Dynamics of brain activity underlying working memory for music in a naturalistic condition

Master Thesis

Iballa Burunat

November 2012

Master's Degree Programme in Music, Mind and Technology

UNIVERSITY OF JYVÄSKYLÄ

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄN YLIOPISTO

Tiedekunta – Faculty Humanities	Laitos – Department Music
Tekijä – Author Iballa Burunat	
Työn nimi – Title Dynamics of brain activity underlying working memory for music in a naturalistic condition	
Oppiaine – Subject Music, Mind and Technology	Työn laji – Level Master degree
Aika – Month and year November 2012	Sivumäärä – Number of pages 70

Tiivistelmä – Abstract

Working memory (WM) is at the core of any cognitive function as it is necessary for the integration of information over time. Despite WM's critical role in high-level cognitive functions, its implementation in the neural tissue is poorly understood. Preliminary studies on auditory WM show differences between linguistic and musical memory, leading to the speculation of specific neural networks encoding memory for music. Moreover, in neuroscience WM has not been studied in naturalistic listening conditions but rather in artificial settings (e.g., n-back and Sternberg tasks). Western tonal music provides naturally occurring motivic repetition and variation, recognizable units serving as WM trigger, thus allowing us to study the phenomenon of motif-tracking in the context of real music. Adopting a modern tango as stimulus, behavioural methods were used to identify the stimulus motifs and build a time-course predictor of WM neural responses. This predictor was then correlated with the participants' functional magnetic resonance imaging (fMRI) signal obtained during a continuous listening condition. Neural correlates related to the sensory processing of a set of musical features were filtered out from the brain responses to music to aid in the exclusive recruitment of executive processes of music-related WM. Correlational analysis revealed a widely distributed network of cortical and subcortical areas, predominantly right-lateralized, responding to the WM condition, including ventral and dorsal areas in the prefrontal cortex, basal ganglia, and limbic areas. Significant subcortical processing areas, active in response to the WM condition, were pruned with the removal of the acoustic content, suggesting these music-related perceptual processing areas might aid in the encoding and retrieval of WM. The pattern of dispersed neural activity indicates WM to emerge coherently from the integration of distributed neural activity spread out over different brain subsystems (motoric-, cognitive- and sensory-related areas of the brain).

Asiasanat – Keywords working memory; cognitive neuroscience; music; musical motifs; functional magnetic resonance imaging (fMRI); correlational analysis; distributed networks;
Säilytyspaikka – Depository
Muita tietoja – Additional information

A mi troika de la felicidad: Rochi, José y Gis

Acknowledgments

In an attempt to answer this question [localization of brain processes], the neurologist Karl Lashley, in a series of experiments beginning around 1920 and running for many years, tried to discover where in its brain a rat stores its knowledge about maze running. In his book The Conscious Brain, Steven Rose describes Lashley's trials and tribulations this way:

> Lashley was attempting to identify the locus of memory within the cortex, and, to do so, first trained rats to run mazes, and then removed various cortical regions. He allowed the animals to recover and tested the retention of the maze-running skills. To his surprise it was not possible to find a particular region corresponding to the ability to remember the way through a maze. Instead all the rats which had had cortex regions removed suffered some kind of impairment, and the extent of the impairment was roughly proportional to the amount of cortex taken off. Removing cortex damaged the motor and sensory capacities of the animals, and they would limp, hop, roll, or stagger, but somehow they always managed to traverse the maze. So far as memory 'as concerned, the cortex appeared to be equipotential, that is, with all regions of equal possible utility. Indeed, Lashley concluded rather gloomily in his last paper "In Search of the Engram", which appeared in 1950, that the only conclusion was that memory was not possible at all.'

<div align="right">D. Hofstadter (1999)</div>

Contents

List of figures & tables

Glossary

3-T: 3 tesla

AC: acoustic components

ACF: autocorrelation function

BA: Brodmann area

BOLD: blood-oxygen-level-dependent;

CDF: cumulative distribution function

CMRO2: metabolic rate of oxygen

CS: cluster size

DFT: discrete Fourier transform

dlPFC: dorsolateral prefrontal cortex

dlPMC: dorsolateral premotor cortex

DoF: degrees of freedom

EEG: electroencephalography

fMRI: functional magnetic resonance imaging

HDR: hemodynamic response

HRF: hemodynamic response function

IFG: inferior frontal gyrus

IFS: inferior frontal sulcus

IPL: inferior parietal lobule

IPS: intraparietal sulcus

LTM: long-term memory

MedFG: medial frontal gyrus

MEG: magnetoencephalography

MFG: middle frontal gyrus

MNI: Montreal Neurological Institute

MTG: middle temporal gyrus

PET: positron emission tomography

PFC: prefrontal cortex

PMd: premotor dorsal

PMv: premotor ventral

PreG : precentral gyrus

pre-SMA: premotor supplementary motor area

rCFB: regional cerebral blood flow

SFG: superior frontal gyrus

SI: statistical image

SMA: supplementary motor area

SMG: superior marginal gyrus

SPL: superior parietal lobule

TAL: Talairach

TMS: transcranial magnetic stimulation

vlPFC: ventrolateral prefrontal cortex

vlPMC: ventrolateral premotor cortex

WM: working memory

1 INTRODUCTION

Working memory (WM) is at the core of any cognitive function as it is necessary for the integration of information over time (Nan, Knösche, Zysset, & Friederici, 2008) helping us make sense of the continuity of our experience of time and of our self. The study of WM is central to understanding how memory and thought work (Wager & Smith, 2003). Despite WM's critical role in high-level cognitive functions, its functioning and mapping in the neural tissue is poorly understood. If we want to comprehend all other aspects of cognition, it is fundamental to first explain how humans store and process information.

Why is the scientific study of WM for music important? Music is ubiquitous and seems to be associated with a distinct brain architecture. In recent years there has been a significant increase in research studies on low- and high-level music processing in the brain, including phenomena such as perception of psycho-acoustic features, performance, and music-driven emotion and memory, aimed to describe and understand the music-brain interaction: how music engages the brain and how it affects cognition in different ways. In addition to being the foundation of cognition, memory is also crucial in emotion, and it should be noted that emotion in music is thought to be one of the major factors that shape how and what we remember (Dolan, 2002). The mnemotechnic power of music is well known: "Beyond the repetitive motions of walking and dancing, music may allow an ability to organize, to follow intricate sequences, or to hold great volumes of information in mind—this is the narrative or mnemonic power of music" (Sacks, 2007). Indeed the engraving, persistent quality of musical memory is of extraordinarily remarkable accuracy, and the study of music-related memory circuits in the brain could illuminate on the intriguingly distinct way in which our selective brains listen to music. Finally it is important to emphasize that the study of how memory encodes music will also tell us about the nature of human memory in general.

Auditory WM has been mainly studied using vocal stimuli and only recently a few studies have started investigating the neural networks engaged in auditory WM for music. Preliminary findings in research on auditory WM show differences between linguistic and musical memory (Deutsch, 1970; Salamé & Baddeley, 1989), leading to the speculation of specific networks encoding memory for music. However, the finding that musical training enhances the performance during verbal tasks (Chan, Ho, & Cheung, 1998) reveals rather overlapping structures for verbal and tonal WM. Similarly, Koelsch, Schulze, Sammler, Fritz, Müller and Gruber (2009) studied auditory WM during rehearsal and storage for syllables and pitches in non-musicians and found that WM for both verbal and tonal information share neocortical, subcortical and cerebellar networks, providing evidence for a high degree of overlap between the functional architecture of verbal and tonal WM. Interestingly, a later study (Schulze, Zysset, Mueller, Friederici, & Koelsch, 2011) revealed specific verbal and

tonal-related WM components only in musicians, suggesting functional plasticity[1] induced by music training. Thus the question about the existence of a specialized memory system for non-phonological information remains open.

In addition to the scarcity of WM studies dealing specifically with music, most experimental settings typically employ simpler materials. Certainly in neuroscience WM has been rarely studied in naturalistic listening situations and rather using artificial target detection tasks (e.g., n-back and Sternberg[2] tasks) with manipulated stimuli, all of which might create mental states not characteristic of brain's behaviour in more natural, attentive situations.

If we consider that humans have evolved in a natural complex auditory scene environment, capable of segregating auditory objects for interaction and survival (Janata, 2002), it is reasonable to believe that in studying music-driven cognitive processes in the brain, more naturalistic approaches are crucial if we aim at a) mapping those functional brain areas engaged in acoustically complex environment-conditioned processing, and b) comparing and supporting the experimental findings resulting from the use of artificially created stimuli against more natural and complex approaches that most reliably replicate the acoustic environments our brains have adapted to.

Thus we used a naturalistic setting, denoting both a) a non-manipulated, complex, real-life music stimulus and b) a natural continuous, free listening condition. In our setting participants attentively listened to the piece from beginning to end, without performing any tasks. This allowed subjects to move away from possible mental states arising from such target detection tasks that may not be characteristic of brain's behaviour in more natural, attentive situations. Such paradigm constitutes an unusual approach as opposed to the usual practice in research studies focusing on auditory processing in the brain (Koelsch et al., 2009, Pallesen, Brattico, Bailey, Korvenoja, Koivisto, Gjedde, & Carlson, 2010; Pereira, Teixeira, Figueiredo, Xavier, Castro, & Brattico, 2011; Brattico, Alluri, Bogert, Jacobsen, Vartiainen, Nieminen, & Tervaniemi, 2011; Levitin & Menon, 2003; Janata, Tillmann, & Bharucha, 2002). Even if still ecologically significant, findings derived from traditional approaches employing artificially controlled musical stimuli would need to be validated against results coming from rich, naturalistic approaches, more representative of the complex auditory phenomena the brain has evolved to respond to.

---

1 Functional plasticity refers to the nervous system's remarkable ability to respond, reorganize and adapt in response to internal and external changes. This ability has important implications for learning (Bellis, 2003). The induced changes may occur as a consequence of very different events such as the normal development and maturation of the organism, the acquisition of new skills (learning), following damage to the nervous system and as a result of sensory deprivation (Shaw, McEachern, & Eachern, 2001). It is influenced by the constant interaction between the individual and his environment. Thus we can think of plasticity as the bridge between brain and behaviour (Gjelsvik, 2008).

2 During an n-back task participants must continuously monitor the identity or location of stimuli that appear sequentially, and indicate, usually by pressing a button, whether the currently presented stimulus has been presented n items before its onset (Owen, McMillan, Laird, & Bullmore, 2005), thus for n > 0 the task requires both maintenance and updating of the last n stimuli in order (Andrade, 2010). Pallesen et al. (2010) used the n-back paradigms in tasks about memorizing octaves of chords, whereby participants had to respond after each stimulus by pressing a button a) whether the octave of the chord matched that of the previous trial (n=1); b) whether the octave matched the chord presented two trials back (n=2). The Sternberg task (Sternberg, 1969) is another test of maintenance in WM, where participants are asked to press a button to indicate that the probe stimulus currently presented was previously displayed in a set of stimuli, usually containing 3-9 elements (Andrade, 2002).

Our goal in the present study was to determine the topography of music-elicited WM using a naturalistic attentive listening condition and a non-manipulated piece of music, and not to determine the specificity of neural networks recruited for musical WM versus verbal WM, as we did not use an analogous verbal condition that would allow drawing such conclusions. Thus the resulting findings do not extend beyond the scope of exploring the functional neuroarchitecture of WM for music in a naturalistic setting in musicians.

We hope that the findings of this study will offer a valuable contribution to the ongoing research on musical WM, and in WM in general, by a) using a naturalistic paradigm, whereby activation of WM-related neural networks is studied by tracing motivic repetition that naturally occurs in Western tonal music, to more traditional approaches, given the scarcity of such functional magnetic resonance imaging (fMRI) studies; b) controlling for the variance accounted by the acoustic features in the music in order to fine-tune the identification of WM function in the brain.

The structure of this thesis is as follows: first, a theoretical background presenting an overview of relevant background concepts related to methodology in the field of cognitive neuroscience used in the present study is provided. This section includes a description of some of the influential psychological and neuroanatomical theories of WM and the challenges they face, as well as a report on fMRI studies related to WM with auditory stimuli with special emphasis on the neuroanatomical findings. This is followed by an exhaustive explanation of the methodological process, including the perceptual and fMRI experiment. Finally, results are reported and discussed.

2 THEORETICAL BACKGROUND

2.1 COGNITIVE NEUROSCIENCE OF MUSIC

Cognitive neuroscience is the interdisciplinary study of human cognition with special emphasis in the neural substrates of cognitive processes, that is, it studies the biological substrates underpinning cognition, the fundamental question about the representation of knowledge in the brain.

In the words of Milner, Squire, and Kandel (1998):

> "...cognitive psychology was concerned not simply with specifying the input and output for a particular behaviour but also with analyzing the process by which sensory information is transformed into perception and action—that is, with evaluating how a stimulus leads to a particular behavioural response. In redirecting scientific attention to mental operations, cognitive psychologists focused on information processing, on the flow of sensory information from sensory receptors to its eventual use in memory and action. It was implicit in the cognitive approach to behaviour that each perceptual or motor act has an internal representation in the brain: a representation of information in patterns of neural activity".

Thus, in cognitive neuroscience particular signals of the nervous system are of interest inasmuch they can be used to explain cognitive functions, and in this sense it is a functional neuroscience. To relate changes in neural activity to specific cognitive functions, the cognitive psychology community has long acknowledged the necessity for the insight of neuroimaging techniques. Particularly, cognitive neuroscience has been long and profoundly concerned with the study of the neuronal mechanisms enabling the storage and retrieval of information about the world, since these bind almost every aspect of information processing (perception, decision making, motor control, emotion, and consciousness [Wilson & Keil, 2001]).

As for the growing interest and development of cognitive neuroscience in the musical domain, the last twenty years have been crucial in this respect. Music is ubiquitous, a human feature, as ancient as homo sapiens, deeply rooted in our biology, with a seemingly distinct and extensive functional neuroarchitecture, capable of inducing vivid, intense emotions, all of which makes it an appealing phenomenon to study different areas of human nature. Sensory-motor mechanisms can be also studied using music since they activate not only when performing, but also when listening to it (Lahav, Saltzman, & Schlaug, 2007; Zatorre, Chen, & Penhune, 2007; Haueisen & Knösche, 2001). Moreover, musicians —experts in the musical domain— exhibit functional and structural changes in the brain, what also has driven music into a device for studying brain plasticity (Hyde, Lerch, Norton, Forgeard, Winner, & Evans, 2009; Gaser & Schlaug, 2003; Schlaug, 2006; Münte, Altenmüller, & Jäncke, 2002).

2.1.1 FMRI AND THE BOLD SIGNAL

Functional magnetic resonance imaging (fMRI) is in neuroscience the prevalent neuroimaging method —with high spatial and medium temporal resolution— that has undergone an explosive growth in recent years, used by researchers and clinicians to image human brain activity in response to given mental tasks, allowing them to assess correlates of brain activity on a time scale of roughly a few seconds.

FMRI does not trace activity from single neurons, but rather activity arising from large population of neurons, and it does this in a non-invasive manner. Series of brain images are acquired during the course of an fMRI experiment, which allows researchers to measure the signal change between those images and make inferences (Lindquist, 2008). Thus fMRI provides a unique perspective on brain function. However, what researchers actually measure is not neuronal activity directly, i.e., changes in electrical potential or in chemical gradients. They use an indirect measure of brain activity given by another physiological marker: metabolic changes associated to neuronal activity. Specifically, fMRI uses blood properties as indices of brain activity. These properties fluctuate according to the metabolic demands of active neurons (the principle of neurovascular coupling: the relationship between changes in cerebral blood flow [CBF] and local neural activity; see Figure 1).
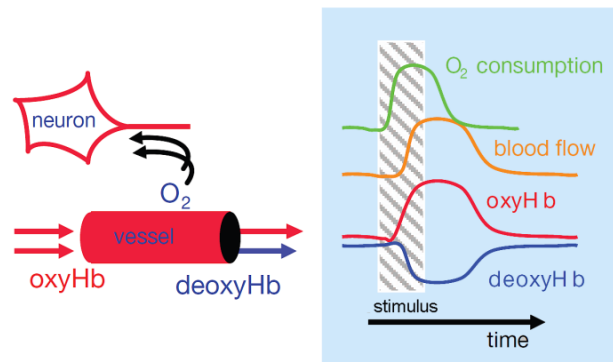
Figure 1. Principle of neurovascular coupling: neuronal activation requires a higher consumption of oxygen which is supplied by blood perfusing the tissue. Concentration of oxygenated and deoxygenated hemoglobins (oxyHb, deoxyHb) are modulated after a neuronal stimulus.

In particular, fMRI measures the change in magnetization between oxygen-rich (oxyhemoglobin) and oxygen-poor blood (deoxyhemoglobin), the so-called BOLD (blood-oxygen-level-dependent) contrast (Huettel, Song, & McCarthy, 2009). The justification of this contrast relies on the principle that neuronal activity demands an increase in energy, supplied through glucose and oxygen by the vascular system to the tissue. Oxygen is carried on molecules of hemoglobin, which has iron in it, and the magnetic properties of hemoglobin change based on whether they have oxygen attached to them (diamagnetic) or not (paramagnetic[3]). Thus the amount of iron flowing into that region

---

[3] Paramagnetic material is only magnetically attracted in presence of an externally applied magnetic field, whereas diamagnetic material is repelled by them (Miessler & Tarr, 2004).

changes, this is, the magnetic and thermodynamic properties of the area change. As a result, there is an increase in concentration of oxyhemoglobin and a decrease of deoxyhemoglobin in the area. Because deoxyhemoglobin has paramagnetic properties, it is precisely its relative decrease what makes it detectable by the scanner (Moridani, 2009). Hence we can see the difference magnetically between the resting state and the active state of the brain. This allows 'watching' the brain in action as it is working.
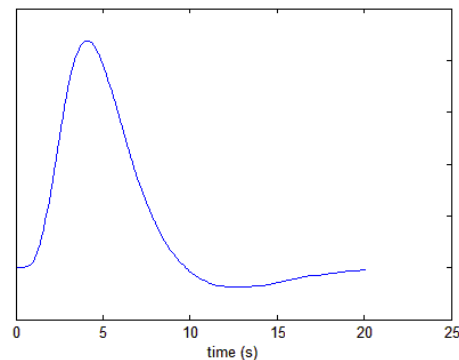
Figure 2. The standard canonical model for the HRF used in fMRI data analysis illustrates the main features of the response.

So when a part of our brains is used, the oxygen requirements in that area increase and the vascular system responds accordingly, although with certain delay. Thus the BOLD response to brief neuronal activity consists of a short onset delay, a rise to a peak after a few seconds, a return to baseline, and a prolonged undershoot (see Figure 2). Sometimes an initial decrease in the BOLD signal is reported due to initial oxygen extraction before increases in blood flow (Huettel et al., 2009). The course of changes in blood flow is called the hemodynamic response (HDR). The HDR has interferences from various sources, and statistical techniques are needed to remove the noise.

A linear system is often used to model the HDR, i.e., the magnitude of the HDR to individual stimuli is assumed to be equal to the summation of independent responses to each stimulus. However, neuronal activity expected very close in time (derived from not sufficiently separated stimuli) leads to a reduced than expected hemodynamic amplitude (known as refractory effect). Inter-stimulus interval should be at least 5-6 seconds, which seems to be the refractory period for many types of stimuli, to guarantee the return of the hemodynamic response to baseline. In the presence of refractory effects a linear model will overestimate the hemodynamic response and reduce effectiveness of experimental analysis (Huettel et al., 2009). Many statistical techniques to analyse fMRI data exist, which aim at producing a spatial map of localized significant signal changes in the brain in response to the task under investigation (Jezzard, Matthews, & Smith, 2001).

In short, the fMRI signal can be described as the underlying neuronal activity expressed through the hemodynamic response, with added noise (Frackowiak, Ashburner, Penny, & Zeki, 2004), and it is actually reflecting fluctuations in oxygen rich blood flow that lag behind the underlying neural activity, since the vascular system is very slow. FMRI can then show what parts of the brain are active over several seconds, what makes fMRI very good at telling where things are happening but not very good at telling when they are happening.

2.1.2 FMRI: PROS AND CONS

Despite currently being the backbone of neuroimaging in cognitive neuroscience, allowing the acquisition of knowledge and insights into brain function, there are some drawbacks associated with the use of fMRI. We mentioned earlier that the HDR is a very slow response, unlike magnetoencephalography (MEG) or electroencephalography (EEG) measurements, with a temporal resolution to the millisecond. Another limitation is the fact that fMRI does not provide evidence of a brain region being essential for a function, although this limitation also applies to EEG and MEG techniques as well. Thus it would require integration with i.e., transcranial magnetic stimulation (TMS) to allow for reversible interference (Jezzard et al., 2001). In addition, the magnitude of the fMRI signal reflects differences across brain regions or even conditions within the same region, a problem that does not derive from the inability to estimate cerebral metabolic rate of oxygen (CMRO2) from the BOLD signal, but to the sensitivity of the HDR to the spatial and temporal sparsity of the activated neuronal population (Logothetis, 2008). Furthermore, fMRI may potentially confuse excitation and inhibition, which complicates the interpretation of fMRI data.

Another downside is noted by Attwell and Iadecola (2002) when highlighting the assumption that HDR is determined by energy use of neuronal populations. It has been recently suggested that the HDR is driven by neurotransmitter-related signalling and not directly by the local energy needs of the brain, because most energy is used to power postsynaptic currents and action potentials rather than presynaptic or glial activity. Another consideration is the cost of fMRI: behavioural experiments are preferred over fMRI if the hypothesis can be addressed by both (Henson, 2005).

