inputs. The frequency corresponding to the beat percept was then inferred from the relative enhancement of one SS-EP compared to the others in the EEG spectrum and compared to the amplitude found at these frequencies in the sound envelope.

However, the phase of the beat-related SS-EPs was not directly investigated (although Figure 5 of Study 2 suggests that a consistent phase of beat was selected across participants in Pattern 1, compared to Pattern 5 for instance; this interpretation is also corroborated by the tapping results shown in Figure 1). Nevertheless, this information is important to understand the mechanisms leading to the sponatenous tendency to select a given frequency and phase consistent across participants when listening to given rhythmic patterns.

In future research, we plan to investigate this point specifically. Two groups of participants having experience with musical notation would be presented with two distinct written interpretations of meter of a rhythmic pattern. According to a given interpretation of the meter induced by the musical notation, all downbeats would coincide with actual sound in the pattern (i.e., non-syncopated rhythm), whereas another written interpretation of the meter would induce a meter at the same frequency than in the first metric induction but at a phase in which one or more downbeats would coincide with silence (i.e., syncopated rhythm). The two groups of participants would then listen to the pattern, after having being presented with the musical notation of either the non-syncopated or the syncopated version of the rhythm. A beat-related SS-EP is expected to be elicited at the same frequency in the two groups, but is expected to differ in amplitude, as feeling a syncopated meter would recruit distinct neural activity. Most importantly, this would also allow comparing the phases of these distinct beat-related SS-EPs, while the rhythmic input and the frequency of the beat would not change.

Nevertheless, time information can be retrieved through several phase analysis procedures (e.g., using autocorrelation calculations, time-frequency analysis using sliding discrete Fourier transforms or wavelets transforms, narrowband filtering centered on the expected frequency of SS-EP, as performed in Study 3, Hilbert transform or from the Fourier transform itself, as performed in Study 4 through inter-trial phase coherence calculation). To this aim, the high signal-to-noise ratio obtained by the SS-EP method may help to recover phase information.

III.3. Study 3: CAPTURING WITH EEG THE NEURONAL ENTRAINMENT AND COUPLING UNDERLYING SENSORIMOTOR SYNCHRONIZATION TO THE BEAT

Study 3 was conducted to test whether the *tight coupling between beat perception and movement* could also be studied using our approach. Participants listened to a periodic sound and tapped with their right or left hand in synchrony to every second beat (i.e., at half the beat frequency). The beat-related and movement-related SS-EPs elicited in the EEG could thus be disentangled, at least partly, based on their distinct frequencies. Importantly, we found that moving in synchrony to the auditory beat modified the internal processing of the auditory input, as compared to a passive listening. Specifically, the auditory activity was enhanced within the hemisphere in which the hand movement was predominantly processed, and time-locked to the actual movement.

If this interpretation is correct, it would be worth testing the causal relationship between the forms of neural sensorimotor coupling as observed in Study 3 and behavior. For example, participants would be asked to perform a similar sensorimotor synchronization task but at various tempi making the task more challenging, although still possible to perform (i.e., slower or faster, to reach the rate limits for sensorimotor synchronization). Modifications in the neural sensorimotor coupling could be observed, thus corroborating the interpretation of the results of Study 3.

Importantly, another aspect that is fundamental to interpret the results of Study 3 is the question of the neural sources of the elicited SS-EPs (as developed in the precedings sections of the Discussion). As mentioned in the study (Section II.2.3.4.5), because of the low spatial resolution of EEG data, these findings have to be taken cautiously. Indeed, alternative interpretation of the complex signature of movement-related activities in the frequency domain (i.e., the generation of responses at harmonic frequencies which would not necessarily have the same scalp topography as the response obtained at 1.2 Hz) cannot be excluded. Future studies based on other methods to sample brain activity such as MEG, fMRI or the invasive recording of LFPs could address this crucial question more adequately.

III.4. Study 4: STEADY-STATE EVOKED POTENTIALS AS AN INDEX OF MULTISENSORY TEMPORAL BINDING

Study 4 investigated whether our approach was suitable to study *multisensory integration in beat perception*. How humans build an integrated representation of beat when it is induced through distinct sensory channels (auditory and visual simultaneously) was explored using long-lasting auditory and visual periodic stimuli. The auditory and visual beats were either temporally congruent (leading to a unified perception of beat) or temporally incongruent (not leading to a unified audiovisual beat percept). Features of the auditory and visual inputs, distinct from the beat, were additionally modulated at distinct frequencies. According to the frequency tagging approach, these additional periodic modulations allowed isolating in the EEG spectrum the SS-EPs elicited by the processing of simultaneously presented auditory and visual stimuli, based on their distinct frequencies. We found that the synchronous audiovisual beats elicited enhanced auditory and visual SS-EPs, as compared to the asynchronous audiovisual beats. Moreover, we found that this increase resulted from increased phase consistency of the SS-EPs across trials.

III.4.1. Cross-frequency modulation studied using frequency tagging. Our results suggest that temporal congruency enhances the processing of multisensory inputs, possibly through an improved dynamic attending. Moreover, they suggest that EEG frequency-tagging with SS-EPs may constitute a suitable tool to index multisensory integration in humans. Finally, the results of Study 4 indicate that periodicities induced within the frequency range of beat and meter but across distinct sensory modalities are susceptible to modulate activities elicited at higher frequencies. In the perspective of further research, it would be worth exploring these mechanisms of cross-frequency and cross-modality modulation more systematically, to determine the frequency range in which these nonlinear responses may occur (e.g., testing cross-frequency modulation between two SS-EPs elicited concomitantly, with the lower frequency lying outside the frequency range of beat and meter perception, to determine whether such cross-frequency modulation is specific to this range or not) (Schroeder and Lakatos, 2008; Ding and Simon, 2009).

III.4.2. Is periodicity special? Studies 1 to 4 have captured neural responses that were interpreted as the support of beat and meter. Moreover, these studies revealed some characteristics thought to be related to the processing of beat and meter percepts: internally-driven neural periodicities; selective neural entrainment to periodicities corresponding to beat and meter abstracted from rhythmic patterns; neural resonant frequencies for beat and meter perception; phase and frequency coupling of the periodic entrainment to the beat with periodic movement-related neural activities supporting sensorimotor synchronization to a periodic input; and finally enhanced entrainment to external periodic inputs when these periodic inputs are perceived as synchronized to each other. Given our results, it is worth wondering whether these effects are specific to periodic contexts or not. That is, could all these mechanisms that were captured within periodic contexts in the present work be also observed in non periodic contexts? If it is the case, periodicity would remain a particularly useful feature, not because it would highlight mechanisms specific to periodic contexts, but rather because it would improve the signal-to-noise ratio and make the identification of the neural responses more objective, based on the prior assumption of the frequency at which the responses would concentrate.

To answer this question, it has been proposed that the brain is biased towards either a “rhythmic” or a “continuous” mode of processing, which depens on the dynamics of task demands (Schroeder and Lakatos, 2008). When there is task-relevant temporal structure that sensory systems can entrain to, lower-frequency oscillations entrain to this temporal structure and become instrumental in sensory processing, by modulating the excitability of the neural population accordingly.

To explain the bias towards rhythmic mode in the nervous system, Schroeder and Lakatos (2008) suggested that such mode would be the most cost-effective to provide sustained monitoring of temporal prediction across long periods of time. Along this line, a periodic context could constitute the most powerful context leading to this rhythmic mode, since periodicity optimizes prediction. By extension, this argument would also explain the reason why we tend to perceive periodicities from sequences of sounds that are not strictly periodic in reality: this perceptual bias toward periodicity allows getting into a rhythmic mode of operation, and optimize it, through dynamic attending processes. Interestingly, the argument of the cost-effectiveness balance to explain a tendency to periodic behavior in a system is similar to those explaining the production of periodic movements for long distance displacement of the body observed in numerous animal species. Indeed, to perform efficient body displacement along time, the organisms adapt a periodic mode, as in locomotion in humans, because it constitutes the most cost-effective pattern when movements need to be produced along time (see Section II.2.1).

IV. GLOSSARY

Several terms pertaining to music, physics or neuroscience are recurrently used in the thesis. Various definitions exist for these terms. This glossary aims at clarifying their use in the present work. The definitions mentioned in the glossary are also discussed within the different sections.

IV.1. Entrainment in physics, biological systems and neuroscience

Oscillation. In physics, an oscillation refers to a *periodic phenomenon*, characterized by its frequency, phase and amplitude. In biological systems, the observed oscillations are rarely strictly periodic. Generally, they are *quasi-periodic*, and their periods, or cycle durations, fluctuate around a mean value.

Oscillation waveforms. The oscillation waveform can vary between two types, the *harmonic oscillator*, whose signal is similar to a sinusoidal waveform, and the *relaxation oscillator*, whose signal is similar to a square wave. Relaxation oscillators are the result of non-linear events such as a sudden capacitor discharge. The frequency of the output signal can be changed by altering the time taken by the capacitor to build up its charge (Pikovsky et al., 2003).

In neural systems, the macroscopic appearance of several rhythmic activities, such as the 5 to 10 Hz oscillations observed in the rat hippocampus, resembles the sinusoid pattern of harmonic oscillators. However, considering the properties of neurons, these oscillations are likely to be relaxation oscillators at the level of the individual cell, with a build-up time required to depolarize the membrane, followed by a discharge time of the action potential, once the threshold of voltage-gated sodium channels have been reached (Buzsaki, 2006).

