SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Department of Psychology

Ph.D. Program in Psychology, Linguistics and Cognitive Neuroscience
Curriculum The Mind, Brain, and Behaviour
Cycle XXXV

The musical experience across perception and emotions: from music entrainment to multi-sensory integration

Doctoral Thesis by:
Francesco De Benedetto
854312

Tutor: Prof. Alice Mado Proverbio
Coordinator: Prof. Simona Sacchi

Academic Year 2021/2022

If death is pretty final

I'm collecting vinyl

I'm gonna DJ at the end of the world

The music could provide the light

you cannot resist

I'm gonna DJ – R.E.M.

Table of Contents

ABSTRACT **IV**

RIASSUNTO **VI**

GENERAL INTRODUCTION **8**

Aims and Hypothesis **10**

Entrainment to Music **12**

Emotions in Music **14**

Emotion Interactions in Music and visual Art 16

Semantics of Music **18**

Semantic Interactions in Music and visual Art 19

Neural Underpinnings of Art Perception **21**

TAGGING THE BEAT OF NATURALISTIC MUSIC IN THE BRAIN **23**

Introduction **24**

Materials and Methods **26**

Participants 26

Auditory Stimuli 27

Behavioural Data 28

M/EEG pre-processing 28

Auditory Steady State Response M/EEG Analysis for naturalistic Music 28

Inter-Subject Coherence Analysis for naturalistic Music 31

Statistical Analysis 32

Results **33**

ASSR and ISC for regular compared to irregular Sound Onsets 33

Correlations with behavioural Measures 36

Discussion **38**
ASSR and ISC for regular and irregular Sound Onsets 38
Perceptual Entrainment to the musical Beat 39
Entrainment and Emotion 40

EMOTION COMPREHENSION IN MUSIC AND VISUAL ART **41**

Introduction **42**

Materials and methods **44**
Participants 44
Stimuli 44
Stimuli Validation 45
Procedure 46
EEG recording and analysis 47

Results **49**
ERPs to artworks 49
ERPs to music pieces 51
ERPs to audio-visual stimuli 52
Source Reconstruction - swLORETA 53

Discussion **56**
ERPs to Artworks 56
ERPs to Music 58
Multimodal Integration 58
Source Reconstruction 59

SEMANTIC INTERACTION BETWEEN MUSIC AND ARTWORKS **61**

Introduction **62**

Materials and Methods **64**
Participants 64
Stimuli 64
Stimuli Validation 66
Procedure 69

EEG recording and analysis 70

Results 71

N2 response (250-350 msec) 72

N400 response (300-500 msec) 73

Source Reconstruction – swLORETA 74

Discussion 76

N2 and conflicting incoming Information 76

N400 and semantic Incongruence 77

Source Reconstruction 78

GENERAL CONCLUSIONS **79**

REFERENCES **87**

Abstract

Music is a complex stimulus that can induce different reactions and influence both our body and mind. This work aimed to shed light on three different aspects of musical experience, going from neural to emotional, and from emotional to cognitive.

When listening to a musical piece, one thing we may notice is the impulse to move to the beat. Entrainment theory describes the process of interaction between independent rhythmical processes. The perception of a beat in a complex auditory signal was found to be related to the emergence of a sustained periodic response in the EEG spectrum. With the first study, we propose a new method for analysing Auditory Steady State Response (ASSR) in naturalistic music, aiming at investigating neural entrainment and its relationship with emotions. We compared musical stimuli with regular and irregular timed sound onset and applied beat tagging analysis on an M/EEG dataset. To access the beat, we grouped sound onsets by two or four. Behavioural evaluations of musical pieces were correlated with the ASSR and inter-subject coherence (ISC) values, to investigate their association. For the regular, but not the irregular onset piece, both EEG ASSR and ISC peaked at the same frequency as the sound onset frequency. When grouping sound onsets, EEG ASSR showed frequency peaks as the sound onset frequency divided by two and by four. Correlating the neural data and the behavioural measures we found little evidence of a close link between the two. Therefore, it can be said that our behavioural results go toward a more complex emotion induction mechanism, in which entrainment is one of the many processes for evoking emotions.

The following step investigated how the emotional content of music affects the perception of emotions experienced during artwork perception. We hypothesized that emotionally incongruent music would negatively affect the comprehension of the emotional content of artworks. ERP data showed that N170 response to visual art was influenced not only by sensory and attentional processes but also by stimulus emotional valence. Both visual Late Positivity (LP) and auditory N400 were modulated in amplitude by the emotional content of the musical stimulus. The multimodal Late Positivity (LP) was modulated by the stimulus emotional congruence, with larger responses to positive than negative congruent pairs, and to congruent than incongruent pairs. This may suggest a role of LP in indexing the degree of aesthetic pleasantness of the stimulation. An swLORETA source reconstruction applied to multimodal LP potentials showed that left MFG, left SFG, right precuneus, and right MTG were the most active brain areas during multimodal artistic simulation. Therefore, these areas possibly underpin the arising of aesthetic sensations, leading to the conclusion that auditory and visual emotional processing may share the same neural mechanisms.

As a final step, we investigated both the capability of music and paintings to convey specific semantic meaning and the interaction across the two sensory modalities. Previous literature have shown that instrumental music can communicate iconic, indexical, and symbolic meanings, similar to language. This study aimed to investigate the effect of background music, semantically congruent or incongruent, on the semantic understanding of artworks. The hypothesis is that listening to incongruent background music would interfere with the semantic recognition of the visual stimulus. The data seem to suggest that both paintings and music can convey specific semantic meanings. Moreover, it was observed how interference in the semantic domain also occurred across sensory modalities, as shown in the modulation of the N400. Finally, these data strengthen the hypothesis of a bidirectional modulation between visual and auditory channels and that semantic interference is a general cognitive mechanism.

Riassunto

La musica è uno stimolo complesso che può influenzare sia il corpo che la mente.

Questo lavoro si pone l'obiettivo di indagare tre aspetti dell'esperienza musicale, andando dal neurale all'emotivo e dall'emotivo al cognitivo. Quando ascoltiamo un brano musicale una delle cose che notiamo è la voglia di muoverci a ritmo. Le teorie sull'entrainment lo descrivono come l'interazione tra processi ritmici indipendenti. La percezione del ritmo all'interno di uno stimolo acustico è associata all'emergere di una risposta periodica nel tracciato EEG. Nel primo studio proponiamo un nuovo metodo per l'analisi delle Auditory Steady State Response (ASSR, risposte uditive evocate) in uno stimolo musicale, allo scopo di studiare l'entrainment neurale e come questo si colleghi alle emozioni. Abbiamo confrontato un brano musicale con una successione di suoni regolare ed uno con una successione irregolare, e applicato delle analisi di beat tagging su un dataset M/EEG. Per rilevare il beat abbiamo raggruppato i suoni in coppie e in quattro. Abbiamo quindi raffrontato le valutazioni comportamentali con i valori di ASSR e Inter-Subject Coherence (ISC, coerenza tra soggetti), per studiarne la relazione. I risultati hanno mostrato che per il brano con successioni di suoni regolari, ma non per quello con successioni irregolari, l'ASSR e l'ISC mostravano picchi nelle stesse frequenze di comparsa dei suoni. Allo stesso modo, quando raggruppati, l'ASSR mostravano picchi corrispondenti alla metà e al quarto delle frequenze di comparsa dei suoni. Le correlazioni tra i dati neurali e quelli comportamentali ci hanno dato pochi risultati circa la loro relazione. Per questo motivo crediamo che la musica induca emozioni attraverso meccanismi complessi, dove l'entrainment è solo uno dei fattori in gioco.

Successivamente è stato indagato come il contenuto emotivo della musica sia influenzato dalla percezione delle emozioni suscitate dai quadri. Abbiamo ipotizzato che emozioni incongruenti tra i due stimoli avrebbero inficiato la comprensione emotiva del quadro. I dati ERP dimostrano che la N170 in risposta ai quadri è influenzata non solo da processi attentivi, ma anche dalla valenza emotiva del quadro. Per quanto riguarda la musica, sia la N400 che la late positivity (LP) sono modulate dal contenuto emotivo dello stimolo. La LP multimodale invece è modulata dalla congruenza emotiva tra gli stimoli, con risposte più ampie per stimoli positivi rispetto ai negativi. Questo suggerisce che la LP possa indicare il grado di piacevolezza estetica dello stimolo. La soluzione inversa swLORETA applicata alla LP multimodale, infine, ha evidenziato l'attivazione dei giri frontale medio e superiore di sinistra, del precuneo e del giro temporale medio di destra. Di conseguenza, queste aree potrebbero indicare la comparsa di una percezione estetica, portandoci alla conclusione che i processi emotivi visivo e uditivo condividano gli stessi substrati neurali.

In conclusione, abbiamo indagato la capacità di musica e quadri di convogliare significati semantici specifici e l'interazione tra questi due tipi di stimoli. La letteratura ha mostrato che, così come il linguaggio, la musica è in grado di comunicare significati simbolici. Questo studio ha lo scopo di investigare l'effetto di una musica di sottofondo, semanticamente congruente o incongruente, sull'elaborazione semantica di un quadro. L'ipotesi è che ascoltare una musica incongruente interferisca con la comprensione semantica dello stimolo visivo. I risultati indicano come sia musica che quadri siano in grado di comunicare contenuti semantici specifici. Inoltre, così come accade per il linguaggio, la discrepanza tra i contenuti emotivi convogliati dalle diverse modalità sensoriali modula la componente N400. In ultimo, questi risultati rafforzano l'ipotesi di una influenza reciproca tra i canali uditivo e visivo e dell'interferenza semantica come meccanismo cognitivo che avviene anche tra diverse modalità sensoriali.



General Introduction

The Oxford Dictionary defines music as "The art or science of combining vocal or instrumental sounds to produce beauty of form, harmony, melody, rhythm, and expressive content". From the physics point of view, it is just air pressure transformed into neural reactions. Then, these responses make the body reacts, from autonomic (see Mojtabavi et al., 2020 for a review) to emotional changes (Cook et al., 2019). Indeed, more and more studies have demonstrated the effect of music on physiological variables such as blood pressure and heart rate, respiration and body temperature (D. S. Ellis & Brighouse, 1952; Gomez & Danuser, 2004; Koelsch & Jancke, 2015; Lorch et al., 1994; A. M. Proverbio, Manfrin, et al., 2015). Noguchi and colleagues (2012) presented participants with two classical music pieces and silence to investigate whether the emotion induced by the music influenced respiratory rate and other autonomic responses in participants. Volunteers reported uncomfortable feelings while listening to Stockhausen's music, compared to Chopin and silence, and their respiration rate increased. Also, music has been proposed as a therapeutical intervention (Mojtabavi et al., 2020) due to its beneficial effect on cognitive, psychosocial, behavioural and motor functions (Brancatisano et al., 2020; de Witte et al., 2020). Several studies, in fact, focused on the long-term effects of music training on cognitive functions (Benz et al., 2016) and on its impact on development (Miendlarzewska & Trost, 2014): from the claims of the Mozart effect (Thompson et al., 2001) to music therapy in the context of cognitive rehabilitation (Thaut, 2010). Last but not least, music is widely used as an emotion regulation strategy (Cook et al., 2019). As a matter of fact, music is also used as a treatment for mood disorders (Raglio, 2015; Sakka & Juslin, 2018b, 2018a). Ray and Götell (2018) investigated the influence of music intervention, as delivered by trained assistants, on depression symptoms and well-being in individuals with dementia. The results showed how music therapy was effective in decreasing depression symptoms. Moreover, musical activities such as singing and music-with-movement helped stabilize the symptoms and increased well-being.

Thus, we can say that music is listened to in various contexts. In our everyday life, we listen to music for example to celebrate ceremonies, for entertainment, to relax, or used just as background sound (Gonzalez et al., 2021; Koelsch, 2011a, 2014a). The psychological effects of background music have been investigated in various situations and for different purposes, such as retail stores (Knoeferle et al., 2017), theatre and film shows (Fröber & Thomaschke, 2019). In their study, Knoeferle and colleagues investigated the choice of music in retail stores. They started hypothesizing that fast, i.e. more arousing music, might alleviate the effect of high social density. They confirmed the hypothesis, adding as a result that social density also has an impact on customers' spending (Knoeferle et al., 2017). Hence, both listening to and playing music engage the brain in different

ways. This richness makes music an ideal tool for investigating the workings of the human brain (Koelsch & Siebel, 2005).

The interest in the emotions aroused by music arose in ancient Greece (Budd, 1985). But did music develop as an evolutionary adaptation, or is it a cultural creation? In recent years, the concept of evolution of musicality has re-emerged as the ability to create and experience music (Savage, 2019). Bharucha and collaborators (2006) theorize a model of music experience where the processing of acoustic and musical structure elicits the processing in domains that do not relate to acoustic or musical structure, namely affect and motion. In other words, part of the musical structure and its cognitive representations operate to elicit experiences of affect and motion. In fact, it has been hinted that the emotions evoked by music are the main reason one chooses to engage in musical experience (Juslin & Laukka, 2004).

Aims and Hypothesis

The main point of this thesis is to understand how the brain engages into music. This topic is approached from three viewpoints, ranging from neural to emotional, and from emotional to cognitive.

The first step of this work, then, was to develop a method to tag the sound onset periodicities (i.e., the frequency of the sound onsets) and the neural entrainment to a naturalistic musical piece in the brain (Study 1). Moving forward, we investigated how emotions evoked by music again integrate, or maybe interfere, with visually evoked emotions (Study 2). Lastly, we assessed the capability of music to convey semantic meanings and how this information integrates with semantic information coming from a different sensory channel (Study 3).

When listening to music, we may find ourselves tapping or moving our feet to the beat of the music. Beat perception is fundamental to experience music. Music contains temporal patterns, and beat perception is the ability to perceive those temporal patterns. The Neural Resonance Theory (NRT) provides a model explaining how an external pulse interacts with brain rhythms inducing its internal representation, resulting in the coupling between the two (Large, 2008). However, to date, most studies on perceptual and rhythmic entrainment to music used pure tones or simplified acoustic stimuli. Indeed, Nozaradan and colleagues (2011) showed how a sustained periodic response in the EEG spectrum emerges from the perception of a beat, that is a steady-state evoked potential (SS-EP) tuned to the beat frequency. In this study, participants were also asked to imagine either a binary or a ternary meter while listening to the tone. The authors found that the imagined meter induced an

additional periodic signal in the EEG. More recently, Tal and collaborators (2017) came to the same conclusion recording MEG while participants were listening to acoustic rhythms. As a result, they found an enhanced neural response at the beat frequency. The study aimed to investigate neural entrainment to a naturalistic musical piece. A novel methodology to investigate neural entrainment was implemented. Also, it has been proposed that entrainment might represent an emotion induction mechanism (Juslin et al., 2010). Since, unlike previous studies, we investigated entrainment with a naturalistic musical stimulus, we would provide more insights into the association between entrainment and emotions.

With the second study, we will focus on emotion comprehension and integration. The study aimed at investigating how the emotional content of music affected the perception of emotions experienced during artwork perception. Previous studies have shown how both music (Brattico, 2019; Eerola & Vuoskoski, 2013; Juslin et al., 2014; Proverbio et al., 2020a; Proverbio et al., 2020b; Trost et al., 2012) and paintings (Gao & Xie, 2018; Melcher & Bacci, 2013; van Paasschen et al., 2015) can convey emotions. To this end, we set up our study expecting the N400 ERP component to be modulated by the emotional meaning of the musical stimulus, as found comparing different emotional valences in language (Proverbio & Piotti, 2021). According to Baumgartner (2006), perceived emotions are enhanced when congruently combined in visual and auditory sensory channels. We also paired music and paintings either congruently or incongruently concerning the evoked emotion. The hypothesis was that listening to a musical piece carrying incongruent emotional content as the emotion induced by a painting, would have interfered with the emotional comprehension of the stimulus as a whole.

As a final step, the aim was to investigate the capability of music and paintings to convey specific semantic meaning and the interaction across the two sensory modalities. The literature has widely studied the relationship between music and language (Besson et al., 2011; Besson & Schön, 2001; Jäncke, 2012; Kraus & Slater, 2015; Slevc, 2012). There is also evidence about how the brain processes music (Brattico et al., 2020; Gonzalez et al., 2021; Völker, 2021; Vuust et al., 2022) and paintings (Pang et al., 2013; Pelowski et al., 2017; Specker et al., 2017; Vartanian & Skov, 2014), separately. Although, less is known about how these two kinds of stimuli interfere and integrate with each other. Hence, the aim of the third study was first to corroborate the capability of both music and paintings to convey specific semantic meaning, and second to investigate the interaction between them. Wöhner and colleagues (2020) tested Chen and Spence's model of cross-modal semantic interactions (2018) and concluded that semantic interference also appears across sensory modalities. In this regard, we expected the N400 to be modulated by the semantic relatedness of the two stimuli. Therefore, here the hypothesis was that listening to a musical piece whose semantic content

was incongruent with respect to the semantic domain of a painting, would have affected the semantic recognition of the visual stimulus.

The project moves from the neural to the cognitive point of view: from entrainment to emotions, and from emotions to semantics.

Entrainment to Music

Two rhythmic processes can interact with one another in a process known as entrainment, whereby they eventually lock into a common phase and/or periodicity.

Music is something that develops through time, and time perception can be influenced by music listening (Noulhiane et al., 2007). A crucial feature of music is rhythm, namely sound periodicity (or sound patterns in time), and its perceptual correlate, namely beat. It has been observed that Beat Perception and Synchronization (BPS) is also present in nonhuman animals (Bregman et al., 2012; Large & Gray, 2015; Patel et al., 2008). Findings from Patel and co-workers are consistent with the "vocal learning and rhythmic synchronization" hypothesis, which proposes that vocal learning is the only requirement for BPS to occur (Patel et al., 2008). Entrainment, however, is more than just synchronization. As defined by Trost and colleagues (2017), entrainment is the process through which two physical or biological systems become synchronized by interacting with each other. In other words, when perceiving a regular rhythmic activity, the phase of internal oscillation converges to that of the external driving force, and the endogenous oscillations become phase-locked to the external rhythm (Deng et al., 2020). Variations in rhythm constitute an immediate way to induce emotional reactions, with faster beats causing physiological arousal and slower ones producing relaxation (Bhatti et al., 2016; Cervellin & Lippi, 2011; Koelsch & Jancke, 2015). Trost and colleagues (2017) have pointed out that there are different levels of entrainment: perceptual, physiological, motor, and social. In their view, these different levels compose a single process, where one level affects the other.

The first level of entrainment is perceptual and consists of integrating auditory information into periodic information, such as, in this case, the musical beat. This process requires an integration of sensory inputs and top-down feedbacks, precisely expectations, to form a cognitive representation of the auditory signal. In perceptual entrainment, temporal expectancies play an important role in orienting attention along the rhythmical structure and so along the musical beat. The Dynamic Attending Theory (DAT; Jones et al., 1989) proposes that only when attentional processes are

entrained to the various periodicities of the musical meter, the periodicities in a musical rhythm can be perceived.

The second level is the autonomic physiological entrainment, which is the inclination for biological rhythms to synchronize with rhythms perceived outside the body. An excellent example of physiological entrainment comes from the use of music in sports, where music is associated with significant beneficial effects on physical performance (see Terry et al., 2020 for a review). Despite several studies showed the effects of musical tempi on heart rate, respiration rate, and blood pressure (Ellis & Thayer, 2010; Proverbio et al., 2015), there is no general consensus about this interaction (see Cervellin & Lippi, 2011 and Koelsch & Jancke, 2015 for a review). Trost and colleagues (2017) proposed that autonomic entrainment influences the peripheral physiological component of the affective response to music, serving as an emotion induction mechanism and consequently as the link between entrainment and emotion.

Moving forward, motor entrainment is something that probably we all have experienced, that is the synchronization of movements with musical rhythm. On the one hand, there is the concept of rhythm, which is the perception of the varying pattern present in music (Honing, 2012); on the other hand, there is the cognitive concept of meter, which is the ability to perceive periodicities that does not necessarily coincide with a sound, as in the case of syncopated rhythms (Nozaradan et al., 2012), where temporal expectation plays a fundamental role. Unlike the Dynamic Attending Theory, Predictive Coding (PC) provides a framework for explaining the process of expectations and predictions, where input (namely rhythm) and predictions (namely meter) are intertwined with each other in a continuous comparison between them (Koelsch et al., 2019). Trost and Vuilleumier suggested that the link between rhythm perception and emotional experience is the involvement of several motor system structures in both contexts (Trost & Vuilleumier, 2013). The more in synchrony with the music, the more pleasant the affective state (Janata et al., 2012).

Finally, social entrainment arises when motor actions of two or more people become coordinated in time (for a review, see Keller et al., 2014; Molnar-Szakacs & Overy, 2006; Overy, 2012; Overy & Molnar-Szakacs, 2009). As proposed by the SAME (Shared Affective Motion Experience) model, when playing music in a social context, music is seen as a sequence of internal and expressive motor actions that synchronize performers and listeners, ending in a sense of empathy and social connection (Overy, 2012). As a result, an additional source of enjoyment for music listening might come from participating in social interactions with either another musicians or listeners (Trost et al., 2017).

Overall, the key brain regions involved in different levels of entrainment indeed mirror the distinction between levels. In other words, brain regions involved depend on the level of the

entrainment process (Trost et al., 2017). As for perceptual entrainment, brain networks implicated in time perception are the same as in rhythm perception (Geiser et al., 2014). Moreover, both rhythm perception and production seem to share similar neural correlates (Chen et al., 2008a, 2008b). When referring to the physiological entrainment, we can mention the Central Autonomic Network proposed by Thayer and Lane (2000), a model of neurovisceral integration in emotion regulation that combines visceral, attentional, and affective information in service of a goal-directed and motivated behaviour. The overlap of regions involved in physiological processes and emotion processing supports the interconnection of these processes (Trost et al., 2017). Finally, the literature provides evidence regarding, again, the intersection of the activation of cortical areas involved in social interactions and brain processes involved in emotion processing (Ménoret et al., 2014; Yun et al., 2012).

To conclude, there is no consensus about the relationship between these different kinds of entrainment or the direction of emotion induction. Despite this, entrainment might play a role in clinical interventions due to the involvement of networks of brain regions implicated both in entrainment processes and in emotional experience. Most notably, presenting rhythmic stimuli has proven to be beneficial to patients with movement disorders but also to healthy elderly people (see Dalla Bella, 2020 for a review). Using music and rhythm as a rehabilitation tool has also proven effective with patients with Parkinson's Disease (PD) (Pacchetti et al., 2000; Stupacher et al., 2022) or individuals with Autism Spectrum Disorders (ASD) (Bharathi et al., 2019). Moreover, music therapy interventions engage patients as much in listening as in movement, resulting in an entrainment process (Trost et al., 2017).

Emotions in Music

When referring to emotions, the first challenge is to define the construct of emotion. According to Scherer (2005), an emotion is "an episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism". In describing emotion, Scherer also distinguishes utilitarian emotions from aesthetic emotions, where the former facilitates our adaptation to an event, and the latter are characterized by the absence of functionality. On the other hand, the aesthetic experience is described as a complex phenomenon involving preferences, emotions, appreciation, and evaluative judgments induced by an object. Referring specifically to music, Brattico and Pearce (2013) define the aesthetic experience of music as "one in which the

individual immerses him- or herself in the music, dedicating his or her attention to perceptual, cognitive and affective interpretation based on the formal properties of the perceptual experience". Hence, this experience includes both cognitive and affective processes that are temporally and spatially defined, namely associated with specific neural networks. Researchers have proposed a model in which the components of aesthetic experience are dynamic modules, evolving over time as a result of distinct neural mechanisms. In their work, Bharucha and collaborators (2006) state that emotions that arise from listening to music are often the reason why we choose to listen to music.

When investigating emotions in music, most studies in the literature utilize a dimensional model for emotion (Russel, 1980), which represents emotions as a combination of valence and arousal dimensions. Unlike a discrete emotion model, which states that all emotions can be derived from a set of finite basic emotions, the dimensional model is probably more used and useful because emotions categories have been modified in the context of music (Eerola & Vuoskoski, 2013). In fact, emotions elicited by music tend to be more positive than everyday-life emotions (Juslin & Laukka, 2004; Zentner et al., 2008).

Recently, Putkinen and colleagues (2021) examined the functional neural basis of music-induced emotions with also the aim of discriminating between neural circuits involved in emotions evoked by music and in emotions elicited by biological events. The authors examined the participants' brain activity with fMRI while they were listening to music pieces evoking four different kinds of emotions (happiness, sadness, fear, and tenderness). They found that auditory and motor cortices exhibit distinct brain representations of music-induced emotions. Specifically, these music-induced emotions activated a network including the auditory cortices, motor regions (SMA, precentral and postcentral gyri, cerebellum), cingulate cortex, insula, and precuneus. Additionally, results show a weak activation of limbic regions, amygdala, prefrontal and orbitofrontal areas involved in non-musical emotion processing (Putkinen et al., 2021). These results partly go against what Koelsch (2014b) states in his review: music affects the activity in core emotion networks, including limbic and para-limbic regions. In his work, Koelsch also describes tension, namely musical expectancy, as a critical feature in eliciting emotions from a musical piece (Koelsch, 2014b).

The role of music expectancy is indeed a contributing factor when considering emotions elicited by music. Krumhansl and Agres (2008) note that the constant changing in expectation during music listening leads to alternating tension to a relaxing state. According to Green and colleagues (2012), exposure positively affects music liking so that previously unfamiliar music is more appreciated after repeated listening. This effect can be explained by the existence of a core tension between knowing and liking. The right amount of familiarity and uncertainty is associated with the possibility of generating positive prediction errors (Salimpoor et al., 2015), namely, better than

anticipated. Steinbeis and co-workers (2006) investigated the effect of harmonic expectancy violations on emotions in musicians and non-musicians. The authors found both physiological response and the ERP component Early Negativity (EN) increased their amplitude with harmonic irregularity. Hence, they suggested that the recognition of an irregularity in music, and therefore the perception of tension, would result in an emotional response. Since no differences between musicians and non-musicians were found, the authors concluded that the emotional experience in response to music was not modulated by expertise. The same concept, meaning music expectancy, can be read in light of the Predictive Coding theory (PC; Friston, 2005), by which the brain compares the sensory input to knowledge gained through experience to predict the causes and sources of its internal states. So, the balance, or the tension, between predictions and upcoming events is a source of emotional responses.

Emotion Interactions in Music and visual Art

Music has often been related to other sensory domains. Several studies have used background music to investigate the modulation of concurrent mental processes. For instance, Baumgartner and colleagues (2006) found that music was able to enhance the emotional experience evoked by visually presented affective pictures, when congruently combined. In their fMRI study, the authors showed participants fearful and sad pictures, either alone or combined with congruent emotional musical pieces. In the combined condition, the authors reported a more intense emotional experience. Results showed increased activation in most brain areas involved in emotion perception, such as the amygdala, medial temporal lobe, ventral frontal cortex, left insula, brainstem nuclei, precuneus, hippocampus and parahippocampus. Although in this study the authors only examined the negative emotional condition, the presentations of emotionally congruent visual and musical stimuli evoke strong emotional feelings and experiences. Also, the authors conclude that pictures presentation alone evoked a more cognitive emotional perception process, while the combined condition elicit a rather automatic experience (Baumgartner et al., 2006).

Also, Timmers and Crook (2014) associated vision with music through priming pictures that elicited happy, sad, or neutral emotions. They hypothesized that a primed emotion would have activated expectations for upcoming musical events, and so functioned as a prime for upcoming musical events. By pairing a musical melody with a priming picture, the authors expected to have two sources of musical expectation: on the one hand, the melody itself and the cross-modal prime on the other hand. The conditions were a happy or a sad coupling between the primed emotion and the upcoming melody. The expectation for the melody was hypothesized to be consistent with the

primed emotion. As a result, Timmers and Crook (2014) found that the emotional content of the visual images was able to influence musical expectations, contributing to and facilitating perception.

Similarly, Logeswaran and Bhattacharya (2009) investigated whether music-elicited emotions might influence subsequent vision-elicited emotional processing. In this study, facial expressions were primed by the previous presentation of either happy or sad music. Behavioural results indicated that when the facial expression was primed by congruent music, the processing of related facial expressions was enhanced. These cross-modal priming effects occurred within 100 msec post-stimulus presentation. The N1 component of ERPs was found to be enhanced for happy, compared to sad, musical priming, suggesting an early cross-modal priming. Again, in a study by Jomori and co-workers (2013), visual evoked magnetic fields (VEF) were recorded while participants were listening to emotional music and watching either a face or a landscape picture. The results indicated a larger M170 related to the negative music condition, with the activation of the inferior temporal (IT) area. As the visual processing in the IT area was modulated by emotional music stimulation, the authors inferred that this area could be considered one of the areas that received top-down and feed-forward regulation.

The literature does not provide much evidence about the interactions between emotions evoked by two artistic stimuli, such as music and artworks. Parrott (1982) investigated the combination of music and paintings on emotional judgments, concluding that the emotional effect of music and painting was additive rather than interactive. The same deduction was made by Limbert and Polzella (1998). They paired music and paintings by style, asked participants to rate their aesthetic experience, and found that listening to matching music while viewing the paintings intensified the aesthetic experience. Marin and Leder (2013) explored how the complexity of a stimulus affects its aesthetic experience, both in paintings and in music, although separately. The authors found a positive association between subjective complexity and arousal, but not between subjective complexity and pleasantness. In the case of artworks, a gender difference was also found: the direction of the association between complexity and pleasantness differed for males and females, being positive for males and negative for females (Marin & Leder, 2013). However, in the case of music, some of the findings may be due to the kind of manipulation the authors applied to control for complexity. To add complexity to the excerpts, the authors summed extra melodic parts, which can lead to greater expressiveness rather than more complexity.

To best match music and paintings, Lee and colleagues (2016) developed an algorithm that extracted the emotions from music and pictures based on their features. As for the paintings, they used a colour-based method to extract emotional adjectives from the image, whereas for the music

they considered acoustic features as dynamic, harmony, or rhythm. As a matter of fact, pairing, congruently or not, visual and musical stimuli is a way to access their mutual influence.

Campos-Bueno and collaborators (2015) set up a study to investigate the reciprocal emotional interaction between artistic stimuli. First, the authors accessed the affective experience of music and paintings separately, on valence and arousal levels. In a second experiment, they investigated the emotional inter-relationship between the two kinds of stimuli. Results from the first experiment show that musical stimuli have higher variability in arousal and valence rating compared to artworks. On the arousal level, low-arousal paintings increase their arousability when paired with positive or negative music. Contrariwise, when low-arousal music interacts with positive or negative painting, its arousability decreases. On the valence level, it is possible that results from the first study biases the authors' conclusion, who assert a leading role for music over visual stimuli. That is, both unpleasant and pleasant music carries the valence of paintings towards the valence of the music itself (Campos-Bueno et al., 2015). This goes against what Yuval-Greenberg and Deouell (2009) found: a visual dominance over the acoustic stimulation, albeit with a different set of stimuli.

In a recent study, Braun Janzen and co-workers (2023) examined the influence of musical emotions over the aesthetic experience, i.e. judgments of emotions and aesthetic value, of a visual artwork in a naturalistic environment. Specifically, visitors of a contemporary art museum experienced an abstract painting by Kandinsky while listening to background music evoking different emotions, or silence. The authors found no effects of background music on beauty, liking, or arousal ratings. Contrariwise, background music influenced participants' judgments of the artwork's valence to the extent that participants who listened to positive-valenced music rated the painting as more pleasant than those who listened to negative-valenced music (Braun Janzen et al., 2023).

Semantics of Music

It is important to highlight how we can refer to semantic processing not only in relation to written or auditory/verbal language but also to other systems of communication, such as music. The human mind constantly and automatically ascribes meaning to acoustic musical information during listening. For this reason, the musical expression has been recognized as a "universal language" to emphasize how the properties that define it turn out to be universally accessible to individuals (Swain & Swain, 1997). Together with language, music is a complex communication system. Probably the main feature that music lacks, with respect to language, is ''semanticity''. Language can be used to convey an unlimited set of discrete, propositional meanings, whereas music cannot (Fitch, 2005).

The processing of meaning plays a fundamental role in linguistics; nonetheless, language is not the only possible mean to convey meaningful information.

In his work, Limb (2006) pointed out the commonalities between music and language, such as the hierarchical structure (syntax/harmony), vocabulary (words/chords and intervals), tonal properties (inflection/timbre), and a temporal clock (prosody/rhythm). However, an essential difference between the musical language and the traditional language can be recognized in the fact that the meaning evoked by the former is less defined than the meaning derived from the latter (Antovic, 2009). This may lead us toward viewing music as a form of expression rather than a form of communication (Slevc & Patel, 2011).

But can we refer to semantics when it comes to music? A large amount of research has focused on the overlapping processes between music and language (see Patel, 2003 for a review). Kraus and Slater (2015) have highlighted the connections and differences between music and language from one another. In their work, the authors emphasised how the integration between the auditory system and other cognitive and sensory systems is the key to make sense of sound in a communication (Kraus & Slater, 2015). Schlenker (2017) states that music also has a semantical structure, although one which is based on different principles from linguistic semantics: in contrast to language, music semantics typically communicates more abstract information, and its informational content is derived using other methods. According to Schlenker (2019), semantics of music is closer to picture semantics than to logical semantics. In the case of music, an auditory percept permits a set of inferences that have a causal role in inferencing the semantic content. Music semantics aggregates inferences coming from two main sources: auditory cognition and tonal properties of music.