When analyzing fMRI time series, statistical packages like SPM facilitate the task but they can be easily misused if their principles are not fully understood. Statistical errors are frequent in fMRI analysis, one of the reasons of the fair amount of contradictory results. Another reason is that there are few attempts at replication (Henson, 2005). The interpretation and conclusion drawn from fMRI results often ignore the actual limitations of the methodology.

Finally, fMRI studies are an area of research very vulnerable to being sensationalised (Caulfield, Rachul, Zarzeczny, & Walter, 2010). Fancy, clean images of brains showing impressive results in form of red blobs with accompanying fitting scientific narrative are very seducing, and lead to think of fMRI as a translucent window, through which we can clearly and unmistakably observe

psychological processes as they happen inside the brain. That is the reason why neuroimaging is dangerously tempting for bad science and has been so often used for unethical commercialization. Logothetis (2008) indicates that many of the limitations of fMRI are not related to physics or poor engineering (and therefore solved by increasing the power of the scanners), but inherently to brain complex circuitry and functional organization, facts ignored by inappropriate designs. It is therefore central to the use of neuroimaging techniques that scientists fully understand their tools and agree on the experimental protocols suitable for fMRI in order to maximize the chances for significant unbiased results. Additionally, not all psychological theories can be confirmed or disconfirmed by the use of fMRI.

It should be noted that, despite all the disadvantages described, fMRI is certainly not the only methodology with limitations. Its non-invasive nature grants its wide use by neuroscientists and medical community. Of special importance is its high spatial resolution that allows locating certain critical areas vey precisely, which helps neurosurgeons minimize side effects when placing implants or removing tumours. Brain mapping is also useful in detecting distinct brain "signatures" that physical injury or some diseases might be identified with, as well as in diagnosing neurodegenerative diseases like Parkinson's and Alzheimer's, and tracking how treatments work. It principally contributes to observing our brains more intimately, while they learn and adapt to the environment.

In this section we have emphasized the limitations over the advantages of fMRI, since knowing the vulnerabilities of the technique is central in deciding how far to go when interpreting the results. To take full advantage of any methodology, researchers should understand its foundations, assumptions and limitations.

2.2   WHAT IS WORKING MEMORY?

2.2.1   SHORT-TERM AND LONG-TERM MEMORY

Memory is the capacity to encode, store and retrieve information (Wilson & Keil, 2001). It was in 1898 when W. James first hypothesized that memory might not be a unitary system. Half a century later the hypothesis of a two-component view (short-term and long-term memory [STM and LTM, respectively]) started to emerge upon empirical evidence (specially the amnesiac syndrome: preserved STM with impaired LTM or vice versa). STM reflects an ability to hold a limited amount of information that remains temporarily accessible and does not cause permanent anatomical or chemical changes in the synapses between nerve cells, whereas LTM persists much longer because it shapes structural changes in the brain (long-term memories result from grouping associated items through connections between neural networks [Squire & Kandel, 2008]). LTM is an extensive store and record of prior events, existing in all theoretical views (Cowan, 2008).

The modal model (Atkinson & Shiffrin, 1968) suggested that STM is crucial for LTM, thus damage to STM mechanisms would impair learning, but such evidence was not found. This motivated the dual-task paradigm advanced by Baddeley and Hitch (1974), where participants performed two distinct STM-dependent tasks, one of which was gradually increased to absorb more STM capacity, with the other task being impaired but never obliterated, thus leading to a reformulation of the STM hypothesis and the proposal of a multicomponent STM (term replaced by WM), one of the most influential models to date, containing an executive centre (limited attentional control) with two independent slave subsystems (a visual-spatial system and speech-based information system [Wilson & Keil, 2001]). The term "working memory" was first used by Miller, Galanter and Pribram (1960) applied to certain aspect of memory used to plan and perform behaviour (i.e., retaining partial results in solving a calculation [Cowan, 2008]). This term was then adopted by Baddeley and Hitch (1974) when demonstrated that only one STM mechanism or store could not account for different STM-related processes.

The two possible ways in which STM and LTM may differ still remains controversial though: a) in duration, whereby STM items decay with time; b) in capacity, whereby there is a limited amount of items that STM can hold (N. Cowan, 2008). Thus within the information-processing capabilities of the memory system, STM represents a bottleneck both in its time limit and in its information capacity (Snyder, 2000), and according to different measurement, its capacity seems to be an average of seven different elements, plus or minus two (Miller, 1956).

2.2.2 SHORT-TERM MEMORY VERSUS WORKING MEMORY

Although STM and WM are used interchangeably in the literature, STM can be thought of as a passive store component capable of temporarily holding certain amount of information, whereas WM extends beyond the mere storage to be conceived as an information manager, as a link between STM and LTM. However, discrepancies about the definition of WM exist in the current literature: it has been defined a) as a STM system applied to cognitive tasks (Engle, Tuholski, Laughlin, 1999; Conway, Kane, Bunting, Hambrick, Wilhelm, Engle, 2005); a) as a multi-component system that stores and manipulates information in STM (Baddeley & Hitch, 1974); (c) as the use of attention to manage STM (Engle, 2002); d) as immediate perceptions together with related activated long-term memories, semi-activated contextual information not in consciousness plus information that has just been in consciousness (Snyder, 2000). In this line (d) other authors' views (Engel & Singer, 2001; Ruchkin, Grafman, Cameron, & Berndt, 2003; Ward, 2003; Cowan, 1999; Ericsson & Kintsch, 1995) regard WM not as a separate system, but as an activated subset of LTM.

From different theoretical views, WM seems to be a type of memory in the short term domain that performs information-processing functions, is more attention demanding, and correlates with cognitive needs: WM includes STM plus extra processing mechanisms that help to make use of STM. Miyake and Shah (1999) propose a unifying definition of WM (p. 450):

"Working memory is those mechanisms or processes that are involved in the control, regulation, and active maintenance of task-relevant information in the service of complex cognition, including novel as well as familiar, skilled tasks. It consists of a set of processes and mechanisms and is not a fixed 'place' or 'box' in the cognitive architecture. It is not a completely unitary system in the sense that it involves multiple representational codes and/or different subsystems. Its capacity limits reflect multiple factors and may even be an emergent property of the multiple processes and mechanisms involved. Working memory is closely linked to LTM, and its contents consist primarily of currently activated LTM representations, but can also extend to LTM representations that are closely linked to activated retrieval cues and, hence, can be quickly reactivated."

Working memory and music

According to Snyder (2000), limitations deriving from our memory capacity influence how humans perceive temporal patterns of events and the boundaries between them, guiding our decisions about how events relate to one another in order to comprehend them as a whole and predict future events. Music that is intended as a communicative tool should respect the structure of memory "even if we want to work against that structure" (p. 3).

WM mechanisms are needed to form a coherent representation of the auditory flow by allowing to retain information over time. However, as an unfolding sequence of temporal events, all music properties are at mercy of the human WM thresholds, limiting our capacity for retention. Segmentation of the undifferentiated auditory stream in smaller units is crucial for this coherence representation. Lerdahl and Jackendoff (1983) emphasized the importance of segmenting the continuous flow of musical events ("when a listener has construed a grouping structure, he has gone a long way towards 'making sense' of it" [p.13]). Identifying musical events is an end product of the ongoing perceiving process which is governed by certain rules to segregate elements (Handel, in Cambouropoulos, 2009, p. 12).

2.3 PSYCHOLOGICAL AND FUNCTIONAL NEUROANATOMICAL THEORIES OF WORKING MEMORY

According to Milner et al. (1998), there are two key components in the study of memory: a) where it is stored and what brain systems are involved (the *systems* problem of memory); and b) how memory is stored at each site (the *molecular* problem of memory). The scope of the present study lies within the systems problem of memory. This represents a difficulty in that strictly psychological theories of WM cannot be tested, unless they associate neuroanatomical loci for their functional modules. In the following we will examine first some psychological theories about WM, to proceed with the neuroanatomical findings and the assumptions derived thereof.

2.3.1 PSYCHOLOGICAL THEORIES

There have been several different models of WM, but Baddeley and Hitch's (1974) tripartite WM model comprising storage and processing components has been extremely influential, allowing to explore cognition and to generate new hypotheses in various fields. Their model has been consolidated and minor changes have been introduced. Baddeley and Hitch's multicomponent model (1974) resulted from experimental work based on dual task studies, founded on the idea that tasks requiring the same processing mechanisms will show interference effects when performed simultaneously. Their experiments showed auditory and visual dissociation, whereby verbal tasks, and not visual, interfered with other verbal tasks in STM, and so WM model was proposed consisting of separate memory systems: (1) visuo-spatial sketchpad (for visual information) and (2) phonological loop (for auditory-linguistic information [Snyder, 2009]). These subsystems have passive storage and active rehearsal mechanisms and are subordinate to a third component: (3) the central executive (thinking and planning), which manipulates the WM information. Without rehearsal, information rapidly decays in a matter of seconds. In this rehearsal process, selective attention is involved. Auditory WM therefore involves attending, listening, processing, storing and recalling (Andrade, 2002). Most recently a (4) episodic buffer has been included in the model (Baddeley, 2000) as a third subsystem necessary to explain cross-domain associations, which is also assumed to act as a link or buffering mechanism between WM and LTM. However, Baddeley's model does not emphasize the contributions of LTM, which would reduce the WM load by organizing and grouping information in WM into a smaller number of units (Miller, 1956; Ericsson & Kintsch, 1995). In 2007, Baddeley introduced a hedonic detector to incorporate emotional processing in his WM model (p. 294) to account for how emotion mediates memory storage.

A number of studies have supported Baddeley's WM model, but some research has also challenged the empirical basis of the model, resulting in some conflicts and inconsistencies. For instance, the hypothesis of modality-specific storage systems for spatial location and verbal recall in WM was contradicted when these systems were equally impaired by articulatory suppression[4] (Jones, Farrand, Stuart, & Morris, 1995). This apparent division between visual and verbal STM would be then a misapprehension (Jones et al., 1995) of the fact that what is being disrupted during the articulatory suppression is serial recall (coding of order information) in verbal memory rather than item information. His model has also been criticized for not specifying whether musical information is subserved by the phonological loop, and for ignoring other sensory modalities (i.e., olfactory, tactile, gustatory) besides visual and auditory buffers.

Berz (1995) studied music-related STM in connection with LTM, plus the influence of background music and speech on the performance of several tasks. He found differences in the nature of both the processing and storage of musical stimuli. According to Berz, most compelling evidence for a separate musical component in Baddeley's multicomponent model of WM is the phenomenon of the

---

4 Method that consists in 'occupying' or 'using' the phonological loop by i.e., repeating a word. By engaging the phonological loop in this way, it can be determined whether other tasks requiring the same loop are inhibited.

unattended music effect, this is, unattended instrumental music would cause same disruptions on verbal performance as would unattended speech if there was one global store for both (Berz, 1995). His model is an extension of Baddeley and Hitch's model, in that he included an additional subsystem separate from the phonological loop accounting for both short-term storage and processing of musical information. He suggests a processing of information held in STM with structures kept in LTM, thus supporting the use of LTM in information and strategies for WM.

Cowan's Embedded-Processes Model (1999) is an attempt to synthesize a number of empirical findings. His WM model relies on five principles: a) WM information derives from LTM, the currently activated subset of LTM, and the activated subset of memory in the focus of attention and awareness; b) there exist different processing limits to different faculties; c) the focus of attention is voluntarily (central executive system) and involuntarily (attentional orienting system) controlled; d) there is habituation, that is, unchanged features of a stimulus over a long time do not elicit awareness since they are of no key importance to the individual; e) awareness increases processing (number of features encoded). Cowan regards WM not a separate system from LTM, but as subset of LTM representations, and as such WM would arise by focusing attention (with a capacity=4 items) onto an unlimited set of activated LTM representations (Cowan, 1995, 2005). He proposes that WM is not limited to one mechanism, and any necessary group of mechanisms at the individual's disposal is likely to be used to retrieve the needed information. According to Cowan, using more than one mechanism is usually less taxing than relying on just one (Cowan, 1999). His idea of WM assumes activation mechanisms, attentional and executive mechanisms, and LTM retrieval mechanisms cooperating to build an effective WM system.

2.3.2 NEUROANATOMICAL THEORIES

2.3.2.1 THE PREFRONTAL CORTEX

Although the mapping of the functional architecture of the WM in the brain is a very intricate current research subject, there is consensus in the research community about the critical involvement of areas in the prefrontal cortex (PFC) in WM functions, as indicated by Fuster's (1987) main results from single unit recordings. The PFC, and particularly the dorsolateral prefrontal cortex (dlPFC), has been consistently found active in tasks requiring executive functions (Kane & Engle, 2002). Two types of WM executive processes have been traditionally distinguished that are predicted to be assigned to different brain structures: a) executive control, i.e., regulation of encoding, strategy selection, manipulation and retrieval of information; b) active maintenance, i.e., keeping information available online. Executive control is thought to be embodied by the PFC; however, consensus regarding the loci of maintenance functions has not been reached so far. Some locate maintenance functions in posterior (parietal) regions of the brain, assigning the PFC a mediator role in controlling the information stored in those posterior areas (Knight & D'Esposito, 2003; Postle & Rypma, 2000; Curtis & D'Esposito, 2003; Postle, 2006). Others, in contrast, hypothesize dorsolateral regions to also

reflect maintenance of online sensory representations per se (Cohen, Perlstein, Braver, Nystrom, Noll, Jonides, & Smith, 1997; Goldman-Rakic, 1987).

2.3.2.2 MATERIAL-FUNCTIONAL DISTINCTION

The organization of the PFC has been studied with regards to the material type (verbal, spatial, and object information without spatial features) stored in WM, but also by executive function or process type (function), allowing then for a polymodal representation of information in overlapping areas. The material versus function distinction derived primarily from Baddeley's influential multicomponent model of WM. Neuroimaging studies addressed how the brain would embody these i.e., verbal and visual subcomponents and whether the hypothesized psychological distinctions (of either functional or material nature), validated in behavioural studies, were respected by brain patterns of activation. However, specialization by material type raises a problem in that distinct material categories may underlie different types of strategies (Courtney, Petit, Maisog, Ungerleider, & Haxby, 1998), which would be confounded, thus creating the illusion of different activation patterns based on material type (Wager & Smith, 2003).

According to a functional organization, superior and ventral frontal cortices are implicated in executive and rehearsal processes respectively. Evidently, this presupposes that the experimental tasks targeted at mapping different functions in the brain must be able to distinguish these non-modality specific processes (Coltheart, 2006a). Concerning the material distinction, spatial and non-spatial WM seems to be coded in dorsolateral and ventrolateral regions respectively (Wager & Smith, 2003). However, results from neuroimaging studies on segregation-by-domain WM are not consistent, or have failed to provide evidence about an organization of the cortex by material type in WM (D'Esposito & Postle, 2002; Stern, Owen, Tracey, Look, Rosen, & Petrides, 2000). An explanation could lie in the designs or analyses employed, susceptible to contamination by stimulus-related variance (Postle, 2006). In opposition, it appears that a topographic distinction (mid-ventrolateral mid-dorsolateral) might exist according to the nature of the WM process rather than the material type (Petrides, Alivisatos, & Evans, 1995). Owen (1997) suggests that conflicting evidence derived from electrophysiological recording studies with monkeys (Funahashi, Bruce, & Goldman-Rakic, 1989, 1990; Wilson, O'Scalaidhe, & Goldman-Rakic, 1993) might be due to either a) a minor task design change producing subtle alterations in processing requirements; or b) a further sensory modality-based specialization within the dorsolateral and ventrolateral systems that an fMRI approach cannot detect, whereas electrophysiological techniques can. In fact, based on his review of neuroimaging results, Owen (1997) indicated that stimulus-modality specialization might occur within the ventrolateral prefrontal cortex (vlPFC), thus segregation by material type and process would be represented across the lateral PFC at different neuroanatomical levels.

2.3.2.3 MOVING AWAY FROM THE SUB-COMPONENT PARADIGM

In the lack of neuroimaging converging evidence to date in support of a sub-component model of memory where different sensory-modalities or functions are represented topographically in the brain tissue, new theories that try to make sense of such disparity of results emphasize the importance of parietal and temporal brain areas specialized in perceptual-processing as necessary mediators for information storage (Jonides, Lacey, & Nee, 2005). Storage that does not recruit perceptual mechanisms would challenge this idea, as pointed out by Jonides (2005). More evidence to move away from this multi-buffer paradigm is, for instance, the fact that inhibition of the dlPFC using TMS or rTMS (Mottaghy, Pascual-Leone, Kemna, Töpper, Herzog, Muller-Gartner, & Krause, 2003; Oliveri, Caltagirone, Filippi, Traversa, Cicinelli, Pasqualetti, & Rossini, 2000; Koch, Oliveri, Torriero, Carlesimo, Turriziani, & Caltagirone, 2005), despite interfering with WM tasks, does not completely break down executive functions, thus supporting the existence of additional areas (either more general or more specific) performing this function (Mottaghy, 2006). For Fuster (2001) executive functions result from the connections between the PFC and LTM representations, and WM arises from the sustained activation of such network. It seems that conscious recall of sensory contents requires information binding to combine the spatially isolated brain areas within this emergent distributed network (Damasio, 1990; Fodor & Pylyshyn, 1988). In relation to this, Crick & Koch (1990) posited that binding and sensory awareness might be intimately related. It also has been proposed that information transfer to WM is mediated by a temporal pattern of synchronous discharges in selected neuronal populations induced by attentional mechanisms, i.e., binding of neural activity by synchronization activates WM (Crick & Koch, 1990). Similarly, Engel and Singer (2001) support the hypothesis that neural synchrony within millisecond precision may be implicated in arousal, perceptual integration, attentional selection and WM. Hence temporal dynamics within neuronal populations seems to be critical for the functioning of WM (Tallon-Baudry, Bertrand, Peronet, & Pernier, 1998; Sarnthein, Petsche, Rappelsberger, Shaw, & Stein, 1998; Engel & Singer, 2001).

In short, the two fundamental obstacles that the standard multi-component model faces are following: a) its aim to locate specialized buffers onto brain tissue, which has been challenged by the prolific incoming empirical evidence, leading to new theories, i.e., WM arises from the coordination of sensory-, representation-, and action-related systems (Postle, 2006); b) the many new dissociations uncovered by neuroimaging research within the postulated sensory buffers, hence demanding the accommodation of an increasing number of specialized subcomponents by the model.

2.3.2.4 SUMMARY

WM is defined as system that facilitates the temporal storage of sensorial information in a limited quantity, that also includes control, regulation and active maintenance of such information. It is one of the most intensively studied subjects in cognitive psychology. Current neuroimaging attempts to map functional multi-component models of WM described in the literature has been to date a

challenging undertaking, thus suggesting WM might not be a dedicated system in a restricted area in the brain anatomy, but a more complex network, a property emerging from the interactions of neural population in highly distributed brain areas (D'Esposito, 2007; Postle, 2006; Fuster, 2001). Studies on executive function reviewed by Collette, Hogge, Salmon and Van der Linden (2006) agree with the hypothesis that executive functions are not restricted to anterior cerebral areas, but disseminated throughout the brain. Also questions of transient or sustained activity are critical in studying executive processes, hence the importance of considering temporal dimension (Collete et al., 2006) when looking at patterns of activity.

Last, a mention about the assumptions of the task designs. In the study of WM, recognition tasks (identification of a previously presented stimulus, i.e., n-back or Sternberg tasks) and recall tasks (actively retrieval to conscious memory of a previously presented stimulus, i.e., largely span task) are widely used in the literature. Both paradigms (recognition and recall) are assumed to trigger the same processes, but this may not be the case.

2.3.3   NEUROIMAGING STUDIES OF MUSIC-RELATED WORKING MEMORY

The functional neuroarchitecture of the phonological loop has so far been explored primarily for language and fMRI studies indicate that Broca's area and premotor areas (premotor supplementary motor area [pre-SMA], supplementary motor area [SMA], ventrolateral premotor cortex [vlPMC] and dorsolateral premotor cortex [dlPMC]) activate significantly during the phonological rehearsal process (Awh, Jonides, Smith, Schumacher, Koeppe, & Katz, 1996; Fiez, Raife, Balota, Schwarz, Raichle, & Petersen, 1996; Gruber & von Cramon, 2003; Paulesu, Frith, & Frackowiak, 1993; Ravizza, Delgado, Chein, Becker, & Fiez, 2004). However, within the auditory WM, differences in the study of linguistic and musical memory have been found. Deutsch (1970) observed a dissociation between tones and phonemes in a WM task with tones, where tones impaired performance to a greater extent than phonemes. Results obtained by Salamé & Baddeley (1989) using a verbal WM task that was interfered by instrumental and vocal music indicated a similar dissociation, namely, that vocal music interfered more with the task than instrumental music. Findings were interpreted by the authors in terms of a phonological store that excluded non speech-like sounds. However, Chan et al. (1998) found that musicians perform better in verbal tasks than non-musicians, which emphasized common underlying processes for verbal and tonal WM.