Oscillation time course. In biology, the term "oscillator" is sometimes used as synonym with *self-sustained oscillator* (Hutcheon and Yarom, 2000). The self-sustained property of oscillators could refer to oscillatory activities (1) spontaneously occurring independently of an external stimulation, i.e., at resting state of the system, (2) persisting after the end of a periodic external stimulation, or (3) triggered by an external stimulus which itself is not periodic. Such self-sustained oscillators have been described in neural systems, for example in the form of neurons presenting spontaneous periodic spiking activity in the absence of any external input, thus hypothesized to act as pacemakers (Hutcheon and Yarom, 2000).

An ideal oscillator can reach a *steady state*, in which the amplitude and phase of the signal does not vary along time (Regan, 1989). However, real oscillators do not exhibit a steady state, due to dampening factors leading to an inevitable loss of energy along time (Pikovsky et al., 2003).

Entrainment. Synonym with *synchronization* or *coupling*. The entrainment phenomenon usually refers in physics to the frequency and phase locking of one oscillator to another oscillator, referred to as the driving oscillator. In biological systems, an accurate frequency locking can be achieved but, in most situations, there is a lag between the phase of the driving oscillator and the phase of the entrained oscillator (Roelfsema et al., 1997).

An illustration of entrainment in neural systems is the steady-state evoked potential (SS-EP), in which a periodic series of stimuli elicits a periodic response in the human EEG, whose frequency and phase are locked to the periodicity of the stimulus. Hence, the SS-EP can be viewed as a neuronal oscillation coupled to an external driving oscillator (Regan, 1989). Importantly, this does not necessarily imply that the neuronal population generating these oscillatory signals is a self-sustained oscillator. Moreover, the phase correlation between co-occurring signals, even with a zero phase lag, does not obligatorily mean that the signals are actually coupled. For example, it cannot be inferred from the co-activation of various brain regions in response to a series of external stimuli that these regions are functionally coupled, although they show correlated timing of activation. In other words, both could be independently coupled to the external stimulus.

Importantly, non oscillatory signals can also phase-lock to each other. Such *time locking* is probably what explains the event-related potentials (ERPs) triggered by a single external event, or a train of non periodic events.

Resonance. A resonant oscillator usually refers to an oscillator entrained by a driving oscillator only within particular frequency bands. The frequency at which the amplitude of the entrained oscillatory signal is maximal constitutes the *resonance frequency* of the oscillator. The frequency band in which the oscillator is entrained by a given input delimits the *frequency-tuning function* of the resonant oscillator, centered on its resonance frequency (see Section I.2.2.3.2 for a discussion on frequency tuning functions obtained from EEG recordings).

Nevertheless, delimiting a frequency-tuning function in neurons cannot be considered as evidence that these neurons are oscillatory in nature. The factors determining the band-pass function, or frequency-tuning function, of these neurons can constrain the activity of these neurons, even if they present a non oscillatory behavior. In turn, if the dynamic activity of one neuron is limited to a narrow frequency range due to narrow bandpass filtering generated by structural constraints of the neuron for instance, it results in an almost periodic activity, even when the neuron is stimulated with white noise (Joris et al., 2004).

Cross-modulation products. Synonym with intermodulation, distortion product, interference or sidebands. Cross-modulation products usually refer to as the non-linear convergence, or convolution, of two distinct oscillatory signals (e.g., fn and fm) in the time domain, leading to the emergence of additional oscillatory signals at frequencies corresponding to fn ± fm and their (sub)harmonic components (Giani et al., 2012) (Fig. IV.1).

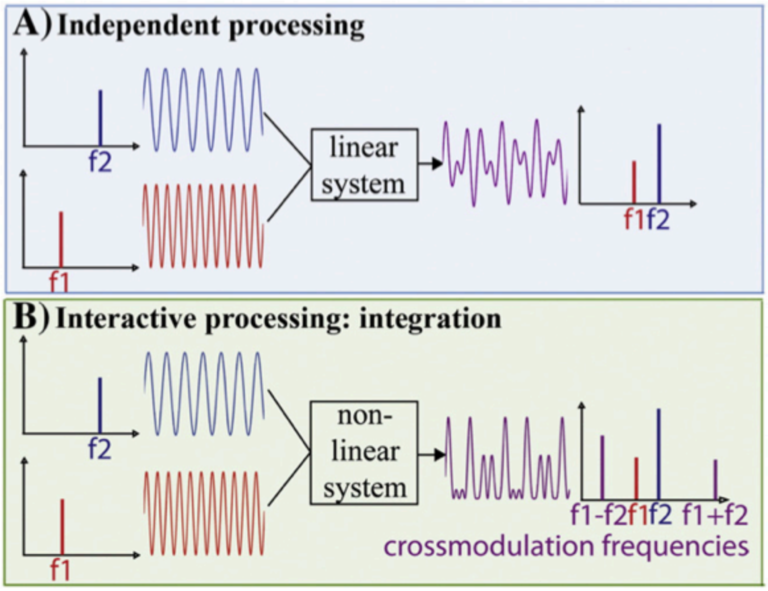


Figure IV.1. From Giani et al. (2012). A. Two frequencies unfolded in a linear system. B. The two frequencies in a non linear system. Their convergence produces crossmodulation frequencies. The signals are represented in the frequency domain (first columns of the left and right panels) and in the time domain (second columns of the left and right panels).

In the human EEG, cross-modulation products can be observed, for example, when subjects listen to a continuous sound whose amplitude has been modulated periodically at two distinct, but close, frequencies. Peaks of activities can emerge at frequencies different from the frequencies of stimulation and their harmonics, but corresponding to the sum or difference of the two stimulation frequencies (and/or their harmonics). The emergence of crossmodulation products is hypothesized to result from the non-linear convergence of the two signals within some neuronal populations (Giani et al., 2012; Regan, 1989; Sutoyo and Srinivasan, 2009).

Importantly, the emergence of crossmodulation products suggests that the SS-EPs cannot be entirely explained by a linear succession of transient evoked potentials. Indeed, a linear sum of two signals elicited at two distinct frequencies simultaneously in the EEG is not expected to generate an additional signal at a frequency corresponding to the sum or difference of the two input frequencies.

Moreover, such processes of non-linear convergence are likely to occur most often between non oscillatory signals, since the inputs received by the convergent neurons are not systematically periodic by nature within the nervous system. Nevertheless, cross-modulation products are easily identified in the EEG for instance when occurring from oscillatory signals, because their frequencies can be predicted based on the fundamental frequencies of stimulation.

IV.2. Musical terms

Tempo. The tempo refers to the velocity of a musical stream. A sequence of sounds can be played at a slow or at a fast tempo. However, within a specific range of tempi, we are perfectly able to recognize a melody played at different tempi (see also Section II.2.3.2).

Rhythm. The rhythm refers to the relative duration of each event within an auditory sequence. Specifically, it refers in music to the timing interval between the sound onsets within a sequence of sounds. By this definition, the rhythm can be considered as a physical property of the auditory stimulus.

Beat. Synonym with pulse, or tactus. The beat usually refers to the perception of a periodicity, and the entrainment to move on this periodicity when listening to certain types of music. In other words, the beat can be considered as the perception of virtual isochronous time intervals occurring even if the music is not periodic in reality (London, 2004) (Fig. IV.2).

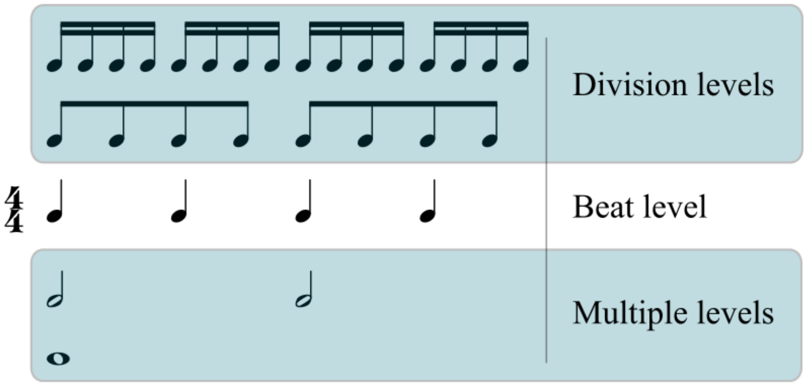


Figure IV.2. Schematic representation of various metric levels (including the beat periodicity) using musical notation (the space between notes are proportional to their duration according to the musical notation). The metric levels can be slower (multiple levels here) or faster (division levels here) than beat frequency.

Meter. Strongly related to the beat, the meter (synonym of metric structure) refers to the integer ratio grouping or subdivision of beat periods into larger or smaller frequencies (Fig. IV.2). If we consider the beat as the fundamental frequency entraining our perceptual system, the various related metric levels can be considered as harmonics or subharmonics of beat frequency. This organization involves a hierarchical order across the beat and the related metric levels, in which the beat would be the most salient periodicity acting as a referent but adaptable timing frame (London, 2004).

Groove. The groove can be viewed as the effectiveness of music to entrain individuals to move to the beat. The phenomenology of groove implies sensorimotor coupling processes that are inherent in the subjective experience of beat and meter in music. This explains the urge to move the body with music (see also Sections II.1.3.1, II.2.2.4.2 and II.2.3.5).

Expectation. Synonym with anticipation, expectancy or prediction. Music could be considered as an unstable equilibrium between expectations induced by various features such as harmony, rhythm, timber dynamics, and violations of these expectations (Huron, 2006). While expectations can build up within a nonperiodic sequence, the periodicity of beat and meter, as reference frame, improves the build-up of expectations by optimizing the predictability of successive events along time.

Coordination. Used as a synonym with entrainment or time coupling. Music leads to the coordination between players and dancers. While movement coordination can build up within a nonperiodic sequence, the periodicity of beat and meter, as reference frame, improves movement coordination between individuals, by optimizing the predictability of successive events along time. Hence, the anticipation of the successive events along the sequence allows compensating the delays due to movement planning and execution.