Several studies have shown that the processes of activation, selection, and integration of semantic meaning involve mainly three brain regions: the left posterior temporal cortex, the left anterior temporal cortex, and the left inferior frontal cortex (Lau et al., 2008). The former plays a role in storing lexical representations (Gold et al., 2006), the second is associated with sentence processing (Humphries et al., 2005), and the latter is involved in semantic processing (Lau et al., 2008). A specific ERP component, the N400 is defined as an electrophysiological index of processing of conceptual semantic information (Friederici & Wartenburger, 2010). It has also been shown to index semantic integration or inhibition (Calma-Roddin & Drury, 2020). Within this work, we will focus on the second of the two main functions of the N400.

Semantic Interactions in Music and visual Art

Visual art can be seen as a mean of transmitting concepts out to the world in a manner independent of language, culture, or economic background. In this way, visual art represents a form

of communication accessible to all (Melcher & Bacci, 2008; van Paasschen et al., 2015). The creation of an artistic work arises from the creativity of the author. This can be described as a non-linear process consisting of two central moments: the definition of what the artist wants to realize, and the translation of this intention into a physical form, that is, the actual work of art (de Pisapia et al., 2016).

A series of studies have been conducted to understand how a viewer's perception of a work of art occurs (Locher et al., 2007). Locher and colleagues outlined a model that describes the relationship between eye movements and visual aesthetics (Locher et al., 2007; Locher, 2006; Nodine et al., 1993). This model consists of two stages: first, a global exploration of the painting to obtain an overall impression and the semantic meaning, and then a focal exploration of pictorial features to satisfy cognitive curiosity and develop aesthetic appreciation. Furthermore, the use of abstract images can lead to the understanding of high-level semantic information. Non-realistic images, in fact, allow for the reproduction of the same complex relationships that distinguish naturalistic scenes, but also allow to generate multiple combinations to explore new types of visual scenes. This led to the conclusion that the semantic understanding of a visual representation is related to three factors: the semantic importance of visual features, the object's saliency, and whether an object is noticed and remembered (Zitnick et al., 2016). Despite this, works of art are often associated with a condition defined as semantic instability: the difficulty of recognition (i.e., images with hidden elements) but the capability to still convey meaning.

Muth and co-workers (2016) investigated the relevance of perceptual challenge to the positive effect of semantic stability on the appreciation of visual stimuli. To do so, the authors manipulated the order of movie episodes and their emerging Gestalt. The results show how experiencing difficulties in reading the Gestalt of a visual scene allows for more rewarding increases in certainty, compared to less visually indeterminate contexts. The authors also interpreted their outcome considering the Predictive Coding (PC; Friston, 2005) theory, stating that the pleasure of predictive progress might be greater in unstable semantic contexts (Muth et al., 2016).

What happens when an input channel is added to vision? Bolivar and collaborators (1994) examined the semantic and formal audio-visual information integration in the listener/viewer. The authors questioned the Congruence – Associationist model proposed by Marshall and Cohen (1988). This model posits that music influences attention based on structural audio-visual similarities, namely the Gestalt principle of similarity and alters meaning of the visual stimulus based on temporal congruity. Thus, Bolivar and co-workers (1994) designed three experiments assessing first the semantic congruency of audio-visual pairs, then temporal congruency, and lastly the overall video meaning. The authors concluded that semantic congruency and the degree of salience in the meaning

play a crucial role in judging temporal congruency. Temporal congruence guides attention on visual aspects, to which music assigns then a meaning. Thus, the degree of temporal congruence controls for the effect of music on interpreting the meaning of the visual stimulus.

In addition to the semantic interactions between music and paintings, previous research investigated the integration of visual and auditory information with regards to the complexity of the incoming information. Recently, Klein and colleagues (2021) proposed a cross-modal homeostasis effect to explain the interaction between music and people's evaluation of a visual scene. The starting point is the existence of an optimum both for auditory stimulation and visual display. The assumption is that simultaneous exposure to visual and musical stimuli negatively affects peoples' processing capacity, increasing the preference of simpler stimuli compared to high-complexity stimuli (Klein et al., 2021). The authors presented participants with artistic pictures while listening to background music and asked participants to rate their liking of the visual stimuli. Both music and pictures varied in their complexity. The findings demonstrate that music modifies the inverted-U relationship between visual complexity and image preference. Specifically, adding the musical stimulus shifts the evaluation towards higher values of simpler visuals. The authors ascribe their results to the mere presence of the music stimuli per se, rather than their complexity.

Neural Underpinnings of Art Perception

On a neural level, viewing paintings activates a distributed network of brain areas. A meta-analysis by Vartanian and Skov (2014) highlights areas in the visual cortex, the occipital lobes, and the inferior temporal cortex as areas involved in processing visual features. In contrast, the visuospatial exploration of the visual stimuli was attributed to the precuneus. Also, the superior temporal gyrus was found to be activated and was related to the conceptual integration of information. Finally, the activation of the anterior insula, the putamen, and the posterior cingulate cortex indicate that structures involved in emotion processing are active when experiencing visual art (Vartanian & Skov, 2014). Several studies have also demonstrated that the neural underpinnings visual art perception are different from the one involved in object recognition (Nadal et al., 2008), meaning that experiencing visual artistic stimuli is more than just an object recognition process. Different models of aesthetic visual perception have been proposed (Chatterjee, 2004; Chatterjee & Vartanian, 2016; Koelsch et al., 2015; Leder & Nadal, 2014; Redies, 2015; Tinio, 2013). By examining these models, the general conclusion is that aesthetic appreciation and judgement are highly complicated and sophisticated cognitive processes involving multiple brain regions (Liu et al., 2017). Boccia and colleagues (2016)

conducted a meta-analysis intending to identify the neural response to the visual aesthetic evaluation (VAE) of artworks. Results suggest that different visual areas contribute differently to an extensive aesthetic experience. Among the most commonly activated areas are the orbitofrontal cortex (OFC), anterior insula, and anterior cingulate cortex (ACC) (Brown et al., 2011), and also the reward circuit and selective activation of central nodes of the default mode network (DMN) (Vessel et al., 2012). Although the aesthetic experience is very subjective, results from Vessel and colleagues show that areas included in the DMN are activated across individuals. According to Cupchik and collaborators (2009), the aesthetic contemplation of an image reveals the activation of a brain network including the left lateral PFC (BA 10), as a top-down control in directing perception toward aesthetic orientation, the insula (BA 13), as the experience of emotion, the left lingual gyrus (BA 19), and right inferior temporal gyrus (BA 37/19), as mnemonic processing (Cupchik et al., 2009).

In this work, we will focus on ERP components relative to emotion and semantic processing. In their review work, Olofsson and colleagues (2008) examined the ERP components elicited by affective pictures by subdividing them into short-, middle-, and long-latency range ERPs. In the short latency ERPs, they reported an N1 response, which is sensitive to physical stimulus factors and indexes early sensory processing. N1 was also found to be greater in response to non-pleasant images. Furthermore, in the long latency range, P300 was found to be sensitive to picture valence (Cano et al., 2009), whereas the Late Positive Potential (LPP) is found to be sensitive to pictures' arousal (Wood & Kisley, 2006). Moreover, we rely on the N400 to access inconsistencies in the concurrent emotional stimulation. In fact, this ERP component has been used for different domains and different associations between domains, such as pictures and words (Nigam et al., 1992), words and graphical signs (Imbir et al., 2018), beat gesture and pitch accent (Morett et al., 2020), music and language (Calma-Roddin & Drury, 2020; Daltrozzo & Schön, 2008), and music and pictures (Zhou et al., 2014). In the same vein, previous research has shown how the ERP component N400 can be used to access semantic and syntactic violations in artworks (Cohn & Paczynski, 2019b; Markey et al., 2019a). Markey and colleagues (2019b) presented participants with surrealist paintings containing semantic or syntactic inconsistencies, edited versions without inconsistencies, and real photographic versions of each painting as control. The authors were able to distinguish between syntactic, i.e. structural, and semantic, i.e. about meaning, inconsistencies. The former affected earlier negative ERP components, namely N300 and N400, whereas the latter affected the later positive component, namely P600.



Study 1:

Tagging the Beat of naturalistic Music in the Brain

Introduction[1]

Music is something that develops through time, and time perception can be influenced by music listening (Noulhiane et al., 2007). In music, sound patterns are organized over time in rhythms, and underlying these rhythms we can perceive a periodic beat, to which we move, tap along or dance. Rhythm is the varying pattern of durations that is audibly present in the music (Honing, 2012), while beat perception (London, 2004) refers to the perceptual correlate of these periodicities, that does not necessarily depend on the physical presence of a sound, as in syncopated rhythms (Nozaradan et al., 2012). Periodicities in sound sequences occur at faster and slower rates, subdividing or grouping the perceived beat at distinct metrical levels (Honing, 2012). Variations in rhythm constitute an immediate way to induce emotional reactions, with faster beats causing physiological arousal and slower ones producing relaxation (Bhatti et al., 2016; Cervellin & Lippi, 2011; Koelsch & Jancke, 2015).

One mechanism that has been hypothesized to explain how rhythm is perceived is neural entrainment. As defined by Trost and colleagues (2017), entrainment is the process through which two physical or biological systems become synchronized by virtue of interacting with each other. In other words, when perceiving a regular rhythmic activity, the phase of internal oscillations converges to that of the external driving force, and the endogenous oscillations become phase-locked to the external rhythms (Deng et al., 2020). In the brain, neural responses synchronize to the frequency of an external auditory input. This process is reflected in the Auditory Steady State Responses (ASSR; Roß et al., 2000), and can be detected both in the electric and magnetic signals of neural transmission in the brain, as measured respectively with electroencephalography (EEG) and magnetoencephalography (MEG). ASSRs do not only synchronize to the fast frequencies of the sounds, but also entrain to the slower frequencies of the amplitude-modulated envelope of the auditory signal, giving us the opportunity to study beat and meter (Lenc et al., 2020; Nozaradan, 2014, 2017). It has been proposed that neural synchronization to a stimulus provides a mechanism for selective attention and integration of information (Schroeder et al., 2008; Schroeder & Lakatos, 2009).

So far, neural entrainment to music has been studied mostly using pure tones or simplified acoustical stimuli. Nozaradan and colleagues (2011) 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,

---

[1] This chapter is based on the paper by De Benedetto, F., Haumann, N. T., Celma-Miralles, A., Vuust, P., Proverbio, A. M., Petersen, B., & Brattico, E., Tagging the onset frequency of naturalistic music in the brain (in preparation).

appearing as a steady-state EP at the beat frequency. EEG recording was used to test their hypothesis that neural entrainment to rhythm and meter can be caught as steady-state evoked potentials (SS-EPs). Since the amplitude envelope of the sound used to induce the beat was not strictly regular, the authors interpret the induced neural periodic response as the mechanism that can modulate attentional and perceptual processes over time (Nozaradan et al., 2011). Again, Nozaradan and collaborators (2012) presented participants with different rhythmic patterns while recording EEG and found that rhythmic stimuli elicited SS-EPs in the EEG spectrum and that the amplitude of the SS-EPs increased at frequencies related to beat and meter, despite there was no changes in the acoustic envelope. More recently, Kaneshiro and coworkers (2020) studied the neural correlation during music listening. They used songs from Bollywood movies, variously modified, as musical stimuli. Results show how when shuffling the measure of the song, brain responses in the time-domain were the most correlated in both ISC (Inter-Subject Correlation) and SRC (Stimulus-Response Correlation). Instead, in the frequency-domain, the coherence was highest at frequencies related to the beat (Kaneshiro et al., 2020).

Entrainment, however, is more than just synchronization. It has been proposed that entrainment may also function as a powerful tool for evoking emotions (Juslin et al., 2010). Trost and colleagues (2017) have pointed out that there are different levels of entrainment: perceptual, physiological, motor, and social. In their view, these different levels compose a single process, where one level affects the other. In this work, we refer specifically to the neural entrainment, that is the process by which neural oscillations becomes synchronized with the periodicities of the auditory environment, as a neural mechanism related to the psychological sound percept.

The experience of music is also based on expectations and predictions. In the domain of rhythm, a clear example is the metrical beat, which is extracted from the distinct ongoing rhythms to create and anticipate upcoming tones occupying strong and weak positions. In this predictive coding framework, bottom-up (input, i.e. rhythm) and top-down (prediction, i.e. meter) processes are completely reliant on one another, and the system requires comparison between the two. If there is any mismatch between input and prediction, the latter is changed to adjust with the first (Vuust & Witek, 2014). Violations in auditory expectancies have been studied using the mismatch negativity or MMN response (Näätänen, 1995; Näätänen et al., 2007), an ERP component elicited when the incoming information differs from predicted neural representations (Winkler, 2007). Some studies show in fact that the brain predicts differently the metrical positions of the beat because distinct event-related responses MMN appear depending on onbeat versus offbeat violations (Bouwer & Honing, 2015). To create this cognitive representation, both top-down and sensory information are needed. In this case, top-down processes refer to musical expectancy and, related to rhythm,

temporal expectancies have a role in orienting attention throughout time (Trost et al., 2017). Indeed, as found in Bolger and colleagues (2013), the metrical structure of an entraining rhythm modulates the directing of attention over time. This is in line with the Dynamic Attending Theory (DAT; Jones et al., 1989), which explains the perception of meter in music in terms of entrainment processes. This theory proposes that the periodicities in a musical rhythm can only be perceived because of entrainment of attentional processes to the simultaneously overlapping periodicities of the musical meter. Moreover, sensitivity to the metrical structure increases when an individual listens to music that induces a certain level of arousal or pleasantness (Trost et al., 2017). This suggests that tempo and pulse clarity may play a part in neural entrainment as well. On the one hand, faster tempo increases arousal and slower tempo decreases arousal in the listener (Husain et al., 2002). On the other hand, the clarity of the pulse contributes to the prediction of the valence of the musical stimulus (Trost et al., 2015). Kaneshiro et al. (2020), however, found a dissociation between neural correlates and behavioural ratings for the non-modified song, that is that unedited tracks were rated as the most pleasant and interesting, but showed less ISC. The authors interpret this as evidence that beat tagging and enjoyment of a piece do not go hand in hand and that engagement may reflect an attentional state. Therefore, there is contradictory evidence between entrainment and emotions.

Here, using EEG (and MEG), we aimed to determine whether neural entrainment to sounds can exist also in a complex, realistic auditory environment, namely in naturalistic musical pieces. To disentangle the relationship between entrainment and emotional states, we correlated neural activity entrained to the quasi-periodic rhythms of the songs with several aesthetic judgments about the songs. For this and in line with Trost et al. (2015), we hypothesized that ASSRs would positively correlate with the aesthetic judgments (specifically, familiarity, liking, pleasure, and arousal) obtained for each song. Yet, the direction of this relationship would remain unaddressed.

Materials and Methods

Participants

This study includes 17 healthy young adults (average age of 25.4 ±3.0 years (9 females). Here we apply beat tagging analyses on a subset of an EEG dataset previously collected to study event-related responses (Haumann et al., under revision). The Research Ethics Committee of the Central Denmark Region approved the project.

Additionally, to test the replicability of the findings, a subset of a larger dataset called "Tunteet" was analysed in the frequency dimension. The analyses of the evoked responses are

reported in Haumann et al. (2018; 2020). This subset of data consists of 48 young participants of average age 28.3 years (SD = 8.6) (23 females, 25 males). The Tunteet database includes combined electroencephalography (EEG) and magnetoencephalography (MEG) recordings, though for the present study, only the MEG gradiometer data showing highest signal-to-noise ratio were reanalysed. These datasets were collected at the Biomag laboratory at Helsinki University Hospital between 2012–2013 and already extensively published (Alluri et al., 2015; Bonetti et al., 2017; Burunat et al., 2015, 2017; Carlson et al., 2015; Criscuolo et al., 2019; Haumann et al., 2016; Kliuchko et al., 2015, 2016, 2018, 2019; Saari et al., 2018).

The Tunteet project was approved by the Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (approval number 315/13/03/00/11, March 2012). All procedures for both datasets were conducted according to the Declaration of Helsinki (BMJ, 1991; 302:1194).

Auditory Stimuli

Data from a previous databases were analysed. Within this work, we focused on two pairs of musical pieces with regular compared to irregular timed sound onsets. The alignment of the sound onsets to a regular beat was estimated with the pulse clarity measure (range: 0-1) from the MIR Toolbox (Lartillot & Toiviainen, 2007). The pieces with regularly timed sound onsets were a Danish popular song Foraarsdag by Annet Linnet (Jeg er jo lige her album, Pladecompagniet, 1988), with a medium pulse clarity of 0.70, and a progressive metal/rock piece, Stream of Consciousness by Dream Theatre (Train of thought album, Elektra Records, 2003), with a medium pulse clarity of 0.35. The music with irregularly timed sound onsets were a tango piece by Astor Piazzolla, Adios Nonino, a studio version with a low pulse-clarity of 0.12, and a live recorded version of the same music composition in Lausanne (The Lausanne Concert album, BMG Music, 1993), with a low pulse clarity of 0.13. The compared Foraarsdag and Adios Nonino studio recording lasted about 4 mins, after which behavioural data were collected. The compared Stream of Consciousness and Adios Nonino live version lasted approximately 8 minutes, followed by collection of behavioural data. Being Adios Nonino a very complex piece of music with several tempo and rhythm changes, we focused on its middle part, the one with a more stable rhythm, from which we could better verify an engagement.

Behavioural Data

Behavioural data for the compared Foraarsdag and Adios Nonino studio recording were the favourite music genre of participants. For the musical pieces Stream of Consciousness and Adios Nonino live version we collected behavioural evaluations, that is ratings of familiarity, liking, pleasure, and arousal. Participants answered to a 5-point Likert scale (from 1 to 5) for each one of the four appraisals. Familiarity, liking, pleasure, and arousal scores were taken into account to assess the aesthetic judgment of the participants with regards to the musical pieces.

M/EEG pre-processing

The EEG and MEG datasets were from the preceding evoked response studies high-pass filtered at 1 Hz and low-pass filtered at 25 Hz in order to suppress low and high frequency noise (Haumann et al., 2018; 2020 under revision). Since the spontaneous beat tapping frequency is approximately at 2 Hz and the fastest perceivable sound onsets intervals are near 20 Hz, the 1-25 Hz range is also suitable for analysing beat and sound onset frequencies. Bad electrode signals (>100μV or flat line, on average 0.2 electrodes) were replaced by interpolation of neighbour electrode signals, and all MEG channels were automatically corrected for any bad channels or interferences outside the MEG helmet with the Elekta Neuromag™ MaxFilter 2.2 Temporal Signal Space Separation (tSSS) software (Taulu & Hari, 2009) . Subsequently, eye movement and heartbeat artifacts were suppressed with the Independent Component Analysis (ICA) Infomax algorithm (Bell & Sejnowski, 1995), and artifactual ICA components were subtracted, on average 1.7 components from the EEG and 2.6 components from the MEG gradiometers. Finally, the EEG electrodes were re-referenced to the average across all electrodes, and all data were down-sampled to 125 Hz sampling rate (Kaneshiro et al., 2020).

Auditory Steady State Response M/EEG Analysis for naturalistic Music

We propose a new method analysing auditory steady state responses (ASSR) to regular sound onsets in naturalistic music. With the ASSR analysis method introduced by (Nozaradan et al., 2011) it is assumed a pulse train of regular sound onsets (with isochronous sound onsets, or constant sound onset asynchrony). However, in real music, the assumed regular sound onsets will not be perfectly regular over time, rather they continuously vary in phase due to micro-timing and tapping speed variation in the music performance (e. g., Kaneshiro et al., 2020; Skaansar et al., 2019). Therefore, we propose a measure of the average steady state response aligned to each individual sound onset. If there is a regular steady state response within a shorter time window, the average M/EEG waveform

time-locked to each sound onset will be phase-locked to adjacent steady state response cycles, and its amplitude can be estimated.

To measure the naturalistic ASSR, the M/EEG data was time-locked to the sound onsets, by reapplying manual sound onset detections from previous publications (Haumann et al., 2020; Haumann et al., under revision). Kaneshiro et al. (2020) applied a moving 5-second time window for analysing beat or sound onset related ASSR oscillations in continuous EEG data, and we applied the same 5 seconds time window for ASSR analysis. Although, instead of moving the time window continuously over time, the time window was iteratively centred at each sound onset. Sound onset trials with residual noise were automatically rejected when the waveform amplitude exceeded 100 μV in the EEG (on average 99.0% of the EEG trials were analysed after rejection) and 400 fT/cm in the MEG gradiometers (on average 95.2% of the MEG gradiometer trials were analysed after rejection). Similar to Nozaradan et al. (2011), to increase the signal-to-noise ratio, the average ASSR was then measured across the time windows time-locked to each sound onset. To convert the average ASSR from the time- to the frequency-domain, Fourier analysis was applied with the same settings as applied by Kaneshiro et al. (2020), with a Hamming window and 1024 point fast Fourier transform with zero-padding equally distributed around the 5 seconds time window.

The proposed new naturalistic ASSR analysis method functions as a time- and frequency-specific bandpass filter, which suppresses any signal outside the frequency and time points of the sound onsets. Also, residual noise in the EEG and MEG waveforms would be bandpass filtered, which results in apparent low amplitude frequency peaks at the frequency of the sound onsets for residual noise (because the noise amplitude is more suppressed at frequencies outside compared to inside the sound onset frequency bins). The ASSR results were therefore corrected for this bandpass filtered residual noise. Residual noise of the same amplitude as the analysed data was estimated by applying the same waveforms as in the main analyses but with randomized order of the time samples (to eliminate the regular ASSR oscillations). As expected, the noise spectra showed low amplitude peaks at the sound onset frequencies, though, at significantly lower amplitudes in comparison to the ASSR spectra. For each participant and music piece, this bandpass filtered noise spectra was then subtracted from the ASSR spectra.

In the ASSR spectra based on the time windows time-locked to each sound onset, any slower beat than the sound onset frequency will also be suppressed due to this band-pass filtering effect. To further enable the detection of possible endogenous beat grouping activity at frequencies slower than the sound onset frequency, the ASSR spectra were additionally analysed with time windows time-locked to every second and every fourth sound onset. If slower beat frequencies at half or furth the rate of the sound onset frequency were present, they would be measurable in the ASSR spectra in

addition to the faster sound onset frequency. However, excluding every second or fourth sound onset trial will also increase the noise level in the ASSR spectra.

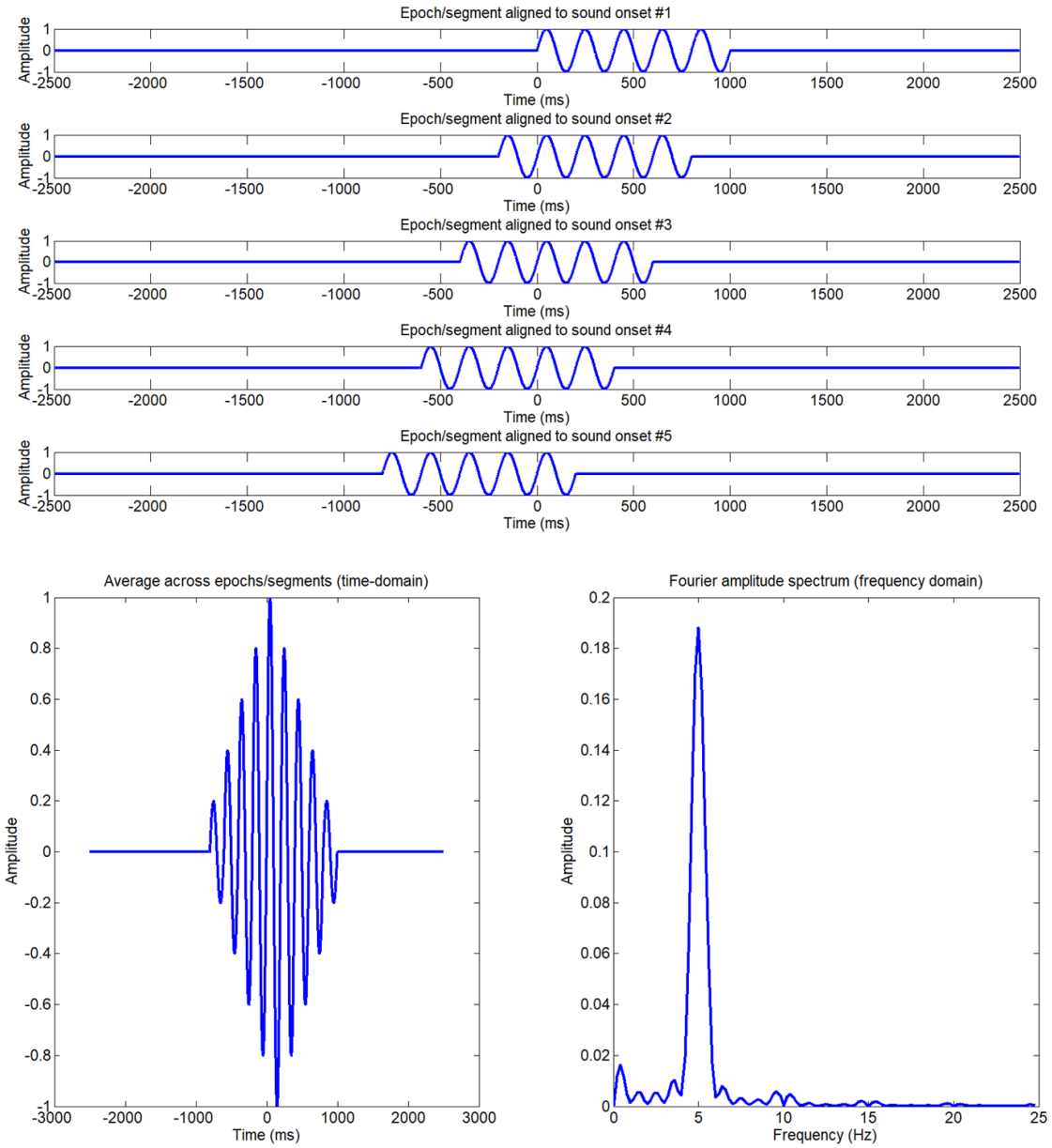


Figure 1.1. (Top) Intra signal-averaging at regular sound onsets in naturalistic music that will result in an auditory steady state response (ASSR) cycle. (Bottom left) Average signal containing an SSR. (Bottom right) A Fourier transform of the frequency of the auditory steady state response.

To circumvent this issue, we applied the solution suggested by Nozaradan et al. (2011), where the mean of the -/+ 0.37 Hz neighbouring frequency bin was subtracted from each frequency bin (taking into consideration the width of the spectral leakage at the main lobe for the Hamming

window). Thereby, the new naturalistic ASSR analysis approach should enable the estimation of both sound onset and beat-related frequencies while suppressing noise in the continuous M/EEG.

To compare the ASSR M/EEG results with sound onset related frequencies in the audio stimuli, the root-mean-squared sound intensity was calculated with the MIR Toolbox (Lartillot & Toiviainen, 2007), and the sound intensity curve was subjected to the same ASSR analysis procedure as the M/EEG datasets.

Inter-Subject Coherence Analysis for naturalistic Music

The inter-subject coherence (ISC) approach to measure sound onset and beat frequency in naturalistic music was applied with the same analysis procedure and settings as in the previous work by Kaneshiro et al. (2020). As a first step, high amplitude residual noise was ignored by replacing waveform samples with voltage level exceeding four standard deviations of the electrode's mean power with NaNs, and this procedure was repeated four times. Reliable Component Analysis (RCA) was then applied to increase the signal-to-noise ratio of the continuous EEG and MEG waveforms (Dmochowski et al., 2012, 2018). RCA estimates spatial filter components (across the EEG or MEG channels) that optimizes inter-subject correlation between the M/EEG waveform across participants. As in previous studies we applied only the first and most reliable component, reliable component 1 (RC1), which is a weighted sum across channels that projects the multi-channel waveforms onto a single time-domain component waveform. Kaneshiro et al. (2020) performed the RCA on each music piece separately and visually verified that the topography of the RC1 was comparable between the music pieces. Computing RCA across the music pieces would provide a higher certainty that the RCA components were the same between the music pieces. For the EEG dataset, it was possible to perform the RCA on the complete dataset, including the EEG waveforms for the two tested music pieces and for all participants. For the MEG gradiometer dataset, the MEG gradiometer computations included a very large data matrix with 204 channels high density sensor array by time samples for two eight minutes pieces by 48 participants, which exceeded the maximum of 16 GB RAM resources available on common hardware. Therefore, for the MEG dataset, as with Kaneshiro et al (2020), it was necessary to perform the RCA on each piece separately, and it was verified that the RC1s were comparable between the two music pieces. For the estimation of the MEG gradiometer topography of the RC1 it was also necessary to temporarily down-sample to 62.5 Hz sampling rate to reduce the number of time samples, though, the 125 Hz sampling rate was restored when the RC1 component waveform was projected from the 125 Hz sampling rate multi-channel array based on the estimated RC1 topography.

The magnitude-squared coherence between participants for each frequency was estimated based on the RC1 waveforms by applying the *mscohere* Matlab function. For the M/EEG RC1 waveforms, the magnitude-squared coherence algorithm measures a correlation (range: 0-1) in spectral power variance over time between a pair of participants. This means that if the spectral power in a specific frequency varies similarly over time between a pair of participants, then the correlation coefficient would be high for that frequency and pair of participants. The default settings were applied, where the magnitude-squared coherence is calculated by converting the RC1 waveforms to time-frequency representations with short-time Fourier analysis, with a Hamming window, 1024 point fast Fourier transform, zero-padding equally distributed around a 5-seconds time window, and 50% overlap between time samples. Finally, the inter-subject-coherence (ISC) was calculated as the average magnitude-squared coherence for each frequency across all pairs of participants (Kaneshiro et al., 2020).

Statistical Analysis

Music with regular compared to irregular Sound Onsets

To test the first hypothesis, that naturalistic music with regularly but not irregularly timed sound onsets results in auditory steady state responses at the frequency of the sound onsets (Kaneshiro et al., 2020). The amplitudes of the ASSR spectra and the coherence values of the ISC were compared between the music pieces with regularly and irregularly timed sound onsets at the expected peak frequencies detected from the sound intensity ASSR spectra. Since the amplitudes were not consistently normally distributed, the Wilcoxon signed rank test was applied to verify that the amplitude and ISC at the expected sound onset peak frequencies were higher for the music with regularly timed sound onsets compared to irregularly timed sound onsets.

Correlation between ASSR/ISC and aesthetic Appreciation

To investigate the association between the aesthetic evaluation of each musical piece and the neural activity, we first assessed the possible relationships with Spearman's correlations. Also, we computed a Mantel test between the individual behavioural measures and the ISC at the frequency peaks. The Mantel test calculates the similarity between two distance matrices. In order to convert the behavioural responses into matrices, we calculated the absolute value of the distance, namely the difference, of each score to the others, resulting in a matrix with values which can fall in a range from 0 (completely dissimilar) to 1 (completely similar). The Mantel test performed a correlation between the matrix of each behavioural measure and the ISC computed on the frequency and time domain of the EEG and MEG data. Mantel test p-values were calculated with 5000 permutations.

Correlation between Onset-grouping and behavioural Measures

Once grouped the sound onsets into two or four to access the beat frequency, we correlated the amplitudes relative to the beat frequencies with the behavioural ratings. As for the Foraarsdag piece, we compared the amplitudes with the preferred genre, whereas for the Stream of Consciousness piece we compared the beat amplitudes with the aesthetic ratings.

Results

ASSR and ISC for regular compared to irregular Sound Onsets

For the EEG dataset, the music piece with regular sound onsets showed three frequency peaks (at 4.76 Hz, 9.64 Hz, and 14.40 Hz) for the sound intensity changes in the audio. Peaks at the same two lowest frequencies (4.76 Hz and 9.64 Hz) were present in the EEG ASSR (Figure 1.2) and of significantly higher amplitude in comparison to the ASSR spectrogram for the music piece with irregular sound onsets (Table 1.1). The frequency peak for the slowest sound onset periodicity (4.76 Hz) was also present in the ISC for the music piece with regular sound onsets and of significantly higher ISC values compared to the ISC for the music piece with irregular sound onsets (Table 1.2).

For the MEG dataset, the music piece with regular sound onsets showed three frequency peaks (at 4.15 Hz, 8.42 Hz, and 16.85 Hz) in the sound intensity curve. The two slowest frequency peaks (4.15 Hz and 8.42 Hz) were observed also in the MEG gradiometer ASSR and the ISC values (Figure 1.2), and the two peaks showed significantly higher ASSR amplitude and ISC values for the music with regular compared to irregular sound onsets.

When analysing the possible effects of grouping by two or by four sound onsets into a slower beat, the music piece with regular sound onsets showed frequency peaks in the EEG ASSR at the slowest sound onset frequency divided by two (2.32 Hz) and divided by four (1.22 Hz). These onset-grouping related frequency peaks were not present in the spectrum for the sound intensity changes in the music piece. Moreover, the music piece with regular sound onsets from the MEG dataset showed a frequency peak in the MEG gradiometer ASSR at the slowest sound onset frequency divided by two (2.08 Hz). Also, this onset-grouping related frequency in the ASSR was not resembled by a peak in the spectrum for the sound intensity changes in the music stimuli.