Using fMRI, Hickok, Buchsbaum and Humphries (2003) investigated verbal (nonsense sentences) and tonal (novel piano melodies) WM in non-musicians and observed the activation of Broca's area and left premotor regions in both conditions. These areas were also found in another experiment with non-musicians conducted by Koelsch et al. (2009) employing sung syllables as stimuli, whereby either their pitch or syllable had to be remembered. Both conditions recruited strikingly similar areas: Broca's area, ventrolateral premotor cortex (vlPMC), dorsal premotor cortex, the planum temporale, inferior parietal lobe, the anterior insula, basal ganglia and thalamus, and the cerebellum, thus

dissuading from the idea of a 'tonal' loop. Schulze et al. (2011) introduced musicians along with non-musicians to explore whether WM for verbal and tonal information rely on the same or different functional neuroarchitecture in both groups. The overlapping topography between the verbal and non-verbal condition resulting from the performance of non-musicians validated the previous findings. Conversely, musicians' performance in the verbal task involved additional structures in both conditions. More specifically, there were a number of core WM structures —with significantly different weightings for both musicians and non-musicians— recruited in both tonal and verbal WM (Broca's area, left premotor cortex, pre-SMA/SMA, left insular cortex, left inferior parietal lobe), plus specific structures were recruited by musicians only during verbal (right insular cortex) or only during tonal WM (right globus pallidus, right caudate nucleus, left cuneus and left cerebellum; plus the right premotor cortex, the left putamen, and the right cerebellum were more strongly engaged than in the verbal condition).

With this different activation pattern for verbal vs. musical WM in musicians, Schulze et al. (2011) provided the first neuroimaging evidence supporting the hypothesis that auditory WM might not be a unitary system, and a distinction exists in musicians. It is known that musical expertise encompasses both anatomical and functional changes in the brain (Münte et al., 2002), and contributes to improved WM performance (Gaab, Gaser, & Schlaug, 2006). Recent neuroimaging evidence provided by Schulze, Mueller, and Koelsch (2011) has supported this view. In their study, musicians exhibited better WM performance for structured (tonal) over non-structured (atonal) sequences as well as differences in neurofunctional footprint, whereas no significant differences between the two conditions were observed in non-musicians, denoting musical proficiency as a facilitating factor for auditory WM.

In summary, neuroimaging methods have been crucial in detecting brain regions actively engaged in memory for music, providing evidence for the existence of specific subcomponents for musical and verbal WM in musicians. Schulze and Koelsch (2012) suggest that although there seems to be a considerable overlap of the underlying mechanisms and neural correlates of verbal and tonal WM, preliminary findings point to the existence of both a tonal and a phonological loop in musicians, proposing that functional plasticity has been induced by musical training. Regarding the core common areas, engagement of similar areas for language and music is reasonable to be expected since, after all, music processing shares some circuitry with spoken language processing, especially when subjects are experts in both music and language systems (Koelsch, Gunter, von Cramon, Zysset, & Lohmann, 2002; Patel, Gibson, Ratner, Besson, & Holcomb, 1998; Patel, 2003; Steinbeis & Koelsch, 2008).

3 METHODOLOGY

3.1 OVERVIEW

In this study the activation of music-related neural networks involved in WM was investigated by tracking the repetition of the two salient motifs in the piece. In music, WM gets activated even if subjects have already a LTM of the piece overall (Jonides, Lewis, Nee, Lustig, Berman, & Moore, 2008). Hence, this did not represent a problem for the study. We predicted the temporal course of HDR to the repetition of the musical motifs. A perceptual test was conducted to fragment the music into small segments, a necessary step for subsequently extracting the relevant motifs thereof and build the WM model about the neural networks underpinning WM for music. Next, the model was correlated against the fMRI brain signal at each voxel in the brain to retrieve the relevant brain areas involved in WM for music. To ensure that the observed activations were due to memory formation and retrieval rather than sensory processing, time series of acoustic features of the music were also added to the model as nuisance regressors (regressors of no interest) to remove the effect of the acoustic content in the brain and isolate the effects of WM processes. In the next sections we first proceed to describe the perceptual experiment aimed at retrieving the motifs in the piece, followed by an in-depth explanation of the materials and methods adopted in the fMRI analysis.

3.2 PERCEPTUAL EXPERIMENT

Segmentation, or phrasing, of the auditory stream is a fundamental process in music perception crucial for making sense of the musical flow due to our limited memory capacity. Listeners automatically detect structural boundaries in the music that correspond to different sized structural units, i.e., movements, sections, themes and motifs. Lerdahl and Jackendoff (1983) regarded these structural divisions groups as "the most basic component of musical understanding".

Thus, to build the WM regressor, the music stimulus needed to be fragmented into small segments, from which relevant motivic material (repetitions and variations) —expected to trigger WM— were extracted to build the WM model. With this aim a perceptual test was carried out where participants' task was to segment the piece in real time as they listened to it.

3.2.1 PARTICIPANTS

A total of 26 participants (female=11; age range=18-56; mean age=30±8 SD; musical training: starting age=11±5 SD years; mean total training=11±8 SD years; mean listening time=7±3 SD hours/week; mean practicing time: 30±25 SD min/day) took part in the experiment. We decided to recruit participants that were at least amateur musicians, who could understand how to perform this music-specific task.

3.2.2 PROCEDURE

The segmentation protocol was conducted on a computer running a Max/MSP[5] environment-based platform. Participants were instructed to (1) fill in a personal background information form; (2) listen to the entire piece of music from beginning to end once without performing the segmentation task; (3) perform the segmentation task (to avoid fatigue effects, listeners did not segment the music from the beginning to the end; instead ten[6] long sections of the music were presented to the participants twice in randomized order); (4) answer general questions about their performance (see Figure 3 for screenshots of the experiment platform). During the perceptual test participants segmented the music while listening to it in real-time. Participants were instructed to click with a computer mouse on a grey button (labelled "CLICK TO PLAY THE MUSIC EXCERPTS") to trigger the playback of one random excerpt at a time and immediately proceed to segment it by clicking on the red button beside it (labelled "CLICK TO SEGMENT THE MUSIC EXCERPTS") every time they heard a boundary in the music, i.e., the end or beginning of a segment. For each segment, clicks were logged as time coordinates (in milliseconds) at which they occurred for further analysis. The segmenting task was defined to the participants in terms of cutting the excerpts into small phrases or motifs where appropriate boundaries were found while the music was playing. The criteria to define these motifs or phrases were up to the participants.
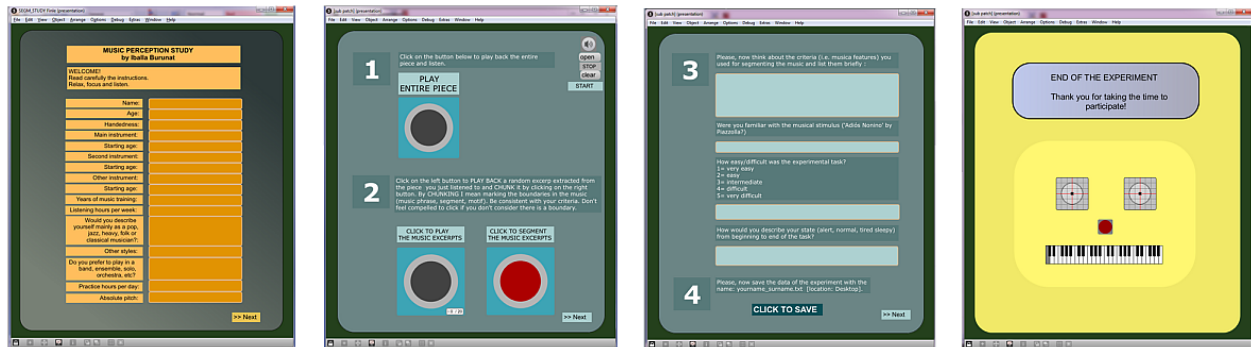
Figure 3. Screens 1-4 of the Max/MSP-based platform. From left to right, the personal background information form; the main experiment screen, with tasks (1) [listening to the entire piece] and (2) [perform the segmentation on the excerpts]; general questions about performance; end-of-the-experiment screen.

3.2.3 RESULTS

There was great variation among participants' number of clicks (mean=60± 38 SD; range=15-137). A greater number of clicks revealed a more fine-grained segmentation within consistent hierarchical boundaries. Participants' clicks (segmentation points) were merged into a one-dimensional array of values (time coordinates in milliseconds). Kernel Density Estimation (KDE) was used to estimate the unknown underlying distribution of the data. KDE is a statistical non-parametric method of estimating the probability density function (PDF) of a random variable (Silverman, 1986) and can

---

5 MAX/MSP 5.1 (Cycling '74, San Francisco, CA 94103, USA)
6 These excerpts correspond to higher level section boundaries of the stimulus starting at (1) 0:00, (2) 1:08, (3) 1:28, (4) 2:23, (5) 2:56, (6) 3:39, (7) 4:36, (8) 5:08, (9) 6:17, (10) 7:07 with an average length of 50.6 seconds each.

be thought of as a smoothed form of a histogram. A non-parametric approach suits the collected data because it is assumed to have a nonstandard distribution.
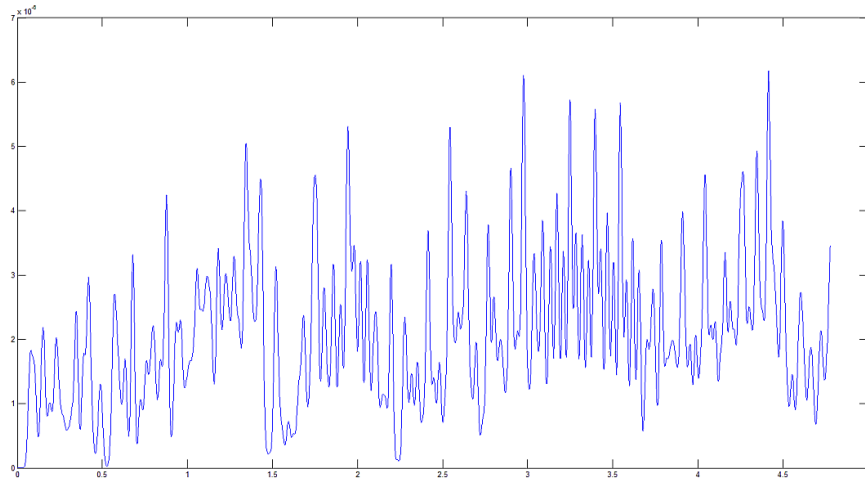
Figure 4. A total of 3289 clicks were recorded in the perceptual test, spanning along a time axis of 478000 equally spaced points (milliseconds of the stimulus excluding the end applause). Important for the KDE is not the kernel function itself (Gaussian, Epanechnikov or quadratic), but the bandwidth selection (Silverman, 1986). In this case, the bandwidth of the kernel-smoothing window was set to one second after several trials.

A KDE was computed for the vector of clicks by placing a kernel function over each observation in the sample (see Figure 4). The estimate was based on a normal kernel function, using a window parameter (width). The kernel bandwidth controls the degree of smoothness and its choice is crucial in density estimation: the wider the bandwidth, the smoother the estimate. If too small, the density estimate becomes spiky (if arbitrarily small, as many spikes as single data points in the set) and bias is reduced with respect to the true density but at expense of a larger variance in the estimates; while if too large, it will oversmooth the estimate which will mask the data structure, reducing the variance at the expense of a larger bias (see APPENDIX A for a detailed explanation). Thus, we want a compromise that minimizes the error between true and estimated density. Considering this, several KDE curves were plotted and a bandwidth of one second seemed a good compromise across participants' clicks. The bandwidth of the kernel-smoothing window was set to 1 second after several trials (see APPENDIX A for Matlab code).

In our data, the density was evaluated at 478000 equally spaced points (corresponding to milliseconds of the entire stimulus excluding the final applause) covering the data range. Next, maxima in the KDE function were extracted (see APPENDIX A) by adjusting different density thresholds. The lower the threshold, the more fine-grained segmentation is obtained (i.e., more subtle boundaries in the music coming to the surface). We applied a threshold that produced sixty-four peaks (segmentation points), which seemed a good compromise for a ~8 minute piece if we are looking for STM chunks.

Using MIR Toolbox's (Lartillot & Toiviainen, 2007) *mirsegment* function, the resulting segmentation points of the analysis can be visualized, as shown in Figure 5.
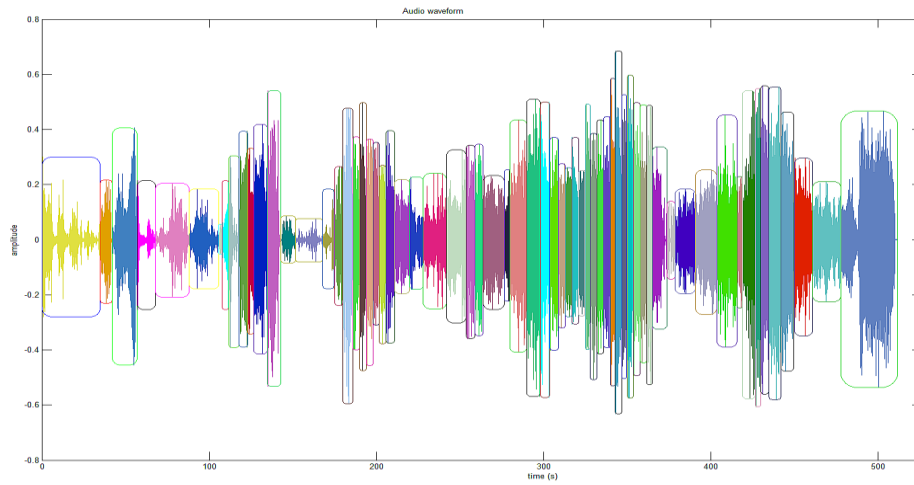
Figure 5. Segmentation of the music according to the participants' KDE, as plotted on the audio waveform using *mirsegment* [MIR Toolbox]). The last segment corresponds to part of the final long cadence plus the applause from the live performance.

A small lag in the participants' responses was found occasionally, since the segmenting task was done in real time. Adios Nonino is a very dynamic piece, full of rubato sections and long cadences, at which different participants, even those possessing high musical expertise, will tend to click along its whole length, either at the beginning, at the middle, or at the end. The piano introduction was hard to segment especially when listening to it for the first time due to its unpredictiveness, and most of the density points belonging to this section did not therefore reach the established threshold for peak-picking. Other non-cadencial, more rhythmical sections seemed easier to chunk, showing high density peaks, which denoted accurately occurring boundary locations.

3.2.4   PARTICIPANTS' SELF-REPORT

At the end of the experiment, participants were asked (a) to briefly list the criteria (e.g. musical features) they used for segmenting, (b) whether they were familiar with the piece, (c) to rate the difficulty of the task [1=very easy; 2=easy; 3=intermediate; 4=difficult; 5=very difficult]; and (d) to describe their alert status from beginning to end of the task (alert, normal, tired, sleepy). Participants rated the difficulty of the task on average as intermediate and difficult. No participants considered the task very easy; four rated it as easy, and two considered the task very difficult. With regard to alertness, most of the participants reported to be in normal or alert state throughout the experiment, although six participants declared to become less alert towards the end. Only four out of 26 participants acknowledged being familiar with the music (hence replicating the frequency observed in the study by Alluri, Toiviainen, Jääskeläinen, Glerean, Sams, and Brattico, 2012). When examining the reported criteria used by the participants to segment the music, assorted responses are

found. A summary of the different segmentation formulas, as reported by participants, follows: rhythm, harmonic rhythm, pauses, key change, melody change, timbral change, repeating patterns, drastic changes in dynamics, melodic contour, expectations in mood change, melody and bass line, different features in different parts of the music, chord progression (functional structure), texture changes, piano bass when acting as accompaniment, and cadences.

3.2.5 MOTIF IDENTIFICATION: MOTIF A AND B

From the sixty-four segments we proceeded to identify the main motifs —and their repetitions and variations— which are the building blocks of Piazzolla's tango, thereby discarding the remaining musical content, which belongs to a more introductory, transitional or developmental nature. In doing so, two salient motifs and their occurrences were identified. Following this observation, we decided to join two segments as they accounted for a single complete instance of one motif. The density value of the segmentation peak between these joined segments had yielded a very low (the 7th lowest value out of 64) at the chosen threshold, which also supported the decision of fusing them together.
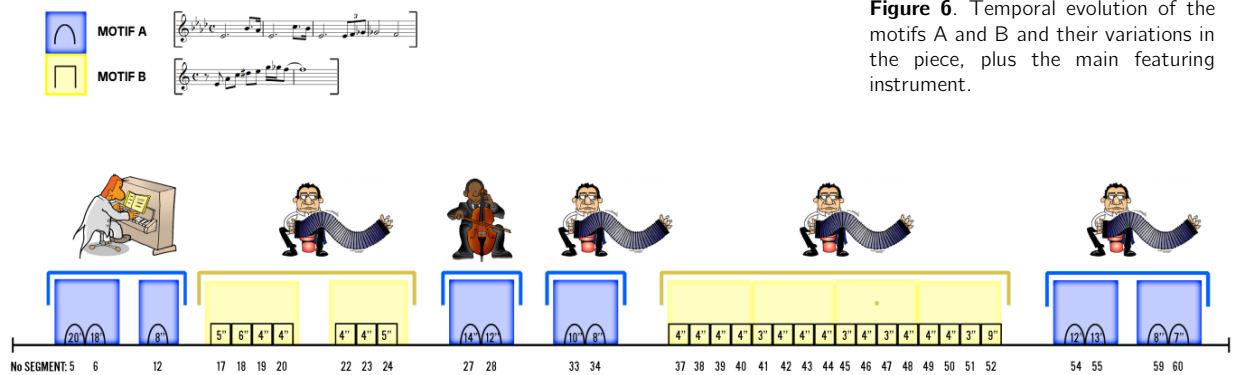
Figure 6. Temporal evolution of the motifs A and B and their variations in the piece, plus the main featuring instrument.

Thus, from the 63 segments, 34 correspond to the two salient motifs of the piece that we have identified as motif A (11 appearances) and motif B (23 appearances; see schema of the sequence of motifs in Figure 6). Motif A is characterized as being slow, rubato, more expressive and cadential, whereas motif B is fast and very rhythmical with a clear beat. With regards to timbre, different occurrences of motif A's melody are featured by the piano, cello and bandoneon, whereas in case of motif B, melody is always featured by the bandoneon. Motif A has hefty tempo changes, so it is expanding and retracting constantly. If we take the extreme case, the first and third instances of motif A have a duration of 20 and 2 seconds, respectively, whereas changes in tempo for motif B are more subtle. Motifs also vary in melodic contour and display ornaments (especially motif A). In regards to tonality, motifs are continuously transposed, i.e., motif A is transposed up by a 2nd minor, 2nd major, 3rd minor, 3rd major, 4th augmented and 7th minor; and down by a 2nd minor.

Motif B is transposed up by a 3rd major, 4th and 5th; and down by a 2nd minor, 3rd minor, 3rd major and 4th.

3.3 FMRI EXPERIMENT

3.3.1 FMRI DATA ACQUISITION

3.3.1.1 PARTICIPANTS

Eleven healthy participants with no history of neurological or psychological disorders, with formal musical training (styles: classical=5, folk=2, pop/rock=4; instruments: string=4, percussion=3, wind=2, keyboard=2; mean starting age=9.1±3.4 SD years; mean total training=16.1±6 SD years; mean practice time=2.5±1.2 SD hours/day) took part in the fMRI experiment (females=5; age range=19-31; mean=23.2±3.7 SD; handedness: 9 right-handed , 2 n/a).

3.3.1.2 STIMULUS MATERIAL

The music piece used for the listening experiment was *Adios Nonino*, an ˜8-minute tango piece by Argentinian composer Astor Piazzolla (1921-1992). The recording is an emblematic performance during Astor Piazzolla's Sexteto European tour (1989-1990) recorded live at MAD (Moulin a Danses), Lausanne, Switzerland on November 4th, 1989[7]. The piece, written in October 1959 while in New York in memory of his father, Vicente "Nonino" Piazzolla, a few days after his father's death, has proven to be one of Piazzolla's most well-known compositions. However, only two of the subjects reported to be familiar with the piece.

Adios Nonino is characterized by two well-defined, salient motifs. Musical recursion is a specific feature of Western tonal classical tradition, in which naturally occurring repetition and variation patterns in the music constitute a distinct attribute, hence allowing the study of motif-tracking in the context of real music. Moreover, *Adios Nonino* is a very challenging music stimulus rich in timbre and modulations, and particularly in this very vibrant and dynamic performance, exhibiting strong shifts in tempo, dynamics, and rhythm, which is in effect representative of a complex auditory scene.

3.3.1.3 TASK SPECIFICATION

Participants were instructed to listen to the stimulus (average sound level=80 dB) with gradient noise attenuating (˜30dB) headphones (plus extra attenuation via cotton inserted in the headset),

---

7 Audio CD (May 23, 2006). Label: Milan Records; ASIN: B000F6ZID8. Performers: Astor Piazzolla (bandoneon); Horacio Malvicino (guitar); Carlos Nozzi (violoncello); Angel Ridolfi (double bass, upright bass); Daniel Binelli (bandoneon); Gerardo Gandini (piano).

and remain still and relaxed as they underwent fMRI scanning while keeping their eyes open. The study protocol proceeded under local ethical committee guidelines.