Regularity. Refers to the recurrence of a pattern along time. The pattern repetition is one of the ingredients ubiquitously used in musical rhythms, at least in popular dance music. The periodic structure of beat and meter constitutes a regular form, whereas all forms of regularities induced by the patterns repetition, are not necessarily periodic.

V. REFERENCES

Adrian ED. 1949. The Sherrington lectures: 1. Sensory integration. University of Liverpool Press, UK.

Appelbaum LG, Wade AR, Pettet MW, Vildavski VY, Norcia AM. 2008. Figure-ground interaction in the human visual cortex. J Vis. 8:8.1-19.

Arnell KM, Jolicœur P. 1999. The attentional blink across stimulus modalities: Evidence for central processing limitations. J Exp Psychol Hum Percept Perform. 25:630-648.

Arom S. 2001. Prolegomena to a biomusicology. In Wallin and Merker. The origins of music. MIT Press, Cambridge, Massachussets, USA.

Aschersleben G. 2002. Temporal control of movements in sensorimotor synchronization. Brain Cogn. 48(1):66-79.

Aschersleben G, Prinz W. 1995. Synchronizing actions with events: the role of sensory information. Percept Psychophys. 57(3):305-17.

Bach M, Meigen T. 1999. Do's and don'ts in Fourier analysis of steady-state potentials. Doc Ophthalmol. 99:69-82.

Bal T, McCormick DA. 1997. Synchronized oscillations in the inferior olive are controlled by the hyperpolarization-activated cation current I(h). J Neurophysiol. 77(6):3145-56.

Bartlett NR, Bartlett SC. 1959. Synchronization of a motor response with an anticipated sensory event. Psychological Review. 66:203-218.

Beek PJ, Peper CE, Daffertshofer A. 2002. Modeling rhythmic interlimb coordination: beyond the Haken-Kelso-Bunz model. Brain Cogn. 48:149-65.

Bell AJ, Sejnowski TJ. 1995. An information-maximization approach to blind separation and blind deconvolution. Neural Comput. 7 :1129-59.

Bendor D, Wang X. 2007. Differential neural coding of acoustic flutter within primate auditory cortex. Nat Neurosci. 10:763-71.

Bendor D, Wang X. 2005. The neuronal representation of pitch in primate auditory cortex. Nature 436(7054):1161-5.

Bendor D, Wang X. 2010. Neural coding of periodicity in marmoset auditory cortex. J. Neurophysiol. 103(4):1809-22.

Berens P. 2009. CircStat: A Matlab Toolbox for Circular Statistics. Journal of Statistical Software. 31:10.

Berger H. 1929. Uber das Elektrenkephalogramm des Menschen. Arch Psychiatr Nervenkr. 87 :527.

Bergeson TR, Trehub SE. 2007. Signature tunes in mothers' speech to infants. Infant Behav Dev. 30(4) :648-54.

Bertelson P. 1999. Ventriloquism: 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, The Netherland. pp. 347–362.

Besle J, Bertrand O, Giard MH. 2009. Electrophysiological (EEG, sEEG, MEG) evidence for multiple audiovisual interactions in the human auditory cortex. Hear Res. 258:143-5.

Bevan MD, Magill PJ, Terman D, Bolam JP, Wilson CJ. 2002. Move to the rhythm: oscillations in the subthalamic nucleus-external globus pallidus network. Trends Neurosci. 25 :525-31.

Bishop G. 1933. Cyclical changes in excitability of the optic pathway of the rabbit. Am J Physiol. 103 :213-224.

Bolton TL. 1894. Rhythm. Am J Psychol. 6:145-238.

Boltz M, Jones MR. 1986. Does rule recursion make melodies easier to reproduce? If not, what does? Cognitive Psychology 18 :389-431.

Boltz MG. 1989b. Rhythm and "good endings": Effects of temporal structure on tonality judgments. Perception & Psychophysics 46 :9-17.

Boltz MG. 1994. Changes in internal tempo and effects on the learning and remembering of event durations. Journal of Experimental Psychology: Learning, Memory, and Cognition 20 :1154-1171.

Bourguignon M, De Tiège X, Op de Beeck M, Pirotte B, Van Bogaert P, Goldman S, Hari R, Jousmäki V. 2011. Functional motor-cortex mapping using corticokinematic coherence. Neuroimage. 55:1475-9.

Bourguignon M, Jousmäki V, Op de Beeck M, Van Bogaert P, Goldman S, De Tiège X. 2012. Neuronal network coherent with hand kinematics during fast repetitive hand movements. Neuroimage 59(2) :1684-91.

Bouwer F, Honing H. 2012. Rhythmic regularity revisited: Is beat induction indeed pre-attentive? Proceedings of the International Conference on Music Perception and Cognition, Thessaloniki, Greece.

Brainard DH. 1997. The Psychophysics Toolbox. Spat Vis. 10:433-6.

Bregman AS. 1990. Auditory Scene Analysis: The Perceptual Organization of Sound. The MIT Press, Cambridge, Massachusetts, USA.

Brochard R, Abecasis D, Potter D, Ragot R, Drake C. 2003. The “tick-tock” of our internal clock: Direct brain evidence of subjective accents in isochronous sequences. Psychol Sci. 14:362-366.

Burger B, Saarikallio S, Luck G, Thompson MR, Toiviainen P. 2012. Emotions Move Us: Basic Emotions in Music Influence People's Movement to Music. In Proceedings of the 12th International Conference on Music Perception and Cognition (ICMPC) / 8th Triennial Conference of the European Society for the Cognitive Sciences of Music (ESCOM), Thessaloniki, Greece.

Busch NA, Dubois J, VanRullen R. 2009. The phase of ongoing EEG oscillations predicts visual perception. J Neurosci. 29: 7869-7876.

Butler MJ. 2006. Unlocking the Groove: Rhythm, Meter, and Musical Design in Electronic Dance Music. Indiana University Press, Indiana, USA.

Buzsáki G, Draguhn A. 2004. Neuronal oscillations in cortical networks. Science 304:1926-9.

Buzsaki G. 2006. Rhythms of the brain. Oxford Univ Press, London, UK.

Calabrese RL. 1998. Cellular, synaptic, network, and modulatory mechanisms involved in rhythm generation. Curr Opin Neurobiol. 8(6):710-7.

Capilla A, Pazo-Alvarez P, Darriba A, Campo P, Gross J. 2011. Steady-state visual evoked potentials can be explained by temporal superposition of transient event-related responses. PLoS One 6(1): e14543.

Chang HW, Trehub SE. 1977. Infants' perception of temporal grouping in auditory patterns. Child Dev. 48(4):1666-70.

Chapin HL, Zanto T, Jantzen KJ, Kelso SJ, Steinberg F, Large EW. 2010. Neural responses to complex auditory rhythms: the role of attending. Front Psychol. 1:224.

Chen JL, Penhune VB, Zatorre RJ. 2008. Listening to musical rhythms recruits motor regions of the brain. Cereb Cortex. 18:2844-54.

Chen JL, Zatorre RJ, Penhune VB. 2006. Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. Neuroimage 32(4):1771-81.

Clayton M. 2001. Music and Meaning. Special issue of British Journal of Ethnomusicology 10(1).

Coles MG. 1989. Modern mind-brain reading: psychophysiology, physiology, and cognition. Psychophysiology 26:251-69.

Colon E, Nozaradan S, Legrain V, Mouraux A. 2012. Steady-state evoked potentials to tag specific components of nociceptive cortical processing. Neuroimage 60(1):571-81.

Colon E, Legrain V, Mouraux A. 2012. Steady-state evoked potentials to study the processing of tactile and nociceptive somatosensory input in the human brain. Neurophysiol Clin. 42(5):315-23.

Cooper G, Meyer LB. 1960. The rhythmic structure of music. University of Chicago Press, Chicago, USA.

Cottereau BR, McKee SP, Ales JM, Norcia AM. 2011. Disparity-tuned population responses from human visual cortex. J Neurosci. 131:954-65.

Csicsvari J, Jamieson B, Wise KD, Buzsáki G. 2003. Mechanisms of gamma oscillations in the hippocampus of the behaving rat. Neuron 37:311-22.

Daffertshofer A, Peper CL, Beek PJ. 2005. Stabilization of bimanual coordination due to active interhemispheric inhibition: a dynamical account. Biol Cybern. 92:101-9.

Danielsen A. 2006. Presence and Pleasure - the funk grooves of James Brown and Parliament. Wesleyan University Press, Middletown, Connecticut, USA.

de Hemptinne C, Nozaradan S, Duvivier Q, Lefèvre P, Missal M. 2007. How do primates anticipate uncertain future events? J Neurosci. 27(16):4334-41.

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.

De Waal FBM. 1988. The communicative repertoire of captive bonobos (Pan paniscus), compared to that of chimpanzees. Behaviour 106:183–251.

Demany L, McKenzie B, Vurpillot E. 1977. Rhythm perception in early infancy. Nature 266(5604):718-9.

Desain P, Honing H. 1999. Computational Models of Beat Induction: The Rule-Based Approach. Journal of New Music Research 28(1):29-42.

Desain P, Honing H. 2003. The formation of rhythmic categories and metric priming. Perception 32(3):341–365.

Ding N, Simon JZ. 2009. Neural representations of complex temporal modulations in the human auditory cortex. J Neurophysiol 102(5):2731-43.

Ding J, Sperling G, Srinivasan R. 2006. Attentional modulation of SSVEP power depends on the network tagged by the flicker frequency. Cereb Cortex 16(7):1016-29.