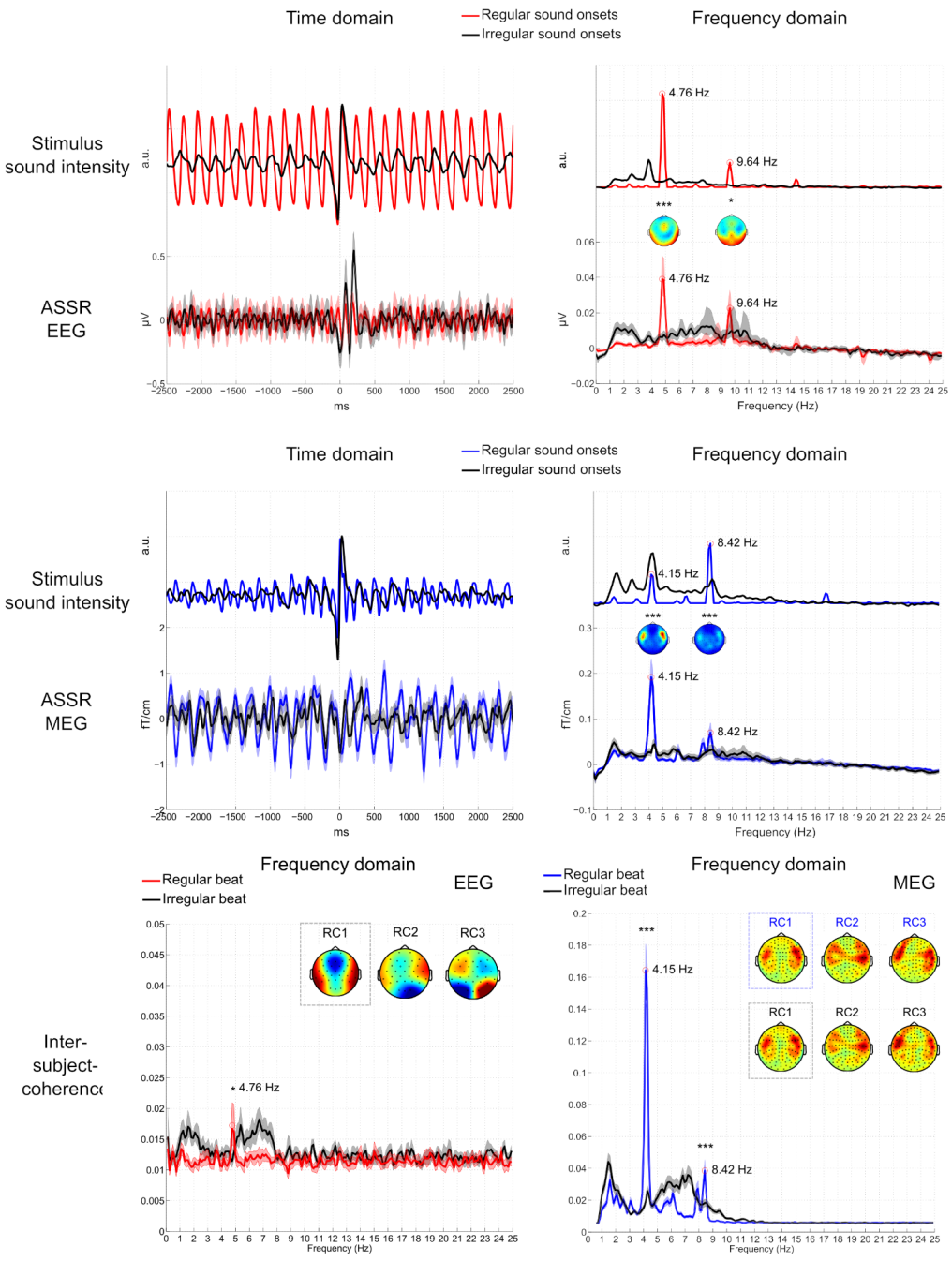


Figure 1.2. ASSR and ISC for regular compared to irregular sound onsets.

Average sound intensity changes and the auditory steady-state evoked responses (ASSR) time-locked to each sound onset in the piece with regular compared to irregular sound onsets. ASSR for the EEG dataset shown in (top left) the time domain (msec) and in (top right) the frequency domain (Hz). ASSR for the MEG dataset shown in (middle left) the time domain (msec) and in (middle right) the frequency domain (Hz).

ISC in the frequency domain (Hz) for the piece with regular compared to irregular sound onsets from the EEG dataset (bottom left) and the MEG dataset (bottom right), and the reliable component 1 (RC1) applied for the ISC analyses.

*Significantly higher ASSR amplitudes and ISC values for regular compared to irregular sound onsets are indicated at the * $p<.05$ and *** $p<.001$ significance level.*

Table 1.1. Regular and irregular sound onset stimuli type are reported in the first and second column respectively. The stimulus frequency refers to the ASSR frequency peaks. Measure = data collection technique; M_{diff} at stimulus frequency = mean difference of amplitude values; N = sample size; p-value = p-values referred to the mean difference of amplitude values; p_{FDR} = false discovery rate for multiple comparisons.

Regular stimulus	Irregular stimulus	Regular stimulus frequency (Hz)	Measure	M_{diff} at stimulus frequency	N	*p-value*	p_{FDR}
Pop song	Tango (a)	4.7607	EEG	0.03 µV	17	<.001	<.001***
		9.6436	EEG	0.02 µV	17	.023	.033*
		14.4043	EEG	-0.00 µV	17	.109	.136
Middle part of tango (a)	Tango (a)	3.7842	EEG	0.01 µV	16	.323	.339
		7.5684	EEG	0.01 µV	16	.002	.003**
Alternative rock/metal	Tango (b)	4.1504	EEG	0.01 µV	40	<.001	<.001***
			MEG mag.	0.49 fT	48	<.001	<.001***
			MEG grad.	0.14 fT/cm	43	<.001	<.001***
		8.4229	EEG	0.00 µV	40	.060	.080
			MEG mag.	0.14 fT	48	<.001	<.001***
			MEG grad.	0.03 fT/cm	43	<.001	<.001***
		16.8457	EEG	0.00 µV	40	.183	.203
			MEG mag.	0.01 fT	48	.160	.188
			MEG grad.	-0.00 fT/cm	43	.546	.546
Middle part of tango (b)	Tango (b)	4.0283	EEG	0.04 µV	40	<.001	<.001***
			MEG mag.	2.21 fT	48	<.001	<.001***
			MEG grad.	0.51 fT/cm	43	<.001	<.001***
		8.0586	EEG	0.03 µV	40	<.001	<.001***
			MEG mag.	1.32 fT	48	<.001	<.001***
			MEG grad.	0.31 fT/cm	43	<.001	<.001***

Table 1.2. Regular and irregular sound onset stimuli type are reported in the first and second column respectively. The stimulus frequency refers to the ISC values peaks. Measure = data collection technique; M_{diff} at stimulus frequency = mean difference of ISC values; N = sample size; p-value = p-values referred to the mean difference of amplitude values; p_{FDR} = false discovery rate for multiple comparisons.

Regular stimulus	Irregular stimulus	Regular stimulus frequency (Hz)	Measure	M_{diff} at stimulus frequency	N	*p-value*	p_{FDR}
Pop song	Tango (a)	4.7607	EEG	0.002	17	.011	.019*
		9.6436	EEG	0.000	17	1.000	1.000
		14.4043	EEG	-0.001	17	.005	.009**
Alternative rock/metal	Tango (b)	4.1504	EEG	0.003	48	<.001	<.001***
			MEG mag.	0.112	48	<.001	<.001***
			MEG grad.	0.149	48	<.001	<.001***
		8.4229	EEG	0.001	48	.104	.157
			MEG mag.	0.013	48	<.001	<.001***
			MEG grad.	0.016	48	<.001	<.001***
		16.8457	EEG	0.000	48	.243	.291
			MEG mag.	0.000	48	.123	.165
			MEG grad.	0.000	48	.430	.469

Correlations with behavioural Measures

Correlation with ASSR/ISC and Aesthetic appreciation

To assess the association between neural activity and aesthetic judgment, we first computed Spearman correlation with both the Stream of Consciousness and the Adios Nonino (the middle part with more regular rhythm).

As for Stream of Consciousness, no significant correlations were found. Whereas, for Adios Nonino we found a positive correlation between the frequency amplitude in the MEG spectrum and the pleasure ratings (Spearman's rho = 0.436; $p < 0.01$). Moreover, in the EEG spectrum, the frequency amplitude correlated positively with both liking (Spearman's rho = 0.434; $p < 0.05$) and pleasure (Spearman's rho = 0.374; $p = 0.05$) rates.

We also performed the Mantel test between the individual behavioural measures and the individual ISC values at the frequency peaks. Comparing these values, we found significant correlations in the Stream of Consciousness piece with the pleasure ratings ($p = 0.02$) and in the middle part of the Adios Nonino with the liking ($p < 0.05$) rates. The complete results for the Mantel test are reported in Table 1.3.

Table 1.3. Mantel test results for each musical stimulus. Familiarity, Liking, Arousal, and Pleasure refer to the behavioural measures. EEG, MEG mag. (magnetometers), and MEG grad. (gradiometers) refer to the data collection technique. Values refer to the p-values of the comparison between behavioural measures and the ISC computed in the time domain.

Alternative Rock/Metal piece					
		Familiarity	Liking	Pleasure	Arousal
Time Domain	EEG ISC	0,67	0,43	0,02 *	0,40
	MEG Grad. ISC	0,97	0,99	0,86	0,93
	MEG Mag. ISC	0,96	0,96	0,60	0,91
Tango piece (b)					
		Familiarity	Liking	Pleasure	Arousal
Time Domain	EEG ISC	0,17	0,48	0,55	0,25
	MEG Grad. ISC	0,27	0,28	0,78	0,33
	MEG Mag. ISC	0,28	0,24	0,73	0,43
Tango (middle part) piece (b)					
		Familiarity	Liking	Pleasure	Arousal
Time Domain	EEG ISC	0,52	0,03 *	0,15	0,14
	MEG Grad. ISC	0,30	0,25	0,65	0,59
	MEG Mag. ISC	0,34	0,59	0,75	0,83

Correlation between onset-grouping and behavioural measures

To access the relationship between behavioural measures and the amplitudes relative to the beat frequencies with the preferred music genre, for the Danish pop song piece (Foraardags), and with the aesthetic ratings for the alternative rock/metal piece (Stream of Consciousness). Also in this case, we found a negative correlation between the amplitude of the beat frequency and the Dansktop music genre (Spearman's rho = 0.540, $p < 0.05$). For the alternative rock/metal piece (Stream of Consciousness), we found no significant correlation.

Discussion

With this study, we aimed to investigate the perceptual entrainment of a naturalistic musical stimulus and to better understand the relationship between entrainment and aesthetic judgments. We compared musical stimuli with regular and irregular timed sound onset. We applied beat tagging analysis on an M/EEG dataset. Here, we propose a new method for analysing Auditory Steady State Response (ASSR) in naturalistic music. Unlike the ASSR analysis proposed by Nozaradan (2011), we measured the average SSR aligned to each sound onset. With a regular SSR, we would have found an M/EEG waveform phase-locked to the ASSR. In addition, since music unfolds over time, we tested our analysis both in the frequency domain and in the time domain, as recommended by recent works (Novembre & Iannetti, 2018; Rajendran & Schnupp, 2019). The proposed new naturalistic ASSR analysis method functions as a time- and frequency-specific bandpass filter, which suppresses any signal outside the frequency and time points of the sound onsets.

ASSR and ISC for regular and irregular Sound Onsets

Here, using electroencephalography and magnetoencephalography, we aimed to determine whether neural entrainment to sounds can exist also in a complex, realistic auditory environment, namely in naturalistic musical pieces. As expected, we found peaks in the EEG ASSR at the same frequencies as the frequency of the sound onset (4.76 Hz and 9.64 Hz). Also, the Inter-Subject Coherence (ISC) showed the same frequency peaks. Matching results were found with the MEG dataset, with frequency peaks in the MEG gradiometer ASSR and ISC values at the same frequencies as the sound intensity changes (4.15 Hz and 8.42 Hz). This is in line with Kaneshiro and colleagues (2020), who found that musical stimuli with basic features led to statistically significant ISC. More recently, Kaneshiro and co-workers (2021) linked the ISC to states of engagement during naturalistic music listening. In this work, participants listened to a cello concert movement lasting for 8 mins, compared to a scrambled version of the same musical piece. EEG ISC analysis investigating the time-varying engagement during music listening showed higher levels of ISC during periods of musical tension in the intact musical stimulus. By contrast, ISC was non-significant for long periods of time apart from the climax of the piece. Considering these findings, since we find different ISC values for the irregular sound onset piece, we may say that in our case each participant perceives differently a more complex stimulus, either rhythmically or emotionally, therefore less ISC emerges. Moreover, Geiser and colleagues (2014) defined global timing as the percept of temporal patterns created by sequential acoustic event. Global timing relies on the listener to incorporate temporal patterns that span several seconds. With a naturalistic musical stimulus, though, micro-timing variations play a

fundamental role in lowering rhythmic predictions. Our participants were both musicians and non-musicians and, given the fact that previous studies demonstrated that expertise modulates rhythm perception (Matthews et al., 2016; Skaansar et al., 2019; Stupacher et al., 2017), we included different skills to extrapolate rhythmic patterns among our participants. It has been suggested, in fact, that the neural brain responses are a reflection of the conscious perception of the musical structure (Nave et al., 2022). This also can be a reason for finding less ISC for the irregular sound onset piece. These differences can be inscribed both in the Dynamic Attending Theory (DAT; Jones et al., 1989) and in the predictive coding framework (Friston, 2005), where we can say that a more complex rhythm make predictions more difficult.

Perceptual Entrainment to the musical Beat

With this work we also aimed to investigate perceptual entrainment to the beat frequency. To access the beat, we grouped sound onsets by two or four. Consequently, EEG ASSR showed frequency peaks as the sound onset frequency divided by two (2.32 Hz) and divided by four (1.22 Hz). The same was found for the MEG gradiometer ASSR (2.08 Hz). Here, as in Nozaradan and collaborators (2012), the magnitude of the ASSR elicited at sound onsets related frequencies was significantly higher as compared to the magnitude of the ASSR elicited at frequencies unrelated to the sound intensity changes. Additionally, the grouping-related frequency does not appear in the sound intensity spectrum, so this may indicate that this activity is not a mere passive representation of the progression of acoustic events. Previous studies demonstrated that cortical oscillations are modulated also by speech processing (Ding & Simon, 2014; Lamminmäki et al., 2014) and its comprehension (Horton & Srinivasan, 2013), attention (de Jong et al., 2010; Lenc et al., 2020) ,or rhythmic imagery task (Celma-Miralles et al., 2016). Meltzer and colleagues (Meltzer et al., 2015) set up two studies to investigate the effect of musical comprehension and attention on ASSR. In the first study, to test for music comprehension, they presented participants with four edited classical pieces of music and their scrambled version, where segments were randomly reorganized. In the second study, to test for attention-mediated effects, the authors asked participants to read a text while ignoring music, or to ignore the text and focus on the music. The findings demonstrate that steady-state response to the beat is influenced by the ability to understand a musical melody, being weaker for the familiar version than for the scrambled version. Also, attention modulated the neural response to music so to the extent that they increase with attention (Meltzer et al., 2015). In our case, participants were instructed to concentrate on the musical stimulus, so this may be a factor that modulated our resulting ASSR. This piece of result is also in line with a recent study by Nave and collaborators (2022), who used musical excerpt with an ambiguous beat pattern to induce beat

perception in the listeners, to investigate subsequent ASSR. The results confirmed that ASSR reflect modulation of neural activity at the same frequency as the beat. The authors interpret their outcomes as an effect of attention and conscious processing on ASSR (Nave et al., 2022). In our opinion, though, future research is necessary to unravel endogenous by exogenous processes.

Entrainment and Emotion

Finally, with this work, we wanted to examine the link between perceptual entrainment and perceived emotions. Several studies have claimed a direct relationship between the two (Deng et al., 2020; Labbé & Grandjean, 2014; Trost et al., 2014, 2017; Vuilleumier & Trost, 2015). Comparing behavioural ratings of aesthetic judgments with ASSR, ISC, and beat tagging values, we found little evidence supporting this remark. Juslin and colleagues (2010) suggest that entrainment processes play a causal role in emotion induction. Neural entrainment is affected by the acoustic properties of the stimulus (Ross & Balasubramaniam, 2014). Likewise, perceptual entrainment has been associated with the familiarity with the musical stimulus (e.g., Kumagai et al., 2017), with the attention and comprehension of the stimulus (Lenc et al., 2020; Meltzer et al., 2015). On the other hand, Labbé and Grandjean (2014) were able to predict music-evoked emotion by means of visceral entrainment, namely felt changes in one's own body, and motor entrainment, namely one's urge to move to the music. Besides, Trost and co-workers (2014) relate entrainment mechanisms with the reward system, which in turn plays a fundamental role in emotion processing (Blood & Zatorre, 2001; Ferreri et al., 2019; Zatorre, 2015). Nonetheless, the same authors (Juslin et al., 2010; Trost & Vuilleumier, 2013) assert that rhythmic entrainment represents just one of the multiple emotion induction mechanisms which might be at play. Indeed, if entrainment and emotions were so tightly linked, we would have coupled more consistently these two processes. Future studies may address this issue specifically by isolating factors that contribute to music-evoked emotions. We saw how neural oscillations are influenced by several factors, such as musical expertise (Stupacher et al., 2017), attention (Lenc et al., 2020), familiarity with the song (Kumagai et al., 2017), or the properties of the song itself (Nave et al., 2022). At the same time, also emotion perception is modulated by musical expertise (Battcock & Schutz, 2022; Castro & Lima, 2014), attention (Kanske et al., 2011; Schupp et al., 2003), and familiarity (Margulis, 2013; Pereira et al., 2011). Hence, the fact that different factors influence both entrainment and emotion might lead to a correlation between the two. Despite previous literature found this connection, this was not the case of the present study. Anyhow, this was not the main focus of this study, therefore future research should address this topic more specifically. Thus, what remains to be disentangled is how entrainment and emotions are linked to one another and the direction of the impact.



Study 2:

Emotion Comprehension in Music and visual Art

Introduction[2]

Several studies have investigated how listening to background music can affect and modulate concurrent mental processes in a multimodal manner (for example Bottiroli et al., 2014; Hanser et al., 2015; Nadon et al., 2021; Proverbio & De Benedetto, 2018). At this regard, Jeong and colleagues (2011) recorded fMRI signals during multimodal perception of auditory (music) and visual (faces) stimuli (audio-visual processing) that might be congruent o incongruent with each other. The results showed how emotional congruence across sensory modalities would enhance activity in auditory regions of the brain. Also, musical stimulation seemed to affect behavioural emotional ratings of the observed faces. These results are similar to the findings of Baumgartner and colleagues (2006b), who presented IAPS fearful and sad pictures combined with emotional music. Results showed that the congruent presentation of emotional pictures and music evoked stronger emotions than the incongruent condition, thus concluding that music was able to enhance the emotional experience evoked by visually presented affective pictures, when congruently combined. Likewise, Jolij and Meurs (2011) found that participants were more sensitive to mood-congruent than incongruent emotional stimuli. In this study, participants had to detect schematic happy and sad faces embedded in noise, while listening to happy or sad music. As a result, they found that participants were more sensitive to mood-congruent than incongruent emotional stimuli. In other words, when listening to happy music they were more accurate in detecting happy faces as compared to when listening to sad music, and vice versa. The authors interpreted these results as the role of expectation in visual processing. Again, Braun Janzen and colleagues (2023) investigated the influence of musical emotions on the emotion and aesthetic perceived value of abstract paintings in a naturalistic environment. The results showed how music valence significantly influenced participants' judgment of the pleasantness of the painting. In this study, visitors of an art museum were freely enjoying a painting by Kandinsky while listening to a positive or a negative background music. The results suggested that participants who listened to positive-valence music tended to rate the painting as more pleasant than those who listened to negative-valence music. This led the authors speculate that the aspects associated with the aesthetic experience of music may influence the aesthetic experience of visual art (Braun Janzen et al., 2023).

More recently, Proverbio and collaborators (2020) compared congruent and incongruent affective audio-visual information (i.e., musical pieces and facial expressions) that conveyed same or

[2] This chapter is based on the paper by De Benedetto, F., Ghiraldini, E., Capizzi, N., Proverbio, A.M., Comparing and combining emotions evoked by music and artworks: an ERP study (submitted to Psychology of Music).

different emotional content, while recording EEG/ERP signals. It was found that several ERP components were modulated by the emotional content of both stimuli. When incongruently associated, audio-visual pairs elicited an enhanced N400 response, probably indicating the detection of a mismatch in the emotional content conveyed by the two sensory modalities (Proverbio et al., 2020). This findings demonstrated how the emotional content of the musical pieces could be clearly distinguished and comprehended by the listener, similarly as facial expressions.

To investigate this matter, in the present study, visual artworks were combined with classical musical pieces. We aimed at investigating how the emotional content of music affected the perception of emotional sensations experienced during artwork perception. Specifically, the purpose was to study the distinctiveness of the emotional meaning that music is able to evoke. We presented pairs of matched/unmatched music and paintings with the hypothesis that emotionally incongruent pairs would negatively affect the overall emotional comprehension of the stimulus. It was expected that the emotional incongruency would be indexed by an N400 response (Kamiyama et al., 2013; see Kutas & Federmeier, 2011 for a review), meaning that the emotional content of both music and pictures had been conveyed, and processed as different. On the other hand, P300 component was expected to be greater in response to congruent pairs (Proverbio et al., 2020). This because affective influences are present from around 100 ms after stimulus onset, as high-arousing pictures elicit a more positive-going P300 over parietal recording sites (Conroy & Polich, 2007). Also, there is a general consensus on an effect of both emotional valence (Herbert et al., 2008; Kissler et al., 2007) and arousing level (Cuthbert et al., 2000; Hajcak et al., 2010) on the P300.

As in the present work we used congruent and incongruent stimulus pairs, we expect the congruent category to be perceived as more positive and the incongruent as more negative. Therefore, we expect congruent stimuli to be more pleasant than incongruent ones, and so elicit larger LP components. Indeed, LP is thought to be larger to more arousing and pleasurable stimuli (Cuthbert et al., 2000). Cuthbert and colleagues (2000) found that emotionally arousing pictures elicited a greater late positive potential, compared to nonaffective stimuli. The authors interpret these changes as a motivational significance, in the sense of an increased investment of attentional resources toward emotional stimuli.

Materials and methods

Participants

Twenty-five (13 females) University students took part in the study; their age ranged between 18 and 35 (M = 21.92, SD = 1.99). All had normal or corrected-to-normal vision and were right-handed (M = 0.78) as assessed by the Oldfield Inventory (Oldfield, 1971). They reported no history of drug abuse or neurological or mental disorders.

The experiment was conducted with each participant's understanding and written consent according to the Declaration of Helsinki (BMJ 1991; 302: 1194) and following the APA ethical standards (1992, American Psychological Association). The Ethical Committee of the University of Milano-Bicocca approved the study (CRIP protocol n. RM-2019-218). Participants were blinded to the purpose of the experiment. All students were granted academic credits for their participation in the study.

Data from 7 subjects were discarded due to excessive recording artifacts, so the analysis were conducted on a sample of 18 subjects (10 females) with mean age 22.11 (SD = 1.81).

Stimuli

Stimuli were 310 artworks and 33 musical excerpts. All the stimuli underwent a validation process aimed at identifying their belonging to one of the 4 emotional categories. Emotional categories were the same for both visual and auditory stimuli, positive and negative in polarity, and high or low in arousal: happiness, relax, fear, sadness.

Visual Stimuli. The paintings selected as stimuli in this study were roughly dated between the fourteenth and twentieth centuries; most of them were realized with oil on canvas. All the artworks displayed were in bright colour, all had size of 500 x 333 pixels, and all equiluminant across experimental categories, as accessed by the ANOVA performed on luminance values [F (4, 188) = 1,97, p = 0,1]. As for the target category, 20 artistic multi-colour paintings depicting series of fishes were used. Stimuli were equiluminant across emotional categories, as proved by an ANOVA performed on individual luminance values.

Auditory Stimuli. All musical fragments were excerpts of real instrumental classical music pieces. Music stimuli expressing Happiness, Sadness, and Fear pieces were validated and used in a previous ERP study (Proverbio et al., 2015). Relax music pieces were selected considering the piece key (minor or major) and its harmonic and rhythmic structure, as it is known to convey specific

emotional information (Schutz, 2017). Hence, relaxing pieces were characterized by slow tempo (e.g., Andante or Allegretto) and major key. Each fragment lasted 25 seconds. To avoid abrupt starting and ending, the first and the last 3 seconds of music were faded in and out (12%). All stimuli were also normalized (-1 Db normalization) and levelled (70 Db levelling) via Audacity software.

Stimuli Validation

Visual Stimuli. Visual stimuli were 75 joyful-, 82 relaxing-, 75 sad-, and 78 fearful pictures. Stimuli were validated both for the emotional content and familiarity. Both assessments were realized via Google Form. The surveys were available online, for both PC and mobile devices. The aim of the first survey, relative to artwork emotional content, was to establish the emotion felt by looking at the stimulus, and its intensity. After receiving instructions, participants were administered, randomly mixed and one at a time, 310 pictures depicting coloured paintings, lasting 1 second each. The task consisted in deciding which emotion, among the 4 presented, best fitted the emotion felt and the degree of this association with the selected category, by means of 3-point Likert Scale, ranging from 1 (not much) to 3 (extremely). To access the distinctiveness across different artworks evoking different emotions, an ANOVA was performed on the mean values of the categorization of each artwork, with emotional category as a factor (happiness, relax, sadness, fear). The significance of the emotion factor [$F(3, 183) = 5.73$; $p < .001$] indicates that the artworks were attributable to the selected emotional category. On the basis of these results the most distinguishable pictures, that is the ones with the highest percentage of correct categorization, were selected.

Subsequently, artworks were also evaluated for their familiarity and the aesthetics value by another group of participants, also using a 3-point Likert Scale. We performed an ANOVA on the mean familiarity values [$F(3, 27) = 0.83$, $p = 0.48$]. The results showed that paintings belonging to the various categories were equally unfamiliar to the subjects. Additionally, an ANOVA on mean values of aesthetic appreciation was carried out, with emotion as a factor [$F(3, 27) = 2.63$, $p = .07$]. Again, the results showed a lack of significance across categories, meaning that the paintings were rated as having no differences in the aesthetic value across different categories.

Auditory Stimuli. The procedure for the validation of musical pieces took place in the EEG laboratory to provide a standardized stimulus administration. All participants listened to the stimuli via a Beats EP noise-cancelling headset (Beats Electronics LLC). After receiving instruction from the researcher, they had to specify which emotion among the 4 presented was more closely associated to the music heard, which was the intensity of this association (on a 3-point Likert Scale from 1 to 3), and how familiar was the musical piece (on a 3-point Likert Scale from 0 to 2). 20 out of 33 musical

stimuli were selected on the basis of this procedure, after discarding unrelated to exceedingly familiar music. As for visual art, individual distinctiveness and familiarity values were subjected to 2 ANOVAs, with emotion as a factor of variability (happiness, relax, sadness, fear). Both ANOVAs (distinctiveness = [F (3, 16) = 3.09, p = 0.06]; familiarity = [F (3, 16) =0.29, p = 0.82]) showed no differences across emotional categories.

Based on the validation, a final stimulus set comprising 48 artworks and 5 musical pieces per category was selected to be used in the EEG/ERP recording.

Procedure

Participants were comfortably seated in an anechoic and electrically shielded cubicle. Visual stimuli were presented on a PC screen located outside the cabin, 120 cm from the participants; auditory stimuli were played through a set of Sennheiser HD202 headphones. Instructions were provided in paper form, and participants were told to keep their gaze fixed on the centre of the screen, where a little dot acted as a fixation point. They were also instructed to sit relaxed and still and to avoid any eye or body movement during the recording session.

The task consisted in responding as accurately and as quickly as possible to artworks depicting fishes. The assignment was non-relevant to the experiment but was aimed at keeping the attention of participants on stimulation. Answers were provided pressing a joypad key with the index finger of the right or left hand, and the response hand was switched across trials. Stimuli were presented with EEvoke v2.2 stimulation software (ANT Neuro, Hengelo, The Netherlands) in random order across 4 mixed blocks. Each block comprised 48 pictures and 4 musical excerpts belonging to all emotional categories, and 4 or 5 target stimuli. Visual stimuli lasted 1,000 msec and the inter-stimulus interval (ISI) ranged between 600-800 msec; the background music lasted for ~25 sec. Each experimental sequence had the duration of ~ 1 min 30 sec and started with the presentation 3 warning signals of 1,000 msec ("Ready", "Set", "Go!") and ended with "Thank you".

Visual and auditory stimuli were paired so as to create emotional congruence or incongruence in the audio-visual stimulation. Incongruent pairs were: happiness-fear, happiness-sadness, relax-fear, relax-sadness, sadness-happiness, sadness-relax, fear-happiness, fear-relax; where the first category refers to the video and the second one refers to the audio stimulation (Figure 2.1).

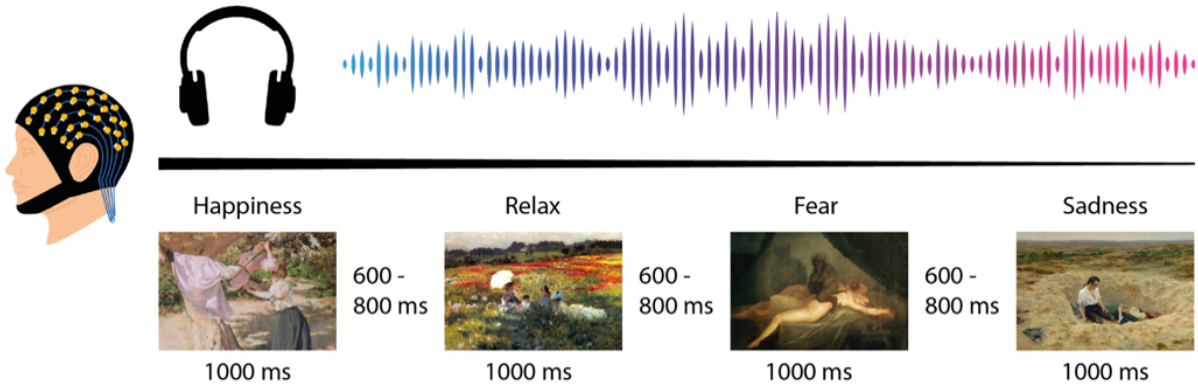


Figure 2.1. Experimental paradigm. Visual stimuli were paired with either congruent or incongruent background music. Both music and artworks emotional categories were Happiness, Relax, Fear, Sadness. ISI lasted 600 – 800 msec. In the figure, you can see an example for each emotional category.

EEG recording and analysis

EEG signals were continuously recorded from 128 sites located according to the international 10-5 system (Oostenveld & Praamstra, 2001) and mounted on a standard EEG cap. Sampling rate of 512 Hz (a band-pass filter of 0.16–70 Hz), horizontal (hEOG) and vertical (vEOG) eye movements were additionally recorded, and linked mastoids that served as the reference lead. Electrode impedance was maintained below 5kΩ. Computerized artifact rejection and manual eye-inspection was performed prior to averaging, in order to remove epochs in which eye movement, muscle-related potentials or amplifier blocking occurred. The artifact rejection criterion was a peak-to-peak amplitude exceeding 40 μV and resulted in a rejection rate of ~ 15%. EEG epochs were synchronized with the onset of the artworks and analysed using EEProbe v2.2 software (ANT Software, Enschede, The Netherlands). Event related potentials (ERPs) were averaged off-line from 100 msec before to 1,000 msec after stimulus onset and averages were off-line filtered (band-pass 0.16-30 Hz). ERPs components were identified and measured with reference to the average baseline voltage over the interval of − 100 to 0 msec at the sites and latencies at which they reached their maximum amplitudes.

As for visual ERP components, mean amplitude of N170, P300, and Late Positivity (LP) ERP components were measured. As for acoustic ERP component N400 was measured. Finally, multimodal Late Positivity (LP) was measured.

N170 response was recorded from posterior sites (Oz, O1, O2) in the 150-180 msec temporal window; P300 responses was recorded from parietal sites (POz, Pz) in the 300-500 msec temporal window; visual LP was recorded from antero-frontal sites (AF3, AF4, AF7, AF8) in the

800-1000 msec temporal window. N400 response was recorded from central-parietal sites (C1, C2, Cz, CPz, CCP1h e CCP2h) in the 400-500 msec temporal window and the multi-modal LP from anterior-frontal (AFF1, AFF2, F1, F2, F7 e F8) sites in the 700-900 msec temporal window. The choice of the electrode clusters and time windows were made based on previous literature (Calma-Roddin & Drury, 2020; Olofsson et al., 2008; Proverbio et al., 2020a; Proverbio et al., 2020b; Proverbio et al., 2020c).