3.3.1.4 FMRI IMAGES

Functional images were acquired at the 3-T scanner (3.0 T Signa VH/I General Electric) at the Advanced Magnetic Imaging (AMI) Centre (Aalto University, Espoo, Finland) with a temporal resolution of 2 seconds. Thirty-three oblique slices (thickness=4 mm, interslice skip=0 mm) providing whole-brain coverage were imaged (echo time=32 ms; flip angle=75°) per participant, with a field of view of 200x200 mm and a matrix size of 64x64, using a single-shot gradient echo planar imaging (EPI) pulse sequence. To measure BOLD changes, each EPI acquisition was run for 256 volumes (the first 4 volumes were later discarded to allow for the magnetization to stabilize) per participant. T1-weighted structural images (field of view=260×260 mm; matrix=256×256; thickness=1 mm; interslice skip=0 mm) were also collected for individual coregistration.

3.3.2 FMRI DATA ANALYSIS

3.3.2.1 FMRI DATA PREPROCESSING

Functional MRI scans were first partly preprocessed in SPM5 (Statistical Parametric Mapping 5)[8]. For each participant low-resolution images were realigned (translation and rotation corrections did not exceed 2 mm and 2^0, respectively) and registered to the corresponding high-resolution T1-weighted structural images. These were in turn registered to the MNI (Montreal Neurological Institute; Evans, Kamber, Collins, & MacDonald, 1994) standard brain template using a 12 parameter affine transformation. Images were next segmented into gray matter, and normalized. Functional images were then blurred to best accommodate anatomical and functional variations across participants as well as to enhance the signal-to-noise by means of spatial smoothing using a 6 mm full-width-at-half-maximum Gaussian filter.

The rest of the fMRI data processing was carried out in Matlab. Following the spatial smoothing operation, the fMRI signal was detrended by means of spline interpolation (anchor points=6) in order to remove low frequency noise, thereby filtering out the slow scanner drift and other potential drift-inducing sources in the fMRI time series like thermal noise from the scanner's field strength, physiological effects like heart-rate, respiration, and basal metabolism fluctuation. Spline interpolation finds the underlying function (trend) between the equally spaced anchor points or nodes in the signal (anchor points are defined to be the middle point of the segment in the time axis and the mean of the time series values in the y axis), thus capturing the non-linearities in the scanner drift and removing the fitted trend from the data. As a result the residual has these low frequencies removed. As well as subtracting the low frequency fluctuations, removal of

---

8 Wellcome Trust Centre for Neuroimaging, London, UK (http://www.fil.ion.ucl.ac.uk/spm).

high-frequency fluctuations in the data is a typical procedure that can be advantageous, aimed at maximizing the signal-to-noise ratio (Jezzard et al., 2001). To this end and following the detrending operation, temporal filtering (Gaussian smoothing with kernel width=5 s) was applied to the fMRI signal.

Movement-related variance components in fMRI time-series (resulting from residual motion artefacts from the registration operation) are another source and one of the most serious confounds of brain analysis (Frackowiak et al., 2004) that can reduce the sensitivity to true activations. To solve this, six variable parameters (vector of x, y, and z translations and rotations about the major axes at each time point describing each voxel displacement relative to the template image [Jezzard et al., 2001]) were treated as nuisance terms and regressed from each voxel's time series. We saved these resulting residuals for further statistical analysis.

A generic model for describing fMRI data introduced by Lindquist (2008) is shown in Eq. 1 for activation in a single voxel in one subject). The assumptions made are a) each volume contains N voxels, b) sampled at T time points, c) repeated for M subjects, and d) present components in the data are task-related BOLD response, nuisance covariates and noise.

$$y_{ij}(t) = \sum_{g=1}^{G} z_{ijg}(t)\gamma_{ijg} + \sum_{k=1}^{K} x_{ijk}(t)\beta_{ijk} + \varepsilon_{ij}(t) \qquad \text{Eq. 1}$$

for i=1,...,N, j=1,...,M, and t=1. Where $z_{ijg}(t)$ is the contribution of nuisance terms at time t with amplitude γ_{ijg}; $x_{ijk}(t)$ is the task-related HDR for the kth condition at time t with amplitude β_{ijk}; and $\varepsilon_{ij}(t)$ the noise process.

3.3.2.2 STATISTICAL MODELLING

In the following we will introduce the statistical analysis method used, followed by an explanation of the processing of the regressor of interest (WM model) used in the correlation analysis and the perceptual experiment used to fragment the music. Questions about orthogonality between the regressor of interest and nuisance regressors are also dealt with, after which we proceed with the intrasubject and subsequent second-level statistical analyses.

3.3.2.2.1 Correlational analysis

Correlational designs are used to uncover functional areas in the brain by quantifying the correspondence between predicted and measured BOLD signal to a stimulus. The underlying idea is that the magnitude of the hemodynamic response changes as a function of cognitive 'load', relying on the assumption of linearity of the BOLD responses initiated by the stimulus (Jezzard et al., 2001).

Thus the shape of the response to the stimulus of interest must be predicted first, and only then correlation coefficients between each voxel time course and the predicted waveform of the effect of interest can be calculated.

For two time courses X (actual) and Y (predicted), the correlation coefficient is computed as

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 \sum(Y - \bar{Y})^2}}$$

Eq. 2

which returns 0 for no correlation, 1 for perfect positive correlation and -1 for perfect negative correlation. This correlation operation is repeated for every voxel time course in the brain. The resulting r values are then thresholded for significance, which will produce a map of significant activity in the brain for the specific condition under investigation.

3.3.2.2.2 Design matrix specification

In fMRI studies the design matrix is the mathematical structure that contains the explanatory variables as they change over time. It is formulated as an n×p matrix, where n is the number of observations over time and p is the number of variables (Frackowiak et al., 2004). Thus columns represent the different factors, effects of interest but also effects of no interest (nuisance regressors), designed to remove the variance associated with unwanted known sources of variability; and the rows specify values for each factor at each observation in time. In our approach we have used a binary explanatory variable (WM) indicating whether the effect is 'on' (an activation) or 'off' (no activation), represented by ones or zeros respectively. Factors can be also continuous functions measuring different levels of the variable, as is the case of our nuisance regressors (movement parameters and acoustic features).

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

Figure 7. Example of a binary design matrix. Columns represent observations over time, rows represent the different variables.

When designing the model, the number of columns in crucial. Overfitting the model with too many regressors reduces the number of available degrees of freedom in the data, which leads to reduced sensitivity to detect any activation (Jezzard et al., 2001). Similarly, correlated factors can confound the validity of the model, since the explained variance is shared between them.

3.3.2.2.3 Building the regressor of interest

WM regressor

To build the hypothesis of WM, we chose from the thirty-four motivic segments resulting from the perceptual experiment (see Figure 8), those thought to trigger WM, namely all occurrences of motif A and B except the first instance thereof. These repetitions and variations constitute the 'on' condition, while the rest of the music is assumed not to elicit WM.
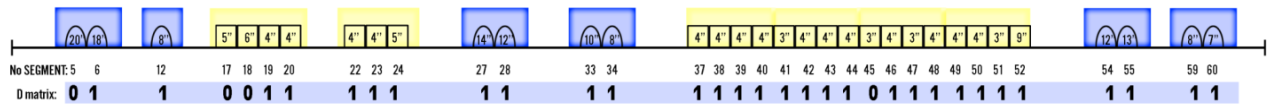
Figure 8. Segments corresponding to motivic material in the piece used to build the WM regressor. For the rest of the music not represented here zeros are assigned.

Segments 45 to 49 were found to significantly deviate from both main motifs' identities. However, given the identical repetitive nature of the sequence, it was included in the WM hypothesis assuming it would trigger WM. Similarly as proceeded previously for motifs A and B, the first motif of this sequence (segment #45) was discarded and the rest was included in the model as part of the 'on' condition. Hence the regressor of interest was defined as a time series vector (sampling rate=1 Hz) of binary values having ones for the 'on' conditions and zeros elsewhere.
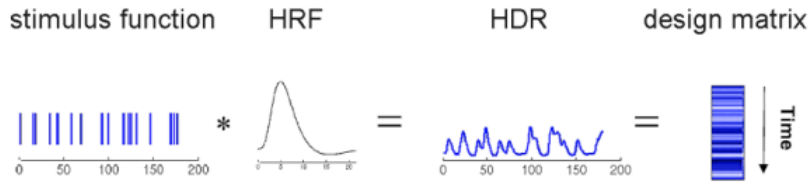
Figure 9. LTI systems typically model the expected HDR as the convolution of a finite impulse function (assumed neuronal activity caused by the task under investigation, usually depicted as a delta function) positioned at event onset with a canonical HRF (impulse response function): $f \otimes g(t) = \int_0^t f(\tau)g(t-\tau)d\tau$. Figure adapted from Lindquist, 2008.

This regressor underwent further processing to be comparable with participants' brain responses prior to correlational analysis. To this end, the regressor time series was convolved with a canonical double-gamma hemodynamic response function (HRF) in order to match the hemodynamic response delay typical of dependent (BOLD) contrast (see Figure 9). In this operation, the WM time series of predicted activations acts as the input of a linear time invariant (LTI) system and the HRF as the

impulse response function. LTI systems are typically used to model the relationship between the stimulus and the HDR (Lindquist, 2008).

Once convolved, the WM regressor —with a sampling rate of 1 Hz (one value per second)— was downsampled to .5 Hz to match the sampling rate of the fMRI scanner. Next, the first 13 samples (26 s) of the WM and AC nuisance regressors were removed. Those corresponded to the first 4 volumes excluded after the fMRI acquisition due to T1 stabilizing effects, and to the subsequent 9 samples (matching the length of the HRF) to avoid artefacts resulting from the convolution operation. Finally the last 12 samples (24 s) of brain volumes and regressors corresponding to the applause from the live performance were also excluded. The final time series had a length of 231 samples.

REP regressor

We proceeded by designing an additional regressor: one that is identical to the WM regressor but contains the excluded first instances of the motifs, in other words, all the motifs are represented in this new regressor (see Figure 10). This condition will be called REP in the following. We hypothesized that the correlates of this new variable do not reflect WM activation (since the first appearances of the motifs —those that would not trigger WM recall— are always 'on'), although we do not know the underpinning phenomenon it is related to. The motivation behind the REP regressor was to perform a subtraction (t-test) WM-REP on the final resulting statistical brain maps (from the correlation analysis with each regressor), which would filter out the insignificant overlap of activated brain areas between them. This might uncover a more nuanced map of significant activations that would expose active areas implicated in WM (those significantly higher and lower z--scores of the WM map over the REP map), thus fine-tuning the results.

1 = WM regressor (binary); 2 = WM regressor (convolved)
3 = REP regressor (binary); 4 = REP regressor (convolved)
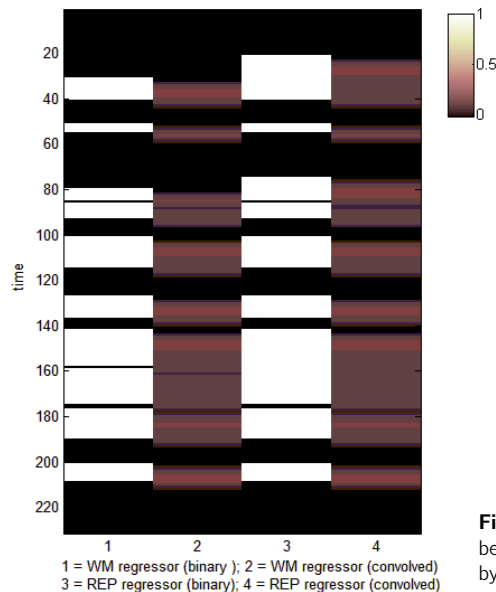
Figure 10. Regressors WM and REP before and after being convolved, side by side.

3.3.2.2.4 AC nuisance regressors

Time courses of perceptually validated acoustic features (in the following referred to as acoustic components, AC) provided by Alluri et el. (2012) were added to the model in a second statistical run as nuisance regressors (see Figure 11) to remove the variation evoked by the acoustic features in the brain signal and best isolate the WM effect, but also to observe the extent to which musical feature processing networks might be involved in or might overlap with music-relevant WM processes in the brain when comparing the results of the two analyses (post-processed acoustic components were provided by the author[9]). The acoustic components (n=6) correspond to the main timbral, tonal, and rhythmic features in the stimulus and are continuous-valued variables (see Alluri, 2012, for an in-depth acoustic description).
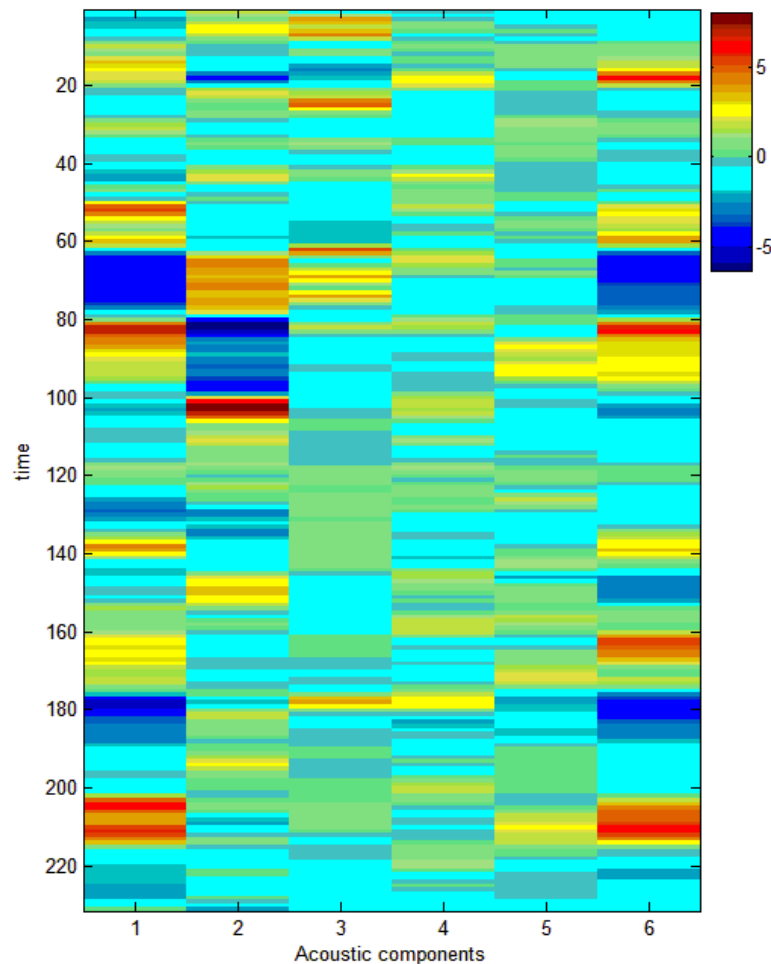
Figure 11. AC nuisance regressors convolved and resampled. They correspond from left to right to fullness, brightness, timbral complexity, key clarity, pulse clarity, and activity.

---

9 Post-processing included convolution with a double-gamma HRF, detrending, and downsampling, to match the fMRI data.

3.3.2.2.5 Orthogonality of regressors

As mentioned in 3.3.2.2.2 (design matrix specification), collinearity between regressors should be avoided for design efficiency, as it would reduce the validity of the analysis preventing the identification of independent effects (Huettel et al., 2009).

Orthogonality between acoustic components and the WM regressor was examined. To do so, Pearson's correlation coefficients (r) were calculated between the WM regressor and each of the acoustic components. These were mostly weak (r range=0.09-0.21), but their significance had to be corrected due to serial correlation, as the data points are not independent from each other. In order to estimate the significance of the r coefficients, the effective degrees of freedom were calculated between each AC and the WM regressor following Pyper and Peterman's (1998) procedure (for a detailed explanation see 3.3.2.2.6 [intra-subject analysis]). The degrees of freedom ranged from 46 to 80 across AC, and were used along with the r coefficients to compute the corrected p-value. Corrected p-values were large (range=0.10-0.46), indicating with confidence that the effect of interest and the individual AC are independent.
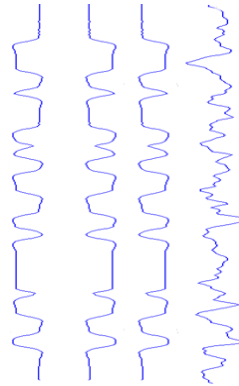
Figure 12. An overdetermined model with linearly dependent variables is rank deficient: the condition of independence among the regressors should be met to ensure that the resulting estimates are independent thus validating the use of the model.

However we needed to test also whether a linear combination of the AC, rather than individual components, could account for our WM variable. Thus we performed multiple regression whereby the WM variable was regressed against the whole set of AC, solving the general linear model

$$Y = X\beta + \alpha$$

Eq. 3

Where the response Y is the WM time series; the set of predictors X are the AC; β is the coefficient estimate (one in this case) of the multiple linear regression of the WM responses on the AC predictors; and α is the intercept. We obtained a moderate r^2 of 0.13, suggesting there is some variance in the WM regressor that can be explained by the AC. Thus some of the signal of interest may be removed, which cannot be avoided using a naturalistic setting, where independence between

the motivic structure and the temporal evolution of the AC cannot be controlled for. However, we can say with certainty that most of the variance in WM cannot be explained by the AC.

The REP regressor was also tested for orthogonality against the AC following the same procedure, by computing Pearson's r correlation between REP and individual AC, which resulted in two AC (*fullness* and *key clarity* [Alluri et al., 2012]) with moderate-low correlation coefficients of r=0.26 (corrected p=0.09) and r=-0.25 (corrected p=0.05) respectively[10]. Therefore the REP condition can be moderately explained by two acoustic components (*fullness* at r=0.26 and *key clarity* at r=-0.25).

3.3.2.2.6 Intra-subject analysis

The following procedure was run for the brain data time series with and without AC regressed out (i.e., with and without the variance accounted by the acoustic components). A non-parametric approach similar to that adopted by Alluri, 2011 was followed in order to analyse the fMRI data series.

We computed Pearson's correlation coefficients for all participants between each voxel time course and the WM regressor. The sampling distribution of Pearson's r coefficients is not normally distributed, becoming increasingly skewed as r deviates from zero. In order to reduce the skewness and make the sampling distribution approximately normal, we convert the r maps into z-scores images using the Fisher's Z-transform (Fisher, 1915), defined by

$$z = \frac{\sqrt{n-3}}{2} \cdot \ln\frac{1+r}{1-r} = \arctan(r)$$

Eq. 4

where z has a Gaussian distribution with zero mean and unit variance and has a standard error of

$$\sigma_z = \frac{1}{\sqrt{N-3}}$$

Eq. 5

where N=number of samples.

SERIAL CORRELATION

The distribution of correlation coefficients is assumed to be based on the number of n independent observations, for which the degrees of freedom are n-2. However in our fMRI time series, this assumption of independency between adjacent data points is not met because of the effect of the HRF and the smoothing operation in the temporal domain, and so the data is serially correlated and

---

10 We also calculated r^2 between REP and the full set of AC, which was slightly higher (r2=0.14) than the obtained r2 (=0.13) with the WM regressor.

the true number of independent samples is lower than the total number of scans. Certainly this overestimation of degrees of freedom (DoF) in the computation of the coefficients underestimates the error variance, which produces inflated test statistics (type I errors [Monti, 2011]). Simply stated, detecting significant correlations between predicted and observed response increases with smoothness, to the point that even in the absence of neuronal response, intrinsic autocorrelations lead to high test statistics (Friston, Jezzard, & Turner, 1994). Therefore these obtained Fisher z-scores had to be corrected due to this problem.

To correct the Fisher z-scores for significance, we first estimated the effective DoF by following a non-parametric procedure described in Pyper and Peterman (1998) based on a Monte-Carlo simulation. The effective degrees of freedom are given by Eq. 6:

$$\frac{1}{df} \approx \frac{1}{N} + \frac{2}{N}\sum_{j}\frac{N-j}{N}\,\rho_{xx}(j)\,\rho_{yy}(j) \qquad \text{Eq. 6}$$

where N is the number of observations, and $\rho_{xx}(j)$ and $\rho_{yy}(j)$ are the autocorrelations of the signals at lag j.

Accordingly, a Monte Carlo simulation (number of trials n=10000) was run, where a randomly chosen voxel time course was used along with the WM time series as inputs x and y in Eq. 6 to obtain the effective DoF. This operation was simulated for each participant for the total number of trials after which estimates were averaged. Thus resulting effective degrees of freedom accounting for the temporal dependence in the time series ranged from 45 to 50 across participants.

Once obtained the DoF across participants, we proceeded to normalize the z-scores as shown in Eq. 7 by dividing them by the standard error (see Eq. 5), where effective degrees of freedom enter in the calculation.

$$corrected\ z = \frac{z}{\frac{1}{\sqrt{df-3}}} \qquad \text{Eq. 7}$$

At this stage for each participant we had a corrected z-score map.

3.3.2.2.7 Group-level analysis

Once the corrected statistical maps at individual level were obtained (first-level analysis), second-level analysis was carried out on these results to produce group maps of significant voxels. The idea is to combine participants' results and produce an "average" activation map. To do so, and following the approach described in Lazar (2008), Fisher's method (1950) was used to obtain the pooled group-level map from the individual p-value images, whereby the group statistic is

$$T_F = -2 \sum_{i=1}^{k} \ln p_i$$

<div align="right">Eq. 8</div>

where p is the p-value associated with the t-statistic from subject i from k individuals. Thus participants' z-scores were transformed to the corresponding p-values, which were pooled across participants using Eq. 7. This resulted in a group t-statistic that under the H_0 follows a Chi-square distribution X^2_{2k}, with 2k degrees of freedom. Using the CDF (cumulative distribution function), we computed the upper tail probabilities at each of the Chi-square t-statistic values with v degrees of freedom defined by $1 - P(X^2 \mid v)$.