Donoghue JP, Sanes JN, Hatsopoulos NG, Gaál G. 1998. Neural discharge and local field potential oscillations in primate motor cortex during voluntary movements. J Neurophysiol. 79(1):159-73.

Draganova R, Ross B, Borgmann C, Pantev C. 2002. Auditory cortical response patterns to multiple rhythms of AM sound. Ear Hear. 23:254-65.

Drake C, Botte MC. 1993. Tempo sensitivity in auditory sequences: Evidence for a multiple look model. Percept Psychophys. 54:277-286.

Drake C, Jones MR, Baruch C. 2000. The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. Cognition 77:251-88.

Drake, C. 1993. Reproduction of musical rhythms by children, adult musicians, and adult nonmusicians. Percept. Psychophys. 53(1), 25-33.

Eagleman DM, Tse PU, Buonomano D, Janssen P, Nobre AC, Holcombe AO. 2005. Time and the brain: how subjective time relates to neural time. J Neurosci. 25(45):10369-71.

Eck D. 2002. Meter Through Synchrony: Processing Rhythmical Patterns with Relaxation Oscillators. Doctoral thesis, Indiana University, USA.

Edelman GM. 1989. Topobiology. Sci Am. 260(5):76-82, 84-6, 88.

Edelman GM. 1993. Neural Darwinism: selection and reentrant signaling in higher brain function. Neuron 10(2): 115-25.

Eggermont JJ. 2001. Between sound and perception: reviewing the search for a neural code. Hear Res. 157:1-42.

Elliott MT, Wing AM, Welchman AE. 2010. Multisensory cues improve sensorimotor synchronisation. Eur J Neurosci. 31:1828-35.

Ellis RJ, Jones MR. 2009. The Role of Accent Salience and Joint Accent Structure in Meter Perception. Journal of Experimental Psychology: Human Perception and Performance 35:264–280.

Engel AK, Fries P, Singer W. 2001. Dynamic predictions: oscillations and synchrony in top-down processing. Nat Rev Neurosci. 2:704-16.

Engstrom DA, Kelso JAS, Holroyd T. 1996. Reaction-anticipation transitions in human perception-action patterns. Human Movement Science 15(6):809-832.

Essens PJ, Povel DJ. 1985. Metrical and nonmetrical representations of temporal patterns. Percept Psychophys. 37:1-7.

Essens PJ. 1986. Hierarchical organization of temporal patterns. Percept Psychophys. 40:69–73.

Fitch WT, Rosenfeld AJ. 2007. Perception and Production of Syncopated Rhythms. Music Percept. 25:43-58.

Fitch WT. 2006. The biology and evolution of music: A comparative perspective. Cognition 100:173-215.

Fitzpatrick P, Schmidt R C, Lockman J J. 1996. Dynamical patterns in the development of clapping. Child Development 67:2691–2708.

Foxe JJ, Schroeder CE. 2005. The case for feedforward multisensory convergence during early cortical processing. Neuroreport 16: 419-23.

Fraisse P. 1956. Les structures rythmiques : étude psychologique. Publications universitaires de Louvain, Louvain, Belgium.

Fraisse P. 1964. The Psychology of Time. Eyre and Spottiswoode, UK.

Fraisse P. 1980. Les synchronisations sensori-motrices aux rythmes. In J. Requin (Ed.), Anticipation et comportement. Editions C.N.R.S., Paris, France.

Fraisse P. 1982. Rhythm and tempo. In The psychology of music (Deutsch D, Ed.), New York Academic Press, NY, USA.

Fries P. 2005. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn Sci. 9:474-80.

Frigo M, Johnson SG. 1998. FFTW: an adaptive software architecture for the FFT. Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing 3:1381–1384.

Fujioka T, Zendel BR, Ross B. 2010. Endogenous neuromagnetic activity for mental hierarchy of timing. J Neurosci 30:3458-66.

Fujioka T, Trainor LJ, Large EW, Ross B. 2009. Beta and gamma rhythms in human auditory cortex during musical beat processing. Ann N Y Acad Sci. 1169:89-92.

Fujisaki W, Nishida S. 2005. Temporal frequency characteristics of synchrony-asynchrony discrimination of audio-visual signals. Exp Brain Res. 166:455-64.

Galambos R, Makeig S, Talmachoff PJ. 1981. A 40-Hz auditory potential recorded from the human scalp. Proc Natl Acad Sci USA. 78:2643-7.

Galambos R. 1982. Tactile and auditory stimuli repeated at high rates (30-50 per sec) produce similar event related potentials. Ann NY Acad Sci. 388:722-8.

Geiser E, Sandmann P, Jäncke L, Meyer M. 2010. Refinement of metre perception--training increases hierarchical metre processing. Eur J Neurosci. 32(11):1979-85.

Gerloff C, Toro C, Uenishi N, Cohen LG, Leocani L, Hallett M. 1997. Steady-state movement-related cortical potentials: a new approach to assessing cortical activity associated with fast repetitive finger movements. Electroencephalogr Clin Neurophysiol. 102:106-13.

Gerloff C, Uenishi N, Nagamine T, Kunieda T, Hallett M, Shibasaki H. 1998. Cortical activation during fast repetitive finger movements in humans: steady-state movement-related magnetic fields and their cortical generators. Electroencephalogr Clin Neurophysiol. 109:444-53.

Gerloff C, Richard J, Hadley J, Schulman AE, Honda M, Hallett M. 1998. Functional coupling and regional activation of human cortical motor areas during simple, internally paced and externally paced finger movements. Brain 121(Pt 8):1513-31.

Gerloff C, Uenishi N, Hallett M. 1998. Cortical activation during fast repetitive finger movements in humans: dipole sources of steady-state movement-related cortical potentials. J Clin Neurophysiol. 15(6):502-13.

Gerry DW, Faux AL, Trainor LJ. 2010. Effects of Kindermusik training on infants' rhythmic enculturation. Dev Sci. 13(3):545-51.

Ghazanfar AA, Schroeder CE. 2006. Is neocortex essentially multisensory? Trends Cogn Sci. 10:278-85.

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 60:1478-89.

Gibbon J. 1977. Scalar expectancy theory and Weber's law in animal timing. Psychological Review. 84(3):279-325.

Gibbon J, Church RM, Meck WH. 1984. Scalar timing in memory. Ann NY Acad Sci. 423:52-77.

Gibson JJ. 1966. The problem of temporal order in stimulation and perception. J Psychol. 62:141-9.

Giraud AL, Poeppel D. 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci. 15(4):511-7.

Glenberg AM, Mann S, Altman L, Forman T. 1989. Modality effects in the coding and reproduction of rhythms. Mem Cogn. 17:373-383.

Grabe E, Low EL. 2002. Durational Variability in Speech and the Rhythm Class Hypothesis. Papers in Laboratory Phonology, 7.

Grahn JA, Brett M. 2007. Rhythm and beat perception in motor areas of the brain. J Cogn Neurosci. 19:893-906.

Grahn JA. 2009. The role of the basal ganglia in beat perception: neuroimaging and neuropsychological investigations. Ann NY Acad Sci. 1169:35-45.

Grahn JA. 2012. Neural mechanisms of rhythm perception: current findings and future perspectives. Top Cogn Sci. 4(4):585-606.

Grahn JA, Brett M. 2009. Impairment of beat-based rhythm discrimination in Parkinson's disease. Cortex 45(1):54-61.

Grahn JA, Henry MJ, McAuley JD. 2011. FMRI investigation of cross-modal interactions in beat perception: audition primes vision, but not vice versa. Neuroimage 54(2):1231-43.

Grahn JA, Rowe JB. 2009. Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. J Neurosci. 29(23):7540-8.

Grondin S. 2001. From physical time to the first and second moments of psychological time. Psychol Bull. 127(1):22-44.

Grondin S, Meilleur-Wells G, Ouellette C, Macar F. 1998. Sensory effects on judgments of short time-intervals. Psychol Res. 61(4):261-8.

Gross J, Timmermann L, Kujala J, Dirks M, Schmitz F, Salmelin R, Schnitzler A. 2002. The neural basis of intermittent motor control in humans. Proc. Natl Acad Sci USA. 99(4):2299-302.

Grube M, Griffiths TD. 2009. Metricality-enhanced temporal encoding and the subjective perception of rhythmic sequences. Cortex 45:72-9.

Grube M, Lee KH, Griffiths TD, Barker AT, Woodruff PW. 2010. Transcranial magnetic theta-burst stimulation of the human cerebellum distinguishes absolute, duration-based from relative, beat-based perception of subsecond time intervals. Front Psychol. 1:171.

Hagen EH, Bryant GA. 2003. Music and dance as a coalition signaling system. Human Nature 14:21-51.

Haig AR, Gordon E. 1998. Prestimulus EEG alpha phase synchronicity influences N100 amplitude and reaction time. Psychophysiology 35:591-5.

Hannon EE. 2009. Perceiving speech rhythm in music: listeners classify instrumental songs according to language of origin. Cognition 111(3):404-10.

Hannon EE, Johnson SP. 2005. Infants use meter to categorize rhythms and melodies: implications for musical structure learning. Cogn Psychol. 50(4):354-77.

Hannon EE, Soley G, Levine RS. 2011. Constraints on infants' musical rhythm perception: effects of interval ratio complexity and enculturation. Dev Sci. 14(4):865-72.

Hannon EE, Soley G, Ullal S. 2012. Familiarity overrides complexity in rhythm perception: a cross-cultural comparison of American and Turkish listeners. J Exp Psychol Hum Percept Perform. 38(3):543-8.

Hannon EE, Trehub SE. 2005. Metrical categories in infancy and adulthood. Psychol Sci. 16(1):48-55.