The mean amplitude values of each ERP component underwent repeated measure ANOVA. For the visual N170, ERPs were computed regardless the auditory stimulation. Two ANOVAs were carried out on the visual N170 component, one with emotion (happiness, relax, sadness, fear) and recording site (O1, Oz, O2) as factors, the other one as emotional valence (positive or negative) and the same recording sites. Likewise, for the visual P300 factors for the first ANOVA were emotion (happiness, relax, sadness, fear) and recording site (POz, Pz); for the second ANOVA were emotional valence (positive or negative) and the same recording site. The last visual ERP component analysed was the LP: factor for the first ANOVA were emotion (happiness, relax, sadness, fear), recording site (AF3, AF4, AF7, AF8), and hemisphere (left or right); while for the second ANOVA were emotional valence (positive or negative), recording site, and hemisphere (left or right).

On the other hand, ERPs for the acoustic N400 were computed regardless the visual stimulation. Similarly to the ERP visual components, we computed the first ANOVA with emotion (happiness, relax, sadness, fear) and recording site (C1, C2, Cz, CPz, CCP1h e CCP2h) as factors, and a second ANOVA with emotional valence (positive or negative) and recording site as factors.

To access the cross-modal interaction between the two different sensory modalities, we analysed the LP mean amplitude values, performing an ANOVA with congruency between the two categories of stimuli as a factor. That is, the three levels of the congruency factor were congruent negative (both visual and auditory stimulations with negative valence), congruent positive (both visual and auditory stimulations with positive valence), and incongruent (different valence between the visual and the auditory stimulation). Factors of the ANOVA were also recording site (AFF1, AFF2, F1, F2, F7 e F8) and hemisphere (left or right). All the ANOVAs were carried out using Statistica software v.10 (Statsoft). Tuckey's post-hoc comparisons were used. The effect size for the statistically significant factors was estimated using partial eta squared ($\eta p2$) and the Greenhouse-Geisser correction was applied to account for non-sphericity of the data.

Standardized weighted Low Resolution Electromagnetic Tomography (swLORETA; Palmero-Soler et al., 2007; Pascual-Marqui et al., 1994, 2002) were performed on the LP component related to the 4 experimental conditions: positive music, negative music, positive artworks, negative artworks. Source space properties were grid spacing (5 mm) and Tikhonov regularization (estimated

SNR = 3). swLORETAs were applied using ASA4 software (ANT Neuro, Hengelo, The Netherlands).

Behavioral data were not statistically analysed, as the fictious task was non-relevant to the experiment, but reaction times and accuracy were collected. The mean value of the accuracy in detecting the target stimuli was 99,72%, meaning that participants were paying attention to the stimulation provided.

Results

Behavioral data were descriptively analysed. Hits percentage was 99.72%, with a mean response time of 539 msec. As target stimuli were just attention-capturing stimuli, no further analysis were carried out.

ERPs to artworks

Visual N170 was recorded at occipital electrodes (O1, Oz, O2), regardless auditory stimulation. The ANOVA performed on the N170 (150-180 msec) mean amplitude values yielded to the significance of the emotion category (happiness, relax, sadness, fear) factor [$F(3, 51) = 3.27$, $p = .02$], with greater amplitude values in response to relaxing ($M = 2.88$ μV, $SD = 0.58$) than fearful stimuli ($M = 3.76$ μV, $SD = 0.53$). Also the electrode factor was shown significant, as amplitude was smaller in O2 ($M = 4.21$ μV, $SD = 0.61$), as compared to both Oz ($M = 2.75$ μV, $SD = 0.60$) and O1 ($M = 3.31$ μV, $SD = 0.55$). Finally, interaction between these two factors was significant [$F(6, 102) = 2.95$, $p = .01$].

An ANOVA on same sites and latencies was performed on the emotional valence factor (positive vs negative), which was shown significant [$F(1, 17) = 6.24$, $p = .02$], as also the electrode factor [$F(2, 34) = 12.88$, $p < .001$] did. N170 was greater in response to positive stimuli ($M = 3.16$ μV, $SD = 0.59$), as compared to negative stimuli ($M = 3.69$ μV, $SD = 0.56$).

Visual P300 (300-500 msec) was recorded at posterior sites (POz, Pz) regardless of the auditory stimulation. The ANOVA performed on the P300 mean amplitude values, yielded to the significance of the emotion category (happiness, relax, sadness, fear) factor [$F(3, 51) = 4.69$, $p = .005$], with greater values in response to sad ($M = 0.81$ μV, $SD = 0.61$) and fearful ($M = 0.64$ μV, $SD = 0.61$) stimuli, as compared to happy ($M = -0.36$ μV, $SD = 0.65$) or relaxing ($M = 0.45$ μV, $SD = 0.69$) stimuli, as confirmed by post-hoc comparisons. Moreover, electrode (POz, Pz) factor was shown significant [$F(1, 17) = 59.86$, $p < .001$], with greater P300 amplitudes at POz ($M = 1.83$ μV, $SD = 0.71$), as compared to Pz site ($M = -1.05$ μV, $DS = 0.55$).

An ANOVA on same sites and latencies was performed on the emotional valence factor (positive vs negative), which was shown significant [F (1, 17) = 6.92, p = .01]. P300 was greater in response to negative (M = 0.73 μV, SD = 0.58) stimuli, as compared to positive (M = 0.04 μV, SD = 0.66) stimuli.

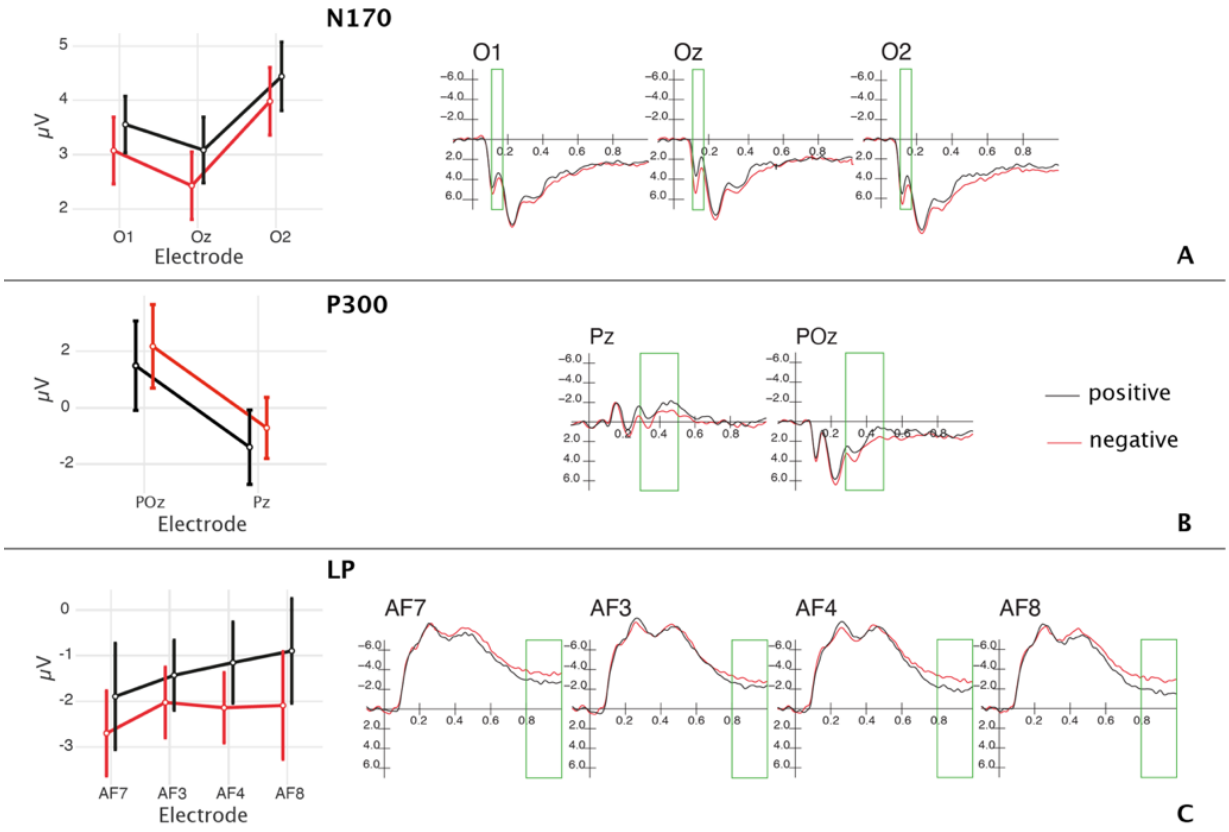


Figure 2.2. (Left) mean area amplitudes of N170 (A), P300 (B) and LP (C) components recorded to visual artworks regardless the auditory stimulation, but as a function of the emotional valence (positive vs negative) of the stimulus. (Right) Grand-average ERP waveforms recorded to emotionally positive and negative paintings. The highlighted time windows correspond to 150 – 180 msec for the N170 (A), 300 – 500 msec for the P300 (B), and 800 – 1000 msec for the LP (C).

Visual LP (800 – 1000 msec), averaged regardless the auditory stimulation, was recorded at antero-frontal sites (AF3, AF4, AF7, AF8). The ANOVA performed on the LP mean amplitude values yielded the significance of the emotion category factor (happiness, relax, sadness, fear) [F (3, 51) = 5.68, p = .001], with greater values in response to happy (M = -1.19 μV, SD = 0.41) stimuli, as compared to sad (M = -2.34 μV, SD = 0.41) or fearful (M = -2.13 μV, SD = 0.34) stimuli, as confirmed by post-hoc comparisons. Additionally, interaction between emotion and electrode factors was also significant [F (1, 17) = 5.38, p = .03].

An ANOVA on same sites and latencies was performed on the emotional valence factor (positive vs negative), which was shown significant [F (1, 17) = 12.95, p = .002]. LP was greater in response to positive (M = -1.34 μV, SD = 0.40) stimuli, as compared to negative (M = -2.24 μV, SD = 0.33) stimuli. Also, the interaction between emotional valence and electrode factors was shown significant [F (1, 17) = 5.38, p = .03].

ERPs to music pieces

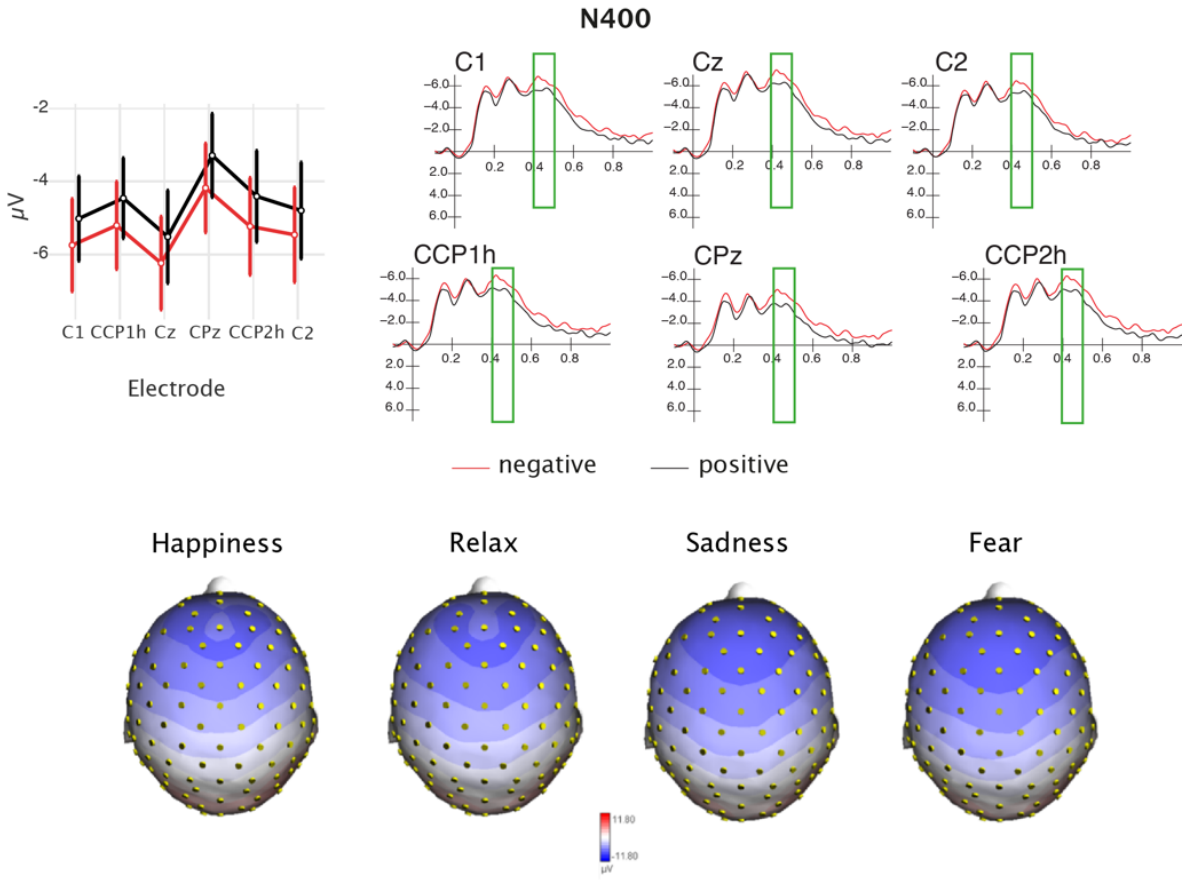


Figure 2.3. (Left) Mean area amplitudes of N400 component recorded to music regardless the visual stimulation, but as a function of the emotional valence of the stimulus. (Right) Grand-average ERP waveforms recorded to emotionally positive and negative music. The highlighted time windows correspond to 400 – 500 msec for the N400 (Bottom) Topographical maps show the voltage distribution of brain potentials recorded in the 400–500 msec time window, as a function of music emotional category.

Acoustical N400 (400 – 500 msec) was recorded at central electrodes (C1, C2, CCP1h, CCP2h, Cz, CPz), regardless the visual stimulation. The ANOVA performed on the N400 mean amplitude values yielded the significance of the emotion (happiness, relax, sadness, fear) factor [F (3, 51) = 3.00, p = .03], with greater values in response to sad (M = -5.46 μV, SD = 0.69) stimuli, as compared to happy (M = -4.33 μV, SD = 0.65) stimuli, as confirmed by post-hoc comparisons.

Furthermore, electrode (C1, C2, CCP1h, CCP2h, Cz, CPz) factor was shown significant [F (5, 85) = 21.64, p < .001]. Mean amplitude was greater in Cz (M = -5.87 μV, SD = 0.58), as compared to other sites. Activation in CPz (M = -3.72 μV, SD = 0.54) was smaller, as compared to other electrodes.

An ANOVA on same sites and latencies was performed on the emotional valence factor (positive vs negative), which was shown significant [F (1, 17) = 5.63, p = .02], with larger N400 responses to negative stimuli (M = -5.34 μV, SD = 0.58), as compared to positive stimuli (M = -4.57 μV, SD = 0.55). The electrode factor [F (5, 85) = 21.64, p < .001] also yielded significance with larger N400 amplitudes in Cz (M = -5.87 μV, SD = 0.58), as compared to other sites.

ERPs to audio-visual stimuli

ERPs to audio-visual stimuli were averaged as a function of the emotional congruence of the audio-visual stimulation, regardless the individual specificity of the emotion evoked by music or artworks. Late Positivity (LP) was recorded at frontal sites (AFF1, AFF2, F1, F2, F7, F8) in 700 – 900 msec latency range. The ANOVA performed on the LP amplitude values yielded the significance of the congruency (congruent positive, congruent negative, incongruent) factor [F (2, 34) = 6.40, p = .004], with greater values in response to congruent and positive (M = -1.61 μV, SD = 0.38) stimulation, as compared to congruent but negative (M = -3.03 μV, SD = 0.48) stimulation, or incongruent (M = -2.26 μV, DS = 0.29) stimulation, as confirmed by post-hoc comparisons. Also, electrode (AFF1, AFF2, F1, F2, F7, F8) factor was shown significant [F (2, 34) = 3.37, p = .04] with grater values at antero-frontal sites (AFF1-AFF2: M = -2.62 μV, DS = 0.35; F7-F8 (M = -1.88 μV, DS = 0.29). Moreover, the significant interaction between congruency and hemisphere factors [F (2, 34) = 10.56, p < .001] and relative post-hoc comparisons showed that the LP was greater in response to congruent and positive stimulation over the right hemisphere (M = -1.37 μV, SD = 0.35) than the left hemisphere (M = -1.85 μV, SD = 0.42). In addition, LP was larger in response to congruent and negative stimulation over the left hemisphere (M = -2.80 μV, SD = 0.43), as compared to the right hemisphere (M = -3.25 μV, SD = 0.56). Finally, an interaction between congruency, electrode, and hemisphere factors [F (4, 68) = 3.29, p = .01] was found significant. More specifically, statistical differences with respect to the congruency of the audio-visual stimulation and between hemispheres were found at frontal (i.e., F1, F2, F7, F8) but not anterior-frontal (AFF1, AFF2) sites. As can be seen in Figure 2.4, differences in congruency factor were larger over the right than left hemisphere, where the LP was more sensitive to the congruency of the audio-visual stimulation.

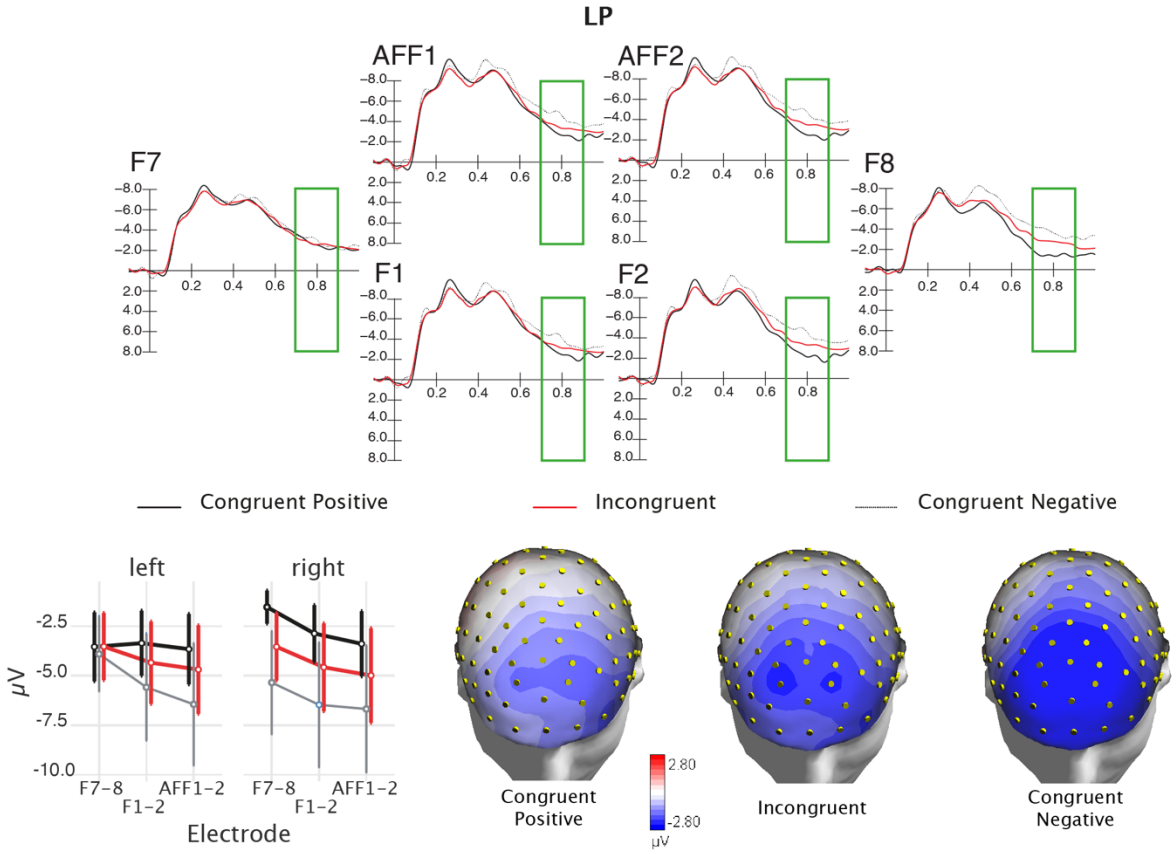


Figure 2.4. (Top) Grand-average ERP waveforms recorded at anterior sites as a function of the congruence between auditory and visual stimuli. The highlighted windows correspond to the 700 – 900 time window. (Bottom left) Mean area amplitude of the LP component recorded as a function of the congruence between auditory and visual stimuli. (Bottom right) Topographical maps show the voltage distribution of brain potentials recorded in the 700 – 900 msec time window.

Source Reconstruction - swLORETA

To locate the possible neural sources of electric potentials recorded in response to the different types of stimuli (positive and negative art, positive and negative music), several swLORETA source reconstructions were performed in the 700 – 900 msec time-window, corresponding to the LP component. For all 4 conditions, the strongest source of activity was localized within the left middle frontal gyrus (BA10) followed by the left superior frontal gyrus (BA8). Common areas of activation were the right middle temporal gyrus (BA21) and the precuneus. Table 2.1 reports a list of all active electromagnetic dipoles that were active in the different types of stimulation.

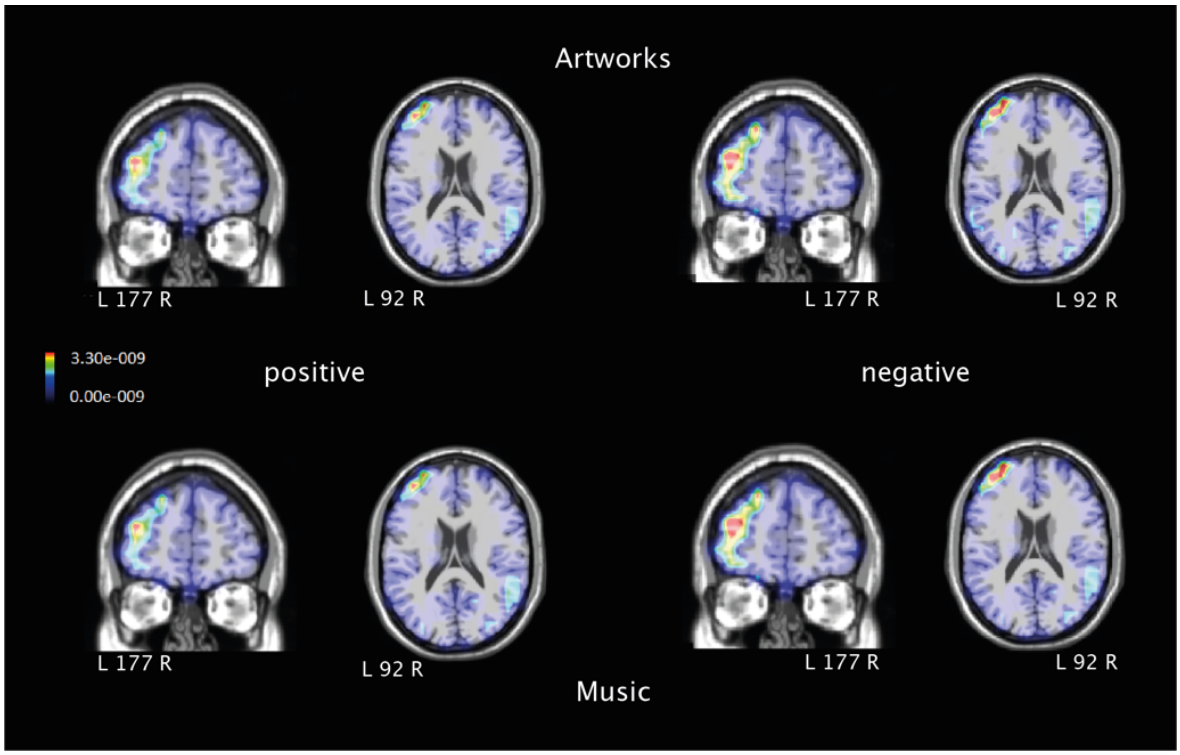


Figure 2.4. swLORETA. Coronal and axial views of the active sources relative to: positive artworks (top left), negative artworks (top right), positive music (bottom left), and negative music (bottom right). In all conditions, the most active brain area was the left Middle Frontal Gyrus (BA10). The source reconstruction was applied to the LP time window (700 – 900 msec).

Table 2.1. Active electromagnetic dipoles (with Talairach coordinates in mm) explaining the surface voltage recorded in the 700 – 900 msec time window according to swLORETA inverse solution applied.

Dipole No.	positive Artworks (magnitude)	negative Artworks (magnitude)	x [mm]	y [mm]	z [mm]	Hem.	Lobe	Gyrus	BA
1022	3,85	4,56	-39,5	51,5	19,5	L	F	Middle Frontal Gyrus	10
970	2,89	4,31	-19,5	41,5	49,5	L	F	Superior Frontal Gyrus	8
118	2,28	2,53	30,5	-68,5	29,5	R	O	Precuneus	31
625	1,72	1,84	50,5	1,5	-40,5	R	T	Middle Temporal Gyrus	21

Dipole No.	positive Artworks (magnitude)	negative Artworks (magnitude)	x [mm]	y [mm]	z [mm]	Hem.	Lobe	Gyrus	BA
1022	3,81	4,62	-39,5	51,5	19,5	L	F	Middle Frontal Gyrus	10
970	3,40	3,73	-19,5	41,5	49,5	L	F	Superior Frontal Gyrus	8
20	3,28	3,76	40,5	-88,5	-10,5	R	O	Inferior Occipital Gyrus	18
118	2,39	2,40	30,5	-68,5	29,5	R	O	Precuneus	31
256	2,39	2,34	50,5	-48,5	19,5	R	T	Superior Temporal Gyrus	22
94	2,23	2,01	-19,5	-78,5	49,5	L	P	Precuneus	19
625	1,60	2,00	50,5	1,5	-40,5	R	T	Middle Temporal Gyrus	21
497	1,60	1,71	-29,5	-18,5	-40,5	L	Limbic	Uncus	20

Discussion

The aim of the present study was to investigate the capability of both music and artworks to convey specific emotions and to explore interactions between these two sensory modalities. Thus, we designed a study pairing paintings and musical excerpts evoking different emotions and examined the processing of the conceptual congruency of the induced emotional content.

We expected that listening to a musical piece inducing an incongruent emotional information with respect to the emotional state induced by a painting would affect the emotional comprehension of the stimulus as a whole. If this was true, we would have found larger LP for negative than positive stimuli, and we expected this component to be right lateralized (Cacioppo et al., 1996; Höfel & Jacobsen, 2007; Olofsson et al., 2008; Proverbio et al., 2020), namely greater. Additionally, we expected that if P300 ERP component was modulated by the arousal of the stimulus, it would have shown greater amplitudes in response to negative stimuli, compared to positive ones (Cuthbert et al., 2000). Finally, we hypothesized that the incongruence of the audio-visual stimulation would elicit an N400 response, as was found in Proverbio and co-workers (A. M. Proverbio & Russo, 2021). Amongst other results, they found how N400 response showed a strong increase in amplitude induced by the emotional incongruence of audio-visual stimuli (regardless of the specific content, e.g., joy, fear, or sadness). N400 has been effectively used to investigate the semantic processing of language (Kutas & Federmeier, 2011), pictures (Barrett & Rugg, 1990), pictures and words (Nigam et al., 1992), sounds and musical gestures (Proverbio et al., 2015a), skilled motor actions (Proverbio et al., 2012), sign language (Proverbio et al., 2015b), words and music (Daltrozzo & Schön, 2008; Goerlich et al., 2011), space and music (e.g., low tones associated with basement, or ascending pitch steps associated with staircase) (Koelsch et al., 2004; Zhou et al., 2014). Overall, the literature suggests the existence of an amodal and shared conceptual system, indexed by N400 response (to incongruity) in different domains, such as language, music, pictures, actions.

ERPs to Artworks

Analysis performed on the visual N170 to visual art showed that it was sensitive to stimulus emotional content and especially polarity, being larger to positive stimuli. Indeed, N170 showed differences between positive and negative stimuli. This is in line with what Olofsson and colleagues (2008) posit in their review work, that is short latency ERP index early sensory processing within the extra-striate visual cortex and are sensitive to physical factors of the stimulus, as well as its sensitivity to selective attention processes. Furthermore, Lee and collaborators (2017) found this component to be sensitive to affective valence and arousal of the stimuli finding greater negativity to positive, as

compared to neutral, stimuli. The same effect was found by Cannas Aghedu and colleagues (2021), where N1 amplitudes in response to positive than negative pictures while listening to romantic music. These results show how early ERP components reflect not only physical characteristic of the stimulus, but also their emotional valence.

P300 amplitudes were analysed in 300-500 msec post-stimulus in response to artworks, regardless the auditory stimulation. In our study, P300 showed larger amplitudes in response to negative stimuli, namely pictures evoking sadness and fear, than to positive stimuli. Consistent with this effect are the results by Bernat and colleagues (2001), showing larger P300 to negative than positive words or sentences. In the same vein, Proverbio and co-workers (2021) found greater P300 responses to sad than happy or frightened faces in a study investigating the effect of emotional incongruity between a phrase and a musical soundtrack. Therefore, the P300 reflects here, as in other cases, the affective valence of the stimulus (Conroy & Polich, 2007; Olofsson et al., 2008), including the emotional distress induced by stimulus observation (Schupp et al., 2004). Also, the major factor of P300 amplitude is task-relevance (Polich & Kokb, 1995); however, in the present study, the visual stimuli were not target stimuli and so non-task related. Therefore, it can be assumed that in addition to the attentional orientation generated by target stimuli (Olofsson et al., 2008), the P300 reflects the salience generated by the stimuli. In our study, we found more negative P300 amplitudes in response to negative than positive stimuli. In other words, negative visual stimuli are able to attract more attention than positive ones.

The last ERP component analysed, relative to visual stimuli, is the Late Positivity (LP). LP amplitudes were analysed in the 800-1000 msec post-stimulus onset, showing greater deflections to positive than to negative stimuli. According to Cuthbert (2000), the LP is modulated by the arousal of the stimuli, towards an increased activation relative to high arousing pictures. Thus, the authors state that late positive wave indicates a selective processing of emotional stimuli (Cuthbert et al., 2000). Also in a study by Yao and colleagues (2016) low-arousal positive words were associated with a reduction of Late positive complex (LPC) amplitudes, compared with high-arousal positive words. This suggests a higher activation in response to arousing stimuli, and this occurs with different types of stimuli. As a matter of fact, the same was found by Proverbio and co-workers in response to music, non-verbal vocalizations, and speech processing (respectively Proverbio et al., 2020a; Proverbio et al., 2020b). These results indicate that the LP is modulated by the valence of the stimuli, regardless the sensory channel.

ERPs to Music

The auditory N400 response was analysed regardless the visual stimulation in the 400-500 msec time window. This ERP component showed larger amplitudes in response to negative than to positive stimuli. A similar pattern of result was found in previous investigations (e.g., Proverbio et al., 2020; Proverbio & Piotti, 2021) , reporting larger N400 responses to negatively valenced voices or instrumental music. N400 was found to be greater also in response to emotionally incongruent pairs as well as to other types of incongruent multimodal information (Goerlich et al., 2011; Kamiyama et al., 2013; Nigam et al., 1992; Zhou et al., 2014). This may indicate that the auditory affective information is processed in part by a common neural mechanism for different kind of stimuli.

Multimodal Integration

To investigate the multimodal integration of visual and auditory stimuli, the Late Positivity (LP) was analysed in the 700-900 msec post-stimulus time window, comparing the congruency and the valence of the audio-visual stimulation. The late positivity is a slow positive ERP wave often described to be modulated by pleasantness (see Hajcak et al., 2010; Olofsson et al., 2008 for a review). The hypothesis was that the stimulus emotional congruence would have led to greater amplitudes, possibly indexing enhanced aesthetic pleasantness (van Peer et al., 2014). Moreover, the increased LP for emotional compared to neutral pictures has been shown to be larger for more intense and arousing stimuli (Cuthbert et al., 2000). A more positive LP has been shown in response also to pleasant, compared to neutral or unpleasant stimuli (Codispoti et al., 2009; Hajcak et al., 2010). Our ERP data showed greater positive potentials to congruent positive affective information than congruent negative and incongruent pairs. This piece of results might suggest that congruent negative pairs were less pleasurable than emotionally conflicting information (i.e., incongruent ones). According to van Peer and colleagues (2014), the processing of the pleasantness of the stimulus is not an independent process, but it is linked with the novelty of the stimulus itself. To better understand this, we can refer to the motivated attention, the notion that emotion directs attention, hence facilitating subsequent processing (Sabatinelli et al., 2005). Through this, we integrate attention and emotion, in line with the model by Markovic (2012), where the role of attention is emphasized in order to reach an aesthetic experience. We can bind this model with what Fröber and Thomaschke (2019) stated, which is that the context can indeed boost the aesthetic experience. As a matter of fact, to watch a film in a theatre, rather than at home, or to view an artwork in a museum, rather than in a laboratory, enhances the aesthetic experience (Fröber &

Thomaschke, 2019; Specker et al., 2017). This is also understandable given that experiencing whichever piece of art in an environment especially designed to appreciate it, can help the viewer focus and better understand the message conveyed. To understand the message, the viewer has also to evaluate the stimulus. This is what Höfel and Jacobsen (2007) investigated, namely whether the aesthetic evaluation of a graphic pattern occurs spontaneously. They asked participant either to view or to contemplate, meaning that participants were asked to assess the aesthetic value, graphic patterns, with no aesthetic-related response needed in either condition. They distinguished between early fronto-central negativity, which reflects processes of aesthetic judgment, and late positivity, which reflects both intentional aesthetic contemplation and aesthetic judgment (Höfel & Jacobsen, 2007). The LP was found to be right lateralized for the aesthetic evaluation, as found in Cacioppo and colleagues (1996). However, since participants were not asked for an overt aesthetic response, Höfel and Jacobsen concluded that the lateralization of the LP occurs even if no judgement was required, and consequently that an aesthetic evaluation of the stimulus appears spontaneously. In the same way, in the present study participants were instructed to simply watch the paintings on the screen without expressing any judgment and the results show a lateralization in the LP, which was larger over the right hemisphere than the left hemisphere.