3.3.2.3 STATISTICAL INFERENCE

The resulting group map was thresholded at a significance level of p=.0005. However, type I errors due to multiple comparisons have to be accounted for when making statistical inferences in a dataset of these dimensions (over 200k data points). The multiple comparisons problem was tackled by means of a cluster-wise significance approach described in Ledberg, Åkerman and Roland (1998) based on a Monte Carlo procedure to obtain an approximation of the cluster size (CS) distribution at a particular significance level, from which the critical CS threshold can be selected (see APPENDIX B for a detailed explanation). Consequently, the obtained corrected critical CS was of 8 voxels for a cluster-defining threshold set at p=.0005 (Z=3.48).

MarsBaR v0.42[11] was used to extract the regions falling under each cluster (the ROIs are based on a manual parcellation of a single brain [colin27] in MNI space). Subsequently, we proceeded to find the number of voxels and the x y z positions for the maximum within each region. The resulting coordinates, in MNI space, were transformed to Talairach space using the first approach described in "The MNI brain and the Talairach atlas" (2012), whereby the SPM95 PET template is mapped onto the MNI-compatible SPM96 template using the SPM96's spatial normalization algorithm with an affine transform[12].

Finally, the labels for each brain region were retrieved using the Talairach JAVA applet (Lancaster, Rainey, Summerlin, Freitas, Fox, Evans, Toga, & Mazziotta, 1997; Lancaster, Woldorff, Parsons, Liotti, Freitas, & Rainey, 2000)

This whole statistical procedure was re-run using the REP regressor for further subtraction as explained in the next section.

---

11 MarsBaR v0.42 (http://marsbar.sourceforge.net)

12 An affine transformation is any transformation that preserves collinearity (i.e., all points lying on a line initially still lie on a line after transformation) and ratios of distances (e.g., the midpoint of a line segment remains the midpoint after transformation [Weisstein, n.d.])

4 RESULTS

Examples of the activation images produced with the WM regressor, in both participants' brain responses including and excluding AC, are plotted in Figure 13. Conservative settings for plotting purposes were used (p=.0001 and CS=20).
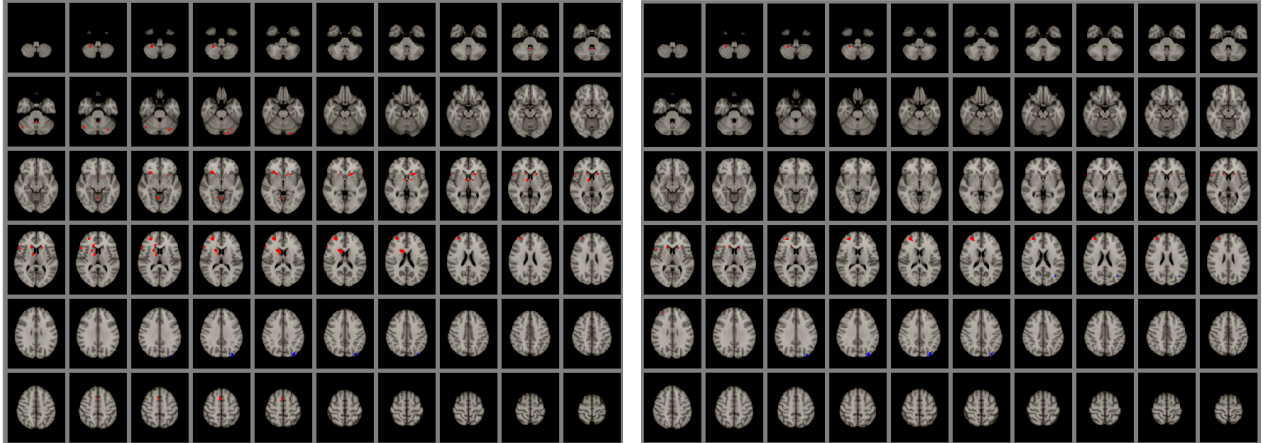
Figure 13. Transversal images showing activation and deactivation to the WM condition in both the brain responses including and excluding AC. A conservative threshold of p=.0001 (z=3.89) and CS ≥ 20 for plotting purposes was used. Voxels are color-coded red and blue for positive and negative correlation respectively.

The first thing we can observe distinctly between the two maps is that whereas the cortical areas reflect mostly similar overlapping activations, it is in the subcortical areas where the difference is evident: when the AC are removed from the brain responses, so is most of the subcortical activation.

At this point we might want to reveal the contributions of each specific state by means of subtracting the two maps. However, by doing straight subtraction on the thresholded binary maps, important information could be lost, i.e., two voxels falling within significant threshold yet with significantly different p-values would yield a subtraction value of zero. Therefore it is recommended to use the continuous map to get a more nuanced subtraction image, and so the method that ensures a higher power would be a Student's t-test on the participant-level z-maps. The t-test is carried out at each voxel across participants yielding a t-test statistic per voxel. The subtraction procedure is explained in detail in APPENDIX D.

Thus we can observe the significant differences in BOLD signal change above threshold in both maps, i.e., eliminate non-significant differences and leave the activation that is significantly higher (or lower for negative correlations) in one map over the other and vice versa.

4.1 Subtraction of WM maps resulting from the AC-inclusive and AC-exclusive brain responses

Using the subtraction procedure described in APPENDIX D we obtained the significantly higher and lower z-scores in the AC-exclusive and AC-inclusive residuals. A transversal view of the overlaid resulting maps is shown in Figure 14 and Figure 15.
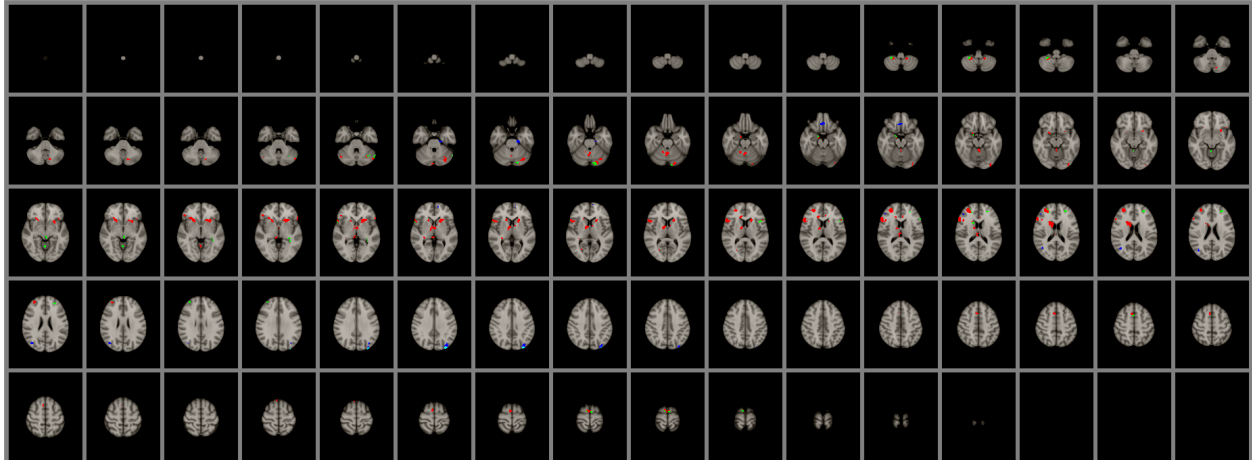
Figure 14. Images obtained after performing the subtraction of WM maps in AC-inclusive vs. AC-exclusive responses. Red indicates significantly higher z-scores (positive correlations) in the AC-inclusive responses, and green indicates significantly higher z-scores in the AC-exclusive responses. Blue and cyan denote significantly lower z-scores (negative correlation) in the AC-inclusive and AC-exclusive responses, respectively. Images were thresholded at p=.0005 (z=3.48), with a critical CS ≥ 8.
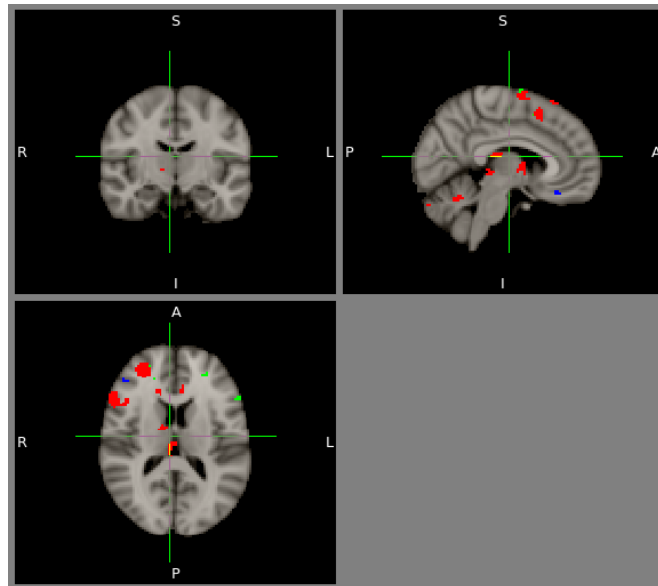
Figure 15. Orthographic projection of the resulting subtraction of WM maps in AC-inclusive vs. AC-exclusive brain responses. Here we can observe the extensive subcortical activation to the WM regressor in the responses including the AC (in red) as opposed to the more reduced significant activation in the responses excluding the acoustic content (in green).

This subtraction map shows the effect of the WM stimulus in the brain with and without AC. As shown in Figure 14, a wide network of cortical and subcortical areas light up responding to the WM condition in the AC-inclusive brain responses, while much of the subcortical activation pattern is removed in the AC-exclusive responses while the cortical activity remains. This reveals that part of the activations present during the WM condition a) are being pruned with the removal of AC, or b) are significantly weaker when compared to the activation map in the AC-inclusive brain responses and so there are not shown in this subtraction. Hence the assumption is that the acoustic correlates might be also subserving WM processing. Figure 16 shows the cortical differences in activation from the two subtractions.
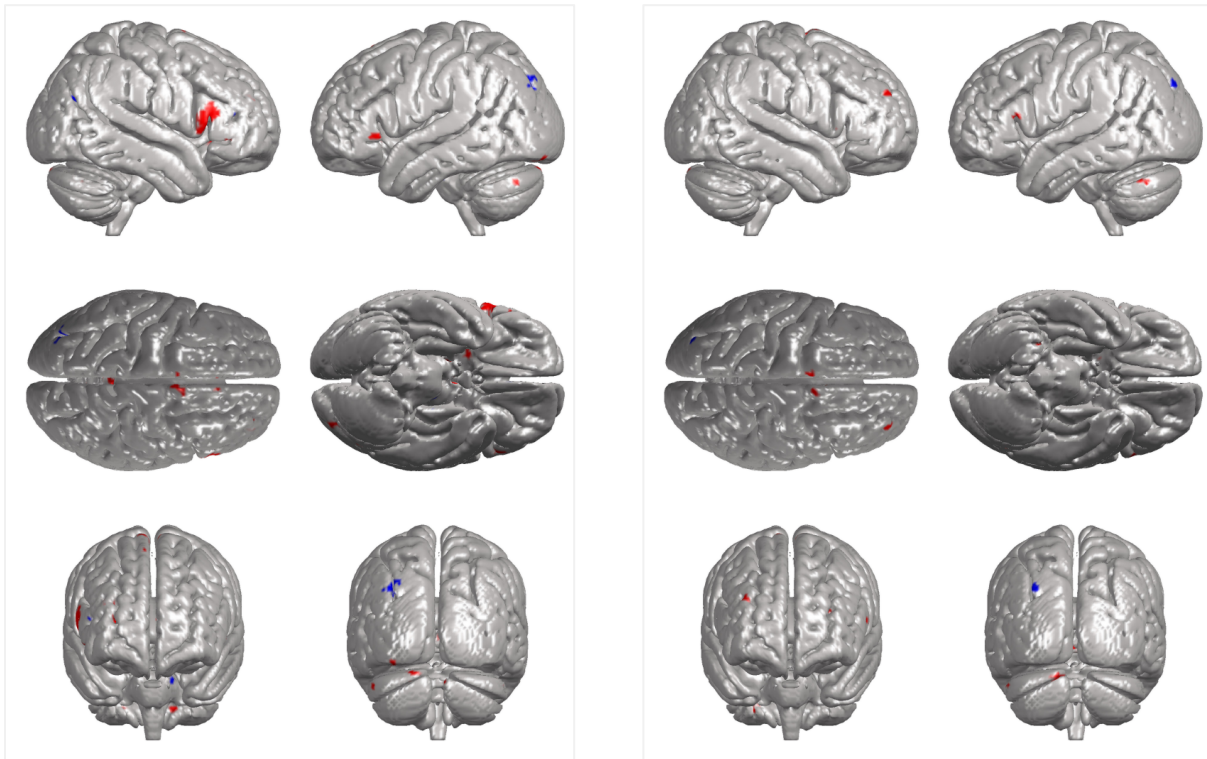
Figure 16. Figures showing the cortical differences in activation between WM in the AC-inclusive responses (left) and in the AC-exclusive responses (right).

4.2 Subtraction of WM vs. REP maps resulting from both AC-inclusive and AC-exclusive brain responses

We performed the same subtraction procedure for the WM and REP maps with the aim of extracting only the significant areas in WM with respect to REP. The results from this subtraction (in both residuals with and without AC regressed out) will be the focus of discussion in the present study, from which we will draw conclusions about WM for music (see Figure 17 and Figure 18 for transversal and orthographic views of the resulting maps, and see Table 1 & Table 2 for the listed clusters and brain regions obtained in these subtractions).
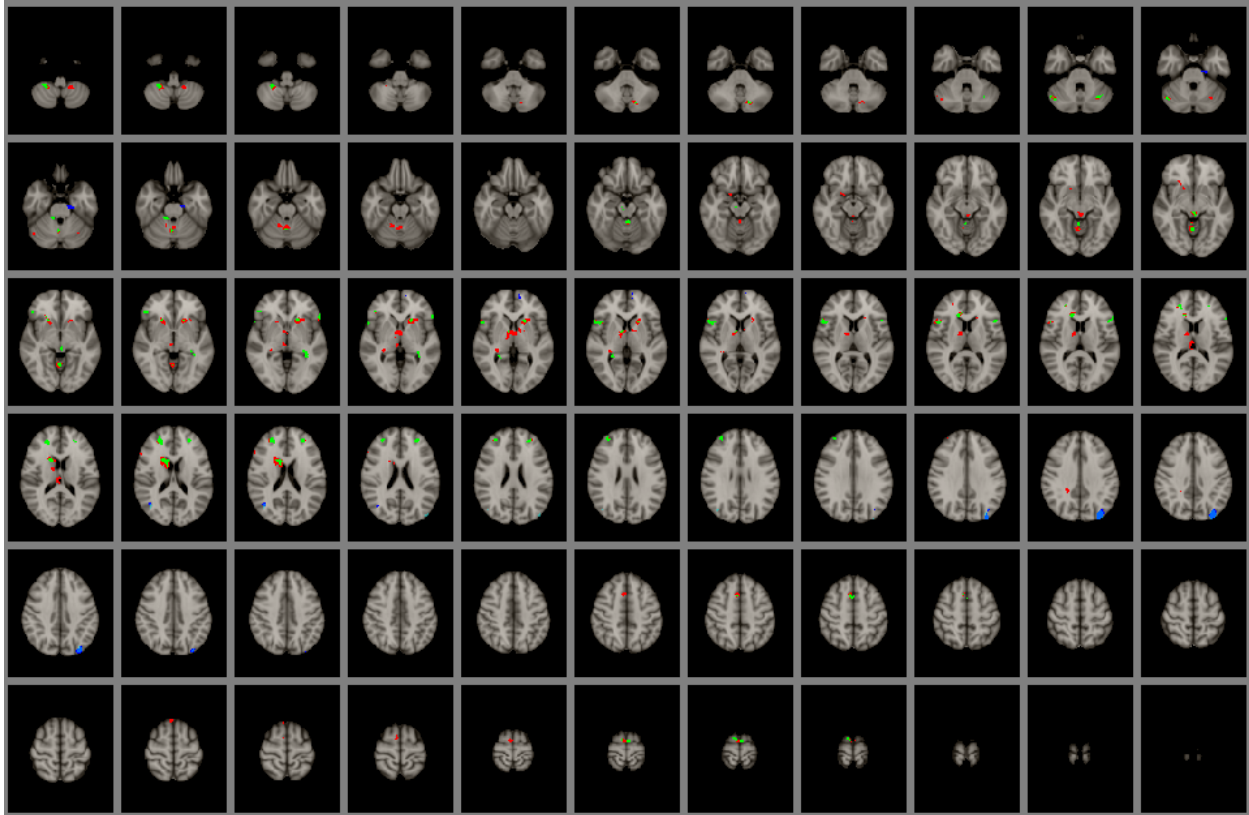
Figure 17. Transversal views of the subtraction WM vs. REP maps in both AC-inclusive and AC-exclusive responses, expected to reflect activations exclusively due to WM. Red indicates significantly higher z-scores (positive correlations) in the AC-inclusive responses, and green indicates significantly higher z-scores in the AC-exclusive responses. Blue and cyan denote significantly lower z-scores (negative correlation) in the AC-inclusive and AC-exclusive responses, respectively. Images were thresholded at p=.0005 (z=3.48), with a critical CS ≥ 8.
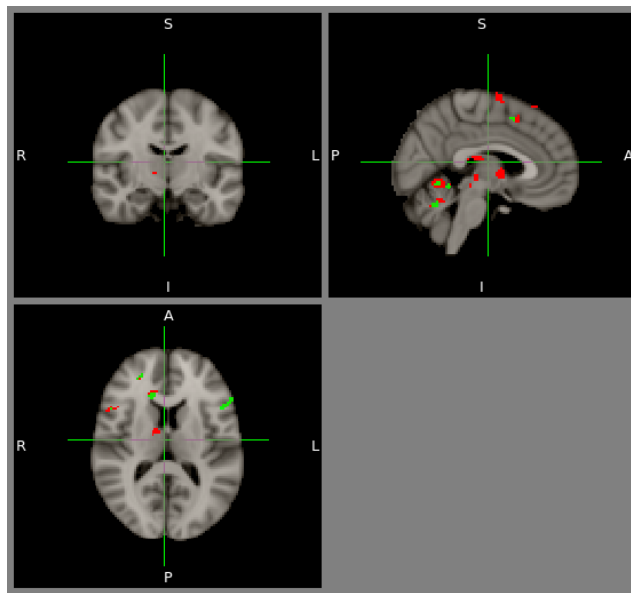
Figure 18. Orthographic projection of the resulting subtraction WM vs. REP maps in both AC-inclusive and AC-exclusive brain responses. As in Figure 15, it can be seen how subcortical activation is pruned when the AC have been regressed out from the brain signal.

Significant WM-related areas in the AC-inclusive brain responses

The t-test results revealed 10 and 19 significant clusters corresponding to positively correlating areas in cerebellar and cerebral regions respectively, and 4 clusters in cerebral areas that correlated negatively with the WM condition. An extensive right-lateralized effect was observed in cerebral as well as cerebellar activations, while predominantly left-lateralized for the deactivations. The WM regressor strongly activated a number of networks in cerebellar regions (right tonsil z=5.35, left semi-lunar lobule z=4.82, right tuber z=4.79, right culmen z=4.54, left tonsil z=4.50), and the largest cluster (k=70) comprised areas of the right declive and culmen. Within the prefrontal cortex (PFC) an extensive number of areas responded to the condition (R k=260; L k=84), with highest peak z-values in the right superior frontal gyrus (SFG; BA10, z=5.09), left inferior frontal gyrus (IFG; BA47, z=4.66), right IFG (BA47, z=4.24; BA45, z=4.16), right medial frontal gyrus (MedFG; BA9, z=4.34), left MedFG (BA6, z=4.45), and right precentral gyrus (PreG; BA44, z=4.61), thus revealing the significant recruitment of the bilateral vlPFC, bilateral dlPFC and (predominantly right) Broca's area.

WM-driven activations were observed in large foci in the right-left MedFG (BA6; k=95) and right PreG (BA44; k=108). Bilateral activation was associated with the IFG (BA47) and the PreG (BA44), considerably greater in the right hemisphere. Similarly greater activation was found in the right MedFG (BA32, BA6 and BA9) compared to its left hemispheric counterpart (BA32 and BA6). A region in BA10 in the left middle frontal gyrus (MFG) was also active, while a localized more medial region within BA10 significantly deactivated. Cortical structures in the right IFG (BA45) and right SFG (BA6 and especially large in BA10) also activated but did not show a contralateral effect. A sizable decrease of activation in the occipital region was seen in the left-hemispheric cuneus prolonging to the superior occipital gyrus (BA19, k=110), and a similar contralateral deactivation co-occurred in a small portion (k=10) of the middle and occipital temporal gyri (BA19). Subcortical structures were seen in the limbic region where a small activated area was observed in the right anterior cingulate (BA24) as well as the subcallosal and parahippocampal gyrus (BA34), whereas a decrease in activation occurred in the contralateral parahippocampal gyrus. Additionally, the right hippocampus and right red nucleus (brainstem) activated during the WM condition. The extensive basal ganglia (putamen, caudate and nucleus of globus pallidus) and thalamus activation to the WM condition were predominantly right-lateralized.

Significant WM-related areas in the AC-exclusive brain responses

In the absence of the AC-dependent activation, the t-test produced 9 and 14 positive correlating clusters in cerebellar and cerebral regions respectively, and other 4 cerebral clusters that correlated negatively with the WM regressor. In the following report of the results we will focus in contrasting the results with the previous t-test (without the removal of the AC-correlating voxels).