Herry C, Bach DR, Esposito F, Di Salle F, Perrig WJ, Scheffler K, Lüthi A, Seifritz E. 2007. Processing of temporal unpredictability in human and animal amygdala. J Neurosci. 27(22):5958-66.

Hogan N, Sternad D. 2007. On rhythmic and discrete movements: reflections, definitions and implications for motor control. Exp Brain Res. 181:13-30.

Honing H. 2012. Without it no music: beat induction as a fundamental musical trait. Ann NY Acad Sci. 1252:85-91.

Hove MJ, Iversen JR, Zhang A, Repp BH. 2012. Synchronization with competing visual and auditory rhythms: bouncing ball meets metronome. Psychol Res. In Press.

Hove MJ, Risen JL. 2009. It's all in the timing: Interpersonal synchrony increases affiliation. Social Cognition 27(6):949-961.

Huron D. 2006. Sweet Anticipation: Music and the Psychology of Expectation. MIT Press, Cambridge, Massachusetts, USA.

Hutcheon B, Yarom Y. 2000. Resonance, oscillation and the intrinsic frequency preferences of neurons. Trends Neurosci 23:216-22.

Iversen JR, Repp BH, Patel AD. 2009. Top-down control of rhythm perception modulates early auditory responses. In Dalla Bella S, Kraus N, Overy K, Pantev C, Snyder JS, Tervaniemi M, et al. (Eds.) The neurosciences and music III: Disorders and plasticity. Ann NY Acad Sci. 1169:58-73.

Iversen JR, Patel AD, Chen Y, Gally JA, Edelman GM. 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, USA.

Iversen JR, Patel AD, Ohgushi K. 2008. Perception of rhythmic grouping depends on auditory experience. J Acoust Soc Am. 124(4):2263-71.

Ivry R. 1997. Cerebellar timing systems. Int Rev Neurobiol. 41:555-73.

Ivry RB, Spencer RM. 2004. The neural representation of time. Curr Opin Neurobiol. 14(2):225-32.

Iyer V. 2002. Embodied Mind, Situated Cognition, and Expressive Microtiming in African-American Music. Music Perception 19(3):387-414.

Janata P, Tomic ST, Haberman JM. 2012. Sensorimotor coupling in music and the psychology of the groove. J Exp Psychol Gen. 141(1):54-75.

Jäncke L, Loose R, Lutz K, Specht K, Shah NJ. 2000. Cortical activations during paced finger-tapping applying visual and auditory pacing stimuli. Cogn Brain Res. 10(1-2):51-66.

Jaques-Dalcroze E. 1920. Le rythme, la musique et l'éducation. Fœtisch, Lausanne, Switzerland.

Jarvis ED. 2004. Learned Birdsong and the Neurobiology of Human Language. Ann NY Acad Sci. 1016:749-777.

Johnson BW, Weinberg H, Ribary U, Cheyne DO, Ancill R. 1988. Topographic distribution of the 40 Hz auditory evoked-related potential in normal and aged subjects. Brain Topogr. 1(2):117-21.

Jones MR, Boltz M. 1989. Dynamic attending and responses to time. Psychol Rev. 96:459-491.

Jones EG, Powell TP. 1970. An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. Brain 93:793-820.

Jones MR. 1976. Time, our lost dimension: Toward a new theory of perception, attention, and memory. Psychological Review 83(5):323-355.

Jones MR, Boltz M, Klein J. 1993. Duration judgments and expected endings. Memory and Cognition 21:646-665.

Jones MR, McAuley JD. 2005. Time judgments in global temporal contexts. Percept Psychophys. 67(3):398-417.

Joris PX, Schreiner CE, Rees A. 2004. Neural processing of amplitude-modulated sounds. Physiol Rev. 84(2):541-77.

Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ. 2000. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. Clin Neurophysiol. 111:1745-58.

Karabanov A, Blom O, Forsman L, Ullén F. 2009. The dorsal auditory pathway is involved in performance of both visual and auditory rhythms. Neuroimage 44(2):480-8.

Kayser J, Tenke CE. 2006. Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: I. Evaluation with auditory oddball tasks. Clin Neurophysiol. 117(2):348-68.

Kayser C. 2009. Phase resetting as a mechanism for supramodal attentional control. Neuron 64:300-2.

Kayser C, Logothetis NK, 2007. Do early sensory cortices integrate cross-modal information? Brain Struct Funct. 212:121-32.

Kayser C, Logothetis NK, Panzeri S. 2010. Visual enhancement of the information representation in auditory cortex. Curr Biol. 20:19-24.

Kayser C, Petkov CI, Logothetis NK. 2008. Visual modulation of neurons in auditory cortex. Cereb Cortex. 18:560-74.

Kiel C, Steven F. 1994. Music Grooves: Essays and Dialogues. University of Chicago Press, Chicago, Illinois, USA.

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. 10:117-25.

Kirschner S, Tomasello M. 2009. Joint drumming: social context facilitates synchronization in preschool children. J Exp Child Psychol. 102(3):299-314.

Klimesch W. 1999. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res Rev. 29(2-3):169-95.

Knight S, Cross I. 2012. Rhythms of persuasion: The perception of periodicity in oratory. Proceedings of the Perspective of rhythm and timing PoRT workshop, Glasgow, UK.

Kopp B, Kunkel A, Müller G, Mühlnickel W, Flor H. 2000. Steady-state movement-related potentials evoked by fast repetitive movements. Brain Topogr. 13:21-28.

Kourtis D, Seiss E, Praamstra P. 2008. Movement-related changes in cortical excitability: a steady-state SEP approach. Brain Res. 1244:113-120.

Krauslis RJ, Liston D, Carello CD. 2004. Target selection and the superior colliculus: goals, choices and hypotheses. Vision Res. 44:1445-1451.

Kumar S, Schönwiesner M. 2012. Mapping human pitch representation in a distributed system using depth-electrode recordings and modeling. J Neurosci. 32(39):13348-51.

Lachaux JP, Rodriguez E, Martinerie J, Adam C, Hasboun D, Varela FJ. 2000. A quantitative study of gamma-band activity in human intracranial recordings triggered by visual stimuli. Eur J Neurosci. 12:2608-22.

Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE. 2008. Entrainment of neuronal oscillations as a mechanism of attentional selection. Science 320(5872):110-3.

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 64:419-30.

Lamme VAF. 2010. Can neuroscience reveal the true nature of consciousness ? Unpublished article.

Langer S. 1942. Philosophy in a New Key. Harvard University Press, Cambridge, MA, USA.

Langner R, Eickhoff SB. 2012. Sustaining Attention to Simple Tasks: A Meta-Analytic Review of the Neural Mechanisms of Vigilant Attention. Psychol Bull. In Press.

Large EW. 2008. Resonating to musical rhythm: Theory and experiment. In Simon Grondin (Ed.) The Psychology of Time. Emerald, West Yorkshire, USA.

Large EW, Jones MR. 1999. The dynamics of attending: How we track time varying events. Psychol Rev. 106:119-159.

Large EW, Kolen JF. 1994. Resonance and the perception of musical meter. Connect Sci. 6:177-208.

Large EW, Snyder JS. 2009. Pulse and meter as neural resonance. The Neurosciences and Music III – Disorders and Plasticity. Ann NY Acad Sci. 1169:46-57.

Large EW, Almonte FV. 2012. Neurodynamics, tonality, and the auditory brainstem response. Ann NY Acad Sci. 1252:E1-7.

Large EW, Fink P, Kelso JAS. 2002. Tracking simple and complex sequences. Psychological Research 66:3-17.

Large EW, Kolen JF. 1999. Resonance and the perception of musical meter. In N. Griffith and P. Todd (Eds.), Musical Networks: Parallel Distributed Perception and Performance. MIT Press. Cambridge, MA, USA.

Lartillot O, Toiviainen P. 2007. A Matlab toolbox for musical feature extraction from audio. Proc of the 10th Int Conference on Digital Audio Effects (DAFx-07).

Lerdahl F, Jackendoff R. 1983. A generative theory of tonal music. Cambridge MIT Press, Cambridge, MA, USA.

Lewis PA, Wing AM, Pope PA, Praamstra P, Miall RC. 2004. Brain activity correlates differentially with increasing temporal complexity of rhythms during initialisation, synchronisation, and continuation phases of paced finger tapping. Neuropsychologia 42(10):1301-12.

Lidji P, Palmer C, Peretz I, Morningstar M. 2011. Listeners feel the beat: entrainment to English and French speech rhythms. Psychon Bull Rev. 18(6):1035-41.

Liégeois-Chauvel C, Peretz I, Babaï M, Laguitton V, Chauvel P. 1998. Contribution of different cortical areas in the temporal lobes to music processing. Brain 121(10):1853-67.

Llinás RR. 1988. The intrinsic electrophysiological properties of mammalian neurons: insights into central nervous system function. Science 242:1654-64.

London J. 2004. Hearing in time: psychological aspects of musical meter. Oxford Univ Press, London, UK.

Longhi E. 2003. The temporal structure of mother-infant interactions in musical contexts. Unpublished doctoral dissertation, University of Edinburgh, UK.

Longuet-Higgins HC, Lee CS. 1982. The perception of musical rhythms. Perception 11:115–128.

Lopes da Silva FH. 2006. Chapter 1 Event-related neural activities: what about phase? Progress in Brain Research. 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.

Lütkenhöner B, Lammertmann C, Simões C, Hari R. 2002. Magnetoencephalographic correlates of audiotactile interaction. Neuroimage 15:509-22.

MacDorman CE. 1962. Synchronization with auditory models with varying complexity. Perceptual and motor skills 1:595-602.

MacDougall HG, Moore ST. 2005. Marching to the beat of the same drummer: the spontaneous tempo of human locomotion. J Appl Physiol. 99:1164-73.