Source Reconstruction

Four different swLORETAs were applied to surface LP potentials in LP latency range (700-900 msec), in order to investigate the neural underpinnings of aesthetic appreciation of music and visual art. The following conditions were analysed: positive visual art, negative visual art, positive music listening, and negative music listening. As for the visual stimulation, in both positive and negative emotional conditions it was found the activation of the left Middle Frontal Gyrus (MFG) and the left Superior Frontal Gyrus (SFG), and the right Precuneus. According to previous literature, these areas would indeed contribute to aesthetic experience. For example, Ticini (2017) found the Orbitofrontal Cortex (OFC) to be associated with object-appraisal mechanisms and the left dorsolateral Prefrontal Cortex (DLPFC) was found to be involved in the network of aesthetic preference (Cattaneo et al., 2014).

As for the auditory stimulation, active electric dipoles in both positive and negative conditions were the SFG, MFG, the Superior Temporal Gyrus (STG), and the Precuneus. All these areas were found to be active in the study by Proverbio and collaborators (2020), where the authors linked this activity to the processing of the emotional content of the stimuli, whether they were linguistic stimuli or in response to digitally converted speech into music. In addition, Proverbio and Piotti (2021) found MFG to be active during music listening, and Tabei (2015) found the precuneus to be

implicated in the evaluation of one's own cognitive emotional changes during music listening. This could indicate that the activation found in the present study reflect the processing of the emotional content within an auditory stimulus.

Brain activation in response to positive and negative stimulation regardless of sensory modality was also investigated. MFG, SFG, Precuneus, and MTG were found to be active in all conditions, thus suggesting that they were engaged in the emotional processing of both visual and musical art, which shared a common neural mechanism. The involvement of frontal areas is in line with what Belin and colleagues (2008) found, that is a right middle temporal activation for processing negative vocalizations and an inferior frontal activation for processing positive vocalizations. The same was found in the study by Proverbio and co-workers (2020) when comparing human vocalization, positive and negative, and violin musical stimuli digitally extracted from them. Again, activity in the middle frontal gyrus has been linked with emotional prosody in speech (Buchanan et al., 2000). Additionally, the activation of MFG (BA10) is described as a key structure implicated in music listening and appreciation (Altenmüller et al., 2014; Burunat et al., 2014; Escoffier et al., 2013; Proverbio & De Benedetto, 2018; see Pando-Naude et al., 2021 for a review). Besides, MFG is also part of the reward network, proving the rewarding and thus pleasurable effect of music (Koelsch, 2020). In this case, this area could be activated in response both the pleasure elicited by artworks (Ferenczi et al., 2016) and by music (Brattico et al., 2020; Koelsch, 2014a), and this can indicate how this area plays a role in both visual and auditory emotional processing.



Study 3:

Semantic Interaction between Music and Artworks

Introduction[3]

One way to convey meaning with music is through evoked emotions (Plourde-Kelly et al., 2021; Proverbio et al., 2020). Nevertheless, emotions are not the only semantic concept the music is able to convey (Koelsch et al., 2004; Sollberger et al., 2003). Koelsch (2011c) have in fact broaden the use of the N400 ERP component also to musical stimuli. Although this ERP component is typically linked to the language domain (Kutas & Federmeier, 2011), more and more recent studies have extended both the application and the interpretation of the N400 (Daltrozzo & Schön, 2008; Kanske et al., 2011; Proverbio & Riva, 2009; Weimer et al., 2019). More specifically, Koelsch and colleagues (2004) investigated both how the N400 ERP component elicited by a word can be modulated by the semantic meaning emerging from the musical information preceding the word, and how the N400 can be directly elicited by music. In their study, participants were presented with a semantic priming paradigm, where the visual target word was presented after either a spoken sentence or a musical excerpt. The results show how the priming effect did not differ across the two conditions. Similarly, Daltrozzo and Schön (2008) developed the same priming paradigm adding a condition in which the linguistic context primed the musical processing. Also in this case, the authors concluded that music can convey semantic concepts. They also included that, given the similar topography of the N400 to both musical and linguistic targets, the conceptual processing of these domains may overlap.

Here, the aim was to access the capability of both music and artworks to convey semantic meaning, and the interaction across the two sensory modalities. In this work, then, we are going to combine these two kinds of stimuli. Thus, is a semantic communication possible also in pictures? In 1990, Batter & Rugg hypothesized that the semantic processing of a visual, pictorial, stimulus would access a different semantic memory system, compared to verbal stimuli (Barrett & Rugg, 1990). In 1996, Carretiè and co-workers supported the idea of specific mechanisms of emotional evaluation of visual stimuli, identifying in the N300 ERP component the more suitable index of a visual emotional stimulation (Carretié et al., 1996). In their more recent work, Franklin and colleagues confirm that the N300 effect is shown primarily on picture stimuli processing, but they ascribe its sensitivity to the semantic similarity of images (Franklin et al., 2007). This, however, strengthens the idea of a

[3] This chapter is based on the paper by De Benedetto, F., Orsenigo, C., Proverbio, A. M., Semantic interactions between music and artworks: an ERP study (in preparation).

semantic concept, and therefore evaluation and elaboration, in pictures. In a recent paper, Abush applies a linguistic semantic approach to pictures. Although different from syntax-driven as in language, the author explains how the semantic and pragmatic interpretation is possible also in the case of pictures and pictorial narratives (Abusch, 2015).

The aim of the present study was to investigate the semantic relationships and the influence of music on paintings' semantics. We paired visual stimuli, namely artworks painted between 1600 and the contemporary era, with auditory stimuli, namely musical excerpts from instrumental classical music pieces. These two categories of stimuli were combined considering their semantic categorization, as a result of the validation process (Proverbio et al., 2022). We hypothesized that listening to a musical piece whose semantic content that is incongruent with the semantic domain of a painting would interfere with the semantic recognition of the visual stimulus. Hence, we matched the artworks with either congruent or incongruent background music. As Wöhner and collaborators (2020) suggest, semantic interference occur also across modalities. The authors tested the model proposed by Chen & Spence (Chen & Spence, 2018) of cross-modal interactions of pictures, sound, and words. This model proposes different processing ways for accessing the semantic system by linguistic and non-linguistic, specifically sounds and pictures, stimuli. The authors set up 2 studies: (i) a picture categorization task and (ii) a word categorization task, both presented with an auditory cue. The main goal of their work was to establish the time-course of audio-visual semantic congruency effects elicited by naturalistic sounds and spoken words on visual picture processing. As a conclusion, Chen & Spence determine that both naturalistic sounds and spoken words enhance the speed of categorization of visual pictures or printed words, when congruent to one another. In addition, the presentation of an auditory cue that was semantically incongruent with the visual target induced an inhibitory effect. Finally, the authors led to the conclusion that this auditory semantic interference on early visual processing can be considered as a cross-modal prediction error (Chen & Spence, 2018). In testing Chen & Spence's interpretation, Whöner and co-workers suggest that, rather than modality-specific processing pathways, the semantic interference is the effect of a competitive selection of amodal semantic-lexical representation involved in both modalities (Wöhner et al., 2020). Additionally, Yuval-Greenberg and Deouell (2009) used a simultaneous presentation of pictures and vocalization of animals, being congruent or incongruent between each other. As a conclusion, the authors posit that the information can facilitate, if congruent, or interfere, if incongruent, with object recognition. Also, visual stimuli influence auditory processing more than vice versa, postulating a visual dominance in object recognition. However, the direction of the effect depends on the reliability of information conveyed.

With this in mind, we designed a study in which we could test both for music semantics and for semantic interference across modalities. To investigate the influence of background music on the categorization of the visual stimuli, we paired the musical excerpts with either semantically congruent or incongruent pictures.

Materials and Methods

Participants

26 University students (14 females) took part to the EEG study as volunteers. Age range was from 18 to 35 years (M = 23.27; SD = 2.93). All students were granted academic credits for their participation to the study. All participants report no history of brain damage, epilepsy, or current medications that could affect the brain electrical activity. Their handedness was accessed through the administration of the Italian version of the Edinburgh Inventory Questionnaire (Oldfield, 1971). The experiment was conducted with the understanding and written consent of each participant according to the Declaration of Helsinki (BMJ, 1991; 302: 1194), and with the approval of the University of Milano-Bicocca Ethics Committee (prot. n. RM-2019-218).

Data from 3 participants were discarded for excessive recording artifacts. Therefore, 23 participants (10 males, 13 females) with age range from 19 to 34 (M = 23.44; SD = 2.86) were included in the EEG data analysis.

Stimuli

After the validation process, 160 artworks and 16 musical pieces were selected. The validation procedure aimed to assign a semantic category to each stimulus, based on its representativeness of that category. Both visual and auditory stimuli were divided into 8 categories: *Lullaby*, *Play*, *Battle*, *Dance*, *Love*, *Tragic Love*, *Relax*, *Requiem*. The pictures have been taken from volumes of Art collections and Museum catalogues, or freely downloaded from Google Images. Most of them were realized with oil painting technique on canvas, and most of them were painted between 1600 and the contemporary era. The final screen size was adjusted, as it was the same (500x330 px; 18.3x13 cm) for all of them.

As for the auditory stimuli, all excerpts were normalized and balanced for intensity via Audacity software and exported in mp3 format. Also, the first and the last 3 seconds were faded in and out.



Lullaby



Play



Battle



Dance



Love



Tragic Love



Relax



Requiem

Figure 3.1. Example of an artwork per semantic category

Stimuli Validation

A semantic category was intended to be assigned to each stimulus as a result of the validation process, depending on how each stimulus was representative of that category. Categories were defined as described in Table 3.1. 20 participants (12 females, 8 males) with age between 18 and 30 (M = 24.4, SD = 3.05) volunteered for this part of the study. All participants had normal hearing, normal (or corrected to normal) vision, and none had suffered from previous or current psychiatric or neurological diseases. The experiment was performed with the understanding and written consent of each participant according to the Declaration of Helsinki (BMJ 1991; 302: 1194) and following the APA ethical standards (1992, American Psychological Association). The Ethical Committee of the University of Milano-Bicocca approved the study.

Table 3.1. Definition of the eight semantic categories applied to both visual and auditory stimuli.

LOVE: Two people in love who display affection and passion (for example by kissing or hugging).
LULLABY: Peacefully sleeping child/children with or without mom
BATTLE: Battle or war scene
GAME: Children playing
TRAGIC LOVE: Love between two people that ended in tragedy (suicide or death of one or both lovers)
DANCE: People dancing
RELAX: People who enjoy some relaxing time in a peaceful environment
REQUIEM: Crucifixion of Christ

Visual stimuli. We selected 273 artworks, 35 for each category. The selection took into account the picture's content (Table 3.2): number of people, animals, and context. We also checked for the absence of any musical instruments in the scene, so as not to create a bias when the visual stimuli would be associated with a piece of music. The task was to choose the category to associate with each painting and evaluate how much the artwork presented was representative of the class; the categories were always visible in the upper part of the screen. The answers were delivered using a 3-point Likert Scale, from 0 (not at all) to 2 (very much). The participants communicated the answers verbally to the researcher. All 273 paintings were presented randomly.

Auditory stimuli. The auditory stimuli were 22 excerpts of instrumental musical tracks. The selection took into consideration the key, minor or major, because it can influence the music's significance (Schutz, 2017). The association with a familiar soundtrack that could bias the semantic judgement was also considered. As it was for the visual stimuli, participants were asked to associate a category to each musical excerpt. They had to answer to a 3-Likert Scale (from 0 to 2) about the intensity of the representativeness. The participants communicated the answers verbally to the researcher.

Table 3.2. Description of artworks' composition and structure.

Category	**# Stimuli**	**Picture's Content**					
		Characters				**Context**	
						Outdoor	**Indoor**
Love	35	Man/Woman Couple 35				18	17
Tragic Love	34	Man/Woman Couple 34				15	19
Battle	35	Army 35		Weapons 35	Animals 22	35	-
Relax	34	Man/Woman Couple 9	Single Ch. 19	Group of People 6	Animals 6	25	9
Requiem	33	Crucified Christ 33	People 26		Animals 5	33	-
Play	34	Children 34			Animals 7	30	4
Lullaby	34	Children 24	Mother and child 10		Animals 5	4	30
Dance	34	Man/Woman Couple 13	Single Ch. 7	Group of People 14		5	23

Figure 3.2 and figure 3.3 show, respectively, the mean of the scores to the Likert Scale for artworks' and music's categorization, enlightening how each category is representative of itself. After the validation process, 160 artworks and 16 musical pieces were selected.

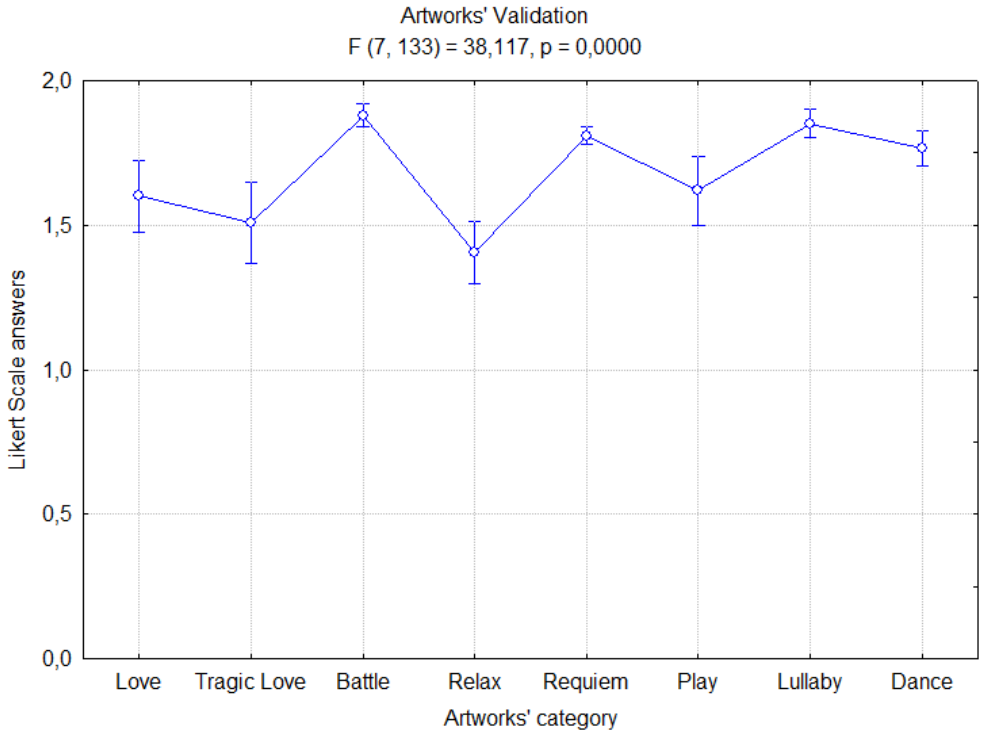


Figure 3.2. Mean of scores (from 0 to 2) to the Likert Scale for Artworks' categorization. All mean values are above 1, showing how each category is representative of itself ($F(7,133) = 38.12$, $p < 0.001$).

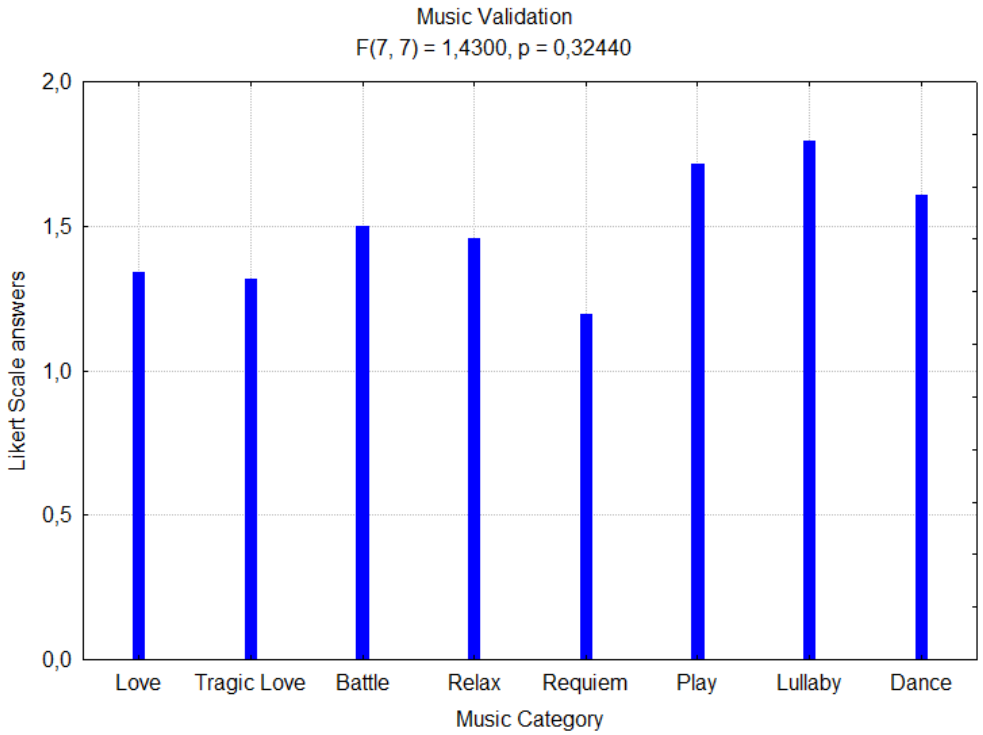


Figure 3.3. Mean of scores (from 0 to 2) to the Likert Scale for Music categorization. All mean values are above 1, showing how each category is representative of itself [$F(7,7) = 1.43$, $p = 0.324$].

Table 3.3. Congruent and Incongruent semantic categories used to create music/artworks stimulus pairs.

CATEGORY	CONGRUENT	INCONGRUENT
LOVE	Love	Battle, Requiem, Play, Lullaby
TRAGIC LOVE	Tragic Love	Relax, Play, Lullaby, Dance
BATTLE	Battle	Relax, Play, Lullaby
RELAX	Relax	Battle, Requiem, Play, Dance
REQUIEM	Requiem	Love, Relax, Play, Lullaby, Dance
PLAY	Play	Love, Tragic Love, Relax, Requiem,
LULLABY	Lullaby	Tragic Love, Battle, Requiem, Dance
DANCE	Dance	Tragic Love, Relax, Requiem, Lullaby

Procedure

Participants were comfortably seated in an electrically and acoustically shielded cubicle. Visual stimuli were presented on a PC screen located outside the cabin, 120 cm from the participants; auditory stimuli were played through a set of Sennheiser HD202 headphones (Sennheiser electronic GmbH & Co. KG). The instructions for the experiment were presented on paper, and participants were told to keep their gaze fixed on the centre of the screen, where a fixation point was. Additionally, they were told to sit relaxed and still while the recording was in progress.

The task consisted in responding as accurately and as quickly as possible to artworks depicting fishes. The assignment was intended to maintain participants' focus on the stimulation and was non-relevant to the experiment. Answers were given by pressing a joypad key with the index finger of the right or left hand, with the answer hand changing between trials. Stimuli were presented with EEvoke v2.2 stimulation software (ANT Neuro, Hengelo, The Netherlands) in a pseudo-randomized order across 4 sequences. Each sequence comprised 40 pictures and 4 musical excerpts, where half of the stimuli were paired congruently and half incongruently, and 4 or 5 target stimuli. Each picture was presented for 1,000 msec at the centre of the screen, after an ISI ranging from 600 to 800 msec

(Figure 3.4). The background music lasted for the whole sequence, which had a duration of ~ 1 min 20 sec.

EEG recording and analysis

EEG data were recorded using a standard EEG cap with 128 electrodes located according to the international 10-5 system (Oostenveld & Praamstra, 2001) using EEProbe (ANT Neuro, Hengelo, The Netherlands). Sampling rate was 512 Hz (a band-pass filter of 0.16–70 Hz), horizontal (hEOG) and vertical (vEOG) eye movements were additionally recorded, and linked mastoids that served as the reference lead. Electrode impedance was maintained below 5kΩ. Computerized artifact rejection and manual eye-inspection was performed prior to averaging, to discard epochs in which eye movement, muscle-related potentials or amplifier blocking occurred. The artifact rejection criterion was a peak-to-peak amplitude exceeding 40 μV and resulted in a rejection rate of ~ 15%. EEG epochs were synchronized with the onset of the artworks and analysed using EEProbe v2.2 software (ANT Software, Enschede, The Netherlands). Event related potentials (ERPs) were averaged off-line from 100 msec before to 1,000 msec after visual stimulus (artwork) onset and averages were off-line filtered (band-pass 0.16-30 Hz).

The mean amplitude area of the N2 ERP component was recorded from frontal-central sites (FCz, Cz, FC1, FC2) in the 250-350 msec time window. The mean amplitude area of the N400 ERP component was recorded also from frontal-central sites (FCz, Cz, FC1, FC2) in the 300-500 msec time window. The electrode clusters and time windows were selected on the basis of previous literature (Calma-Roddin & Drury, 2020; Franklin et al., 2007; Goerlich et al., 2011; Maidhof et al., 2010; Proverbio et al., 2020). Although well-supported by the literature, some authors argue that using such non-independent selection criteria may set a limit by exposing the study to spurious effects (Kriegeskorte et al., 2009). A repeated-measures ANOVA was applied to both N2 and N400 mean amplitude values. Factors of the rmANOVA were congruency (congruent or incongruent) and distinctiveness (high or low).

All the ANOVAs were carried out using Statistica software v.10 (Statsoft). Tuckey's post-hoc comparisons were used. The effect size for the statistically significant factors was estimated using partial eta squared (ηp2) and the Greenhouse-Geisser correction was applied to account for non-sphericity of the data.

A standardized weighted low-resolution electromagnetic tomography (swLORETA; Palmero-Soler et al., 2007; Pascual-Marqui et al., 2002) was applied on the ERP Difference Wave relative to N400 component, subtracting ERPs to Incongruent stimuli from ERPs to Congruent stimuli. Source

space properties were grid spacing (5 mm) and Tikhonov regularization (estimated SNR = 3). To compute the swLORETA ASA4 software was used (ANT Neuro, Hengelo, The Netherlands).

Behavioral data were not statistically analysed, as the task was non-relevant to the aim of the experiment, but reaction times and accuracy were collected. The mean value of the accuracy in answering to the target stimuli was 99,57%, meaning that participants were paying attention to the stimulation provided.

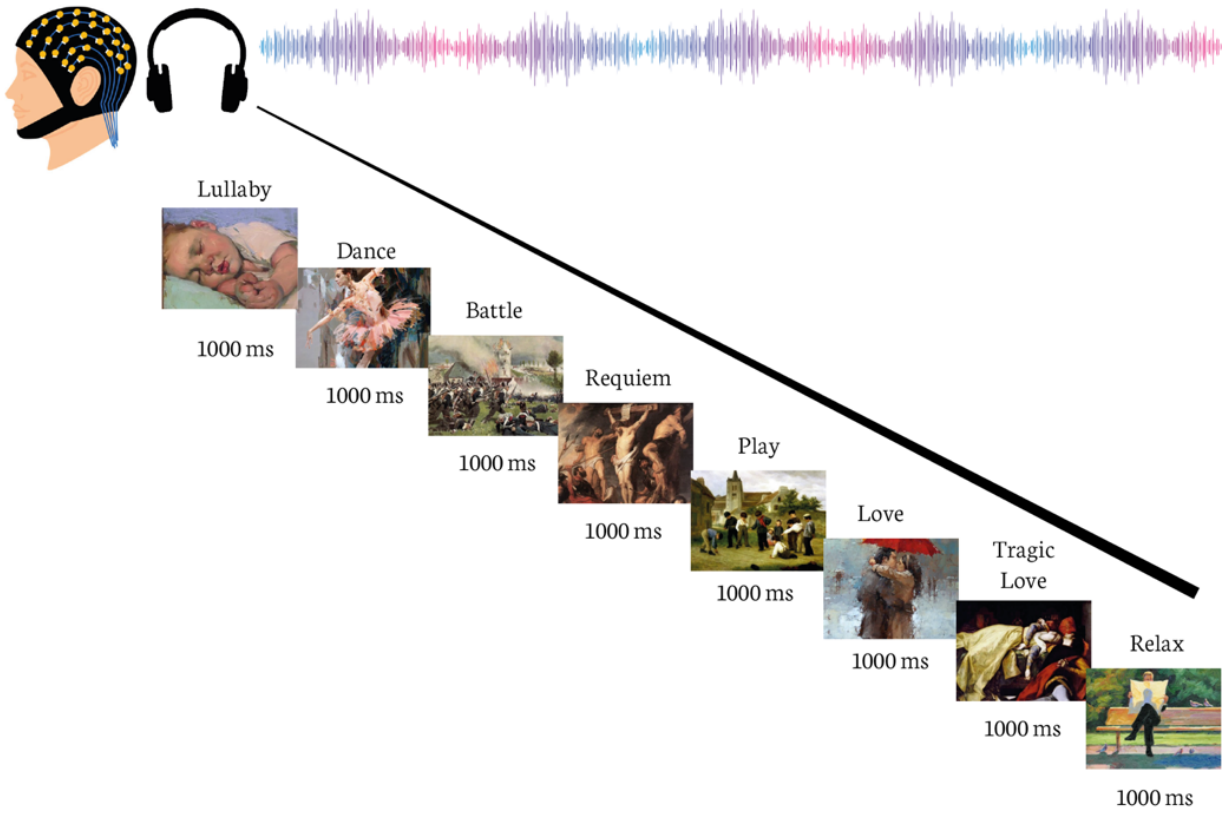


Figure 3.4. Time sketch of the experimental paradigm. Visual stimuli were paired with either congruent or incongruent background music. Both music and artworks semantics categories considered were Lullaby, Play, Battle, Dance, Love, Tragic Love, Relax, Requiem. ISI lasted 600 - 800 msec. In the figure, you can see an example for each semantic category.

Results

Behavioral data were descriptively analysed. Hits percentage was 99.57%, with a mean response time of 528.6 msec. As target stimuli were just attention-capturing stimuli, no further analysis was carried out.

N2 response (250-350 msec)

N2 component was analysed in the 250-350 msec time window, over frontal-central sites (FCz, Cz, FC1, FC2). The ANOVA performed on the mean amplitude values yielded the significant interaction between the congruency and the distinctiveness factors [F (1, 22) = 15. 99, p < .001; Fig. 3.5]. Also, the analysis revealed the significant effect of distinctiveness [F (1, 22) = 21.16, p < .001], with greater amplitudes in response to high distinctive stimuli (M = -5.76 μV, SD = 0.81) than low distinctive stimuli (M = -6.56 μV, SD = 0.80). Lastly, results show the significant effect of the electrode factor [F (3, 66) = 8.84, p < .001], with greater negativities over central electrode Cz (M = -5.80 μV, SD = 0.83), compared to frontal-central sites, namely FCz (M = -4.64 μV, SD = 0.54). Moreover, Figure 3.5 shows the topographical distribution of the voltage in response to congruent and incongruent stimuli with high and low distinctiveness.

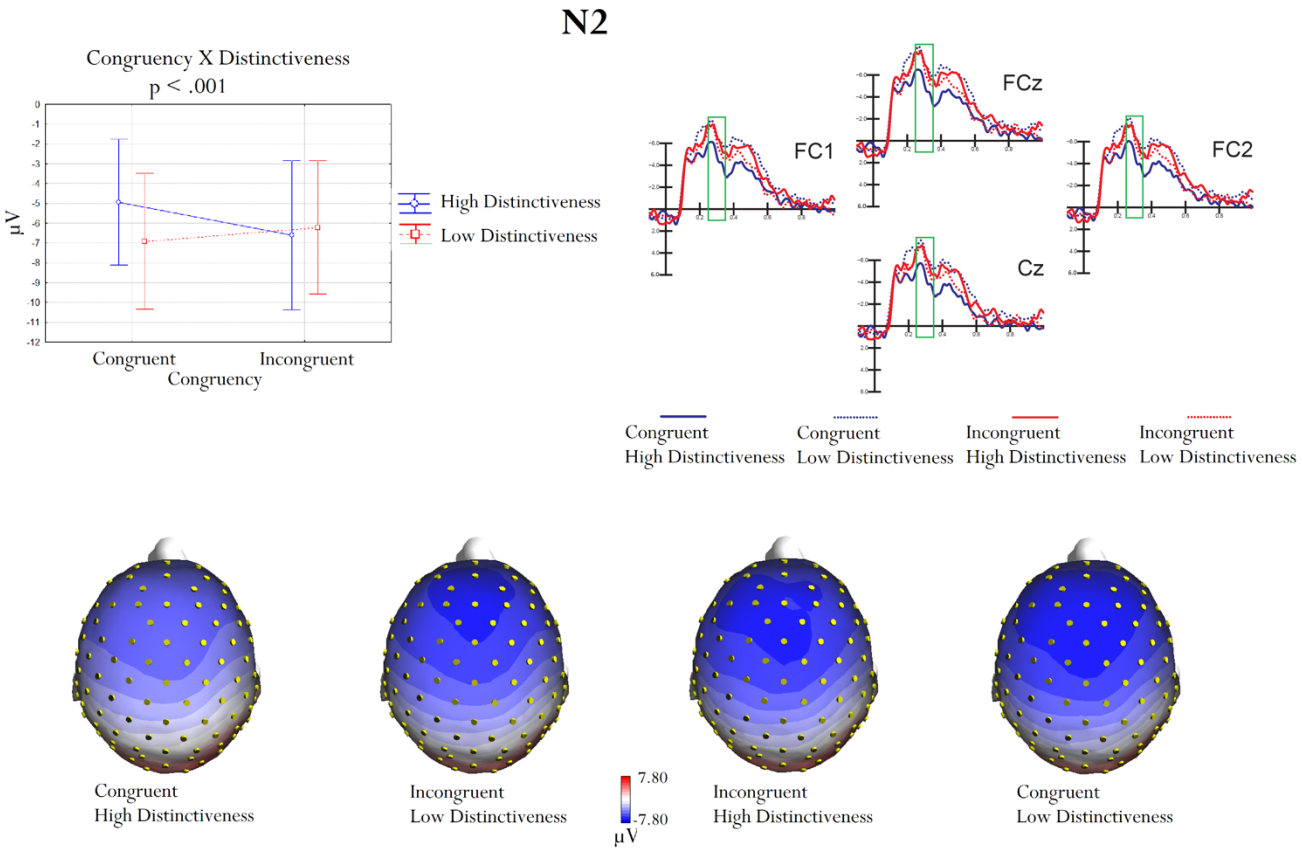


Figure 3.5. (Upper left) mean area amplitudes of N2 component recorded to the interaction between the congruency and the distinctiveness factors in the 250-350 msec time window. (Upper right) Grand-average ERP waveforms relative to the N2 component. The highlighted time windows correspond to 250 – 350 msec. (Bottom) Topographical maps show the voltage distribution of brain potentials recorded in the 250–350 msec time window, as a function of the distinctiveness of the multi-modal stimulation.

N400 response (300-500 msec)

N400 component was analysed in the 300-500 msec time window, over frontal-central sites (FCz, Cz, FC1, FC2). The ANOVA performed on the mean amplitude values yielded the significant interaction between the congruency and the distinctiveness factors [$F(1, 22) = 32.77$, $p < .001$; Fig. 3.6]. Furthermore, the analysis revealed the significant effect of distinctiveness [$F(1, 22) = 10.75$, $p = .003$], with greater amplitudes in response to high distinctive stimuli ($M = -4.64$ μV, $SD = 0.54$) than low distinctive stimuli ($M = -5.38$ μV, $SD = 0.52$). Finally, results show the significant effect of the electrode factor ($F(3, 66) = 13.53$, $p < .001$), with greater negativities over central electrode Cz ($M = -4.59$ μV, $SD = 0.54$), compared to frontal-central sites, namely FCz ($M = 5.45$ μV, $SD = 0.54$). Moreover, Figure 3.6 shows the topographical distribution of the voltage in response to congruent and incongruent stimuli with high and low distinctiveness.