Overall bilateral decrease in activation was observed in cerebellar as well as cerebral regions in the absence of the AC-correlating voxels, with a number of intersecting loci, as it was expected, plus a few new activated areas. Within the PFC, left medial and superior regions in the frontal gyri were reduced (BA10, BA44), and right PreG (BA44) and SFG (BA6) were also drastically reduced in number of activated voxels. The previous activation in BA47 found active in the left hemisphere, as well as the right BA45, BA32 and BA19 were absent in this new activation map. However, increased activation was observed in the left MedFG (BA32), as well as in the right IFG (BA47) and MedFG (BA9). Moreover, the t-test revealed new active areas in the left IFG (BA45) and MedFG (BA9). The decreased activation in the right occipital cortex (middle temporal gyrus [MTG], BA19) was absent. The limbic structures were reduced to the right parahippocampal gyrus (BA19 and BA30). A left temporal region in the hippocampus was found active, and areas belonging to the right subgyral area and right middle temporal gyrus (BA19 and BA39) were involved in decreased activation. Activation previously found in the brainstem was suppressed. Sublobar regions in the globus pallidus and thalamus were absent in this map, and other activations in the caudate and putamen were considerable reduced. However, a sizable (k=76) new area, right insula (BA13), arose in the map.

RESULTS

Table 1.

#	REGION	H (R/L)	BA	k	X	Y	Z	MAX z-value
POSITIVE CORRELATIONS								
1	Cerebellar tonsil	L		21	-15	-48	-47	4.50
2	Cerebellar tonsil	R		57	22	-46	-45	5.35
3	Cerebellar semi-lunar lobule	L		37	-11	-73	-39	4.82
4	Cerebellar tuber	R		33	38	-67	-30	4.79
5	Cerebellar pyramis	L		3	-25	-65	-30	3.83
	Cerebellar tonsil	L		12	-27	-65	-32	4.16
6	Cerebellar nodule	L		26	-1	-56	-24	4.11
	Cerebellar nodule	R		23	3	-60	-24	4.05
7	Cerebellar dentate	R		18	12	-52	-24	4.12
8	Cerebellar lingual region	R		13	-1	-44	-17	3.91
	Cerebellar culmen	R		7	3	-48	-10	3.81
9	Parahippocampal gyrus	R	34	5	15	1	-14	4.18
	Subcallosal gyrus	R	34	1	20	4	-12	3.62
10	Cerebellar lingual region	L		2	-1	-38	-15	3.98
11	Cerebellar declive	R		7	6	-58	-12	3.99
	Cerebellar culmen	R		69	3	-60	-10	4.54
12	Caudate head	R		4	17	18	-3	4.05
	Lentiform nucleus of putamen	R		43	19	16	-5	4.90
13	Caudate head	L		30	-17	14	4	4.54
	Lentiform nucleus of putamen	L		63	-18	16	4	4.93
14	Red nucleus of brainstem	R		4	4	-25	-3	3.89
15	Inferior frontal gyrus	R	47	16	41	30	-1	4.21
16	Sub-gyral hippocampus	R		7	26	-36	0	3.79
17	Caudate body	R		94	10	6	16	5.35
	Thalamus	R		16	10	-13	1	3.90
18	Inferior frontal gyrus	L	47	14	-50	20	-1	4.66
19	Precentral gyrus	R	44	108	47	16	9	4.61
20	Anterior nucleus of thalamus	L		1	-6	1	5	3.48
	Lentiform nucleus of globus pallidus	L		9	-11	-3	1	4.36
21	Precentral gyrus	L	44	20	-45	14	9	4.16
22	Anterior cingulate	R	24	1	8	28	12	3.48
23	Superior frontal gyrus	R	10	43	27	43	20	5.09
	Medial frontal gyrus	R	9	12	22	41	14	4.34
24	Thalamus	R		1	4	-23	11	3.53
25	Inferior frontal gyrus	R	45	11	48	20	17	4.16
26	Middle frontal gyrus	L	10	21	-27	41	21	4.21
27	Medial frontal gyrus	L	32	4	1	10	44	3.73
	Medial frontal gyrus	R	32	43	4	10	46	4.12
28	Superior frontal gyrus	R	6	8	3	26	56	4.24
29	Medial frontal gyrus	L	6	25	-3	-5	61	4.45
	Medial frontal gyrus	R	6	70	6	-3	59	4.29
NEGATIVE CORRELATIONS								
1	Parahippocampal gyrus	L	28	12	-17	-21	-26	4.40
2	Medial frontal gyrus	L	10	13	-10	55	4	4.03
3	Middle temporal gyrus	R	19	10	36	-63	12	3.80
4	Superior occipital gyrus	L	19	108	-31	-83	29	5.25
	Cuneus	L	19	2	-31	-81	31	3.97

Table 2.

#	REGION	H (R/L)	BA	k	X	Y	Z	MAX z-value
POSITIVE CORRELATIONS								
1	Cerebellar tonsil	R		40	22	-46	-45	5.12
2	Cerebellar semi-lunar lobule	L		10	-13	-75	-39	4.14
3	Cerebellar tonsil	L		11	-27	-65	-32	3.85
4	Cerebellar pyramis	R		14	40	-67	-32	3.84
5	Cerebellar nodule	L		3	1	-60	-24	3.56
	Cerebellar nodule	R		7	3	-62	-26	3.79
6	Cerebellar anterior lobe	R		9	12	-40	-27	4.09
7	Cerebellar lingual region	L		2	-4	-48	-19	3.65
	Cerebellar lingual region	R		7	1	-46	-18	3.84
8	Cerebellar culmen	L		2	1	-48	-12	3.49
	Cerebellar culmen	R		8	3	-50	-12	3.70
9	Cerebellar culmen	L		19	1	-60	-10	3.98
10	Caudate head	R		1	17	18	-3	3.57
	Lentiform nucleus of putamen	R		13	20	16	-5	4.20
11	Inferior frontal gyrus	R	47	24	41	30	-1	4.35
12	Hippocampus	L		22	-29	-36	-4	4.01
13	Caudate head	L		9	-17	14	4	4.00
	Lentiform nucleus of putamen	L		16	-18	14	4	4.12
14	Insula	R	13	77	48	12	6	4.16
	Precentral gyrus	R	44	8	47	16	8	3.87
15	Inferior frontal gyrus	L		13	-50	22	-1	4.22
16	Parahippocampal gyrus	R	19	2	22	-44	-1	4.05
	Parahippocampal gyrus	R	30	5	20	-48	-1	3.93
17	Precentral gyrus	L	44	12	-43	14	9	4.32
	Inferior frontal gyrus	L	45	23	-52	20	11	4.23
18	Superior frontal gyrus	R	10	46	27	43	21	4.82
	Medial frontal gyrus	R	9	20	22	41	14	4.13
19	Medial frontal gyrus	L	9	45	-24	39	18	4.13
	Superior frontal gyrus	L	10	4	-24	43	20	3.53
20	Caudate body	R		41	10	6	16	4.21
21	Superior frontal gyrus	L	6	3	-1	4	47	3.84
	Medial frontal gyrus	L	32	13	1	6	46	3.81
22	Superior frontal gyrus	R	12	29	8	-1	63	3.99
23	Medial frontal gyrus	L	6	22	-3	-5	61	4.21
NEGATIVE CORRELATIONS								
1	Middle occipital gyrus	L	19	5	-40	-85	16	4.18
2	Middle temporal gyrus	R	19	7	38	-75	19	3.69
	Middle temporal gyrus	R	39	1	40	-73	15	3.54
3	Cuneus	L	19	89	-29	-83	27	4.91

Table 1 & Table 2. Subtraction (t-test) results for WM versus REP maps in both AC-inclusive (Table 1) and AC-exclusive (Table 2) brain responses. Results reflect significantly higher (activation) and lower (deactivation) z-scores for the WM condition. The significance threshold was set to p=.0005 (z=3.48) and the CS ≥ 8 voxels. The table reports the numbered clusters under which the falling regions are detailed, with information about the hemisphere location (H), cluster size (k; i.e., number of voxels in the activated cluster), Brodmann area (BA), Talairach coordinates (TAL) and their respective peak z-scores.

5 DISCUSSION

The aim of this study was to detect WM-driven activations in the neural tissue resulting from the continuous listening of a real piece of music. The whole brain of eleven healthy participants was imaged during a naturalistic listening and their BOLD responses were recorded. We predicted the temporal evolution of WM activations in the brain triggered by the repetition of the musical motifs in the piece. The expected hemodynamic response (HDR) to stimulus-driven activation was correlated against the actual BOLD signal at each voxel at participant level. To ensure the activations were due to WM, time courses of perceptually validated acoustic features were regressed out from the brain responses. By means of a subtraction (t-test) we compared the maps obtained using the brain responses including and excluding the acoustic correlates in the brain to see extent to which acoustic sensory processing interferes with the WM condition. As a result we clearly observed extensive subcortical activity being removed when filtering out the acoustic processing correlates in the brain. Similarly, when we included the first instance of each motif, and we subtracted it to conditions including only motif repetitions, considerable subcortical processing was pruned or significantly decreased when filtering out the AC-dependent voxels. This might expose dependence of part of WM encoding on the acoustic processing regions, and thus WM-related processing might rely on those overlapping AC-dependent networks. The significant remaining cerebrocortical activation in the AC-filtered map was expected to reveal more relevant areas for WM processing, as they do not intersect with the AC-processing in the brain. As for cortical activity, a rightward asymmetry comprising ventral and dorsal areas in the prefrontal cortex was observed. In the next section the results of the analysis will be discussed in detail.

Distributed activity in the brain derived from the attentive listening task was observed. We could attribute this high degree of distributedness to the fact that WM is a complex, high-level cognitive process. The assumption that a high-level mental process has a high degree of distributedness has support in the literature: distributed networks are known to support cognition, responding even during simple tasks (Curtis & D'Esposito, 2003). In addition, results reviewed by Collete et al. (2006) on WM indicate that executive functioning relies on a distributed cerebral network rather than in a restricted set of anterior cerebral areas (...). Furthermore, the use of a naturalistic paradigm represents an added complexity: "Perception of and action in a complex environment usually requires the parallel processing of information related to different objects or events that have to be kept apart to allow sensory segmentation and goal-directed behaviour" (Engel & Singer, 2001). Thus, as investigated in this study, WM for music is subserved by a number of cortical and subcortical brain structures distributed over the prefrontal cortex, limbic and paralimbic areas and cerebellum.

5.1 NEUROANATOMY OF WORKING MEMORY FOR MUSIC

The discussion will focus first in the results derived from the brain responses including the acoustic components, and differences between these AC-inclusive brain responses and the AC-exclusive responses will be subsequently commented.

Correlational analysis revealed mainly right-lateralized cerebellar responses to the WM regressor (tonsil, culmen, lingual, nodule, dentate, pyramis). In a study by Leung & Alain (2011) right cerebellar activity in the tonsil, culmen and pyramis was revealed in a WM task relevant to location and category. This cerebellar activation might be in connection with temporal processing of the stimuli used (Mathiak, Hertrich, Grodd, & Ackermann, 2004). In Alluri et al. (2012) increased activation in cerebellar areas (declive, uvula and pyramis) was found to correlate with high values of fullness and activity in the Piazzolla stimulus. Thus neural substrate underlying these acoustic features might aid in the WM encoding of the music. A large number of prefrontal cortical regions activated in response to the WM predictor. Regions in the vlPFC and dlPFC (bilateral inferior [BA47] and right superior [BA10] frontal gyri respectively) correlated positively, areas that within the PFC seem to be critical for WM functions (Kane & Engle, 2002), with the IFG reflecting WM aspects in the integration of information over time (Nan et al., 2008). Activity in BA47 has in the context of music been linked with the processing of musical temporal structure, specifically with extracting the correct temporal information in a sequence (Chen et al. 2008). Another recruited area in the IFG was BA45 (right hemisphere). Area BA45 has been found to be part of a vlPFC node in a dissociative ventral-dorsal WM model (Petrides, 2005), whereby the vlPFC node mediates in active information retrieval and encoding, and the dlPFC manipulates and monitors information. It has also been associated with verbal and non-verbal retrieval from LTM (Petrides, 1996). In our study, area BA44 in the PreG responded strongly right lateralized. The role of the left BA44 corresponding to the Broca's area (covering the pars triangularis and pars opercularis) as a syntactic processor or as a WM resource in sentence comprehension has been debated over the last 30 years and remains controversial (Rogalsky, Matchin, & Hickok, 2008). In the context of music, fMRI studies (Koelsch & Siebel, 2005) using chord sequence and melody paradigms have linked music syntactic processing with a predominantly right activation in the pars opercularis. It has been likewise observed active in WM for pitch (Zatorre, Evans, & Meyer, 1994; Koelsch & Siebel, 2005) as well as in music-driven rhythmic tasks or while playing music (Peretz & Zatorre, 2003). In addition, left-hemispheric activation of Broca's area has responded to musical imagery involving semantic retrieval (Halpern & Zatorre, 1999).

Regions in the bilateral MedFG (BA6), right SFG (BA6) and left MedFG (BA9) were found active in response to the WM regressor. These areas have been observed active while continuously updating information and maintaining temporal order in WM (Wager & Smith, 2003). The left MedFG (BA32) also activated in response to WM, which in storage tasks has been found to be frequently active for increased demands in selective attention to features of a stimulus to be stored in WM

(Wager & Smith, 2003). In Alluri et al. (2012) the left MedFG (BA6) was seen active with low levels of activity in the music.

Active cerebral subcortical structures were recruited in the right hippocampus, right parahippocampal gyrus (BA34) and a small region in the right subcallosal gyrus (BA34). The hippocampus has been known to be crucial in the formation of new long-term memories based on episodic or autobiographical events (Cohen & Eichenbaum, 1993; Squire & Butters, 1984) as well as a novelty detector (VanElzakker, Fevurly, Breindel, & Spencer, 2008), and in WM it has an important role during maintenance tasks (Nichols, Kao, Verfaellie, & Gabrieli, 2006). The parahippocampal gyrus is also important in spatial memory, specifically for the encoding and recognition of scenes (Aguirre, Detre, Alsop, & D'Esposito, 1996). A region in the left parahippocampal gyrus (BA28) was observed to correlate negatively in the presence of the AC-correlating voxels. Interestingly, Blood and Zatorre (2001) observed regional cerebral flood flow (rCBF) decreases in the left hippocampus with increasing chills intensity. Hence this left-localized inhibition might be associated to habituation of emotional responses triggered by the motifs once they became familiar upon being coded in WM. Two other alternative explanations were posited by Peretz and Zatorre, (2001): a) pleasure might be maximized not only by activating the limbic system, but also by simultaneously decreasing activity in brain structures associated with negative emotions; b) decreases in this area might occur in relation to anticipation of the chills response than to the chills response itself, as suggested in two studies by Ketter, Andreason, George, Lee, Gill, Parekh, Willis, et al. (1996) and Breiter, Gollub, Weisskoff, Kennedy, Makris, Berke, Goodman, et al., (1997), where amygdala decreases correlated with ratings of 'craving' rather than rush. In any case, the modulation of the limbic system seems to be a complex matter. In the context of acoustic feature processing in the brain, decreased pulse clarity was found to be associated with activation in the left hippocampus (Alluri et al., 2012). The recruitment of this limbic area in connection to decrease of pulse clarity was interpreted by the authors to be associated with the tension caused by the perceived lack of pulse clarity. This tension may be pleasurable, thus activating emotion-related brain areas.

The putamen remained bilaterally and extensively activated. The role of the putamen has been primarily related to motor skills (DeLong, Alexander, Georgopoulos, Crutcher, & Mitchell, 1984), but also in implicit and motivational learning (Packard & Knowlton, 2002). Interestingly, the left putamen was involved in both the phonological and tonal loop, but more strongly in the latter, exclusively in musicians in Schulze's (2011) study on verbal and tonal WM. Likewise, in Pallesen et al. (2010) musicians showed larger BOLD responses than non-musicians in the right putamen (along with other areas) during a musical chord WM task. In acoustic feature processing in the brain, the right putamen was active in response brightness, whereas the left putamen correlated negatively with pulse clarity in the music (Alluri et al., 2012). Another subcortical activation was observed in the caudate region (mainly right-lateralized). As part of the basal ganglia, the caudate has been shown

to be involved not only in motoric functions, but also in learning and memory (Packard & Knowlton, 2002).

Areas in the midbrain, and anterior cingulate (BA24), left nucleus of globus pallidus, and predominantly right thalamus positively correlated with the WM regressor. Thalamic activity has been frequently linked to perceptual and cognitive tasks and particularly to the alert component in attention (Wager & Smith, 2003), whereas the anterior cingulate has been reported to cooperate with the dorsolateral prefrontal cortex for attention shifting (Osaka, Osaka, Kondo, Morishita, & Fukuyama, 2004).

A sized area that correlated negatively in response to the WM regressor was the left-right BA19 (specifically localized in the left cuneus, left superior occipital gyrus and right MTG), an area involved in visual processing. In reference to this deactivated area, Levitin & Menon (2003) reported greater activation in a similar region in the cuneus and superior and middle occipital gyri among other nearby areas in the presence of scrambled music compared to non-scrambled music. From this finding the authors did not infer this activation as reflecting processing of scrambled music, since many of these regions are known to be deactivated during both auditory and visual processing (Laurienti, Burdette, Wallace, Yen, Field, & Stein, 2002). In the study carried out by Alluri et al. (2012), the left-right MTG (with leftward bias) seemed to be implicated in the perceptual processing of timbral features of the Piazzolla piece that was used in the present study. However the only regions in the MTG in our study showed negative correlations with the WM regressor. This might imply that this musical set of features were not significant for WM to encode the motivic material in the music.

Comparison with the results in the AC-exclusive responses

Here we will discuss some differences with the results obtained from the correlational analysis when the AC were included as nuisance regressors, thus being removed from the brain responses. Whereas activity in area BA47 (IFG) was exhibited bilaterally when the AC-correlating voxels remained in the data, this activation was right lateralized in the AC-exclusive brain responses. Conversely, BA45 in the IFG was left-lateralized, but right-lateralized in the AC-inclusive responses. For BA44 (PreG) there was activation in both hemispheres but with decreased extension compared to the responses including the acoustic content. The active regions found in the AC-inclusive responses in the bilateral MedFG (BA6), right SFG (BA6) and left MedFG (BA9) were also found when AC-correlating voxels were excluded from the analysis. However, as previously mentioned, fewer cerebral subcortical structures were recruited. Subcortical active regions were found within the left hippocampus and right parahippocampal gyrus (BA19 and BA30). The putamen, bilaterally and extensively activated in the AC-inclusive brain responses, decreased considerably in size once the AC time courses were extracted from the data. The mainly right lateralized caudate region found in the AC-inclusive responses was present in the AC-exclusive responses but smaller in volume. The

midbrain and anterior cingulate (BA24), left nucleus of globus pallidus, and thalamus were not recruited, that had been active in the AC-inclusive brain responses. The negatively correlating area BA19 in response to the WM regressor in the AC-inclusive responses was also found to approximately overlap in the AC-exclusive brain responses (recruiting left cuneus, left middle occipital gyrus, right MTG extending to a small portion in BA39). Finally, a large portion of the right insula (BA13) was observed to respond to the WM regressor once the AC in the data were excluded, which was not recruited in the presence of the AC-correlating voxels. This brain structure functions as an integral hub activating in response to a number of functions of different nature (cognitive, emotional, and regulatory), and has recently been implicated in critical cognitive control and attentional processes (in salient-event detection) aiding different brain regions to properly respond to salient stimuli (Menon & Uddin, 2010). Alluri et al. (2012) found the bilateral insula (BA13) active to decreased pulse clarity in the music.

5.2 HEMISPHERIC SPECIALIZATION

Whereas studies on verbal WM have found support for a left hemispheric pattern (Wager & Smith, 2003), in our study we observed a rightward bias in response to the music-related WM effect especially pronounced in the presence of the acoustic-related content in the data. Likewise, right-hemispheric predominance has been implicated in a stream segmentation study in natural settings which engaged a ventral network (including the recruitment of vlPFC [BA47, 44/45] for detecting salient events, and dorsal network (including dlPFC [BA9]) for maintaining attention and updating WM (Sridharan, Levitin, Chafe, Berger, & Menon, 2007). These same regions were observed in the present study to be active in music-driven WM, and in fact it seems sensible to consider these segmentation processes as sharing brain circuits with WM, since stream segmentation into perceptually meaningful chunks is a necessary requirement for WM encoding. Furthermore, tempo-tracking structures in music are bilateral, whereas similar structures in speech are predominantly left-lateralized (Levitin & Tirovolas, 2009). This rightward trend has been dominant in the literature on memory-related experiments with music (Zatorre et al., 1994; Griffiths, Johnsrude, Dean, & Green, 1999).