Madison G. 2006. Experiencing Groove Induced by Music: Consistency and Phenomenology. Music Percept. 24:201-208.

Makeig S. 2002. Response: event-related brain dynamics - unifying brain electrophysiology. Trends Neurosci. 25:390.

Makeig S, Bell AJ, Jung TP, Sejnowski TJ. 1996. Independent component analysis of electroencephalographic data, In: D. Touretzky, M. Mozer and M. Hasselmo (Eds). Advances in Neural Information Processing Systems 8:145-151, MIT Press, Cambridge, MA, USA.

Manganotti P, Gerloff C, Toro C, Katsuta H, Sadato N, Zhuang P, Leocani L, Hallett M. 1998. Task-related coherence and task-related spectral power changes during sequential finger movements. Electroencephalogr. Clin Neurophysiol. 109(1):50-62.

Marder E, Bucher D. 2001. Central pattern generators and the control of rhythmic movements. Curr Biol. 11(23):R986-96.

Matell MS, Meck WH. 2000. Neuropsychological mechanisms of interval timing behavior. Bioessays 22(1):94-103.

Mates J. 1994. A model of synchronization of motor acts to a stimulus sequence: I. Timing and error corrections. Biological Cybernetics 70:463–473.

Mauk MD, Buonomano DV. 2004. The neural basis of temporal processing. Ann Rev Neurosci. 27:307-40.

May PJ, Tiitinen H. 2010. Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. Psychophysiology 47(1):66-122.

McAdams S. 1989. Introduction: The many faces of human cognition in musical research and practice. Contemporary Music Review 4:1-7.

McAuley JD. 2010. Tempo and rhythm. In M.R. Jones et al. (eds.), Music Perception, Springer Handbook of Auditory Research 36, USA.

McAuley JD, Jones MR. 2003. Modeling effects of rhythmic context on perceived duration: a comparison of interval and entrainment approaches to short-interval timing. J Exp Psychol Hum Percept Perform. 29(6):1102-25.

McAuley JD, Jones MR, Holub S, Johnston HM, Miller NS. 2006. The time of our lives: life span development of timing and event tracking. J Exp Psychol Gen. 135(3)-348-67.

Merker B, Madison G, Eckerdal P. 2009. On the role and origin of isochrony in human rhythmic entrainment. Cortex 45:4-17.

Michon JA. 1964. Studies on subjective duration. I. Differential sensitivity in the perception of repeated temporal intervals. Acta Psychol Amst. 22:441-50.

Michon JA. 1978. The making of the present: A tutorial review. In J. Requin (Ed.), Attention and performance VII (pp.89-109). Hillsdale, NJ, Erlbaum.

Middleton FA, Strick PL. 2000. Basal ganglia and cerebellar loops: motor and cognitive circuits. Brain Res Rev. 31(2-3):236-50.

Mithen S J. 2005. The Singing Neanderthals: the Origins of Music, Language, Mind and Body. Harvard University Press, Cambridge, MA, USA.

Miyake A. 2001. Individual differences in working memory: introduction to the special section. J Exp Psychol Gen. 130(2):163-8.

Morgan ST, Hansen JC, Hillyard SA. 1996. Selective attention to stimulus location modulates the steady-state visual evoked potential. Proc Natl Acad Sci USA. 93:4770-4.

Mouraux A, Iannetti GD. 2008. Across-trial averaging of event-related EEG responses and beyond. Magn Reson Imaging 26:1041-54.

Mouraux A, Iannetti GD, 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. 31:6079-87.

Müller K, Schmitz F, Schnitzler A, Freund HJ, Aschersleben G, Prinz W. 2000. Neuromagnetic correlates of sensorimotor synchronization. J Cogn Neurosci. 12(4):546-55.

Näätänen R, Kujala T, Winkler I. 2011. Auditory processing that leads to conscious perception: a unique window to central auditory processing opened by the mismatch negativity and related responses. Psychophysiology 48(1):4-22.

Näätänen R, Paavilainen P, Rinne T, Alho K. 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin Neurophysiol. 118(12):2544-90.

Nakata T, Trehub SE, Mitani C, Kanda Y, Shibasaki A, Schellenberg EG. 2005. Music recognition by Japanese children with cochlear implants. J Physiol Anthropol Appl Human Sci. 24(1):29-32.

Navarro Cebrian A, Janata P. 2010. Electrophysiological correlates of accurate mental image formation in auditory perception and imagery tasks. Brain Res. 1342:39-54.

Nelken I, Rotman Y, Bar Yosef O. 1999. Responses of auditory-cortex neurons to structural features of natural sounds. Nature 397(6715):154-7.

Nelson K. 1985. The art of reciting the Qur'an. University of Texas Press, Austin, USA.

Nozaradan S, Peretz I, Missal M, Mouraux A. 2011. Tagging the neuronal entrainment to beat and meter. J Neurosci. 31:10234-40.

Nozaradan S, Peretz I, Mouraux A. 2012. Steady-state evoked potentials as an index of multisensory temporal binding. NeuroImage 60:21-8.

Nozaradan S, Peretz I, Mouraux A. 2012. Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. J Neurosci. 32(49):17572-81.

Nunez PL, Srinivasan R. 2006. Electric Fields of the Brain: The Neurophysics of EEG. 2nd Edition, Oxford University Press, New York, USA.

Obleser J, Herrmann B, Henry MJ. 2012. Neural Oscillations in Speech: Don't be Enslaved by the Envelope. Front Hum Neurosci. 6:250.

Oostendorp TF, van Oosterom A. 1989. Source parameter estimation in inhomogeneous volume conductors of arbitrary shape. IEEE Trans Biomed Eng. 36(3):382-91.

Oostenveld R, Fries P, Maris E, Schoffelen JM. 2011. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci. 2011:156869.

Osman A, Albert R, Ridderinkhof KR, Band G, van der Molen M. 2006. The beat goes on: rhythmic modulation of cortical potentials by imagined tapping. J Exp Psychol Hum Percept Perform. 32(4):986-1005.

Pablos Martin X, Deltenre P, Hoonhorst I, Markessis E, Rossion B, Colin C. 2007. Perceptual biases for rhythm: The Mismatch Negativity latency indexes the privileged status of binary vs non-binary interval ratios. Clin Neurophysiol. 118:2709-15.

Palmer C, Krumhansl CL. 1990. Mental representations for musical meter. J Exp Psychol Human 16:728-741.

Parncutt R. 1994e. Language, Music, and Mind. MIT Press, Cambridge, MA, USA.

Patel AD. 2006. Musical rhythm, linguistic rhythm, and human evolution. Music Perception 24:99-104.

Patel AD. 2008. Music, Language, and the Brain. Oxford University Press, NY, USA.

Patel AD, Daniele JR. 2003. An empirical comparison of rhythm in language and music. Cognition 87(1):B35-45.

Patel AD, Iversen JR, Bregman MR, Schulz I. 2009. Experimental evidence for synchronization to a musical beat in a nonhuman animal. Current Biology 19:827-830.

Patel AD, Iversen JR, Chen Y, Repp BH. 2005. The influence of metricality and modality on synchronization with a beat. Exp Brain Res. 163(2):226-38.

Penel A. 2000. Variations temporelles dans l’interprétation musicale : processus perceptifs et cognitifs. Unpublished PhD dissertation. Université of Paris 6, France.

Peretz I. 1990. Processing of local and global musical information by unilateral brain-damaged patients. Brain 113(4):1185-205.

Peretz I, Coltheart M. 2003. Modularity of music processing. Nat Neurosci. 6(7):688-91.

Petrini K, Russell M, Pollick F. 2009. When knowing can replace seeing in audiovisual integration of actions. Cognition 110:432-9.

Pfordresher PQ. 2003. Auditory feedback in music performance: Evidence for a dissociation of sequencing and timing. J Exp Psychol Hum Percept Perform. 29(5):949-64.

Pfurtscheller G, Lopes da Silva FH. 1999. Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin Neurophysiol. 110:1842-57.

Phillips-Silver J, Aktipis CA, Bryant GA. 2010. The ecology of entrainment: Foundations of coordinated rhythmic movement. Music Percept. 28:3-14.

Phillips-Silver J, Trainor LJ. 2005. Feeling the beat: movement influences infant rhythm perception. Science 308:1430.

Phillips-Silver J, Trainor LJ. 2007. Hearing what the body feels: auditory encoding of rhythmic movement. Cognition 105:533-46.

Phillips-Silver J, Keller PE. 2012. Searching for roots of entrainment and joint action in early musical interactions. Front Hum Neurosci. 6-26.

Phillips-Silver J, Toiviainen P, Gosselin N, Piché O, Nozaradan S, Palmer C, Peretz I. 2011. Born to dance but beat deaf: a new form of congenital amusia. Neuropsychologia 49(5):961-9.

Phillips-Silver J, Trainor LJ. 2008. Vestibular influence on auditory metrical interpretation. Brain Cogn. 67(1):94-102.

Pikovsky A, Rosenblum M, Kurths J. 2001. Synchronization: A Universal Concept in Nonlinear Sciences. Cambridge University Press, MA, USA.

Pinker S. 1997. How the Mind Works. Norton, USA.

Plourde G. 2006. Auditory evoked potentials. Best Pract Res Clin Anaesthesiol. 20:129-39.

Pollok B, Müller K, Aschersleben G, Schnitzler A, Prinz W. 2004. The role of the primary somatosensory cortex in an auditorily paced finger tapping task. Exp Brain Res. 156:111-7.

Pollok B, Gross J, Schnitzler A. 2006. How the brain controls repetitive finger movements. J Physiol. 99(1):8-13.