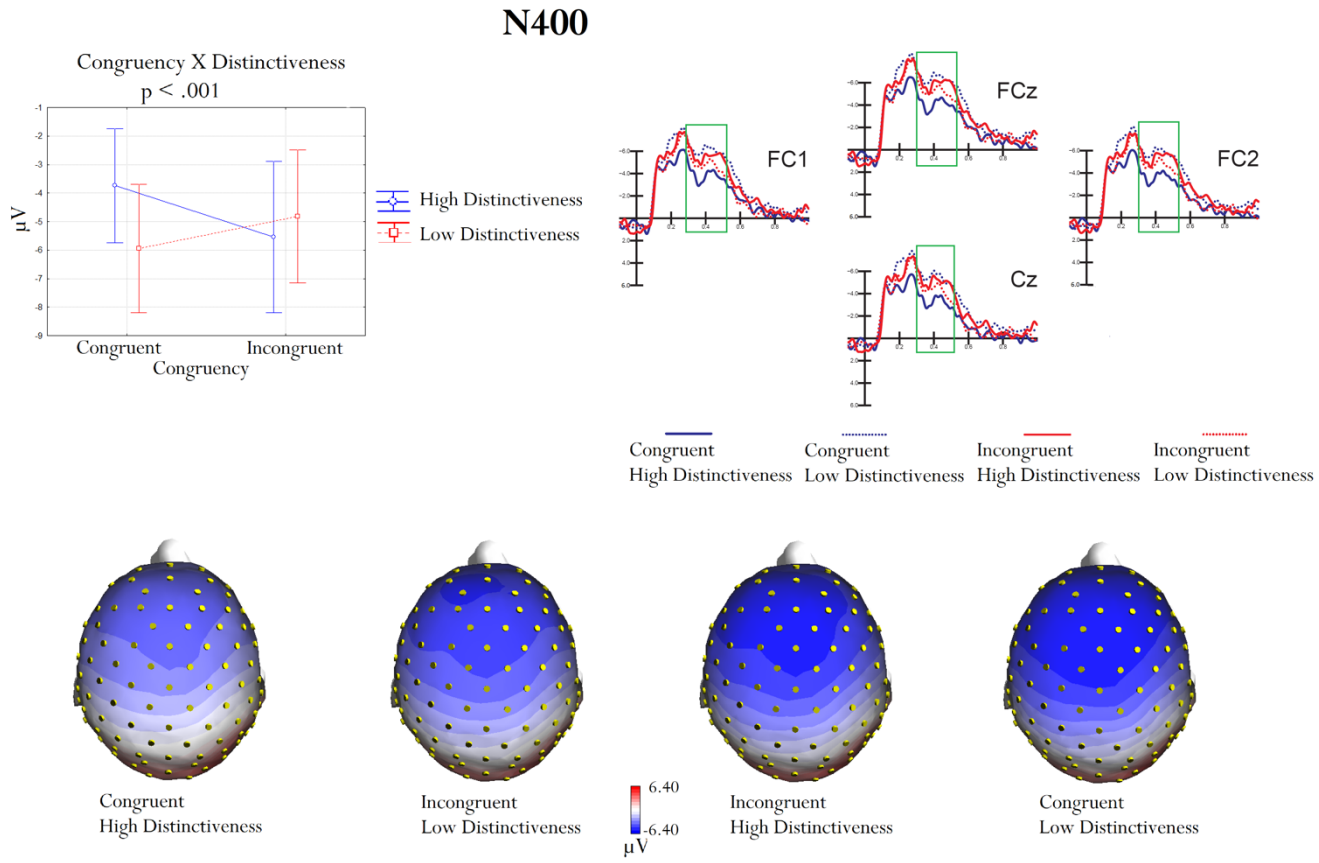


Figure 3.6. (Upper left) mean area amplitudes of N400 component recorded to the interaction between the congruency and the distinctiveness factors in the 300-500 msec time window. (Upper right) Grand-average ERP waveforms relative to the N400 component. The highlighted time windows correspond to 300 – 500 msec. (Bottom) Topographical maps show the voltage distribution of brain potentials recorded in the 300–500 msec time window, as a function of the distinctiveness of the multi-modal stimulation.

Source Reconstruction – swLORETA

The swLORETA inverse solution was applied on the Difference Wave obtained by subtracting ERP waveforms elicited by Congruent stimuli from that induced by Incongruent stimuli in the time window between 300 and 500 msec. To detect the intracortical generators of the N400 component (Fig. 3.7). The most active electromagnetic dipoles were the left Superior Frontal Gyrus (SFG; BA 8) and the right Inferior Temporal Gyrus (ITG; BA 20), amongst other areas. Table 3.4. reports a list of all active electromagnetic dipoles.

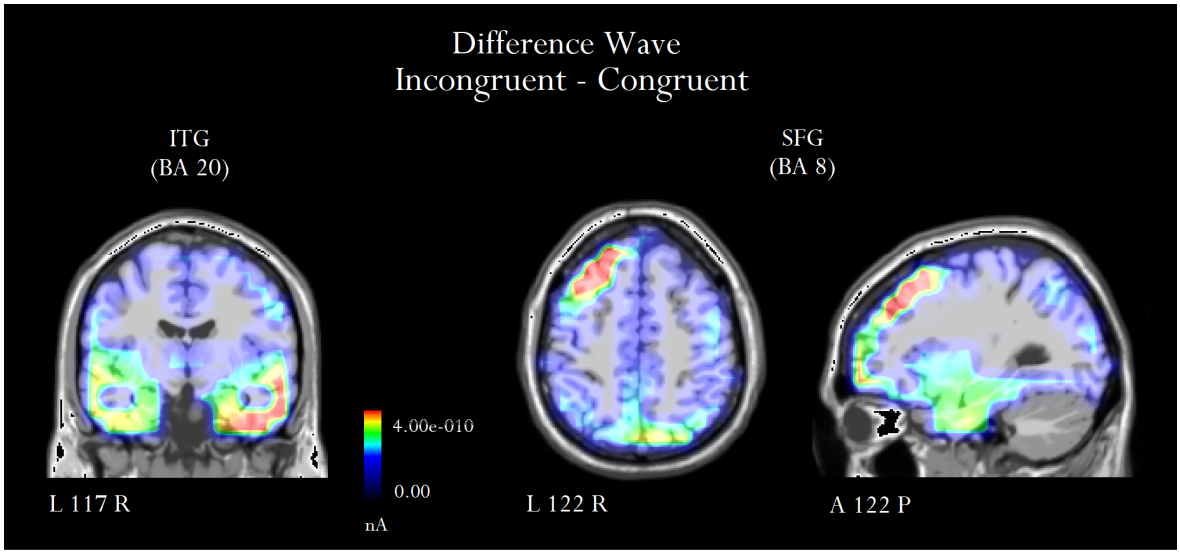


Figure 3.7. swLORETA. Coronal, sagittal, and axial views of the active sources relative to the difference wave obtained by subtracting ERP waveforms elicited by Congruent stimuli from that induced by Incongruent stimuli in the time window between 300 and 500 msec. The most active electromagnetic dipoles were the left Superior Frontal Gyrus (SFG; BA 8) and the right Inferior Temporal Gyrus (ITG; BA 20).

Table 3.4. Active electromagnetic dipoles with Talairach coordinates (in mm) and localization according to swLORETA inverse solution applied on the Difference Wave computer subtracting by subtracting ERP waveforms elicited by Congruent stimuli from that induced by Incongruent stimuli in the 300 - 500 msec time window.

Dipole no.	Magnitude	X [mm]	Y [mm]	Z [mm]	Hem.	Lobe	Gyrus	BA
920	5,97	-29,5	31,5	49,5	L	F	Superior Frontal Gyrus	8
526	4,99	60,5	-8,5	-30,5	R	T	Inferior Temporal Gyrus	20
374	4,14	-69,5	-38,5	-0,5	L			
1047	4,07	-29,5	61,5	-10,5	L	F	Superior Frontal Gyrus	10
65	3,86	40,5	-78,5	39,5	R	O	Superior Occipital Gyrus	19
78	3,77	10,5	-78,5	49,5	R	P	Precuneus	19
1041	3,57	-9,5	61,5	-20,5	L	F	Superior Frontal Gyrus	10
172	3,53	60,5	-58,5	-20,5	R	O	Fusiform Gyrus	37
1032	2,95	20,5	61,5	29,5	R	F	Superior Frontal Gyrus	9
59	2,75	-39,5	-88,5	9,5	L	O	Medial Occipital Gyrus	18
325	2,56	50,5	-38,5	59,5	R	P	Inferior Parietal Lobule	40
285	2,41	-9,5	-48,5	29,5	L	Limbic	Posterior Cingulate	31
581	2,27	-9,5	-8,5	69,5	L	F	Superior Frontal Gyrus	6

Discussion

The aim of the present study was to investigate the neural processes of perception and comprehension of the semantic content of art and music, through the simultaneous presentation of pictorial works and instrumental musical pieces. The visual and auditory stimuli were matched to be either congruent or incongruent in terms of the semantic meaning conveyed. Specifically, the starting hypothesis was that both music and artworks are capable of carrying semantic meanings, and that the incongruence between the two kinds of stimuli would have been indexed by the N400 ERP component. To test for this semantic relationship, we paired either congruently or incongruently music and paintings. As a matter of fact, we found that when presenting incongruent information between the two sensory modalities, N400 response is enhanced.

The semantic processing has been investigated in relation to various contexts. Rather than the language domain (Federmeier, 2007; Kellenbach et al., 2000; Kutas & Hillyard, 1980; Lau et al., 2008), here we refer to semantics in music and in visual arts. With this study, we wanted to broaden the capability of the N400 ERP component, which is mainly used to access incongruences and violations in the language domain, to also index semantic incongruity across sensory modalities. This allowed us on the one hand to test that both visual and auditory stimuli were able to convey semantic meanings, and on the other hand to verify that inconsistencies between modalities were detected. We also focused on an earlier ERP component, the N2, because it is often defined as an index of conflict monitoring (see Folstein & van Petten, 2008 for a review).

N2 and conflicting incoming Information

We analysed an early ERP component, that is the N2. This is a negative wave peaking between 200 and 350 msec after stimulus. In the literature, the N2 is frequently described as sensitive to novelty and mismatch (Folstein & van Petten, 2008). This ERP component is also well-established as an index of cognitive control (Buzzell et al., 2014) and conflict monitoring (van Veen & Carter, 2002). Indeed, in the present study the N2 was modulated by the semantic relationship between the auditory and the visual stimuli, eliciting greater amplitudes in response to incongruent pairs of stimuli. However, in this case, we relate to what Zitnick and collaborators (2016) state, that is that the extraction of semantic meaning from an object starts with a coarse processing of information. Thus, here the modulation of the N2 might reflect the starting of this gross process, namely detecting a conflict during information processing and the consequent need to resolve, or at least monitoring with higher level processes, the on-going conflict. This is in line with the conclusions drawn by Van Veen and Carter (2002b). In their review work, the authors report that the fronto-central N2

reflects conflict recognition in the incoming information, being larger to those conditions that involve the most conflict. Our results show how as early as 250 msec after stimulus onset, even with artistic stimuli as incoming information, a semantic conflict is detected.

N400 and semantic Incongruence

The fact that the N400 shows smaller amplitudes in the congruent condition can be attributed to the fact that in this condition it is possible to find facilitation in the processing of subsequent incoming information. Since in this case music is a background stimulus, this could act as a context that pre-activates the semantic representation, thus resulting in a facilitation of the processing of semantics conveyed by the picture (Federmeier, 2007; Kutas & Federmeier, 2000). Similarly, in the musical domain, Koelsch and collaborators (2004) found that the N400 priming effect in the music domain did not differ from that in the language domain, with respect to latency, scalp distribution, neural sources or amplitude. They presented participants with visual target words after hearing either a spoken sentence or a musical excerpt. Target words that were semantically unrelated to prime sentences elicited a larger N400 than did semantically related prime and sentences. Thus, the authors concluded stating that music can influence the processing of words and it can also prime representations of meaningful concepts.

In the same vein, also in the visual domain we find the N400 response to be modulated by the coherence of primed and subsequent image. Giglio and colleagues (2013) showed participants physically feasible but subtly incongruent actions images to test if the N400 would have indexed the semantic processing also in the action-perception domain, other than the verbal domain. The authors theorize that the N400 not only reflects the incongruence detection, but also the effect of the context comprehension. Again, a study by Cohn and co-workers (2012) implied that, when presenting strings of images, a combination of narrative structure and semantic relatedness can facilitate semantic processing of upcoming images, as reflected by the increase activation of the N400 both when images lacked semantics and narrative structure. As occurs in language, the literature provides evidence that the processing of visual events is modulated by the predictability of an event with regards to its prior context, and that coherent visual sequences allow for the build-up of meaning across the course of the sequence (Cohn & Paczynski, 2019a).

These studies show how the N400 is capable of indexing semantic incongruency in different domains. Thus, to broaden these findings we addressed also the possibility of the N400 to be modulated by the semantic relatedness also across sensory modalities. As a matter of fact, our results show greater amplitudes of the N400 when presenting semantically unrelated music and artworks. This was also more enhanced when the distinctiveness of the category was higher, which means that

the categories were easier to recognize. These confirms the ability of both our kinds of artistic stimuli to deliver semantic meaning.

Source Reconstruction

To locate the neural substrates of the N400, an swLORETA was computed on the Difference Wave obtained subtracting ERPs to Congruent stimuli from ERPs to Incongruent stimuli, in the same temporal window. What the source reconstruction has shown is the activation of the Superior Frontal Gyrus (SFG; BA 8) and the Inferior Temporal Gyrus (ITG; BA20), areas that are accounted amongst the neuronal generators of the N400 (Baggio & Hagoort, 2011; Lau et al., 2008; Silva-Pereyra et al., 2003). Lau and colleagues (2008) used a psycholinguistic model of semantic processing to guide their interpretation of the pattern of activated regions associated with contextual semantic manipulations that lead to N400 effects. We found the activation of SFG, considered to be involved in different functions, from motor task to working memory and attention (du Boisgueheneuc et al., 2006). Li and colleagues (2013) were able to investigate SFG at a subregional level and understand the functional complexity of this area. The authors found that BA8 manly contributes in visuo-spatial processing and working memory performance. Also, SFG is involved in conflict monitoring and error detection functions (Li et al., 2013). During our study, participants heard a background music while watching artistic pictures belonging to either a congruent or an incongruent semantic category. Thus, visuo-spatial processing and conflict monitoring are both implicated in our task and can explain the activation shown in our source reconstruction.

We also found the activation of ITG during the task. In the study by Visser and collaborators (2012) it is noted the activation in inferior temporal regions was higher for picture than words processing. In our case, the activation of this area might be due to artworks processing. Previous functional neuroimaging studies of explicit lexical-semantic decision have also reported activation of a network of frontotemporal brain regions, including lateral and inferior temporal cortices (Gold et al., 2006). ITG is also known to play its role in several functions involved in our study, such as complex visual perception and, most importantly, semantic multimodal sensory integration (Lin et al., 2020). Thus, the activation of this area during our task is likely to be due to this very function. We presented complex visual stimuli, and participants had to integrate semantic information incoming from visual and auditory channels. Furthermore, as in Waberski and colleagues (2001), we found a right hemispheric dominance of source activity related to the induced mismatch between the concurrent stimulation.



General Conclusions

The main goal of this dissertation is to understand how music listening affects the brain. The topic is handled from three different angles: neural, emotional, and semantic. Specifically, the first study was aimed at investigating the relationship between neural entrainment and music-evoked emotions. The next step was to examine how emotions induced by music interact with those induced by artistic visual stimuli. With the third study, we substantiated the capability of music to deliver semantic content, and how this blends with other semantic incoming information.

With the first study, we proposed a new method for analysing auditory steady-state responses (ASSR) to regular sound onsets in naturalistic music. In real music, the sound onsets vary continuously rather than being perfectly regular. For this reason, we proposed a measure of the average steady state response aligned to each sound onset. Given a regular steady-state response within a limited time range, the average M/EEG waveform time-locked to each sound onset is phase-locked to adjacent steady state response cycles. Hence, we estimated the amplitude of the average ASSR. We also tested for correlation between the ASSR and the aesthetic judgments obtained for each song. Our results showed how music pieces with regular sound onsets present frequency peaks at the same frequencies in ASSR, both in the EEG and the MEG signal. Frequency peaks were also significantly higher and showed higher Inter-Subject Coherence (ISC) values for music piece with regular sound onsets than for music with irregular sound onsets. To access the beat frequency, we grouped sound onsets by two or by four. Also in this case, our results showed that EEG and MEG ASSR peaked at that same frequency. From these results, we can imply that the brain engages in music listening by synchronizing its wave with the sound intensity changes and the musical beat. Additionally, we investigated the relationship between neural entrainment and aesthetic judgments. Neural entrainment is considered a valuable emotion induction mechanism, in which the temporal structure of music and physiological activity are linked (Juslin et al., 2010). Despite this, we found unsatisfactory evidence of this link. We could not confirm Juslin and colleagues' hypothesis because we did not find a consistent association between the two measures.

The second study investigated whether the emotional comprehension of visual art was affected by listening to background music. We found that early perceptual processes eliciting visual N170 were sensitive to both physical and emotional characteristics of the stimuli. At later latencies, also the LP response was modulated by the emotional valence of the stimulus, possibly indicating that negative pictures elicit a higher level of arousal, hence attracting more attention and deeper processing of the emotional stimulus, compared to positive pictures. As for the auditory stimulation, the N400 showed higher activations in response to negative than to positive stimuli, as found comparing positive and negative spoken phrases (Proverbio et al., 2020). This goes toward the

hypothesis of a shared neural mechanism for processing affective information, irrespective of the sensory channel. Combining the two kinds of stimuli, the multimodal late positivity possibly indexed the degree of aesthetic appreciation of a multimodal perceptual experience. Therefore, it is hypothesized that perceiving congruent negative pairs were less pleasurable than incongruent ones. Since the LP reflects the levels of the pleasantness of the stimulus, our results show a pattern where positive congruent pairs are the most pleasant, while incongruent pairs are the least pleasant. In between, we find the incongruent pairs, which, evoking mixed emotions, are likely to result in a neutral stimulus. Lastly, the source reconstruction highlighted Middle Frontal Gyrus (MFG) and Superior Frontal Gyrus (SFG) as the most active area in all conditions, going again toward a shared neural mechanism for processing emotional information in both auditory and visual domains.

As for the last study, we presented participants with artworks and background music belonging to either the same or different semantic categories. The aim was to investigate whether music was able to convey specific semantic meaning and how this interacts with semantic information incoming from a different sensory channel. Then, we analysed two ERP components, one was chosen because is known to be sensitive to conflict monitoring, that is the N2; the other one, the N400, was known to be indexing semantic incongruence. On the one hand, we found greater N2 amplitudes concerning incongruent pairs of stimuli than congruent ones. This indicates that a semantic conflict is detected already at early latencies, despite using complex artistic stimuli and music as incoming information. On the other hand, the N400 was analysed. Indeed, our results indicate how the N400 response is enhanced when presenting semantically unrelated information between the two sensory modalities. The source localization computed in the same time window as the N400 highlighted the role of SFG, implicated in error detection and conflict monitoring functions, and Inferior Temporal Gyrus (ITG), involved in visual perception and semantic integration across modalities. If previous literature found the capability to detect semantic information delivered by music (Daltrozzo & Schön, 2009; Koelsch, 2011b; Koelsch et al., 2004; Painter & Koelsch, 2011) and artworks (Amoruso et al., 2013; Cohn & Paczynski, 2019b; Markey et al., 2019a; Proverbio & Riva, 2009), we added the integration of the semantic information incoming from different sensory channels.

In conclusion, we developed a novel methodology to investigate neural entrainment, demonstrated that music affects emotional content of artworks, and that music can convey specific semantic meanings.

From a neural point of view, it has been advocated that the more the brain is synchronized with musical stimulus, the more it will induce emotions (Trost & Vuilleumier, 2013). On the one hand, we corroborated the fact that the brain engages into rhythmic stimuli by resonating at the same frequency as the endogenous stimulus, in this case music. Results from our first study show how the

brain entrains with music having a regular beat, but not with music having irregular beat. On the other hand, our findings do not go in the direction of entrainment as a fundamental emotion induction mechanism, but rather one of many processes that contribute to such a complex phenomenon as experiencing emotion. As proposed by Trost and colleagues (2017), entrainment may be a mechanism that includes different levels, one interacting with the other, to deliver a compound experience of emotion. As defined by Scherer in the Component Process Model (2005), emotion is an episode of interconnected changes in the states of at least five components, each communicating and associated with a different function. In line with this model, each level of entrainment might parallel a component of the same process (Trost et al., 2017). This is also in line with the theory of Cespedes-Guevara and Eerola (2018), by which acoustical cues present in music can be related to physiological changes. These changes develop into an emotional percept by integrating with information of past knowledge, context, and the current psychological state. Moreover, Cespedes-Guevara and Eerola state that their theory complements the theory by Flaig and Large (2014). Flaig and Large propose that music connects with brain's activity which, in turn, modulates the listener's reactions, allowing to predict core affects of emotions. However, the studies presented here both confirm and refute the theory of Cespedes-Guevara and Eerola (2018). In study 2 we saw how music is capable of conveying discrete emotions, which corroborates the constructionist approach. On the other hand, results from the first study reject the close link between neuro-dynamics and affective states. Nonetheless, music-evoked emotions are delivered both by the composer, through structural cues (Gabrielsson, 2016), and by the performer, through acoustic cues (Juslin, 2000). Ultimately, performers' cue utilization have to match the listeners' cue utilization, in order to make the listener feel emotions. Thus, it may be that in our case a link between neuro-dynamics and affective states is missing because emotions are manipulated, and understood, differently. Juslin and Lindström (2010) developed a model to explore the relative contributions of composer cues and performer cues, and their interactions, to induce emotions in the listener. Several studies have suggested that musical cues work in an additive manner in order to induce different emotions (e.g., Eerola et al., 2013; Scherer & Oshinsky, 1977), a theory recently confirmed by Grimaud and collaborators (2022). In their recent work, the authors carried out three studies that show how the listener himself utilizes musical cues to perceive different emotional meanings.

Here, we adopted neural responses using an implicit paradigm to explore the emotions perceived by the listener. Indeed, our results show how the N400 ERP component, which has been consistently associated with the language domain, is also modulated by the emotional valence of the musical stimulus (Study 2). As well, we found the same to component to be an index of the

congruency between semantic information across sensory channels (Study 3). As a consequence, we might say that emotional processing in music share specific neural mechanism with affective language processing. This is in line with what Proverbio and Piotti (2021) proposed, that is an integrated model about how emotional information is processed, irrespective the type of sound, namely speech or music.

This brings us to the matter of music evolution. Over the past years, various theories have been provided, from Pinker's "music as a cheesecake" (1997), which states that music has no adaptive function, to the "Mixed Origin of Music" by Altenmüller (2013), according to which the early roots of music may lie in an ancient affective signalling system. The latter also posits that music is more than affective or emotional communication. Brown (2000) has in fact hypothesized the "musilanguage" model, by which music and language have evolved simultaneously. According to this model, music and language evolved as specialization from a common stage. This concept comes from the common features of the two systems, which are supposed to have evolved first, being neither musical nor linguistic (Brown, 2000). Similarly, Reybrouck and Podlipniak (2019) proposed that music and language have the same affective substrate and may share a common ancestral mechanism for communicative interaction, that would be affective vocalizations. This helps us explain the neural activations we found in study 2, where frontal areas were found to active in response to musical affective stimulation. The same was found in Belin and colleagues (Belin et al., 2008) when comparing affective vocalizations.

In the same vein, both music and speech are dynamic stimuli, meaning that sound changes continuously over time. Thus, speech and music are interpreted according to their expressive information, as well as their sound structure. The major difference lies in the acoustical cues utilized to infer meaning from the two (Reybrouck & Podlipniak, 2019). With our third study we demonstrated that music can convey specific semantic meaning, and that the same occurs for artistic pictures. We also explored the interaction and the integration of semantic information coming from different sensory channels. Our results highlighted the activation of SFG in integrating semantic information coming from different sensory channels. The same area has found to be involved in linguistic semantic processing (Lau et al., 2008). Again, this leads us to confirm the overlapping of neural substrates between music and language.

Another thing to take into account is whether there is a universal component about music, specifically about meanings that music can convey. According to Thompson and Balkwill (2010) cross-cultural music cognition is the study of similarities and differences in cognitive processes and emotional experiences for music across cultures. Cross-cultural music cognition can be used to discriminate between universal and culture-specific factors of forms of music understanding. The

similarities can be explained referring to system constraints, whereas the differences referring to ontogenetic processes rather than innate factors (Thompson & Balkwill, 2010). At this regard, we mention again neural entrainment and the Beat Perception and Synchronization Hypothesis (Patel, 2006). Since beat perception is fundamental for both music perception and production, synchronization to the beat might in turn constitute a central factor in music perception. Patel and colleagues (2008) reported evidence of synchronization to a beat and adjustment to varying tempo in non-human animals, going toward an evolutionary, hence universal, adaptation for music. Morrison and collaborators (2003) studied differences in neural activity in response to culturally familiar and unfamiliar music, finding that listening to different musical styles showed the same activations in the listeners, unlike what happens for language. The authors interpreted their results speculating that the listener may apply his familiar and culturally related rules to comprehend music in a way that is satisfactory enough. Since this process leads to a sufficient understanding of a style of music different from the one of the listener, this would mean that music is more universally understood than language (Morrison et al., 2003). As a matter of fact, Laukka and co-workers (2013) asked professional musicians, coming from different musical tradition, to perform conveying eleven emotions (happiness, sadness, anger, fear, affection, peacefulness, humor, longing, solemnity, spirituality, and neutral) to listeners of different cultures. Results showed that cues used by performers to encode emotions were decoded by listeners also across cultures, although the match between musicians' and listeners' uses of cues was better within the same cultural context. Notwithstanding, recognizing an emotion evoked by a piece of music does not necessarily mean to experience that emotion (Juslin & Laukka, 2004). Moreover, it is possible that emotion classifications used to define emotions in Western music do not always apply to music of other cultures (Laukka et al., 2013). To bypass this issue, Beier and colleagues (2022) explored chill response, as an indirect measure of felt emotion, to familiar and unfamiliar musical styles. The main result was that the number of chills did not vary in response to styles of music of different cultures. Also, the authors demonstrated that some psychoacoustic properties of music communicate emotion cross-culturally, as they were associated with chills across all three styles of music. This led the authors to conclude that similar acoustic events can induce strong emotional responses across cultures (Beier et al., 2022). Therefore, the affective expression in music may depend on a combination of universal and culture-specific factors (Laukka et al., 2013). Beier and co-workers also insert their findings into the Predictive Coding (PC; Friston, 2005) framework, by which music that is neither overly predictable nor completely unpredictable will provide listeners the maximum pleasure (Koelsch et al., 2019; Vuust & Witek, 2014). We ascribe our results from our first study to the same framework,

according to which to create a rhythmic percept, the brain predicts upcoming events on the bases of previous experience.

This brings us again to neural entrainment, closing our circle and bridging neural, emotional and cognitive processing of musical experience.

References

Abusch, D. (2015). Possible worlds semantics for pictures. *The Wiley Blackwell Companion to Semantics*, *December*, 30–51. https://doi.org/10.1201/9781439864302-2

Alluri, V., Brattico, E., Toiviainen, P., Burunat, I., Bogert, B., Numminen, J., & Kliuchko, M. (2015). Musical expertise modulates functional connectivity of limbic regions during continuous music listening. *Psychomusicology: Music, Mind, and Brain*, *25*(4), 443–454. https://doi.org/10.1037/pmu0000124

Altenmüller, E., Kopiez, R., & Grewe, O. (2013). A contribution to the evolutionary basis of music:lessons from the chill response. In *Evolution of emotional communication* (pp. 313–335). Oxford University Press.

Altenmüller, E., Siggel, S., Mohammadi, B., Samii, A., & Münte, T. F. (2014). Play it again, Sam: Brain correlates of emotional music recognition. *Frontiers in Psychology*, *5*(FEB), 1–8. https://doi.org/10.3389/fpsyg.2014.00114

Amoruso, L., Gelormini, C., Aboitiz, F., Alvarez González, M., Manes, F., Cardona, J., & Ibanez, A. (2013). N400 ERPs for actions: Building meaning in context. *Frontiers in Human Neuroscience*, *FEB*. https://doi.org/10.3389/fnhum.2013.00057

Antović, M. (2009). Towards the Semantics of Music: The Twentieth Century. *Language & History*, *52*(1), 119–129. https://doi.org/10.1179/175975309x452003

Baggio, G., & Hagoort, P. (2011). The balance between memory and unification in semantics: A dynamic account of the N400. *Language and Cognitive Processes*, *26*(9), 1338–1367. https://doi.org/10.1080/01690965.2010.542671

Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain and Cognition*, *14*(2), 201–212. https://doi.org/10.1016/0278-2626(90)90029-N

Battcock, A., & Schutz, M. (2022). Emotion and expertise: how listeners with formal music training use cues to perceive emotion. *Psychological Research*, *86*(1), 66–86. https://doi.org/10.1007/s00426-020-01467-1

Baumgartner, T., Lutz, K., Schmidt, C. F., & Jäncke, L. (2006). The emotional power of music: How music enhances the feeling of affective pictures. *Brain Research*, *1075*(1), 151–164. https://doi.org/10.1016/j.brainres.2005.12.065

Beier, E. J., Janata, P., Hulbert, J. C., & Ferreira, F. (2022). Do You Chill When I Chill? A Cross-Cultural Study of Strong Emotional Responses to Music. *Psychology of Aesthetics, Creativity, and the Arts*, *16*(1), 74–96. https://doi.org/10.1037/aca0000310

Belin, P., Fecteau, S., Charest, I., Nicastro, N., Hauser, M. D., & Armony, J. L. (2008). Human cerebral response to animal affective vocalizations. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1634), 473–481. https://doi.org/10.1098/RSPB.2007.1460

Bell, A. J., & Sejnowski, T. I. (1995). An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation*, *7*(6), 1129–1159.

Benz, S., Sellaro, R., Hommel, B., & Colzato, L. S. (2016). Music makes the world go round: The impact of musical training on non-musical cognitive functions-a review. *Frontiers in Psychology*, *6*(JAN), 1–5. https://doi.org/10.3389/fpsyg.2015.02023

Bernat, E., Bunce, S., & Shevrin, H. (2001). Event-related brain potentials differentiate positive and negative mood adjectives during both supraliminal and subliminal visual processing. *International Journal of Psychophysiology*, *42*(1), 11–34. https://doi.org/10.1016/S0167-8760(01)00133-7

Besson, M., Chobert, J., & Marie, C. (2011). Language and music in the musician brain. *Linguistics and Language Compass*, *5*(9), 617–634. https://doi.org/10.1111/j.1749-818X.2011.00302.x

Besson, M., & Schön, D. (2001). Comparison between language and music. *Annals of the New York Academy of Sciences*, *930*, 232–258. https://doi.org/10.1111/j.1749-6632.2001.tb05736.x

Bharathi, G., Jayaramayya, K., Balasubramanian, V., & Vellingiri, B. (2019). The potential role of rhythmic entrainment and music therapy intervention for individuals with autism spectrum disorders. In *Journal of Exercise Rehabilitation* (Vol. 15, Issue 2, pp. 180–186). Korean Society of Exercise Rehabilitation. https://doi.org/10.12965/jer.1836578.289

Bharucha, J. J., Curtis, M., & Paroo, K. (2006). Varieties of musical experience. *Cognition*, *100*(1), 131–172. https://doi.org/10.1016/j.cognition.2005.11.008

Bhatti, A. M., Majid, M., Anwar, S. M., & Khan, B. (2016). Human emotion recognition and analysis in response to audio music using brain signals. *Computers in Human Behavior*, *65*, 267–275. https://doi.org/10.1016/j.chb.2016.08.029

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences*, *98*(20), 11818–11823. https://doi.org/10.1073/pnas.191355898

Boccia, M., Barbetti, S., Piccardi, L., Guariglia, C., Ferlazzo, F., Giannini, A. M., & Zaidel, D. W. (2016). Where does brain neural activation in aesthetic responses to visual art occur? Meta-analytic evidence from neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, *60*, 65–71. https://doi.org/10.1016/j.neubiorev.2015.09.009

Bolger, D., Trost, W., & Schön, D. (2013). Rhythm implicitly affects temporal orienting of attention across modalities. *Acta Psychologica*, *142*(2), 238–244. https://doi.org/10.1016/j.actpsy.2012.11.012

Bolivar, V. J., Cohen, A. J., & Fentress, J. C. (1994). Semantic and formal congruency in music and motion pictures: Effects on the interpretation of visual action. *Psychomusicology: A Journal of Research in Music Cognition*, *13*(1–2), 28–59. https://doi.org/10.1037/h0094102

Bonetti, L., Haumann, N. T., Vuust, P., Kliuchko, M., & Brattico, E. (2017). Risk of depression enhances auditory Pitch discrimination in the brain as indexed by the mismatch negativity. *Clinical Neurophysiology*, *128*(10), 1923–1936. https://doi.org/10.1016/j.clinph.2017.07.004

Bottiroli, S., Rosi, A., Russo, R., Vecchi, T., & Cavallini, E. (2014). The cognitive effects of listening to background music on older adults: Processing speed improves with upbeat music, while memory seems to benefit from both upbeat and downbeat music. *Frontiers in Aging Neuroscience*, *6*(OCT), 1–7. https://doi.org/10.3389/fnagi.2014.00284

Bouwer, F. L., & Honing, H. (2015). Temporal attending and prediction influence the perception of metrical rhythm: evidence from reaction times and ERPs. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01094

Brancatisano, O., Baird, A., & Thompson, W. F. (2020). Why is music therapeutic for neurological disorders? The Therapeutic Music Capacities Model. *Neuroscience and Biobehavioral Reviews*, *112*, 600–615. https://doi.org/10.1016/j.neubiorev.2020.02.008

Brattico, E. (2019). The Empirical Aesthetics of Music. In *The Oxford Handbook of Empirical Aesthetics*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198824350.001.0001

Brattico, E., Brusa, A., Dietz, M. J., Jacobsen, T., Fernandes, H. M., Gaggero, G., Toiviainen, P., Vuust, P., & Proverbio, A. M. (2020). Beauty and the brain: Investigating the neural and musical attributes of beauty during a naturalistic music listening experience. *BioRxiv*, *25*(1), 1–9.