5.3 CONTRAST WITH PREVIOUS STUDIES

When contrasting with music-related studies in WM, we did find somewhat significant intersecting areas, although we failed to detect other equally significant areas within our results. For instance, we found some of the common core regions involved in tonal and verbal WM in musicians reported by Schulze (2011), namely Broca's area, right cerebellum and left putamen, but did not locate others, such as the left-right premotor dorsal/premotor ventral (PMd/PMv), left inferior parietal lobule (IPL), pre-SMA/SMA, left insula, left inferior frontal sulcus (IFS), left intra parietal sulcus/superior parietal lobule (IPS/SPL). From the subcomponent regions specific to the so called "tonal loop" in

musicians (which included left cuneus, right globus pallidus, right caudate nucleus, and left cerebellum), we identified in our study the right caudate region and left cerebellum (we did find activation in the left globus pallidus, instead of its right-sided homologue). The involvement of the planum temporale and superior marginal gyrus (SMG) has been also reported in the literature (Gaab, Gaser, Zaehle, Jancke, & Schlaug, 2003; Zatorre et al., 1994) in pitch WM memory tasks, but were not present in our study. Similarly, the STG seems to be important in short-term auditory retention (Zatorre & Samson, 1991; Gaab et al., 2003; Zatorre et al., 1994; Schulze et al., 2011), but we did not find this region active.

As for the results in the brain responses excluding the acoustic correlates, the same regions reported to intersect with Schulze's tonal loop were found, namely, Broca's area, cerebellum, caudate region and putamen. However the recruitment of these areas was smaller than in the AC-inclusive responses. From the subcortical areas, the putamen was also found active but extensively pruned, and in the the case of the globus pallidus, completely removed from the map.

6 CONCLUSIONS AND FURTHER RESEARCH

We studied music-related WM in musicians using a naturalistic non-standard procedure: a) participants' task was to listen attentively to a piece of music while their responses were recorded, instead of performing auditory-cued tasks; b) an unusually complex stimulus was used containing strong shifts in tempo, timbre, dynamics, tonality and rhythm, more representative of the complex auditory scene environment our brains have evolved to respond to. In the present study the effect of interest —WM activation in the brain— is assumed to be elicited by similar, repeated music material in the piece.

We showed that a naturalistic approach to study WM is viable by using Western tonal music, which provides naturally occurring motivic repetition and variation —recognizable units— serving as WM trigger, thus facilitating the phenomenon of motif-tracking in the context of real music without the need for artificially manipulated stimuli. To reveal the activated networks in the brain, we correlated participants' BOLD signal with the expected hemodynamic response employing non-parametric procedures to tackle specific problems avoiding "black box" approaches. We decided to filter out perceptual correlates of a set of musical features from the data, expecting this would aid to exclusively uncover the executive processes of WM. However, significant subcortical processing active in response to the WM condition in the AC-inclusive brain responses was pruned with the removal of the acoustic content, suggesting these areas, involved with perceptual processing of acoustic features, aid in the encoding and retrieval of WM.

The results derived from the pooled participant map revealed a widely distributed network of cortical and subcortical areas, predominantly right-lateralized, responding to the WM model, some of which had been previously reported in the literature on WM for non-verbal stimuli (cerebellum, prefrontal cortices, and motor-related areas). Additionally —and interestingly— limbic and paralimbic structures were recruited in our study in response to the WM regressor that had not been found in previous studies. Hence the contribution of different perceptual areas seems to be relevant to mnemonic processing.

Due to the characteristics of our WM regressor, this study assumed transient effects derived from the regressor's predictions. Thus sustained activity subserving WM was not included in the study. Continuous and transient patterns of neural activity might suggest different functional roles (Collette et al., 2006). Additionally, the similarity between the motif presentation and subsequent repetitions could be perceptually quantified (i.e., have participants rate the similarity of pairwise comparisons [first motif vs. successive appearances], allowing for parametrically varying degrees of similarity. The predicted response would then be modelled as a function of the continuous values representing different levels of memory load (which reflect fluctuations in HDR intensity) smoothed with a HRF, rather than specified by a binary-valued vector for states 'on' and 'off'.

To deepen into questions of brain specialization, musicians' responses could be contrasted against non-musicians'. There is evidence that confers musicians an enhanced ability to retrieve, monitor, and chunk information over non-musicians (Chen et al., 2008), crucial processes for WM efficiency which could be detected in the brain dynamics. For instance, we might see not also differences in asymmetry across groups for WM, but also within groups across different conditions, i.e., due to different levels of complexity in the music. For instance, although musicians tend to rely on left-sided brain for music processing in a greater extent than non-musicians (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2005), possibly explained by a more consciously learned or analytic approach to the musical input, complex music has been reported to drive even trained musicians into strongly using their 'right brain' (Vollmer-Haase, Finke, Hartje, & Bulla-Hellwig, 1998; McGilchrist, 2010). Vollmer-Haase's explanation is the increased WM requirement to analyse this complex musical material, while McGilchrist proposes it may lie in the perceptual "new" experience of this music on different hearings, due to the impossibility of attending to all parts as a whole.

6.1 WORKING MEMORY AS AN EMERGENT TEMPORAL INTEGRATION MECHANISM

To date very often neurovascular effects do not produce the findings that would support or disprove psychological theories, because a) postulates of psychological nature might not prove easy to support by mapping function onto brain tissue; b) psychological hypotheses do not normally provide a localizationist account of the different functional modules represented in them (Yonelinas, 2002; Heathcote, 2003; Cohen & Squire, 1980; Schacter & Tulving, 1994; Baddeley, 1992). Associating distinct brain loci with functions A and B is no support for a theory if it does not predict this result, or, in other words, given two independent processes A and B, it does not follow that different brain areas will be activated by them. Hence the assumptions deriving from two different fMRI results do not necessarily challenge a theory about the mind that is not strictly neuroanatomical in nature. Clearly this does not render the finding in any way meaningless, but can say little if the purpose is solely to discard or support a psychological theory that lacks neuroanatomical considerations (Coltheart, 2006a, 2006b).

However, in the light of an theory of WM as an non-modular, emergent property arising from the interaction of a highly distributed neural network advanced by some researchers (Postle, 2006; Hazy, Frank, & O'Reilly, 2006; D'Esposito, Postle, & Rypma, 2000; Collete et al., 2006, Miyake & Shah, 1999), we can make some observations, namely, that filtering out the acoustic features from the music in order to find the underlying WM main executive function might be the wrong approach to studying music-driven WM networks, and the only purpose of this subtraction operation, whereby the acoustic correlates are removed, is to observe the significant implication of the subcortical sensory subsystems co-responsible for WM retrieval, filtered out with the exclusion of the acoustic correlates that are necessary for the encoding of the memoranda they represent.

Thus WM's executive centre might then fundamentally be an emergent property arising from the ad-hoc activation of the perceptual neural circuitry that is stimulus-dependent plus areas in the prefrontal cortex, as posited by Fuster (1995), instead of consisting of a fixed 'place' in the cognitive architecture. The PFC, in cooperation with the other brain areas, seems to subserve the common underlying process to all function: temporal integration, a uniform requirement of all WM tasks, as well as of many others (Duncan & Owen, 2000), thus the neural substrate of WM would be as widely distributed as the LTM that supports it (Fuster, 2001).

6.2 MOVING AWAY FROM THE LOCALIZATIONIST APPROACH

The problem of studying complex fMRI data is inherently one of statistical choice and therefore one about the assumptions made. Neuroimaging techniques are an illuminating method for looking at the brain enabling to map the topography related to brain function. However, inferences about recruitment of specific areas only tell about their engagement, and not requirement, in the process under investigation. Additionally, the very nature of executive processes has rendered very difficult for neuroimaging methods to produce homogeneous results as to what their underpinning neural correlates are (Collete et al., 2006). WM, as a complex top-down cognitive process, is dependent on an intricate and widely distributed network. Attempting to separate the individual processes being measured in such distributed circuitry becomes a challenging task, and so the use of subtraction paradigms aiming to isolate the effect of interest has its limitations (McGilchrist, 2010). Moreover, the spatial and temporal resolution inherent to functional neuroimaging is not optimal, particularly when studying memory and attention, that would allow the detection of topographic differences within anatomical regions related to specific memoranda or the temporal evolution of neural excitability at cellular level (Fuster, 2001), because not only the localization of function, but its temporal pattern of activity, is crucial to best characterize function.

The configuration and behaviour of interdependent connections between distant areas seems to be crucial in understanding cognitive functions. Thus more meaningful results could be obtained using multivariate fMRI analysis techniques that model interactions between different regions (Curtis & D'Esposito, 2003), attempting to understand brain function in a connectionist context, which are critical in showing plasticity in the brain. Thus we move from the localizationist approach (localization of information) to functional integration (assessment of patterns of information representation). For instance, phase synchronization of neural sources can reveal information exchange among them, and their timing can be a factor in establishing a causal relationship between the sources (Handy, 2009).

Furthermore, the use of a naturalistic setting to study musical WM using the motivic material as our WM trigger should motivate a similar naturalistic approach to studying verbal WM. Insofar as musical motifs are the iterated, recognizable musical structures upon which a piece is built that

render it coherent, a linguistic discourse is characterized by a theme or subject matter embodied by recurring ideas across the text.

We are just now beginning to unravel the systems problem of WM for music, and so how WM is formed in the brain remains an open issue. This study points in the direction of exploring the functional brain topology of music-elicited WM in a first-time naturalistic listening condition in musicians. In view of the recruitment of hippocampal regions within the present study, we cannot ignore another component to memory: emotion, which is thought to play an influential role in the mnemonic power of music as to how and what we remember. Evidence on the robust integration of musical, memory and emotion, i.e., in autobiographical memories, has been reported (Janata, 2009; Ashley & Luce, 2004; Eschrich, Münte, & Altenmüller, 2008; Jäncke, 2008). However, research on how memory interacts with cognitive and emotional processes is still at an elementary stage.

7 REFERENCES

Aguirre, G. K., Detre, J. A., Alsop, D. C., & D'Esposito, M. (1996). The parahippocampus subserves topographical learning in man. *Cereb. Cortex*, 6, 823–829.

Alluri, V., Toiviainen, P., Jääskeläinen, I. P., Glerean, E., Sams, M., & Brattico, E. (2012). Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *NeuroImage*, 59(4), 3677 – 3689.

Andrade, J. (2002). *Working Memory in Perspective*. Hove: Psychology Press.

Ashley, R., & Luce, K. (2004). Music, autobiographical memory, and emotion. *Emotion*, 451–452.

Atkinson, R., & Shiffrin, R. (1968). Human memory: a proposed system and its control processes. In K. Spence & J. Spence (Eds.), *The Psychology of Learning and Motivation: Advances in Research and Theory. Vol 2. New York: Academic Press* (Vol. 2, pp. 89–195).

Attwell, D., & Iadecola, C. (2002). The neural basis of functional brain imaging signals. *Trends in neurosciences*, 25(12), 621–5.

Awh, E., Jonides, J., Smith, E., Schumacher, E., Koeppe, R., & Katz, S. (1996). Dissociation of Storage and Rehearsal in Verbal Working Memory: Evidence From Positron Emission Tomography. *Psychological Science*, 7(1), 25 – 31.

Baddeley, A. (2007). *Working memory, thought, and action. Oxford University Press*. Oxford University Press.

Baddeley, A. (1992). Is working memory working? The Fifteenth Bartlett Lecture. *Quarterly Journal of Experimental Psychology*, 44, 1–31.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423.

Baddeley, A, & Hitch, G. J. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation Vol. 8* (pp. 47–89).

Bellis, T. J. (2003). *Assessment & Management of Central Auditory Processing Disorders in the Educational Setting: From Science to Practice* (p. 532). Cengage Learning.

Berz, W. L. (1995). Working Memory in Music: A Theoretical Model. *Music Perception 12, 3*.

Biba, M., Esposito, F., Ferilli, S., Mauro, N. Di, Basile, T. M. A., & Orabona, V. (2007). Unsupervised Discretization Using Kernel Density Estimation. *Proceedings of the 20th international joint conferencIe on Artifical intelligence* (pp. 696–701).

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 of the United States of America*, 98(20), 11818.

Brattico, E., Alluri, V., Bogert, B., Jacobsen, T., Vartiainen, N., Nieminen, S., & Tervaniemi, M. (2011). A Functional MRI Study of Happy and Sad Emotions in Music with and without Lyrics. *Frontiers in psychology*, 2, 308.

Breiter, H. C., Gollub, R. L., Weisskoff, R. M., Kennedy, D. N., Makris, N., Berke, J. D., Goodman, J. M., et al. (1997). Acute Effects of Cocaine on Human Brain Activity and Emotion. *Neuron*, 19(3), 591–611.

Cambouropoulos, E. (2009). How similar is similar? *Musicae Scientiae. Discussion Forum 4B*, 7–24.

Caulfield, T., Rachul, C., Zarzeczny, A., & Walter, H. (2010). Mapping the coverage of neuroimaging research. *SCRIPTed*, 7(3), 421–428.

Chan, A. S., Ho, Y. C., & Cheung, M. C. (1998). Music training improves verbal memory. *Nature*, 396(6707), 128.

Cohen, J., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task, *nature*(nature).

Cohen, N., & Squire, L. (1980). Preserved learning and retention of pattern analysing skills in amnesia. *Science*, 210, 207– 209.

REFERENCES

Cohen, N. J., & Eichenbaum, H. (1993). Memory, Amnesia and the Hippocampal System. *Cambridge, MA: MIT Press*.

Collette, F., Hogge, M., Salmon, E., & Van der Linden, M. (2006). Exploration of the neural substrates of executive functioning by functional neuroimaging. *Neuroscience*, 139(1), 209–21.

Coltheart, M. (2006a). What has functional neuroimaging told us about the mind (so far)? *Cortex*, 42 (3), 323–331.

Coltheart, M. (2006b). Perhaps functional neuroimaging has not told us anything about the mind (so far). *Cortex*, 42, 422–7.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: a methodological review and user's guide. *Psychon. Bull. Rev.*

Courtney, S., Petit, L., Maisog, J., Ungerleider, L., & Haxby, J. (1998). An area specialized for spatial working memory in human frontal cortex. *Science*, 279(5355), 1347–1351.

Cowan, N. (1999). An Embedded-Processes Model of Working Memory. In A. Miyake & P. Shah (Eds.), *Models of Working Memory. Mechanisms of Active Maintenance and Executive Control* (pp. 63–100). Cambridge University Press.

Cowan, N. (2005). *Working memory capacity* (p. 264). Psychology Press.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in brain research*, 169(07), 323–38.

Cowan, N. (1995). (1995). *Attention and memory: An integrated framework*. Oxford: OxfordUniversity Press.

Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Semin Neurosci*, 2, 263–275.

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences*, 7(9), 415–423.

Damasio, A. R. (1990). Synchronous activation in multiple cortical regions: a mechanism for recall. *Semin. Neurosci.*, 2, 287–296.

DeLong, M. R., Alexander, G. E., Georgopoulos, A. P., Crutcher, M. D., Mitchell, S. J., & Richardson, R. T. (1984). Role of basal ganglia in limb movements. - Abstract. *Human Neurobiology*, 2(4), 235–24.

Deutsch, D. (1970). Tones and numbers: specificity of interference in immediate memory. *Science*, 168, 1604–5.

Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science (New York, N.Y.)*, 298(5596), 1191–4.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neuroscience*, 10, 475–483.

D'Esposito, M. (2007). From cognitive to neural models of working memory. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 362(1481), 761–72.

D'Esposito, M., & Postle, B. R. (2002). The organization of working memory function in lateral prefrontal cortex: Evidence from event-related functional MRI. In D. T. Stuss & R. Knight (Eds.), *Principles of Frontal Lobe Function* (pp. 168–187). Oxford: Oxford University Press.

D'Esposito, M., Postle, B. R., & Rypma, B. (2000). Prefrontal cortical contributions to working memory: evidence from event-related fMRI studies. *Exp. Brain Res.*, 133, 3–11.

Ebisuzaki, W. (1997). A Method to Estimate the Statistical Significance of a Correlation When the Data Are Serially Correlated. *Journal of Climate*, 10(2), 2147–2153.

Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in cognitive sciences*, 5(1), 16–25.

Engle, R. W. (2002). Working memory capacity as executive attention. *Curr. Dir. Psychol. Sci*, 11.

Engle R. W., Tuholski S. W., Laughlin J. E., Conway, A. R. A (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Experimental Psychology: General.*

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211–245.

Eschrich, S., Münte, T. F., & Altenmüller, E. O. (2008). Unforgettable film music: the role of emotion in episodic long-term memory for music. *BMC neuroscience*, 9, 48.

REFERENCES

Evans, A., Kamber, M., Collins, D., & MacDonald, D. (1994). An MRI-based probabilistic atlas of neuroanatomy. In S. D. Shorvon (Ed.), *Magnetic Resonance Scanning and Epilepsy* (pp. 263–274). New York: Plenum Press.

Fiez, J. A., Raife, E. A., Balota, D. A., Schwarz, J. P., Raichle, M. E., & Petersen, S. E. (1996). A positron emission tomography study of the short-term maintenance of verbal information. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 16(2), 808–22.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10(4), 507–521.

Fisher, R. A. (1950). *Statistical Methods for Research Workers, 11th edn.* Oliver and Boy, London.

Fodor, J. & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: a critical analysis. In S. Pinker & J. Mehler (Eds.), Cambridge, MA: The MIT Press.

Frackowiak, R. S. J., Ashburner, J. T., Penny, W. D., & Zeki, S. (2004). *Human Brain Function, Second Edition* (p. 1144). Academic Press.

Friston, K. J., Jezzard, P., & Turner, R. (1994). Analysis of Functional MRI Time-Series. *In Vivo*, 171, 153–171.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2005). Automatic encoding of polyphonic melodies in musicians and nonmusicians. *J. Cogn. Neurosci.*, 17, 1578–1592.

Funahashi, S., Bruce, C., & Goldman-Rakic, P. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of neurophysiology*, 61, 331–349.

Funahashi, S., Bruce, C., & Goldman-Rakic, P. (1990). Visuospatial coding in primate prefrontal neurons revealed by oculomotor paradigms. *Journal of neurophysiology*, 63, 814–831.

Fuster, J. M. (1987). Single-Unit Studies of the Prefrontal Cortex. In E. Perecman (Ed.), *The Frontal Lobes Revisited* (pp. 109–120). IRBN Press, New York.

Fuster, J. M. (1995). *Memory in the Cerebral Cortex: An Empirical Approach to Neural Networks in the Human and Nonhuman Primate*. Cambridge, MA: MIT Press.

Fuster, J. M. (2001). The Prefrontal Cortex — An Update : Time Is of the Essence. *Neuron*, 30, 319–333.

Gaab, N, Gaser, C., & Schlaug, G. (2006). Improvement-related functional plasticity following pitch memory training. *Neuroimage*, 31(1), 255–263.

Gaab, N., Gaser, C., Zaehle, T., Jancke, L., & Schlaug, G. (2003). Functional anatomy of pitch memory—an fMRI study with sparse temporal sampling. *NeuroImage*, 19(4), 1417–1426.

Gaser, C., & Schlaug, G. (2003). Brain Structures Differ between Musicians and Non-Musicians, 23(27), 9240–9245.

Gjelsvik, B. E. B. (2008). *The Bobath Concept In Adult Neurology* (p. 237). Thieme.

Goldman-Rakic, P. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In Bethesda (Ed.), *Handbook of Physiology: The Nervous System. Vol 5 (1)* (pp. 373–417). Am. Physiol. Soc.

Griffiths, T.D., Johnsrude, I., Dean, J.L., Green, G. G. R. (1999). A common neural substrate for the analysis of pitch and duration pattern in segmented sound? *Neroreport*, 10, 3825–3830.

Gruber, O., & von Cramon, D. Y. (2003). The functional neuroanatomy of human working memory revisited. *NeuroImage*, 19(3), 797–809.

Halpern, A.R., Zatorre, R. J. (1999). When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex*, 9, 697–704.

Handy, T. (Ed.). (2009). *Brain Signal Analysis. Advances in Neuroelectric and Neuromagnetic Methods*. The MIT Press, Cambridge.

Haueisen, J., & Knösche, T. R. (2001). Involuntary motor activity in pianists evoked by music perception. *Journal of cognitive neuroscience*, 13(6), 786–92.

Hazy, T., Frank, M., & O'Reilly, R. (2006). Banishing the homunculus: Making working memory work. *Neuroscience*, 139, 105–118.

Heathcote, A. (2003). Item recognition memory and the receiver operating characteristic. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 29, 1210–1230.

REFERENCES

Henson, R. (2005). What can functional neuroimaging tell the experimental psychologist? *The Quarterly journal of experimental psychology. A, Human experimental psychology*, 58(2), 193–233.

Hickok, G., Buchsbaum, B., & Humphries, C. (2003). Auditory – Motor Interaction Revealed by fMRI: Speech , Music , and Working Memory in Area Spt. *Journal of Cognitive Neuroscience*, 673 – 682.

Huettel, S. A., Song, A. W., & McCarthy, G. (2009). *Functional magnetic resonance imaging (2nd Edition). Magnetic Resonance Imaging*. Sunderland, MA: Sinauer Associates.

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009). Musical training shapes structural brain development. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29(10), 3019–25.

Janata, P. (2009). The neural architecture of music-evoked autobiographical memories. *Cerebral Cortex*, 19(11), 2579.

Janata, P., Tillmann, B., & Bharucha, J. J. (2002). Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cognitive, affective & behavioral neuroscience*, 2(2), 121–40.

Jezzard, P., Matthews, P. M., & Smith, S. M. (Eds.). (2001). *Functional MRI : An Introduction to Methods*. Oxford University Press.

Jones, D., Farrand, P., Stuart, G., & Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of experimental psychology. Learning, memory, and cognition*, 21(4), 1008–18.