Pollok B, Müller K, Aschersleben G, Schmitz F, Schnitzler A, Prinz W. 2003. Cortical activations associated with auditorily paced finger tapping. Neuroreport 14(2):247-50.

Pöppel E. 1997. A hierarchical model of temporal perception. Trends Cogn Sci. 11(2):56-61.

Potter DD, Fenwick M, Abecasis D, Brochard R. 2009. Perceiving rhythm where none exists: event-related potential (ERP) correlates of subjective accenting. Cortex 45(1):103-9.

Povel DJ, Essens PJ. 1985. Perception of temporal patterns. Music Perception 2:411-441.

Povel DJ, Okkerman H. 1981. Accents in equitone sequences. Perception and Psychophysics 3:565-572.

Praamstra P, Turgeon M, Hesse CW, Wing AM, Perryer L. 2003. Neurophysiological correlates of error correction in sensorimotor-synchronization. Neuroimage 20:1283-97.

Praamstra P, Kourtis D, Kwok HF, Oostenveld R. 2006. Neurophysiology of implicit timing in serial choice reaction-time performance. J Neurosci. 26(20):5448-55.

Pressing J. 1998. Error Correction Processes in Temporal Pattern Production. J Math Psychol. 42:63-101.

Pressing J. 2002. Black Atlantic Rhythm: Its Computational and Transcultural Foundations. Music Perception 19(3):285-310.

Provasi J, Bobin-Bègue A. 2003. Spontaneous motor tempo and rhythmical synchronization in 2 1/2- and 4-year-old children. International Journal of Behavioral Development 27(3):220–231.

Purcell DW, Ross B, Picton TW, Pantev C. 2007. Cortical responses to the 2f1-f2 combination tone measured indirectly using magnetoencephalography. J Acoust Soc Am. 122:992-1003.

Rao SM, Harrington DL, Haaland KY, Bobholz JA, Cox RW, Binder JR. 1997. Distributed neural systems underlying the timing of movements. J Neurosci. 17(14):5528-35.

Rao SM, Mayer AR, Harrington DL. 2001. The evolution of brain activation during temporal processing. Nat Neurosci. 4:317–323.

Raymond JE, Shapiro KL, Arnell KM. 1992. Temporary suppression of visual processing in an RSVP task: an attentional blink? J Exp Psychol Hum Percept Perform. 18(3):849-60.

Read HL, Winer JA, Schreiner CE. 2002. Functional architecture of auditory cortex. Curr Opin Neurobiol. 12(4):433-40.

Regan D. 1989. Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine. Elsevier, New York, USA.

Regan D. 1966. Some Characteristics of Average Steady-State and Transient Responses Evoked by Modulated Light. Electroenceph Clin Neurophysiol. 20:238-48.

Regan D. 1972. Cortical Evoked Potentials. Adv Behav Biol. 5:177-92.

Regan D, Heron JR. 1969. Clinical investigation of lesions of the visual pathway: a new objective technique. J Neurol Neurosurg Psych. 32(5):479–83.

Repp BH. 2005. Sensorimotor synchronization: A review of the tapping literature. Psychon B Rev. 12:969-992.

Repp BH. 2006. Rate limits of sensorimotor synchronization. Advances in Cogn Psy. 2:163-181.

Repp BH. 2010. Do metrical accents create illusory phenomenal accents? Atten Percept Psychophy. 72:1390-403.

Repp BH. 1992. Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's “Traumerei”. Journal of the Acoustical Society of America 92(5):2546-2568.

Repp BH. 1997. Expressive timing in a Debussy prelude: A comparison of student and expert pianists. Musicae Scientiae 1(2):257-268.

Repp BH. 2000. Compensation for subliminal timing perturbations in perceptual-motor synchronization. Psychological Research 63:106-128.

Repp BH. 2001. Processes underlying adaptation to tempo changes in sensorimotor synchronization. Human Movement Science 20:277-312.

Repp BH. 2002. The embodiment of musical structure: effects of musical context on sensorimotor synchronization with complex timing patterns. In W. Prinz & B. Hommel (Eds), Common mechanisms in perception and action: Attention and Performance XIX (pp. 245-265). Oxford University Press, Oxford, UK.

Repp BH. 2003. Rate limits in sensorimotor synchronization with auditory and visual sequences: The synchronization threshold and the benefits and costs of interval subdivision. Journal of Motor Behavior 35(4):355-370.

Repp BH. 2005. Rate Limits of On-Beat And Off-Beat Tapping With Simple Auditory Rhythms: 2. The Roles of Different Kinds of Accent. Music Perception 23(2):165-187.

Repp BH. 2006. Musical synchronization. In Music, Motor Control and the Brain, E. Altenmüller, M. Wiesendanger, J. Kesselring (Eds.), Oxford University Press, London, UK, Chapter 4, pp. 55-76.

Repp BH. 2007. Hearing a melody in different ways: Multistability of metrical interpretation, reflected in rate limits of sensorimotor synchronization. Cognition 102, 434-454.

Repp BH, Keller PE. 2004. Adaptation to tempo changes in sensorimotor synchronization: Effects of intention, attention, and awareness. The Quarterly Journal of Experimental Psychology 57:499-521.

Repp BH, Penel A. 2002. Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. Journal of Experimental Psychology: Human Perception and Performance 28(5):1085-1099.

Repp BH, Penel A. 2004. Rhythmic movement is attracted more strongly to auditory than to visual rhythms. Psychological Research 68:252-270.

Roelfsema PR, Engel AK, König P, Singer W. 1997. Visuomotor integration is associated with zero time-lag synchronization among cortical areas. Nature 385(6612):157-61.

Rossignol S, Frigon A. 2011. Recovery of locomotion after spinal cord injury: some facts and mechanisms. Ann Rev Neurosci. 34:413-40.

Saupe K, Schröger E, Andersen SK, Müller MM. 2009. Neural mechanisms of intermodal sustained selective attention with concurrently presented auditory and visual stimuli. Front Hum Neurosci 3:58.

Schaal S, Sternad D, Osu R, Kawato M. 2004. Rhythmic arm movement is not discrete. Nat Neurosci. 7(10):1136-43.

Schaefer RS, Vlek RJ, Desain P. 2010. Decomposing rhythm processing: electroencephalography of perceived and self-imposed rhythmic patterns. Psychol Res. DOI 10.1007/s00426-010-0293-4.

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. 198:137-51.

Scherg M, Berg P. 1991. Use of prior knowledge in brain electromagnetic source analysis. Brain Topogr. 4(2):143-50.

Schnupp J, Nelken I, King A. 2010. Auditory neuroscience: Making sense of sound. The MIT Press, Cambridge, MA, USA.

Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A. 2008. Neuronal oscillations and visual amplification of speech. Trends Cogn Sci. 12:106-13.

Schroeder CE, Lakatos P. 2009. Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci. 32(1):9-18.

Schubotz RI, Friederici AD, von Cramon DY. 2000. Time perception and motor timing: a common cortical and subcortical basis revealed by fMRI. Neuroimage 11(1):1-12.

Schwartze M, Keller PE, Patel AD, Kotz SA. 2011. The impact of basal ganglia lesions on sensorimotor synchronization, spontaneous motor tempo, and the detection of tempo changes. Behav Brain Res. 216(2):685-91.

Sekuler R, Sekuler AB, Lau R. 1997. Sound alters visual motion perception. Nature 385:308.

Senkowski D, Talsma D, Grigutsch M, Herrmann CS, Woldorff MG. 2007, Good times for multisensory integration: Effects of the precision of temporal synchrony as revealed by gamma-band oscillations. Neuropsychologia 45:561-71.

Senkowski D, Talsma D, Herrmann CS, Woldorff MG. 2005. Multisensory processing and oscillatory gamma responses: effects of spatial selective attention. Exp Brain Res. 166:411-26.

Shah AS, Bressler SL, Knuth KH, Ding M, Mehta AD, Ulbert I, Schroeder CE. 2004. Neural Dynamics and the Fundamental Mechanisms of Event-related Brain Potentials. Cereb Cortex 14:476-483.

Shams L, Kamitani Y, Shimojo S. 2000. Illusions. What you see is what you hear. Nature 408(6814):788.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. 1995. Speech recognition with primarily temporal cues. Science 270(5234):303-4.

Singer W. 1999. Neuronal synchrony: a versatile code for the definition of relations? Neuron 24(1):49-65.

Sirota A, Montgomery S, Fujisawa S, Isomura Y, Zugaro M, Buzsáki G. 2008. Entrainment of neocortical neurons and gamma oscillations by the hippocampal theta rhythm. Neuron 60:683-97.

Snyder JS, Large EW. 2005. Gamma-band activity reflects the metric structure of rhythmic tone sequences. Cognitive Brain Res. 24:117-126.

Soley G, Hannon EE. 2010. Infants prefer the musical meter of their own culture: a cross-cultural comparison. Dev Psychol. 46(1):286-92.

Spence C, Driver J. (Eds.) 2004. Crossmodal Space and Crossmodal Attention. Oxford University Press, Oxford, UK.

Stein BE, Meredith MA. 1993. Merging of the Senses. MIT Press, Cambridge, MA, USA.

Stein BE, Wallace MT. 1996. Comparisons of cross-modality integration in midbrain and cortex. Prog Brain Res. 112:289-299.

Steriade M. 2006. Grouping of brain rhythms in corticothalamic systems. Neuroscience 137(4):1087-106.

Stern W. 1900. Das Psychisch Tempo. In Uber psychologie der individuellen differezen. Leipzig, Barth, Germany.

Stevens CJ. 2012. Music perception and cognition: a review of recent cross-cultural research. Top Cogn Sci. 4(4):653-67.