Brattico, E., & Pearce, M. (2013). The neuroaesthetics of music. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(1), 48–61. https://doi.org/10.1037/a0031624

Braun Janzen, T., de Oliveira, B., Ventorim Ferreira, G., Sato, J. R., Feitosa-Santana, C., & Vanzella, P. (2023). The effect of background music on the aesthetic experience of a visual artwork in a naturalistic environment. *Psychology of Music*, *51*(1), 16–32. https://doi.org/10.1177/03057356221079866

Bregman, M. R., Iversen, J. R., Lichman, D., Reinhart, M., & Patel, A. D. (2012). A method for testing synchronization to a musical beat in domestic horses (Equus ferus caballus). *Empirical Musicology Review*, *7*(3–4), 144–156.

Brown, S. (2000). The "Musilanguage" Model of Music Evolution. In *Theories of Music Origin* (pp. 271–300).

Brown, S., Gao, X., Tisdelle, L., Eickhoff, S. B., & Liotti, M. (2011). Naturalizing aesthetics: Brain areas for aesthetic appraisal across sensory modalities. *NeuroImage*, *58*(1), 250–258. https://doi.org/10.1016/j.neuroimage.2011.06.012

Buchanan, T. W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., & Jancke, L. (2000). Recognition of emotional prosody and verbal components of spoken language: an fMRI study. In *Cognitive Brain Research* (Vol. 9). www.elsevier.comrlocaterbres

Budd, M. (1985). *Music and the emotions : the philosophical theories*. Routledge & Kegan Paul. https://www.routledge.com/Music-and-the-Emotions-The-Philosophical-Theories/Budd/p/book/9780415087797

Burunat, I., Alluri, V., Toiviainen, P., Numminen, J., & Brattico, E. (2014). Dynamics of brain activity underlying working memory for music in a naturalistic condition. *Cortex*, *57*, 254–269. https://doi.org/10.1016/j.cortex.2014.04.012

Burunat, I., Brattico, E., Puoliväli, T., Ristaniemi, T., Sams, M., Toiviainen, P., & Snyder, J. (2015). Action in perception: Prominent visuo-motor functional symmetry in musicians during music listening. *PLoS ONE*, *10*(9). https://doi.org/10.1371/journal.pone.0138238

Burunat, I., Tsatsishvili, V., Brattico, E., & Toiviainen, P. (2017). Coupling of action-perception brain networks during musical pulse processing: Evidence from region-of-interest-based independent component analysis. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00230

Buzzell, G. A., Fedota, J. R., Roberts, D. M., & McDonald, C. G. (2014). The N2 ERP component as an index of impaired cognitive control in smokers. *Neuroscience Letters*, *563*, 61–65. https://doi.org/10.1016/j.neulet.2014.01.030

Cacioppo, J. T., Jr., S. L. C., & Gardner, W. L. (1996). Attitudes to the Right: Evaluating Processing is Associated with Lateralized Late Positive Event-Related Brain Potentials. *Personality and Social Psychology Bullettin*, *22*(12), 1205–1219. https://doi.org/10.1177/01461672962212002

Calma-Roddin, N., & Drury, J. E. (2020). Music, Language, and The N400: ERP Interference Patterns Across Cognitive Domains. *Scientific Reports*, *10*(1), 1–14. https://doi.org/10.1038/s41598-020-66732-0

Campos-Bueno, J. J., DeJuan-Ayala, O., Montoya, P., & Birbaumer, N. (2015). Emotional Dimensions of Music and Painting and their Interaction. *The Spanish Journal of Psychology*, *18*, E54. https://doi.org/10.1017/sjp.2015.53

Cannas Aghedu, F., Sarlo, M., Zappasodi, F., Acevedo, B. P., & Bisiacchi, P. S. (2021). Romantic love affects emotional processing of love-unrelated stimuli: An EEG/ERP study using a love induction task. *Brain and Cognition*, *151*(February), 105733. https://doi.org/10.1016/j.bandc.2021.105733

Cano, M. E., Class, Q. A., & Polich, J. (2009). Affective valence, stimulus attributes, and P300: Color vs. black/white and normal vs. scrambled images. *International Journal of Psychophysiology*, *71*(1), 17–24. https://doi.org/10.1016/j.ijpsycho.2008.07.016

Carlson, E., Saarikallio, S., Toiviainen, P., Bogert, B., Kliuchko, M., & Brattico, E. (2015). Maladaptive and adaptive emotion regulation through music: A behavioral and neuroimaging study of males and females. *Frontiers in Human Neuroscience*, *9*(AUGUST). https://doi.org/10.3389/fnhum.2015.00466

Carretié, L., Iglesias, J., García, T., & Ballesteros, M. (1996). N300, P300 and the emotional processing of visual stimuli. *Electroencephalography and Clinical Neurophysiology*, *103*(2), 298–303. https://doi.org/10.1016/S0013-4694(96)96565-7

Castro, S. L., & Lima, C. F. (2014). Age and musical expertise influence emotion recognition in music. *Music Perception*, *32*(2), 125–142. https://doi.org/10.1525/MP.2014.32.2.125

Cattaneo, Z., Lega, C., Gardelli, C., Merabet, L. B., Cela-Conde, C. J., & Nadal, M. (2014). The role of prefrontal and parietal cortices in esthetic appreciation of representational and abstract art: A TMS study. *NeuroImage*, *99*, 443–450. https://doi.org/10.1016/j.neuroimage.2014.05.037

Celma-Miralles, A., de Menezes, R. F., & Toro, J. M. (2016). Look at the beat, feel the meter: Top-down effects of meter induction on auditory and visual modalities. *Frontiers in Human Neuroscience*, *10*(MAR2016). https://doi.org/10.3389/fnhum.2016.00108

Cervellin, G., & Lippi, G. (2011). From music-beat to heart-beat: A journey in the complex interactions between music, brain and heart. *European Journal of Internal Medicine*, *22*(4), 371–374. https://doi.org/10.1016/j.ejim.2011.02.019

Cespedes-Guevara, J., & Eerola, T. (2018). Music communicates affects, not basic emotions - A constructionist account of attribution of emotional meanings to music. *Frontiers in Psychology*, *9*(FEB). https://doi.org/10.3389/fpsyg.2018.00215

Chatterjee, A. (2004). Prospects for a cognitive neuroscience of visual aesthetics. *Bull. Psychol.*, *4*, 56–60. https://doi.apa.org/get-pe-doi.cfm?doi=10.1037/e514602010-003

Chatterjee, A., & Vartanian, O. (2016). Neuroscience of aesthetics. *Annals of the New York Academy of Sciences*, *1369*(1), 172–194. https://doi.org/10.1111/nyas.13035

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008a). Moving on Time: Brain Network for Auditory-Motor Synchronization is Modulated by Rhythm Complexity and Musical Training. *Journal of Cognitive Neuroscience*, *20*(2), 226–239.

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008b). Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex*, *18*(12), 2844–2854. https://doi.org/10.1093/cercor/bhn042

Chen, Y. C., & Spence, C. (2018). Audiovisual semantic interactions between linguistic and nonlinguistic stimuli: The time-courses and categorical specificity. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(10), 1488–1507. https://doi.org/10.1037/xhp0000545

Codispoti, M., Mazzetti, M., & Bradley, M. M. (2009). Unmasking emotion: Exposure duration and emotional engagement. *Psychophysiology*, *46*(4), 731–738. https://doi.org/10.1111/j.1469-8986.2009.00804.x

Cohn, N., & Paczynski, M. (2019a). Structural complexity in visual narratives: Theory, brains, and cross-cultural diversity. *Narrative Complexity and Media: Experiential and Cognitive Interfaces*, 174–199. http://www.visuallanguagelab.com/P/2019.NCM.NC.pdf

Cohn, N., & Paczynski, M. (2019b). The Neurophysiology of Event Processing in Language and Visual Events. *The Oxford Handbook of Event Structure*, *May*, 623–637. https://doi.org/10.1093/oxfordhb/9780199685318.013.26

Cohn, N., Paczynski, M., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2012). (Pea)nuts and bolts of visual narrative: Structure and meaning in sequential image comprehension. *Cognitive Psychology*, *65*(1), 1–38. https://doi.org/10.1016/j.cogpsych.2012.01.003

Conroy, M. A., & Polich, J. (2007). Affective valence and P300 when stimulus arousal level is controlled. *Cognition and Emotion*, *21*(4), 891–901. https://doi.org/10.1080/02699930600926752

Cook, T., Roy, A. R. K., & Welker, K. M. (2019). Music as an emotion regulation strategy: An examination of genres of music and their roles in emotion regulation. *Psychology of Music*, *47*(1), 144–154.

Criscuolo, A., Bonetti, L., Särkämö, T., Kliuchko, M., & Brattico, E. (2019). On the association between musical training, intelligence and executive functions in adulthood. *Frontiers in Psychology*, *10*(JULY). https://doi.org/10.3389/fpsyg.2019.01704

Cupchik, G. C., Vartanian, O., Crawley, A., & Mikulis, D. J. (2009). Viewing artworks: Contributions of cognitive control and perceptual facilitation to aesthetic experience. *Brain and Cognition*, *70*(1), 84–91. https://doi.org/10.1016/j.bandc.2009.01.003

Cuthbert, B. N., Schupp, H. T., Margaret M., B., Niels, B., & Lang, P. J. (2000). Brain potentials in affective picture processing: covariation with autonomic arousal and affective report. *Biological Psychology*, *52*, 95–111. https://doi.org/10.1016/0370-2693(92)91779-9

Dalla Bella, S. (2020). The use of rhythm in rehabilitation for patients with movement disorders. *Music and the Aging Brain*, 383–406. https://doi.org/10.1016/b978-0-12-817422-7.00015-8

Daltrozzo, J., & Schön, D. (2008). Conceptual Processing in Music as Revealed by N400 Effects on Words and Musical Targets. *Journal of Cognitive Neuroscience*, *21*(10), 1882–1892.

Daltrozzo, J., & Schön, D. (2009). Is conceptual processing in music automatic? An electrophysiological approach. *Brain Research*, *1270*, 88–94. https://doi.org/10.1016/j.brainres.2009.03.019

de Jong, R., Toffanin, P., & Harbers, M. (2010). Dynamic crossmodal links revealed by steady-state responses in auditory-visual divided attention. *International Journal of Psychophysiology*, *75*(1), 3–15. https://doi.org/10.1016/j.ijpsycho.2009.09.013

de Pisapia, N., Bacci, F., Parrott, D., & Melcher, D. (2016). Brain networks for visual creativity: A functional connectivity study of planning a visual artwork. *Scientific Reports*, *6*. https://doi.org/10.1038/srep39185

de Witte, M., Spruit, A., van Hooren, S., Moonen, X., & Stams, G. J. (2020). Effects of music interventions on stress-related outcomes: a systematic review and two meta-analyses. *Health Psychology Review*, *14*(2), 294–324. https://doi.org/10.1080/17437199.2019.1627897

Deng, C., Tong, J., Deng, X., Zhang, Z., & Qin, Y. (2020). Emotion Recognition Positively Correlates with Steady-state Visual Evoked Potential Amplitude and Alpha Entrainment. *Neuroscience*, *434*, 191–199. https://doi.org/10.1016/j.neuroscience.2020.01.045

Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience*, *8*(MAY). https://doi.org/10.3389/fnhum.2014.00311

Dmochowski, J. P., Ki, J. J., DeGuzman, P., Sajda, P., & Parra, L. C. (2018). Extracting multidimensional stimulus-response correlations using hybrid encoding-decoding of neural activity. *NeuroImage*, *180*, 134–146. https://doi.org/10.1016/j.neuroimage.2017.05.037

Dmochowski, J. P., Sajda, P., Dias, J., & Parra, L. C. (2012). Correlated components of ongoing EEG point to emotionally laden attention - a possible marker of engagement? *Frontiers in Human Neuroscience*, *MAY 2012*. https://doi.org/10.3389/fnhum.2012.00112

du Boisgueheneuc, F., Levy, R., Volle, E., Seassau, M., Duffau, H., Kinkingnehun, S., Samson, Y., Zhang, S., & Dubois, B. (2006). Functions of the left superior frontal gyrus in humans: A lesion study. *Brain*, *129*(12), 3315–3328. https://doi.org/10.1093/brain/awl244

Eerola, T., Friberg, A., & Bresin, R. (2013). Emotional expression in music: Contribution, linearity, and additivity of primary musical cues. *Frontiers in Psychology*, *4*(JUL). https://doi.org/10.3389/fpsyg.2013.00487

Eerola, T., & Vuoskoski, J. K. (2013). A review of music and emotion studies: Approaches, emotion models, and stimuli. *Music Perception*, *30*(3), 307–340. https://doi.org/10.1525/MP.2012.30.3.307

Ellis, D. S., & Brighouse, G. (1952). Effects of Music on Respiration- and Heart-Rate. In *Source: The American Journal of Psychology* (Vol. 65, Issue 1).

Ellis, R. J., & Thayer, J. F. (2010). Music and autonomic nervous system (dys)function. *Music Perception*, *27*(4), 317–326. https://doi.org/10.1525/mp.2010.27.4.317

Escoffier, N., Zhong, J., Schirmer, A., & Qiu, A. (2013). Emotional expressions in voice and music: Same code, same effect? *Human Brain Mapping*, *34*(8), 1796–1810. https://doi.org/10.1002/hbm.22029

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505. https://doi.org/10.1111/j.1469-8986.2007.00531.x

Ferenczi, E. A., Zalocusky, K. A., Liston, C., Grosenick, L., Warden, M. R., Amatya, D., Katovich, K., Mehta, H., Patenaude, B., Ramakrishnan, C., Kalanithi, P., Etkin, A., Knutson, B., Glover, G. H., & Deisseroth, K. (2016). Prefrontal cortical regulation of brainwide circuit dynamics and reward-related behavior. *Science*, *351*(6268). https://doi.org/10.1126/science.aac9698

Ferreri, L., Mas-Herrero, E., Zatorre, R. J., Ripollés, P., Gomez-Andres, A., Alicart, H., Olivé, G., Marco-Pallarés, J., Antonijoan, R. M., Valle, M., Riba, J., & Rodriguez-Fornells, A. (2019). Dopamine modulates the reward experiences elicited by music. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(9), 3793–3798. https://doi.org/10.1073/pnas.1811878116

Fitch, W. T. (2005). The biology and evolution of music: A comparative perspective. *Cognition*, *100*, 173–215. https://doi.org/10.1196/annals.1360.004

Flaig, N. K., & Large, E. W. (2014). Dynamic musical communication of core affect. *Frontiers in Psychology*, *5*(MAR). https://doi.org/10.3389/fpsyg.2014.00072

Folstein, J. R., & van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, *45*(1), 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x

Franklin, M. S., Dien, J., Neely, J. H., Huber, E., & Waterson, L. D. (2007). Semantic priming modulates the N400, N300, and N400RP. *Clinical Neurophysiology*, *118*(5), 1053–1068. https://doi.org/10.1016/j.clinph.2007.01.012

Friederici, A. D., & Wartenburger, I. (2010). Language and brain. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(2), 150–159. https://doi.org/10.1002/wcs.9

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Fröber, K., & Thomaschke, R. (2019). In the Dark Cube: Movie Theater Context Enhances the Valuation and Aesthetic Experience of Watching Films. *Psychology of Aesthetics, Creativity, and the Arts*. https://doi.org/10.1037/aca0000295

Gabrielsson, A. (2016). The Relationship between Musical Structure and Perceived Expression. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford Handbook of Music Psychology* (2nd ed.). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198722946.013.18

Gao, Y., & Xie, L. (2018). Aesthetics - Emotion Mapping Analysis of Music and Painting. *2018 1st Asian Conference on Affective Computing and Intelligent Interaction, ACII Asia 2018*. https://doi.org/10.1109/ACIIAsia.2018.8470361

Geiser, E., Walker, K. M. M., & Bendor, D. (2014). Global timing: A conceptual framework to investigate the neural basis of rhythm perception in humans and non-human species. *Frontiers in Psychology*, *5*(MAR). https://doi.org/10.3389/fpsyg.2014.00159

Giglio, A. C. A., Minati, L., & Boggio, P. S. (2013). Throwing the banana away and keeping the peel: Neuroelectric responses to unexpected but physically feasible action endings. *Brain Research*, *1532*, 56–62. https://doi.org/10.1016/j.brainres.2013.08.017

Goerlich, K. S., Witteman, J., Aleman, A., & Martens, S. (2011). Hearing feelings: Affective categorization of music and speech in Alexithymia, an ERP study. *PLoS ONE*, *6*(5). https://doi.org/10.1371/journal.pone.0019501

Gold, B. T., Balota, D. A., Jones, S. J., Powell, D. K., Smith, C. D., & Andersen, A. H. (2006). Dissociation of automatic and strategic lexical-semantics: Functional magnetic resonance imaging evidence for differing roles of multiple frontotemporal regions. *Journal of Neuroscience*, *26*(24), 6523–6532. https://doi.org/10.1523/JNEUROSCI.0808-06.2006

Gomez, P., & Danuser, B. (2004). Affective and physiological responses to environmental noises and music. *International Journal of Psychophysiology*, *53*(2), 91–103. https://doi.org/10.1016/j.ijpsycho.2004.02.002

Gonzalez, A., Santapau, M., & Gonzales, J. J. (2021). EEG Analysis during Music Perception. In H. Nakano (Ed.), *IntechOpen* (p. 13). https://doi.org/DOI: 10.5772/intechopen.94574

Green, A. C., Bærentsen, K. B., Stodkilde-Jorgensen, H., Roepstorff, A., & Vuust, P. (2012). Listen, learn, like! Dorsolateral prefrontal cortex involved in the mere exposure effect in music. *Neurology Research International*, *2012*. https://doi.org/10.1155/2012/846270

Hajcak, G., MacNamara, A., & Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: an integrative review. *Developmental Neuropsychology*, *35*(2), 129–155.

Hanser, W. E., Mark, R. E., Zijlstra, W. P., & Vingerhoets, A. J. J. M. (2015). The effects of background music on the evaluation of crying faces. *Psychology of Music*, *43*(1), 75–85. https://doi.org/10.1177/0305735613498132

Haumann, N. T., Kliuchko, M., Vuust, P., & Brattico, E. (2018). Applying acoustical and musicological analysis to detect brain responses to realistic music: A case study. *Applied Sciences (Switzerland)*, *8*(5). https://doi.org/10.3390/app8050716

Haumann, N. T., Lumaca, M., Kliuchko, M., Santacruz, J. L., Vuust, P., Brattico, E., Haumann, N. T., Lumaca, M., Kliuchko, M., Santacruz, J. L., & Vuust, P. (2020). Extracting human cortical responses to sound onsets and acoustic feature changes in real music, and their relation to event rate. *Brain Research*, *1754*(2021), 147248. https://doi.org/10.1016/j.brainres.2020.147248

Haumann, N. T., Parkkonen, L., Kliuchko, M., Vuust, P., & Brattico, E. (2016). Comparing the Performance of Popular MEG/EEG Artifact Correction Methods in an Evoked-Response Study. *Computational Intelligence and Neuroscience*, *2016*. https://doi.org/10.1155/2016/7489108

Herbert, C., Junghofer, M., & Kissler, J. (2008). Event related potentials to emotional adjectives during reading. *Psychophysiology*, *45*(3), 487–498. https://doi.org/10.1111/j.1469-8986.2007.00638.x

Höfel, L., & Jacobsen, T. (2007). Electrophysiological indices of processing aesthetics: Spontaneous or intentional processes? *International Journal of Psychophysiology*, *65*(1), 20–31. https://doi.org/10.1016/j.ijpsycho.2007.02.007

Honing, H. (2012). Without it no music: Beat induction as a fundamental musical trait. *Annals of the New York Academy of Sciences*, *1252*(1), 85–91. https://doi.org/10.1111/j.1749-6632.2011.06402.x

Horton, C., & Srinivasan, R. (2013). Suppression of competing speech through entrainment of cortical oscillations. *J Neu-Rophysiol*, *109*, 3082–3093. https://doi.org/10.1152/jn.01026.2012.-People

Humphries, C., Love, T., Swinney, D., & Hickok, G. (2005). Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. *Human Brain Mapping*, *26*(2), 128–138. https://doi.org/10.1002/hbm.20148

Husain, G., Thompson, W. F., & Schellenberg, E. G. (2002). Effects of Musical Tempo and Mode on Arousal, Mood, and Spatial Abilities. *Music Perception*, *20*(2), 151–171. https://doi.org/10.1525/mp.2002.20.2.151

Imbir, K. K., Jurkiewicz, G., Duda-Goławska, J., Pastwa, M., & Zygierewicz, J. (2018). The N400/FN400 and lateralized readiness potential neural correlates of valence and origin of words' affective connotations in ambiguous task processing. *Frontiers in Psychology*, *9*(OCT), 1–18. https://doi.org/10.3389/fpsyg.2018.01981

Janata, P., Tomic, S. T., & Haberman, J. M. (2012). Sensorimotor coupling in music and the psychology of the groove. *Journal of Experimental Psychology: General*, *141*(1), 54–75. https://doi.org/10.1037/a0024208

Jäncke, L. (2012). The relationship between music and language. *Frontiers in Psychology*, *3*(APR). https://doi.org/10.3389/fpsyg.2012.00123

Jeong, J. W., Diwadkar, V. A., Chugani, C. D., Sinsoongsud, P., Muzik, O., Behen, M. E., Chugani, H. T., & Chugani, D. C. (2011). Congruence of happy and sad emotion in music and faces modifies cortical audiovisual activation. *NeuroImage*, *54*(4), 2973–2982. https://doi.org/10.1016/j.neuroimage.2010.11.017

Jolij, J., & Meurs, M. (2011). Music Alters visual perception. *PLoS ONE*, *6*(4). https://doi.org/10.1371/journal.pone.0018861

Jomori, I., Hoshiyama, M., Uemura, J., Nakagawa, Y., & Hoshino, A. (2013). Effects of emotional music on visual processes in inferior temporal area. *Cognitive Neuroscience*, *4*(April 2013), 21–30.

Jones, M. R., Boltz, M., Antons, C., Butler, D., Johnson, W., Kidd, G., Marsh, K., Marshbum, E., Michon, J., Pratt, M., Pugh, K., Ralston, J., & van Vreden, W. (1989). Dynamic Attending and Responses to Time. *Psychological Review*, *96*(3), 459–491.

Juslin, P., Liljeström, S., Västfjäll, D., & Lundqvist, L. (2010). *How does music evoke emotions? Exploring the underlying mechanisms.* https://psycnet.apa.org/record/2010-02543-022

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(6), 1797–1812. https://doi.org/10.1037/0096-1523.26.6.1797

Juslin, P. N., Harmat, L., & Eerola, T. (2014). What makes music emotionally significant? Exploring the underlying mechanisms. *Psychology of Music*, *42*(4). https://doi.org/10.1177/0305735613484548

Juslin, P. N., & Laukka, P. (2004). Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening. *Journal of New Music Research*, *33*(3), 217–238. https://doi.org/10.1080/0929821042000317813

Juslin, P. N., & Lindström, E. (2010). Musical expression of emotions: Modelling listeners' judgements of composed and performed features. *Music Analysis*, *29*(1–3), 334–364. https://doi.org/10.1111/j.1468-2249.2011.00323.x

Kamiyama, K. S., Abla, D., Iwanaga, K., & Okanoya, K. (2013). Interaction between musical emotion and facial expression as measured by event-related potentials. *Neuropsychologia*, *51*(3), 500–505. https://doi.org/10.1016/j.neuropsychologia.2012.11.031

Kaneshiro, B., Nguyen, D. T., Norcia, A. M., Dmochowski, J. P., & Berger, J. (2020). Natural music evokes correlated EEG responses reflecting temporal structure and beat. *NeuroImage*, *214*. https://doi.org/10.1016/j.neuroimage.2020.116559

Kaneshiro, B., Nguyen, D. T., Norcia, A. M., Dmochowski, J. P., & Berger, J. (2021). Inter-Subject EEG Correlation Reflects Time-Varying Engagement with Natural Music. *BioArXiv*. https://doi.org/10.1101/2021.04.14.439913

Kanske, P., Plitschka, J., & Kotz, S. A. (2011). Attentional orienting towards emotion: P2 and N400 ERP effects. *Neuropsychologia*, *49*(11), 3121–3129. https://doi.org/10.1016/j.neuropsychologia.2011.07.022

Kellenbach, M. L., Wijers, A. A., & Mulder, G. (2000). Visual semantic features are activated during the processing of concrete words: event-related potential evidence for perceptual semantic priming. *Cognitive Brain Research*, *10*, 67–75. www.elsevier.com/locate/bres

Keller, P. E., Novembre, G., & Hove, M. J. (2014). Rhythm in joint action: Psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1658). https://doi.org/10.1098/rstb.2013.0394

Kissler, J., Herbert, C., Peyk, P., & Junghofer, M. (2007). *Buzzwords Early Cortical Responses to Emotional Words During Reading*.

Klein, K., Melnyk, V., & Völckner, F. (2021). Effects of background music on evaluations of visual images. *Psychology and Marketing*, *38*(12), 2240–2246. https://doi.org/10.1002/mar.21588

Kliuchko, M., Brattico, E., Gold, B. P., Tervaniemi, M., Bogert, B., Toiviainen, P., & Vuust, P. (2019). Fractionating auditory priors: A neural dissociation between active and passive experience of musical sounds. *PLoS ONE*, *14*(5). https://doi.org/10.1371/JOURNAL.PONE.0216499

Kliuchko, M., Heinonen-Guzejev, M., Monacis, L., Gold, B. P., Heikkilä, K. v., Spinosa, V., Tervaniemi, M., & Brattico, E. (2015). The association of noise sensitivity with music listening, training, and aptitude. *Noise and Health*, *17*(78), 350–357. https://doi.org/10.4103/1463-1741.165065

Kliuchko, M., Heinonen-Guzejev, M., Vuust, P., Tervaniemi, M., & Brattico, E. (2016). A window into the brain mechanisms associated with noise sensitivity. *Scientific Reports*, *6*. https://doi.org/10.1038/srep39236

Kliuchko, M., Puoliväli, T., Heinonen-Guzejev, M., Tervaniemi, M., Toiviainen, P., Sams, M., & Brattico, E. (2018). Neuroanatomical substrate of noise sensitivity. *NeuroImage*, *167*, 309–315. https://doi.org/10.1016/j.neuroimage.2017.11.041

Knoeferle, K. M., Paus, V. C., & Vossen, A. (2017). An Upbeat Crowd: Fast In-store Music Alleviates Negative Effects of High Social Density on Customers' Spending. *Journal of Retailing*, *93*(4), 541–549. https://doi.org/10.1016/j.jretai.2017.06.004

Koelsch, S. (2011a). Toward a neural basis of music perception - a review and updated model. *Frontiers in Psychology*, *2*(JUN), 1–20. https://doi.org/10.3389/fpsyg.2011.00110

Koelsch, S. (2011b). Towards a neural basis of processing musical semantics. *Physics of Life Reviews*, *8*(2), 89–105. https://doi.org/10.1016/j.plrev.2011.04.004

Koelsch, S. (2011c). Towards a neural basis of processing musical semantics. *Physics of Life Reviews*, *8*(2), 89–105. https://doi.org/10.1016/j.plrev.2011.04.004

Koelsch, S. (2014a). Brain correlates of music-evoked emotions. *Nature Reviews Neuroscience*, *15*(3), 170–180. https://doi.org/10.1038/nrn3666

Koelsch, S. (2014b). Brain correlates of music-evoked emotions. In *Nature Reviews Neuroscience* (Vol. 15, Issue 3). https://doi.org/10.1038/nrn3666

Koelsch, S. (2020). A coordinate-based meta-analysis of music-evoked emotions. *NeuroImage*, *223*(January). https://doi.org/10.1016/j.neuroimage.2020.117350

Koelsch, S., Jacobs, A. M., Menninghaus, W., Liebal, K., Klann-Delius, G., von Scheve, C., & Gebauer, G. (2015). The quartet theory of human emotions: An integrative and neurofunctional model. *Physics of Life Reviews*, *13*, 1–27. https://doi.org/10.1016/j.plrev.2015.03.001

Koelsch, S., & Jancke, L. (2015). Music and the heart. *European Heart Journal*, *36*(44), 3043–3048. https://doi.org/10.1093/eurheartj/ehv430

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature Neuroscience*, *7*(3).

Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences*, *9*(12), 578–584. https://doi.org/10.1016/j.tics.2005.10.001

Koelsch, S., Vuust, P., & Friston, K. (2019). Predictive Processes and the Peculiar Case of Music. *Trends in Cognitive Sciences*, *23*(1), 63–77. https://doi.org/10.1016/j.tics.2018.10.006

Kraus, N., & Slater, J. (2015). Music and language: Relations and disconnections. In *Handbook of Clinical Neurology* (Vol. 129, pp. 207–222). https://doi.org/10.1016/B978-0-444-62630-1.00012-3

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540. https://doi.org/10.1038/nn.2303

Krumhansl, C. L., & Agres, K. R. (2008). Musical expectancy: The influence of musical structure on emotional response. *Behavioral and Brain Sciences*, *31*(5), 584–585. https://doi.org/10.1017/S0140525X08005384

Kumagai, Y., Arvaneh, M., & Tanaka, T. (2017). Familiarity affects entrainment of EEG in music listening. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00384

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Labbé, C., & Grandjean, D. (2014). Musical emotions predicted by feelings of entrainment. *Music Perception*, *32*(2), 170–185. https://doi.org/10.1525/MP.2014.32.2.170

Lamminmäki, S., Parkkonen, L., & Hari, R. (2014). Human Neuromagnetic Steady-State Responses to Amplitude-Modulated Tones, Speech, and Music. *Ear and Hearing*, *35*(4), 461–467.

Large, E. W. (2008). Resonating to Muscial Rhythm. In *Psyhology of Time* (pp. 189–231). Emerald.

Large, E. W., & Gray, P. M. (2015). Spontaneous tempo and rhythmic entrainment in a bonobo (Pan paniscus). *Journal of Comparative Psychology*, *129*(4), 317–328. https://doi.org/10.1037/com0000011

Lartillot, O., & Toiviainen, P. (2007). A Matlab toolbox for musical feature extraction from audio. *International Conference on Digital Audio Effects*, *237*, 244.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933. https://doi.org/10.1038/nrn2532

Laukka, P., Eerola, T., Thingujam, N. S., Yamasaki, T., & Beller, G. (2013). Universal and Culture-Specific Factors in the Recognition and Performance of Musical Affect Expressions. *Emotion*, *13*(3), 434–449. https://doi.org/10.1037/a0031388.supp

Leder, H., & Nadal, M. (2014). Ten years of a model of aesthetic appreciation and aesthetic judgments: The aesthetic episode - Developments and challenges in empirical aesthetics. *British Journal of Psychology*, *105*(4), 443–464. https://doi.org/10.1111/bjop.12084

Lee, J. Y., Lindquist, K. A., & Nam, C. S. (2017). Emotional granularity effects on event-related brain potentials during affective picture processing. *Frontiers in Human Neuroscience*, *11*(March), 1–18. https://doi.org/10.3389/fnhum.2017.00133

Lee, T., Lim, H., Kim, D. W., Hwang, S., & Yoon, K. (2016). System for matching paintings with music based on emotions. *SA 2016 - SIGGRAPH ASIA 2016 Technical Briefs*. https://doi.org/10.1145/3005358.3005366

Lenc, T., Keller, P. E., Varlet, M., & Nozaradan, S. (2020). Attention affects overall gain but not selective contrast at meter frequencies in the neural processing of rhythm. *BioRxiv*. https://doi.org/10.1101/2020.09.23.309443

Li, W., Qin, W., Liu, H., Fan, L., Wang, J., Jiang, T., & Yu, C. (2013). Subregions of the human superior frontal gyrus and their connections. *NeuroImage*, *78*, 46–58. https://doi.org/10.1016/j.neuroimage.2013.04.011

Limb, C. J. (2006). Structural and functional neural correlates of music perception. *Anatomical Record - Part A Discoveries in Molecular, Cellular, and Evolutionary Biology*, *288*(4), 435–446. https://doi.org/10.1002/ar.a.20316

Limbert, W. M., & Polzella, D. J. (1998). Effects of Music on the Perception of Paintings. *Empirical Studies of the Arts*, *16*(1), 33–39. https://doi.org/10.2190/v8bl-gbjk-tlfp-r321

Lin, Y. H., Young, I. M., Conner, A. K., Glenn, C. A., Chakraborty, A. R., Nix, C. E., Bai, M. Y., Dhanaraj, V., Fonseka, R. D., Hormovas, J., Tanglay, O., Briggs, R. G., & Sughrue, M. E. (2020). Anatomy and White Matter Connections of the Inferior Temporal Gyrus. *World Neurosurgery*, *143*, e656–e666. https://doi.org/10.1016/j.wneu.2020.08.058

Liu, J., Lughofer, E., & Zeng, X. (2017). Toward model building for visual aesthetic perception. *Computational Intelligence and Neuroscience*, *2*, 1–13. https://doi.org/10.1155/2017/1292801

Locher, P. J. (2006). The usefulness of eye movement recordings to subject an aesthetic episode with visual art to empirical scrutiny. *Psychology Science*, *48*(2), 106–114.