Sutoyo D, Srinivasan R. 2009. Nonlinear SSVEP responses are sensitive to the perceptual binding of visual hemifields during conventional 'eye' rivalry and interocular 'percept' rivalry. Brain Res. 28:1251245-55.

Talsma D, Doty TJ, Strowd R, Woldorff MG. 2006. Attentional capacity for processing concurrent stimuli is larger across sensory modalities than within a modality. Psychophysiology 43:541-9.

Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG. 2010. The multifaceted interplay between attention and multisensory integration. Trends Cogn Sci. 14:400-10.

Teki S, Grube M, Kumar S, Griffiths TD. 2011. Distinct neural substrates of duration-based and beat-based auditory timing. J Neurosci. 31:3805-12.

Teki S, Grube M, Griffiths TD. 2011. A unified model of time perception accounts for duration-based and beat-based timing mechanisms. Front Integr Neurosci. 5:90.

Tervaniemi M, Saarinen J, Paavilainen P, Danilova N, Näätänen R. 1994. Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity. Biol Psychol. 38(2-3):157-67.

Thul E, Toussaint G. 2008. Rhythm complexity measures: A comparison of mathematical models of human perception and performance. Proc of the 9th Int Conference on Music Information Retrieval. Philadelphia, Pennsylvania, USA.

Tierney AT, Russo FA, Patel AD. 2011. The motor origins of human and avian song structure. Proc Natl Acad Scie USA. 108:15510-15515.

Todd NP, Cody FW. 2000. Vestibular responses to loud dance music: a physiological basis of the "rock and roll threshold"? J Acoust Soc Am. 107(1):496-500.

Todd NP, Lee CS, O'Boyle DJ. 2002. A sensorimotor theory of temporal tracking and beat induction. Psychol Res. 66(1):26-39.

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. 72:289-98.

Toiviainen P, Luck G, Thompson M. 2010. Embodied meter: hierarchical eigenmodes in music-induced movement. Music Percept. 28:59-70.

Toiviainen P, Luck G, Thompson M. 2009. Embodied metre: hierarchical eigenmodes in spontaneous movement to music. Cogn Process. 10(2):325-7.

Toiviainen P, Snyder JS. 2003. Tapping to Bach: Resonance-Based Modeling of Pulse. Music Perception 21:43–80.

Tononi G, Srinivasan R, Russell DP, Edelman GM. 1998. Investigating neural correlates of conscious perception by frequency-tagged neuromagnetic responses. Proc Natl Acad Sci USA. 95:3198-203.

Trainor LJ, Gao X, Lei JJ, Lehtovaara K, Harris LR. 2009. The primal role of the vestibular system in determining musical rhythm. Cortex 45(1):35-43.

Trehub SE, Thorpe LA. 1989. Infants' perception of rhythm: categorization of auditory sequences by temporal structure. Can J Psychol. 43(2):217-29.

Van der Tweel LH, Spekreijse H. 1969. Signal transport and rectification in the human evoked-response system. Ann NY Acad Sci. 156:678–695.

van Noorden L. 1975. Temporal coherence in the perception of tone sequences. Unpublished doctoral dissertation, Eindhoven University of Technology, The Netherland.

van Noorden L, Moelants D. 1999. Resonance in the perception of musical pulse. J New Music Res. 28:43-66.

Varela FJ, Toro A, John ER, Schwartz EL. 1981. Perceptual framing and cortical alpha rhythm. Neuropsychologia 19:675-86.

Varela FJ, Lachaux JP, Rodriguez E, Martinerie J. 2001. The brainweb: phase synchronization and large-scale integration. Nat Rev Neurosci. 2(4):229-39.

Vatakis A, Spence C. 2006. Audiovisual synchrony perception for music, speech, and object actions. Brain Res. 1111:134-42.

Velasco MJ, Large EW. 2011. Pulse detection in syncopating rhythms using neural oscillators. Proc of the 12th Annual Conference of the Int Soc for Music Information Retrieval. In press.

Velasco MJ, Large EW. 2012. Pulse induction in pseudorandom sequences measured with pulse attribution and tapping rate. Society for Neuroscience Abstracts, 38:910.01.

Vialatte FB, Maurice M, Dauwels J, Cichocki A. 2010. Steady-state visually evoked potentials: focus on essential paradigms and future perspectives. Prog Neurobiol. 90(4):418-38.

Victor JD, Zemon V. 1985. The human visual evoked potential: analysis of components due to elementary and complex aspects of form. Vision Res. 25(12):1829-1842.

Vorberg D, Wing AM. 1995. Modeling variability and dependence in timing. In Heuer H, Keele SW (Eds.) Handbook of perception and action. Academic Press, London, UK.

Vorberg D, Wing AM. 1996. Modelling variability and dependence in timing. In H. Heuer and S.W. Keele (Eds.), Handbook of perception and action: Vol. 3. Motor skills, pp. 181–261, London Academic Press, London, UK.

Vos PG. 1973. Waarneming van metrische toon reeksen. Nijmegen, Stichting Studentenpers, The Netherland.

Vroomen J, Keetels M. 2010. Perception of intersensory synchrony: a tutorial review. Atten Percept Psychophys. 72:871-84.

Vuust P, Ostergaard L, Pallesen KJ, Bailey C, Roepstorff A. 2009. Predictive coding of music--brain responses to rhythmic incongruity. Cortex 45(1):80-92.

Vuust P, Roepstorff A, Wallentin M, Mouridsen K, Østergaard L. 2006. It don't mean a thing... Keeping the rhythm during polyrhythmic tension, activates language areas (BA47). Neuroimage 31(2):832-41.

Wallin JEW. 1911. Experimental Studies of Rhythm and Time. Psychological Review 18:100-119.

Wallin NL, Merker B, Brown S. (Eds.) 2000. The origins of music. MIT Press, Cambridge, MA, USA.

Wang Y, Ding N, Ahmar N, Xiang J, Poeppel D, Simon JZ. 2012. Sensitivity to temporal modulation rate and spectral bandwidth in the human auditory system: MEG evidence. J Neurophysiol. 107(8):2033-41.

Wang XJ, Buzsáki G. 1996. Gamma oscillation by synaptic inhibition in a hippocampal interneuronal network model. J Neurosci. 16(20):6402-13.

Warren JE, Wise RJ, Warren JD. 2005. Sounds do-able: auditory-motor transformations and the posterior temporal plane. Trends Neurosci. 28(12):636-43.

Welch RB, Warren DH. 1980. Immediate perceptual response to intersensory discrepancy. Psychol Bull. 88:638-67.

White O, Bleyenheuft Y, Ronsse R, Smith AM, Thonnard JL, Lefèvre P. 2008. Altered gravity highlights central pattern generator mechanisms. J Neurophysiol. 100(5):2819-24.

Wile D, Balaban E. 2007. An Auditory Neural Correlate Suggests a Mechanism Underlying Holistic Pitch Perception. PLoS ONE 2(4): e369.

Williams PE, Mechler F, Gordon J, Shapley R, Hawken MJ. 2004. Entrainment to video displays in primary visual cortex of macaque and humans. J Neurosci. 24:8278-88.

Wilson SJ, Pressing JL, Wales RJ. 2002. Modelling rhythmic function in a musician post-stroke. Neuropsychologia 40(8):1494-505.

Wing AM, Kristofferson AB. 1973b. The timing of interresponse intervals. Perception & Psychophysics 13:455–460.

Winkler I, Háden GP, Ladinig O, Sziller I, Honing H. 2009. Newborn infants detect the beat in music. Proc Natl Acad Sci USA. 106(7):2468-71.

Witek M. 2012. Groove Experience: Emotional and Physiological Responses to Groove-Based Music. Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009), Jyväskylä, Finland.

Witek M, Clarke E, Wallentin M, Hans M, Kringelbach L, Vuust P. 2012 "…and I feel good!" Ratings, fmri-recordings and motion-capture measurements of body movements and plesure in response to groove. Proceedings of the 12th International Conference on Music Perception and Cognition (ICMPC) / 8th Triennial Conference of the European Society for the Cognitive Sciences of Music (ESCOM), Thessaloniki, Greece.

Witt ST, Laird AR, Meyerand ME. 2008. Functional neuroimaging correlates of finger-tapping task variations: an ALE meta-analysis. Neuroimage 42(1):343-56.

Woldorff MG, Gallen CC, Hampson SA, Hillyard SA, Pantev C, Sobel D, Bloom FE. 1993. Modulation of early sensory processing in human auditory cortex during auditory selective attention. Proc Natl Acad Sci USA. 90(18):8722-6.

Wolpert DM, Miall RC, Kawato M. 1998. Internal models in the cerebellum. Trends Cogn Sci. 2(9):338-47.

Woodrow H. 1932. The effect of rate of sequence upon the accuracy of synchronization. J Exp Psychol. 15:357-379.

Zampini M, Shore DI, Spence C. 2003. Audiovisual temporal order judgments. Exp Brain Res. 152:198-210.

Zatorre RJ, Chen JL, Penhune VB. 2007. When the brain plays music: auditory-motor interactions in music perception and production. Nat Rev Neurosci. 8(7):547-58.

Zbikowski L. 2004. Modelling the Groove: Conceptual Structure and Popular Music. Journal of the Royal Musical Association 129(2):272–97.

Zemon V, Ratliff F. 1984. Intermodulation components of the visual evoked potential: Responses to lateral and superimposed stimuli. Biol Cybern. 50:401–408.

Zentner M, Eerola T. 2010. Rhythmic engagement with music in infancy. Proc Natl Acad Sci USA. 107(13):5768-73.

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.

Zion Golumbic EM, Poeppel D, Schroeder CE. 2012. Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective. Brain Lang. In press.