Locher, P., Krupinski, E. A., Mello-Thoms, C., & Nodine, C. F. (2007). Visual interest in pictorial art during an aesthetic experience. *Spatial Vision*, *21*(2), 55–77. www.brill.nl/sv

Logeswaran, N., & Bhattacharya, J. (2009). Crossmodal transfer of emotion by music. *Neuroscience Letters*, *455*(2), 129–133. https://doi.org/10.1016/j.neulet.2009.03.044

London, J. (2004). Hearing in Time: Psychological Aspects of Musical Meter. *Journal of Music Theory*, *48*(2), 325–336. https://about.jstor.org/terms

Lorch, C. A., Lorch, V., Diefendorf, A. O., & Earl, P. W. (1994). Effect of Stimulative and Sedative Music on Systolic Blood Pressure, Heart Rate, and Respiratory Rate in Premature Infants. *Journal of Music Therapy*, *XXXI*(2), 105–118. https://academic.oup.com/jmt/article/31/2/105/977263

Maidhof, C., Vavatzanidis, N., Prinz, W., Rieger, M., & Koelsch, S. (2010). Processing Expectancy Violations during Music Performance and Perception: An ERP Study. *Journal of Cognitive Neuroscience*, *22*(10), 2401–2413. https://doi.org/10.1162/jocn.2009.21332

Margulis, E. H. (2013). Repetition and emotive communication in music versus speech. *Frontiers in Psychology*, *4*(APR). https://doi.org/10.3389/fpsyg.2013.00167

Marin, M. M., & Leder, H. (2013). Examining Complexity across Domains: Relating Subjective and Objective Measures of Affective Environmental Scenes, Paintings and Music. *PLoS ONE*, *8*(8). https://doi.org/10.1371/journal.pone.0072412

Markey, P. S., Jakesch, M., & Leder, H. (2019a). Art looks different – Semantic and syntactic processing of paintings and associated neurophysiological brain responses. *Brain and Cognition*, *134*(May), 58–66. https://doi.org/10.1016/j.bandc.2019.05.008

Markey, P. S., Jakesch, M., & Leder, H. (2019b). Art looks different – Semantic and syntactic processing of paintings and associated neurophysiological brain responses. *Brain and Cognition*, *134*(August 2018), 58–66. https://doi.org/10.1016/j.bandc.2019.05.008

Marković, S. (2012). Components of aesthetic experience: Aesthetic fascination, aesthetic appraisal, and aesthetic emotion. *I-Perception*, *3*(1), 1–17. https://doi.org/10.1068/i0450aap

Marshall, S. K., & Cohen, A. J. (1988). Effects of Musical Soundtracks on Attitudes toward Animated Geometric Figures. *Music Perception*, *6*(1), 95–112.

Matthews, T. E., Thibodeau, J. N. L., Gunther, B. P., & Penhune, V. B. (2016). The Impact of Instrument-Specific Musical Training on Rhythm Perception and Production. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00069

Melcher, D., & Bacci, F. (2008). The visual system as a constraint on the survival and success of specific artworks. *Spatial Vision*, *21*(3–5), 347–362. www.brill.nl/sv

Melcher, D., & Bacci, F. (2013). Perception of emotion in abstract artworks: A multidisciplinary approach. In *Progress in Brain Research* (1st ed., Vol. 204). Copyright © 2013 Elsevier B.V. All rights reserved. https://doi.org/10.1016/B978-0-444-63287-6.00010-5

Meltzer, B., Reichenbach, C. S., Braiman, C., Schiff, N. D., Hudspeth, A. J., & Reichenbach, T. (2015). The steady-state response of the cerebral cortex to the beat of music reflects both the comprehension of music and attention. *Frontiers in Human Neuroscience*, *9*(AUGUST), 1–11. https://doi.org/10.3389/fnhum.2015.00436

Ménoret, M., Varnet, L., Fargier, R., Cheylus, A., Curie, A., des Portes, V., Nazir, T. A., & Paulignan, Y. (2014). Neural correlates of non-verbal social interactions: A dual-EEG study. *Neuropsychologia*, *55*(1), 85–97. https://doi.org/10.1016/j.neuropsychologia.2013.10.001

Micallef Grimaud, A., & Eerola, T. (2022). Emotional expression through musical cues: A comparison of production and perception approaches. *PLOS ONE*, *17*(12), e0279605. https://doi.org/10.1371/journal.pone.0279605

Miendlarzewska, E. A., & Trost, W. J. (2014). How musical training affects cognitive development: Rhythm, reward and other modulating variables. *Frontiers in Neuroscience*, *8 JAN*. https://doi.org/10.3389/fnins.2013.00279

Mojtabavi, H., Saghazadeh, A., Valenti, V. E., & Rezaei, N. (2020). Can music influence cardiac autonomic system? A systematic review and narrative synthesis to evaluate its impact on heart rate variability. *Complementary Therapies in Clinical Practice*, *39*. https://doi.org/10.1016/j.ctcp.2020.101162

Molnar-Szakacs, I., & Overy, K. (2006). Music and mirror neurons: from motion to 'e'motion. *Social Cognitive and Affective Neuroscience*, *1*(3), 235–241. https://doi.org/10.1093/scan/nsl029

Morett, L. M., Landi, N., Irwin, J., & McPartland, J. C. (2020). N400 amplitude, latency, and variability reflect temporal integration of beat gesture and pitch accent during language processing. *Brain Research*, *1747*(March), 147059. https://doi.org/10.1016/j.brainres.2020.147059

Morrison, S. J., Demorest, S. M., Aylward, E. H., Cramer, S. C., & Maravilla, K. R. (2003). fMRI investigation of cross-cultural music comprehension. *NeuroImage*, *20*(1), 378–384. https://doi.org/10.1016/S1053-8119(03)00300-8

Muth, C., Raab, M. H., & Carbon, C. C. (2016). Semantic stability is more pleasurable in unstable episodic contexts. On the relevance of perceptual challenge in art appreciation. *Frontiers in Human Neuroscience*, *10*(FEB2016). https://doi.org/10.3389/fnhum.2016.00043

Näätänen, R. (1995). The Mismatch Negativity: A Powerful Tool for Cognitive Neuroscience. *Ear & Hearing*, *16*(1), 6–18.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Nadal, M., Àngel Capó, M., Rosselló, J., & Cela-Conde, C. J. (2008). Towards a framework for the study of the neural correlates of aesthetic preference. *Spatial Vision*, *21*(5), 379–396. www.brill.nl/sv

Nadon, É., Tillmann, B., Saj, A., & Gosselin, N. (2021). The Emotional Effect of Background Music on Selective Attention of Adults. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.729037

Nave, K. M., Hannon, E. E., & Snyder, J. S. (2022). Steady state-evoked potentials of subjective beat perception in musical rhythms. *Psychophysiology*, *59*(2). https://doi.org/10.1111/psyp.13963

Nigam, A., Hoffman, J. E., & Simons, R. F. (1992). N400 to Semantically Anomalous Pictures and Words. *Journal of Cognitive Neuroscience*, *4*(1), 15–22. https://doi.org/10.1162/jocn.1992.4.1.15

Nodine, C. F., Locher, P. J., & Krupinski, E. A. (1993). *The Role of Formal Art Training on Perception and Aesthetic Judgment of Art Compositions*. *26*(3), 219–227. https://about.jstor.org/terms

Noguchi, K., Masaoka, Y., Satoh, K., Katoh, N., & Homma, I. (2012). Effect of Music on Emotions and Respiration. *Universal Journal of Medical Science*, *24*(1), 69–75.

Noulhiane, M., Mella, N., Samson, S., Ragot, R., & Pouthas, V. (2007). How Emotional Auditory Stimuli Modulate Time Perception. *Emotion*, *7*(4), 697–704. https://doi.org/10.1037/1528-3542.7.4.697

Novembre, G., & Iannetti, G. D. (2018). Tagging the musical beat: Neural entrainment or event-related potentials? *Proceedings of the National Academy of Sciences of the United States of America*, *115*(47), E11002–E11003. https://doi.org/10.1073/pnas.1815311115

Nozaradan, S. (2014). Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1658). https://doi.org/10.1098/rstb.2013.0393

Nozaradan, S. (2017). Musical rhythm embedded in the brain: Bridging music, neuroscience, and empirical aesthetics. In *Enhancing Participation in the Arts in the EU: Challenges and Methods* (pp. 99–113). Springer International Publishing. https://doi.org/10.1007/978-3-319-09096-2_7

Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *Journal of Neuroscience*, *31*(28), 10234–10240. https://doi.org/10.1523/JNEUROSCI.0411-11.2011

Nozaradan, S., Peretz, I., & Mouraux, A. (2012). Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *Journal of Neuroscience*, *32*(49), 17572–17581. https://doi.org/10.1523/JNEUROSCI.3203-12.2012

Oldfield, R. C. (1971). THE ASSESSMENT AND ANALYSIS OF HANDEDNESS: THE EDINBURGH INVENTORY. *Neuropsychologia*, *9*, 97–113.

Olofsson, J. K., Nordin, S., Sequeira, H., & Polich, J. (2008). Affective picture processing: An integrative review of ERP findings. *Biological Psychology*, *77*(3), 247–265. https://doi.org/10.1016/j.biopsycho.2007.11.006

Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, *112*(4), 713–719. https://doi.org/10.1016/S1388-2457(00)00527-7

Overy, K. (2012). Making music in a group: Synchronization and shared experience. *Annals of the New York Academy of Sciences*, *1252*(1), 65–68. https://doi.org/10.1111/j.1749-6632.2012.06530.x

Overy, K., & Molnar-Szakacs, I. (2009). Being together in time: Musical experience and the mirror neuron system. *Music Perception*, *26*(5), 489–504. https://doi.org/10.1525/mp.2009.26.5.489

Pacchetti, C., Mancini, F., Aglieri, R., Fundarˋo, C., Fundarˋo, F., Martignoni, E., & Nappi, G. (2000). Active Music Therapy in Parkinson's Disease: An Integrative Method for Motor and Emotional Rehabilitation. *Psychosomatic Medicine*, *62*, 386–393.

Painter, J. G., & Koelsch, S. (2011). Can out-of-context musical sounds convey meaning? An ERP study on the processing of meaning in music. *Psychophysiology*, *48*(5), 645–655. https://doi.org/10.1111/j.1469-8986.2010.01134.x

Palmero-Soler, E., Dolan, K., Hadamschek, V., & Tass, P. A. (2007). swLORETA: A novel approach to robust source localization and synchronization tomography. *Physics in Medicine and Biology*, *52*(7), 1783–1800. https://doi.org/10.1088/0031-9155/52/7/002

Pando-Naude, V., Patyczek, A., Bonetti, L., & Vuust, P. (2021). Revisiting top-down and bottom-up brain processes: a systematic review and ALE meta-analysis of musical studies. *BioRxiv*, 1–39. https://doi.org/10.1101/2021.04.26.441417

Pang, C. Y., Nadal, M., Müller-Paul, J. S., Rosenberg, R., & Klein, C. (2013). Electrophysiological correlates of looking at paintings and its association with art expertise. *Biological Psychology*, *93*(1), 246–254. https://doi.org/10.1016/j.biopsycho.2012.10.013

Parrott, A. C. (1982). Effects of Paintings and Music, both alone and in combination, on emotional judgments. *Perceptual and Motor Skills*, *54*, 635–641.

Pascual-Marqui, R. D., Esslen, M., Kochi, K., & Lehmann, D. (2002). Functional imaging with low-resolution brain electromagnetic tomography (LORETA): A review. *Methods and Findings in Experimental and Clinical Pharmacology*, *24*(SUPPL. C), 91–95.

Pascual-Marqui, R. D., Michel, C. M. M., Lehmann, D., Pascual-Marquia, R. D., D.Lehmannb, & Michel, C. M. M. (1994). Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *Int J Psychophysiol*, *18*(1), 49–65. https://doi.org/10.1016/0167-8760(84)90014-X

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674–681. https://doi.org/10.1038/nn1082

Patel, A. D. (2006). Musical rhythm, linguistic rhythm, and human evolution. *Music Perception*, *24*(1), 99–104. https://doi.org/10.1525/mp.2006.24.1.99

Patel, A. D., Iversen, J. R., Bregman, M. R., Schulz, I., & Schulz, C. (2008). Investigating the human-specificity of synchronization to music. *Proceedings Od the Intl. Conf. on Music Perception and Cognition*, 1–5.

Pelowski, M., Markey, P. S., Forster, M., Gerger, G., & Leder, H. (2017). Move me, astonish me… delight my eyes and brain: The Vienna Integrated Model of top-down and bottom-up processes in Art Perception (VIMAP) and corresponding affective, evaluative, and neurophysiological correlates. *Physics of Life Reviews*, *21*, 80–125. https://doi.org/10.1016/j.plrev.2017.02.003

Pereira, C. S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S. L., & Brattico, E. (2011). Music and Emotions in the Brain: Familiarity Matters. *PLoS ONE*, *6*(11). https://doi.org/10.1371/journal.pone.0027241

Pinker, S. (1997). *How the Mind works*. W.W. Norton & Co.

Plourde-Kelly, A. D., Saroka, K. S., & Dotta, B. T. (2021). The impact of emotionally valenced music on emotional state and EEG profile: Convergence of self-report and quantitative data. *Neuroscience Letters*, *758*(June), 136009. https://doi.org/10.1016/j.neulet.2021.136009

Polich, J., & Kokb, A. (1995). Cognitive and biological determinants of P300: an integrative review. In *Biological Psychology* (Vol. 41).

Proverbio, A. M., Attardo, L., Cozzi, M., & Zani, A. (2015). The effect of musical practice on gesture/sound pairing. *Frontiers in Psychology*, *6*(APR), 1–7. https://doi.org/10.3389/fpsyg.2015.00376

Proverbio, A. M., Camporeale, E., & Brusa, A. (2020). Multimodal Recognition of Emotions in Music and Facial Expressions. *Frontiers in Human Neuroscience*, *14*(February), 1–14. https://doi.org/10.3389/fnhum.2020.00032

Proverbio, A. M., Crotti, N., Manfredi, M., Adorni, R., & Zani, A. (2012). Who needs a referee? How incorrect basketball actions are automatically detected by basketball players' brain. *Scientific Reports*, *2*. https://doi.org/10.1038/srep00883

Proverbio, A. M., & de Benedetto, F. (2018). Auditory enhancement of visual memory encoding is driven by emotional content of the auditory material and mediated by superior frontal cortex. *Biological Psychology*, *132*, 164–175. https://doi.org/10.1016/j.biopsycho.2017.12.003

Proverbio, A. M., de Benedetto, F., & Guazzone, M. (2020). Shared neural mechanisms for processing emotions in music and vocalizations. *European Journal of Neuroscience*, *April*, 1–21. https://doi.org/10.1111/ejn.14650

Proverbio, A. M., Gabaro, V., Orlandi, A., & Zani, A. (2015). Semantic brain areas are involved in gesture comprehension: An electrical neuroimaging study. *Brain and Language*, *147*, 30–40. https://doi.org/10.1016/j.bandl.2015.05.002

Proverbio, A. M. M., & de Benedetto, F. (2018). Auditory enhancement of visual memory encoding is driven by emotional content of the auditory material and mediated by superior frontal cortex. *Biological Psychology*, *132*, 164–175. https://doi.org/10.1016/j.biopsycho.2017.12.003

Proverbio, A. M., Manfrin, L., Arcari, L. A., de Benedetto, F., Gazzola, M., Guardamagna, M., Nasi, V. L., & Zani, A. (2015). Non-expert listeners show decreased heart rate and increased blood pressure (fear bradycardia) in response to atonal music. *Frontiers in Psychology*, *6*(OCT), 1–13. https://doi.org/10.3389/fpsyg.2015.01646

Proverbio, A. M., Nasi, V. L., Arcari, L. A., de Benedetto, F., Guardamagna, M., Gazzola, M., & Zani, A. (2015). The effect of background music on episodic memory and autonomic responses: Listening to emotionally touching music enhances facial memory capacity. *Scientific Reports*, *5*(October), 1–12. https://doi.org/10.1038/srep15219

Proverbio, A. M., & Piotti, E. (2021). Common neural bases for processing speech prosody and music: An integrated model. *Psychology of Music*, 1–16. https://doi.org/10.1177/03057356211050117

Proverbio, A. M., & Riva, F. (2009). RP and N400 ERP components reflect semantic violations in visual processing of human actions. *Neuroscience Letters*, *459*(3), 142–146. https://doi.org/10.1016/j.neulet.2009.05.012

Proverbio, A. M., & Russo, F. (2021). Multimodal recognition of emotions in music and language. *Psychology of Music*, 1–15.

Proverbio, A. M., Santoni, S., & Adorni, R. (2020). ERP Markers of Valence Coding in Emotional Speech Processing. *IScience*, *23*(3), 100933. https://doi.org/10.1016/j.isci.2020.100933

Proverbio, A. M., Spirito, D., & de Benedetto, F. (2022). The interactive effects of conjoined music and visual art processing. *PsyArXiv*.

Putkinen, V., Nazari-Farsani, S., Seppälä, K., Karjalainen, T., Sun, L., Karlsson, H. K., Hudson, M., Heikkilä, T. T., Hirvonen, J., & Nummenmaa, L. (2021). Decoding Music-Evoked Emotions in the Auditory and Motor Cortex. *Cerebral Cortex*, *31*(5), 2549–2560. https://doi.org/10.1093/cercor/bhaa373

Raglio, A. (2015). Effects of music and music therapy on mood in neurological patients. *World Journal of Psychiatry*, *5*(1), 68. https://doi.org/10.5498/wjp.v5.i1.68

Rajendran, V. G., & Schnupp, J. W. H. (2019). Frequency tagging cannot measure neural tracking of beat or meter. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(8), 2779–2780. https://doi.org/10.1073/pnas.1820020116

Ray, K. D., & Götell, E. (2018). The use of music and music therapy in ameliorating depression symptoms and improving well-being in nursing home residents with dementia. *Frontiers in Medicine*, *5*(OCT), 1–6. https://doi.org/10.3389/fmed.2018.00287

Redies, C. (2015). Combining universal beauty and cultural context in a unifying model of visual aesthetic experience. *Frontiers in Human Neuroscience*, *9*(APR), 1–20. https://doi.org/10.3389/fnhum.2015.00218

Reybrouck, M., & Podlipniak, P. (2019). Preconceptual spectral and temporal cues as a source of meaning in speech and music. *Brain Sciences*, *9*(3). https://doi.org/10.3390/brainsci9030053

Roß, B., Borgmann, C., Draganova, R., Roberts, L. E., & Pantev, C. (2000). A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones. *The Journal of the Acoustical Society of America*, *108*(2), 679–691.

Ross, J. M., & Balasubramaniam, R. (2014). Physical and neural entrainment to rhythm: Human sensorimotor coordination across tasks and effector systems. In *Frontiers in Human Neuroscience* (Vol. 8, Issue AUG). Frontiers Media S. A. https://doi.org/10.3389/fnhum.2014.00576

Russel, J. A. (1980). A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178.

Saari, P., Burunat, I., Brattico, E., & Toiviainen, P. (2018). Decoding Musical Training from Dynamic Processing of Musical Features in the Brain. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-19177-5

Sabatinelli, D., Bradley, M. M., Fitzsimmons, J. R., & Lang, P. J. (2005). Parallel amygdala and inferotemporal activation reflect emotional intensity and fear relevance. *NeuroImage*, *24*(4), 1265–1270. https://doi.org/10.1016/j.neuroimage.2004.12.015

Sakka, L. S., & Juslin, P. N. (2018a). Emotion regulation with music in depressed and non-depressed individuals. *Music &Science*, *1*, 205920431875502. https://doi.org/10.1177/2059204318755023

Sakka, L. S., & Juslin, P. N. (2018b). Emotional reactions to music in depressed individuals. *Psychology of Music*, *46*(6), 862–880. https://doi.org/10.1177/0305735617730425

Salimpoor, V. N., Zald, D. H., Zatorre, R. J., Dagher, A., & McIntosh, A. R. (2015). Predictions and the brain: How musical sounds become rewarding. *Trends in Cognitive Sciences*, *19*(2), 86–91. https://doi.org/10.3389/fnhum.2017.00168

Savage, P. E. (2019). Cultural evolution of music. *Palgrave Communications*, *5*(1). https://doi.org/10.1057/s41599-019-0221-1

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695–729. https://doi.org/https://doi.org/10.1177/1754073912445818

Scherer, K. R., & Oshinsky, J. S. (1977). Cue Utilization in Emotion Attribution from Auditory Stimuli 1. In *Motivation and Emotion* (Vol. 1, Issue 4).

Schlenker, P. (2017). Outline of music semantics. *Music Perception*, *35*(1), 3–37. https://doi.org/10.1525/MP.2017.35.1.3

Schlenker, P. (2019). Prolegomena to Music Semantics. *Review of Philosophy and Psychology*, *10*(1), 35–111. https://doi.org/10.1007/s13164-018-0384-5

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, *32*(1), 9–18. https://doi.org/10.1016/j.tins.2008.09.012

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences*, *12*(3), 106–113. https://doi.org/10.1016/j.tics.2008.01.002

Schupp, H. T., Junghöfer, M., Öhman, A., Weike, A. I., Stockburger, J., & Hamm, A. O. (2004). The facilitated processing of threatening faces: An ERP analysis. *Emotion*, *4*(2), 189–200. https://doi.org/10.1037/1528-3542.4.2.189

Schupp, H. T., Junghöfer, M., Weike, A. I., & Hamm, A. O. (2003). Attention and emotion: An ERP analysis of facilitated emotional stimulus processing. *NeuroReport*, *14*(8), 1107–1110. https://doi.org/10.1097/00001756-200306110-00002

Schutz, M. (2017). Acoustic constraints and musical consequences: Exploring composers' use of cues for musical emotion. *Frontiers in Psychology*, *8*(NOV), 1–10. https://doi.org/10.3389/fpsyg.2017.01402

Silva-Pereyra, J., Rivera-Gaxiola, M., Aubert, E., Bosch, J., Galán, L., & Salazar, A. (2003). N400 during lexical decision tasks: A current source localization study. *Clinical Neurophysiology*, *114*(12), 2469–2486. https://doi.org/10.1016/S1388-2457(03)00248-7

Skaansar, J. F., Laeng, B., & Danielsen, A. (2019). Microtiming and mental effort: Onset asynchronies in musical rhythm modulate pupil size. *Music Perception*, *37*(2), 111–133. https://doi.org/10.1525/MP.2019.37.2.111

Slevc, L. R. (2012). Language and music: Sound, structure, and meaning. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*(4), 483–492. https://doi.org/10.1002/wcs.1186

Slevc, L. R., & Patel, A. D. (2011). Meaning in music and language: Three key differences. Comment on "Towards a neural basis of processing musical semantics" by Stefan Koelsch. *Physics of Life Reviews*, *8*(2), 110–111. https://doi.org/10.1016/j.plrev.2011.05.003

Sollberger, B., Reber, R., & Eckstein, D. (2003). Musical chords as affective priming context in a word-evalutation task. *Music Perception*, *20*(3), 263–282.

Specker, E., Tinio, P. P. L., & van Elk, M. (2017). Do you see what i see? An investigation of the aesthetic experience in the laboratory and museum. *Psychology of Aesthetics, Creativity, and the Arts*, *11*(3), 265–275. https://doi.org/10.1037/aca0000107

Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological, and neural responses. *Journal of Cognitive Neuroscience*, *18*(8), 1380–1393. https://doi.org/10.1162/jocn.2006.18.8.1380

Stupacher, J., Matthews, T. E., Pando-Naude, V., Foster Vander Elst, O., & Vuust, P. (2022). The sweet spot between predictability and surprise: musical groove in brain, body, and social interactions. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.906190

Stupacher, J., Wood, G., & Witte, M. (2017). Neural entrainment to polyrhythms: A comparison of musicians and non-musicians. *Frontiers in Neuroscience*, *11*(APR). https://doi.org/10.3389/fnins.2017.00208

Swain, J. P., & Swain, J. (1997). *Musical Languages* (WW Norton & Company). https://scholar.google.it/scholar?hl=it&as_sdt=0%2C5&q=Swain%2C+J.P.+%281997%29.+Musical+Languages.+New+York%3A+W.W.+Norton+%26+Company.+&btnG=

Tabei, K. I. (2015). Inferior Frontal Gyrus Activation Underlies the Perception of Emotions, while Precuneus Activation Underlies the Feeling of Emotions during Music Listening. *Behavioural Neurology*, *2015*. https://doi.org/10.1155/2015/529043

Tal, I., Large, E. W., Rabinovitch, E., Wei, Y., Schroeder, C. E., Poeppel, D., & Golumbic, E. Z. (2017). Neural entrainment to the beat: The "missing-pulse" phenomenon. *Journal of Neuroscience*, *37*(26), 6331–6341. https://doi.org/10.1523/JNEUROSCI.2500-16.2017

Taulu, S., & Hari, R. (2009). Removal of magnetoencephalographic artifacts with temporal signal-space separation: demonstration with single-trial auditory-evoked responses. *Human Brain Mapping*, *30*(5), 1524–1534. https://doi.org/10.1002/hbm.20627

Terry, P. C., Curran, M. L., Karageorghis, C. I., & Martin, O. v. (2020). Effects of Music in Exercise and Sport: A Meta-Analytic Review. *Psychological Bulletin*, *146*(2), 91–117. https://doi.org/10.1037/bul0000216.supp

Thaut, M. H. (2010). Neurologic music therapy in cognitive rehabilitation. *Music Perception*, *27*(4), 281–285. https://doi.org/10.1525/mp.2010.27.4.281

Thayer, J. F., & Lane, R. D. (2000). A model of neurovisceral integration in emotion regulation and dysregulation. *Journal of Affective Disorders*, *61*, 201–216. www.elsevier.com/locate/jad

Thompson, W. F., & Balkwill, L.-L. (2010). Cross-cultural similarities and differences. In P. N. Juslin & J. Sloboda (Eds.), *Handbook of Music and Emotion: Theory, Research, Applications* (pp. 755–788). Oxford University Press.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2001). Arousal, Mood, and The Mozart Effect. *Psychological Science*, *12*(3), 248–251. https://doi.org/10.1111/1467-9280.00345

Ticini, L. F. (2017). The role of the orbitofrontal and dorsolateral prefrontal cortices in aesthetic preference for art. *Behavioral Sciences*, *7*(2). https://doi.org/10.3390/bs7020031

Timmers, R., & Crook, H. (2014). Affective priming in music listening: Emotions as a source of musical expectation. *Music Perception*, *31*(5), 470–484. https://doi.org/10.1525/MP.2014.31.5.470

Tinio, P. P. L. (2013). From artistic creation to aesthetic reception: The mirror model of art. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(3), 265–275. https://doi.org/10.1037/a0030872

Trost, W., Ethofer, T., Zentner, M., & Vuilleumier, P. (2012). Mapping aesthetic musical emotions in the brain. *Cerebral Cortex*, *22*(12), 2769–2783. https://doi.org/10.1093/cercor/bhr353

Trost, W., Frühholz, S., Cochrane, T., Cojan, Y., & Vuilleumier, P. (2015). Temporal dynamics of musical emotions examined through intersubject synchrony of brain activity. *Social Cognitive and Affective Neuroscience*, *10*(12), 1705–1721. https://doi.org/10.1093/scan/nsv060

Trost, W., Frühholz, S., Schön, D., Labbé, C., Pichon, S., Grandjean, D., & Vuilleumier, P. (2014). Getting the beat: Entrainment of brain activity by musical rhythm and pleasantness. *NeuroImage*, *103*, 55–64. https://doi.org/10.1016/j.neuroimage.2014.09.009

Trost, W., Labbé, C., & Grandjean, D. (2017). Rhythmic entrainment as a musical affect induction mechanism. *Neuropsychologia*, *96*, 96–110. https://doi.org/10.1016/j.neuropsychologia.2017.01.004

Trost, W., & Vuilleumier, P. (2013). *Rhythmic entrainment as a mechanism for emotion induction by music: a neurophysiological perspective*.

van Paasschen, J., Bacci, F., & Melcher, D. P. (2015). The influence of art expertise and training on emotion and preference ratings for representational and abstract artworks. *PLoS ONE*, *10*(8). https://doi.org/10.1371/journal.pone.0134241

van Peer, J. M., Grandjean, D., & Scherer, K. R. (2014). Sequential unfolding of appraisals: EEG evidence for the interaction of novelty and pleasantness. *Emotion*, *14*(1), 51–63. https://doi.org/10.1037/a0034566

van Veen, V., & Carter, C. S. (2002). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Psychology and Behavior*, *77*, 477–482.

Vartanian, O., & Skov, M. (2014). Neural correlates of viewing paintings: Evidence from a quantitative meta-analysis of functional magnetic resonance imaging data. *Brain and Cognition*, *87*(1), 52–56. https://doi.org/10.1016/j.bandc.2014.03.004

Vessel, E. A., Gabrielle Starr, G., & Rubin, N. (2012). The brain on art: Intense aesthetic experience activates the default mode network. *Frontiers in Human Neuroscience*, *6*(MARCH 2012), 1–17. https://doi.org/10.3389/fnhum.2012.00066

Visser, M., Jefferies, E., Embleton, K. v, & Ralph, M. A. L. (2012). *Both the Middle Temporal Gyrus and the Ventral Anterior Temporal Area Are Crucial for Multimodal Semantic Processing: Distortion-corrected fMRI Evidence for a Double Gradient of Information Convergence in the Temporal Lobes*. http://mitprc.silverchair.com/jocn/article-pdf/24/8/1766/1778176/jocn_a_00244.pdf

Völker, J. (2021). Measuring emotional music experience: Spreading activation and BRECVEMA mechanisms. *Psychology of Music*. https://doi.org/10.1177/03057356211010224

Vuilleumier, P., & Trost, W. (2015). Music and emotions: From enchantment to entrainment. *Annals of the New York Academy of Sciences*, *1337*(1), 212–222. https://doi.org/10.1111/nyas.12676

Vuust, P., Heggli, O. A., Friston, K. J., & Kringelbach, M. L. (2022). Music in the brain. *Nature Reviews Neuroscience*, *23*(5), 287–305. https://doi.org/10.1038/s41583-022-00578-5

Vuust, P., & Witek, M. A. G. (2014). Rhythmic complexity and predictive coding: A novel approach to modeling rhythm and meter perception in music. *Frontiers in Psychology*, *5*(SEP). https://doi.org/10.3389/fpsyg.2014.01111

Waberski, D., Kreitschmann-Andermahr, I., Kawohl, W., Darvas, F., Ryang, Y., Gobbele, R. Â., & Buchner, H. (2001). Spatio-temporal source imaging reveals subcomponents of the human auditory mismatch negativity in the cingulum and right inferior temporal gyrus. *Neuroscience Letters*, *308*, 107–110. www.elsevier.com/locate/neulet

Weimer, N. R., Clark, S. L., & Freitas, A. L. (2019). Distinct neural responses to social and semantic violations: An N400 study. *International Journal of Psychophysiology*, *137*(October 2018), 72–81. https://doi.org/10.1016/j.ijpsycho.2018.12.006

Winkler, I. (2007). Interpreting the Mismatch Negativity. *Interpreting MMN Journal of Psychophysiology*, *21*(4), 147–163. https://doi.org/10.1027/0269-8803.21.3.147

Wöhner, S., Jescheniak, J. D., & Mädebach, A. (2020). Semantic interference is not modality specific: Evidence from sound naming with distractor pictures. *Quarterly Journal of Experimental Psychology*, *73*(12), 2290–2308. https://doi.org/10.1177/1747021820943130

Wood, S., & Kisley, M. A. (2006). The negativity bias is eliminated in older adults: Age-related reduction in event-related brain potentials associated with evaluative categorization. *Psychology and Aging*, *21*(4), 815–820. https://doi.org/10.1037/0882-7974.21.4.815

Yao, Z., Yu, D., Wang, L., Zhu, X., Guo, J., & Wang, Z. (2016). Effects of valence and arousal on emotional word processing are modulated by concreteness: Behavioral and ERP evidence from a lexical decision task. *International Journal of Psychophysiology*, *110*, 231–242. https://doi.org/10.1016/j.ijpsycho.2016.07.499

Yun, K., Watanabe, K., & Shimojo, S. (2012). Interpersonal body and neural synchronization as a marker of implicit social interaction. *Scientific Reports*, *2*. https://doi.org/10.1038/srep00959

Yuval-Greenberg, S., & Deouell, L. Y. (2009). The dog's meow: Asymmetrical interaction in cross-modal object recognition. *Experimental Brain Research*, *193*(4), 603–614. https://doi.org/10.1007/s00221-008-1664-6

Zatorre, R. J. (2015). Musical pleasure and reward: Mechanisms and dysfunction. *Annals of the New York Academy of Sciences*, *1337*(1), 202–211. https://doi.org/10.1111/nyas.12677

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions Evoked by the Sound of Music: Characterization, Classification, and Measurement. *Emotion*, *8*(4), 494–521. https://doi.org/10.1037/1528-3542.8.4.494

Zhou, L., Jiang, C., Delogu, F., & Yang, Y. (2014). Spatial conceptial associations between music and pictures as revealed by N400 effect. *Psychophysiology*, *51*, 520–528.

Zitnick, C. L., Vedantam, R., & Parikh, D. (2016). Adopting abstract images for semantic scene understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(4), 627–638. https://doi.org/10.1109/TPAMI.2014.2366143