Jonides, J., Lacey, S. C., & Nee, D. E. (2005). Processes of Working Memory in Mind and Brain. *Society*, 14(1), 2–5.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. a, Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual review of psychology*, 59, 193–224. 4

Jäncke, L. (2008). Music, memory and emotion. *Journal of biology*, 7(6), 21.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychonomic bulletin & review*, 9(4), 637–71.

Ketter, T. A., Andreason, P. J., George, M. S., Lee, C., Gill, D. S., Parekh, P. I., Willis, M. W., et al. (1996). Anterior Paralimbic Mediation of Procaine-Induced Emotional and Psychosensory Experiences. *Archives of General Psychiatry*, 53(1), 59.

Knight, R. ., & D'Esposito, M. (2003). Lateral prefrontal syndrom: a disorder of executive control. In D'Esposito (Ed.), *Neurological Foundations of Cognitive Neuroscience* (pp. 259–279). MIT Press.

Koch, G., Oliveri, M., Torriero, S., Carlesimo, G., Turriziani, P., & Caltagirone, C. (2005). rTMS evidence of different delay and decision processes in a fronto-parietal neuronal network activated during spatial work-ing memory. *Neuroimage*, 24, 34–39.

Koelsch S, Gunter TC, von Cramon DY, Zysset S, Lohmann G, F. A. (2002). Bach speaks: a cortical 'language-network' serves the processing of music. *Neuroimage*, 17, 956–966.

Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., & Gruber, O. (2009). Functional architecture of verbal and tonal working memory: an FMRI study. *Human brain mapping*, 30(3), 859–873.

Koelsch, S., & Siebel, W. a. (2005). Towards a neural basis of music perception. *Trends in cognitive sciences*, 9(12), 578–84.

Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 27(2), 308–14.

Lancaster, J. L., Rainey, L. H., Summerlin, J. L., Freitas, C. S., Fox, P. T., Evans, A. C., Toga, A. W., et al. (1997). Automated labeling of the human brain: a preliminary report on the development and evaluation of a forward-transform method. *Human brain mapping*, 5(4), 238–42.

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., Kochunov, P. V, et al. (2000). Automated Talairach atlas labels for functional brain mapping. *Human brain mapping*, 10(3), 120–31.

Lartillot, O., & Toiviainen, P. (2007). A Matlab Toolbox for Musical Feature Extraction From Audio. *International Conference on Digital Audio Effects, Bordeaux*.

REFERENCES

Laurienti, P. J., Burdette, J. H., Wallace, M. T., Yen, Y.-F., Field, A. S., & Stein, B. E. (2002). Deactivation of sensory-specific cortex by cross-modal stimuli. *Journal of Cognitive Neuroscience*, 14, 420–429.

Lazar, N. A. (2008). *The statistical analysis of functional MRI data*. Springer, New York.

Ledberg, A., Åkerman, S., & Roland, P. E. (1998). Estimation of the Probabilities of 3D Clusters in Functional Brain Images, 128(8), 113–128.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: The MIT Press.

Leung, A. W. S., & Alain, C. (2011). Working memory load modulates the auditory "What" and "Where" neural networks. *NeuroImage*, 55(3), 1260–9.

Levitin, D J, & Tirovolas, A. K. (2009). Current advances in the cognitive neuroscience of music. *Annals of the New York Academy of Sciences*, 1156(The Year in Cognitive Neuroscience 2009), 211–231.

Levitin, D. & Menon, V. (2003). Musical structure is processed in "language" areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. *NeuroImage*, 20(4), 2142–52.

Lindquist, M. (2008). The Statistical Analysis of fMRI Data. *Statistical Science*, 23(4), 439–464.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, 453(7197), 869–78.

Mathiak, K., Hertrich, I., Grodd, W., Ackermann, H. (2004). Discrimination of temporal information at the cerebellum: functional magnetic resonance imaging of nonverbal auditory memory. *Neuroimage*, 21, 154–162.

McGilchrist, I. (2010). *The Master and His Emissary: The Divided Brain and the Making of the Western World*. Yale University Press.

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain structure & function*, 214(5-6), 655–67.

Miessler, G. L., & Tarr, D. A. (2004). *Inorganic Chemistry. 3rd Ed.* Pearson/Prentice Hall publisher.

Miller, G., Galanter, E., & Pribram, K. (1960). *Plans and structure of behavior*. New York, NY: Holt, Rinehart & Winston.

Miller, G. A. (1956). The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychological Review*.

Milner, B., Squire, L. R., & Kandel, E. R. (1998). Cognitive Neuroscience and the Study of Memory. *Cell*, 20, 445–468.

Miyake, A., & Shah, P. (1999). *Models of working memory: mechanisms of active maintenance and executive control*. Cambridge University Press.

Monti, M. M. (2011). Statistical Analysis of fMRI Time-Series: A Critical Review of the GLM Approach. *Frontiers in human neuroscience*, 5(March), 28.

Moridani, M. K. (2009). Functional brain imaging with use of a new and powerful neuroimaging technique, (June), 173–176.

Mottaghy, F. (2006). Interfering with working memory in humans. *Neuroscience*, 139(1), 85–90.

Mottaghy, F., Pascual-Leone, A., Kemna, L., Töpper, R., Herzog, H., Muller-Gartner, H., & Krause, B. (2003). Modulation of a brain- behavior relationship in verbal working memory by rTMS. *Brain Res Cogn Brain Res*, 15, 241–249.

Münte, T. F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature reviews. Neuroscience*, 3(6), 473–8.

Nan, Y., Knösche, T. R., Zysset, S., & Friederici, A. D. (2008). Cross-cultural music phrase processing: an fMRI study. *Human brain mapping*, 29(3), 312–28.

Nichols, E. A., Kao, Y., Verfaellie, M., & Gabrieli, J. D. E. (2006). Working Memory and Long-Term Memory for Faces : Evidence From fMRI and Global Amnesia for Involvement of the Medial Temporal Lobes. *Hippocampus*, 616, 604–616.

Oliveri, M., Caltagirone, C., Filippi, M., Traversa, R., Cicinelli, P., Pasqualetti, P., & Rossini, P. (2000). Paired transcranial magnetic stimulation protocols reveal a pattern of inhibition and facilitation in the human parietal cortex. *J Physiol*, 529, 461–468.

REFERENCES

Osaka, N., Osaka, M., Kondo, H., Morishita, M., Fukuyama, H., & Shibasaki, H. (2004). The neural basis of executive function in working memory: an fMRI study based on individual differences. *NeuroImage*, 21, 623–31.

Owen, A. M. (1997). The Functional Organization of Working Memory Processes Within Human Lateral Frontal Cortex: The Contribution of Functional Neuroimaging. *Neuroscience*, 9, 1329–1339.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Human brain mapping*, 25(1), 46–59.

Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the Basal Ganglia. *Annual review of neuroscience*, 25, 563–93.

Pallesen, K. J., Brattico, E., Bailey, C. J., Korvenoja, A., Koivisto, J., Gjedde, A., & Carlson, S. (2010). Cognitive control in auditory working memory is enhanced in musicians. (E. Warrant, Ed.)*PloS one*, 5(6).

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature neuroscience*, 6(7), 674–81.

Patel, A.D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P.J. (1998). Processing syntactic relations in language and music: an event-related potential study. *J. Cogn. Neurosci*, 10, 717–733.

Paulesu, E., Frith, C. D., & Frackowiak, R. S. (1993). The neural correlates of the verbal component of working memory. *Nature*, 362(6418), 342–5.

Pereira, C. S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S. L., & Brattico, E. (2011). Music and Emotions in the Brain: Familiarity Matters. (J. Pillai, Ed.)*PLoS ONE*, 6(11), e27241.

Peretz, I., & Zatorre, R. J. (2003). *The cognitive neuroscience of music*. Oxford: Oxford University Press.

Petrides, M. (1996). Specialized systems for the processing of mnemonic information within the primate frontal cortex. *Phil. Trans. R. Soc. B*, 351, 1455–1462.

Petrides, M., Alivisatos, B., & Evans, A. C. (1995). Functional activation of the human ventrolateral frontal cortex during mnemonic retrieval of verbal information. *Proc. Natl Acad. Sci. USA*, 92, 5803–5807.

Petrides, M. (2005). Lateral prefrontal cortex: architectonic and functional organization. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 360(1456), 781–95.

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, 139(1), 23–38.

Pyper, B. J., & Peterman, R. M. (1998). Comparison of methods to account for autocorrelation in correlation analyses of fish data, 2140, 2127–2140.

Ravizza, S. M., Delgado, M. R., Chein, J. M., Becker, J. T., & Fiez, J. A. (2004). Functional dissociations within the inferior parietal cortex in verbal working memory. *NeuroImage*, 22(2), 562–573.

Rogalsky, C., Matchin, W., & Hickok, G. (2008). Broca's area, sentence comprehension, and working memory: an fMRI Study. *Frontiers in human neuroscience*, 2(October), 14.

Ruchkin, D. S., Grafman, J., Cameron, K., & Berndt, R. S. (2003). Working memory retention systems: a state of activated long-term memory. *The Behavioral and brain sciences*, 26(6), 709–28; discussion 728–77.

Sacks, O. (2007). *Musicophilia: Tales of Music and the Brain* (p. 425). Picador.

Salamé, P., & Baddeley, A. (1989). Effects of background music on phonological short-term memory. *The Quarterly Journal of Experimental Psychology A Human Experimental Psychology*, 41(1), 107–122.

Sarnthein, J., Petsche, H., Rappelsberger, P., Shaw, G., & Stein, A. von. (1998). Synchronization between prefrontal cortex during human working memory. *Proc Natl Acad Sci USA*, 95, 7092–7096.

Schacter, D., & Tulving, E. (1994). *Memory Systems*. Cambridge, MA: MIT Press.

Schlaug, G. (2006). The Brain of Musicians. *Annals of the New York Academy of Sciences*, 930(1), 281–299.

Schulze, K., & Koelsch, S. (2012). Working memory for speech and music. *Annals of the New York Academy of Sciences*, 1252(1), 229–36.

Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., & Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Human brain mapping*, 32(5), 771–83.

Shaw, C., McEachern, C., & Eachern, J. (2001). *Toward a Theory of Neuroplasticity* (p. 479). Psychology Press.

Silverman, B. (1986). *Density estimation for statistics and data analysis*. London ;;New York: Chapman and Hall.

REFERENCES

Snyder, B. (2000). *Music and memory : an introduction*. Cambridge Mass.: MIT Press.

Snyder, B. (2009). Memory for music. *The Oxford handbook of music psychology* (pp. 107–121). Oxford: Oxford University Press.

Squire, L., & Butters, N. (Eds.). (1984). *The neuropsychology of memory*. New York: Guilford.

Squire, Larry, & Kandel, E. (2008). *Memory: From Mind to Molecules* (p. 256). Roberts and Company Publishers.

Sridharan, D., Levitin, D. J., Chafe, C. H., Berger, J., & Menon, V. (2007). Neural Dynamics of Event Segmentation in Music : Converging Evidence for Dissociable Ventral and Dorsal Networks. *Neuron*, 521–532.

Steinbeis, N., & Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex*, 18, 1169–1178.

Stern, C., Owen, A., Tracey, I., Look, R., Rosen, B., & Petrides, M. (2000). Activity in ventrolateral and mid-dorsolateral prefrontal cortex during nonspatial visual working memory processing: evidence from functional magnetic resonance imaging. *NeuroImage*, *11(5, Pt.*, 392–399.

Sternberg, S. (1969). Memory scanning: mental processes revealed by reaction time experiments. *American Scientist*, 57, 421–457.

Tallon-Baudry, C., Bertrand, O., Peronet, F., & Pernier, J. (1998). Induced γ-band activity during the delay of a visual short-term memory task in humans. *J Neurosci*, 18, 4244–54.

The MNI brain and the Talairach atlas. (2012). *CBU Imaging Wiki*. Retrieved November 30, 2012, from http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach

VanElzakker, M., Fevurly, R. D., Breindel, T, & Spencer, R. L. (2008). Environmental novelty is associated with a selective increase in Fos expression in the output elements of the hippocampal formation and the perirhinal cortex. *Learning & memory (Cold Spring Harbor, N.Y.)*, 15(12), 899–908.

Vollmer-Haase, J., Finke, K., Hartje, W., & Bulla-Hellwig, M. (1998). Hemispheric dominance in the processing of J. S. Bach fugues: a transcranial Doppler sonography (TCD) study with musicians. *Neuropsychologia*, 36(9), 857–867.

Wager, T. D., & Smith, E. E. (2003). Neuroimaging studies of working memory: a meta-analysis. *Cognitive, affective & behavioral neuroscience*, 3(4), 255–74.

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences*, 7(12), 553–559.

Weisstein, E. W. (n.d.). Affine Transformation. From MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/AffineTransformation.html. Wolfram Research, Inc.

Wilson, F., O'Scalaidhe, S., & Goldman-Rakic, P. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science*, 260, 1955–1958.

Wilson, R. A., & Keil, F. C. (Eds.). (2001). *The MIT encyclopedia of the cognitive sciences*. MIT Press.

Yonelinas, A. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46, 441–517.

Zatorre, R. J, Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience*, 8(7), 547–558.

Zatorre, R. J, Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *Journal of Neuroscience*, 14(4), 1908.

Zatorre, R. J., & Samson, S. (1991). Role of the right temporal neocortex in retention of pitch in auditory short-term memory. *Brain*, 114(6), 2403–2417.

APPENDIX A

KDE - KERNEL DENSITY ESTIMATION

Given a random sample x_1, x_2,..., x_n, i=1, 2,..., n, drawn from an distribution with unknown density $f(x)$, the kernel density estimator of f is defined by the function $\hat{f}_h(x)$ as follows:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

Eq.9

where K is called the kernel, an estimator of the density in x by counting the frequency of other points X_i in the same bin as x and weighting them differently depending on their distance from x. The density value of f in x from points X_i depends on how much closer X_i points are to x (the closer to x, the more weight they have), a property fulfilled by kernel functions (usually a PDF that integrates to 1 and is nonnegative); The factor of $1/h$ is so that $\hat{f}_h$ will integrate to 1; and $h > 0$ is the bandwidth of the kernel (Biba et al., 2007).

MATLAB CODE

```
% KDE for the perceptual data, where x is the vector of density values
% evaluated at the points in xi. 478000 is the number of equally points
% (miliseconds) that cover the range of the data, and the width of the
% kernel is 1000 ms.

[x,xi]=ksdensity(data, 0:100:478000,'width', 1000)

% THRESHOLDING & PEAK FINDING

% Obtains the indices in the data set x, where x reaches a peak, i.e., a
local
% maximum. Peaks are found by testing for a negative second derivative at
each   % of the values
ind=find(diff(sign(diff(x)))==-2);

% density threshold=23*10⁻⁷ that retrieves 64 peaks
kde=ind(find(x(ind)>0.0000023));

% obtains segmentation points (peaks) in seconds (rounded-up)
kde=round(kde1/10);
```

APPENDIX B

CLUSTER-SIZE ESTIMATION

In this method by Ledberg et al. (1998), we use the brain statistical images (signal-SI), obtained from the correlational analysis between the WM regressor and participants' brain signal, to generate noise statistical images (noise-SI) that are similar to the signal-SIs in spatial spectral properties except they contain no signal of interest. To generate noise-SIs we followed the random-phase approach in Ebisuzaki (1997; see APPENDIX C). The idea behind is to derive a confidence interval from random series generated using to the original series, and having, therefore, similar properties.

Generation of simulated noise images

We compute the ACF of these noise-SIs as a measurement of spatial dependency in order to generate noise images with the same spatial correlation as the noise-SI (the ACF is not estimated in the signal-SI because the presence of a signal might affect the spatial dependency in the SI). These noise-SIs are transformed into pseudo-normal distributed images, and their ACF is used in turn to generate normally distributed images. To generate images with the same ACF as the pseudo-normal noise-SIs, first an image U of uncorrelated normally distributed random variables is generated. Then we convolve this image with a filter kernel K giving an image S with the desired ACF as follows:

$$S(u) = U(u) * K(u)$$
<div align="right">Eq. 10</div>

To generate the filter kernel K, Ledberg et. al (1998) used the convolution and correlation theorems of Fourier theory, from which following expression is derived for K:

$$K = IFT|FT(P)|$$
<div align="right">Eq. 11</div>

where *IFT and FT* denote the inverse Fourier transform and the Fourier transform respectively, and P represents the pseudo-normal transformed noise-SI.

S is then normalized to zero mean and unit variance, so now it is a normally distributed image that has the same ACF as the pseudo normal transformed noise-SI. Variance and bias of the ACF might influence K, and an error in K leads to errors in the estimation of probability distribution of cluster sizes. Therefore it is crucial to ensure that S has the same ACF that the original noise-SIs. Thus K was estimated as the mean of several ACF kernels obtained in a Monte Carlo simulation (number of runs=110 times [10 per participant]).

Estimation of the distribution of the cluster size statistic

A cluster is defined to be a set of above-threshold connected voxels ("connected" meaning having at least one side in common [6-connectivity]). We aim at estimating the probability that a cluster of size ≥ v (cluster size [CS] is the number of voxels it contains) has occurred by chance. The probability distribution of cluster sizes above a given threshold is derived from the 100 generated normally distributed random images using Eq. 10, and subsequently used to estimate probabilities for the clusters detected in the statistical images generated by testing the hypothesis.

To estimate the distribution of the cluster sizes we generated 100 Noise-SIs using the equation above. As a result from all the noise-SIs, the distribution of cluster sizes of 8 voxels is obtained for p=.0005 (Z=3.48).
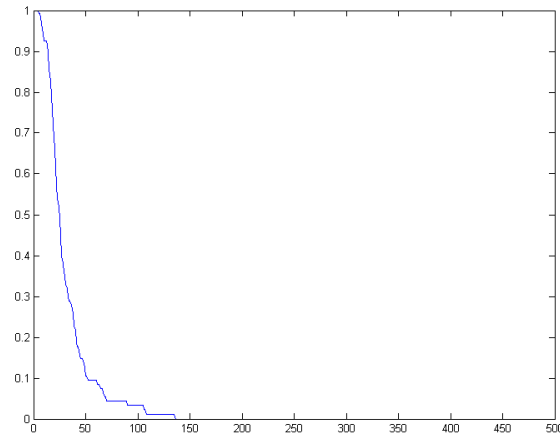
Figure 19. Critical cluster threshold distribution at p=.0005. The x axis represents the number of voxels and y axis the probability it occurs at the given threshold.

APPENDIX C

RANDOM-PHASE TEST

The random-phase test is a nonparametric test which attempts to create random time series that could have come from PA, the population from which A is a member. The statistical test is based on generating a large number of random series with the same power spectra as the first series but with random phases in the Fourier modes. Specifically, the participant's brain signal is correlated with the phase-shifted WM regressor, which produces a randomly generated image identical to the signal-SI in spatial spectral properties except it shows no stimulus-dependent activations. It is important therefore to preserve the fundamental properties of the original series. Ebisuzaki's procedure is a way of "resampling" in the frequency domain, which preserves the power spectrum and not the distribution of values of the original series. Thus the resampled series retain the same autocorrelation as the original series.

The procedure, described as follows (Ebisuzaki, 1997), was applied to the WM regressor for further correlation with the participant's brain signal to generate the noise-SIs:

1. The discrete Fourier transform (DFT) of the WM regressor $(\tilde{a}_i)$ is computed by Eq.12, where $\delta_k = 0$ except for $k = 0$, $N/2$ (even N) in which case $\delta_k = 1$:

$$\tilde{a}_k = \frac{2 - \delta_k}{N} \sum_{j=0}^{N-1} a_j e^{2\pi jk/N} \qquad \text{Eq.12}$$

2. Let $\tilde{r}_0 = 0, \tilde{r}_k = |\tilde{a}_k| \exp(i\theta_k)$ for $0 < k < N/2$, and $\tilde{r}_{N/2} = 2^{1/2} |\tilde{a}_{N/2}| \cos(\theta_{N/2})$ for even N, where θ_k is a uniform random variable from $[0, 2\pi)$. Here we create a Fourier series with the original (WM) series' power spectrum but with random phases.

3. Finally we compute the IFT of $\tilde{r}_k$ (see Eq.13), which gives r_i a random series with the same autocorrelation as the WM series. The power spectrum of the randomly generated and WM original time series is the same, except the phases of r_i are random.

$$r_j = Re \sum_{k=0}^{n} \tilde{r}_k e^{-2\pi ijk/N} \qquad \text{Eq.13}$$

where $n = N/2$ or $\frac{N-1}{2}$ for even and odd N respectively.

APPENDIX D

SUBTRACTION (t-test) BETWEEN ACTIVATION MAPS

A Student's t-test was performed to subtract one map to the other. By using a t-test the difference in means is weighted by the standard deviation in the significant voxels from the two maps being contrasted, thus assigning high t-values to large differences with small standard deviations, and low t-values to small differences with large standard deviations. The t-score is calculated per voxel time series X across participants, where subscripts 1 and 2 represent the two maps being compared, using the formula:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}}$$

Eq.14

where

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

Eq.15

and s_p^2 is the pooled variance

$$s_p^2 = \frac{\sum(X_1 - \bar{X}_1)^2 + \sum(X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2}$$

Eq.16

The map of p-values cannot be used in the t-test, because p-values are not normally distributed so they violate the basic assumption of parametric statistics. Therefore the t-test was performed on the participants' z-score maps. Specifically, since we aimed at finding significantly different positive and negative z-scores, z-values were entered into a right- and a left-tailed t-tests for significantly higher and significantly lower z-values (alpha=0.05) respectively, one voxel at a time, which yielded a t-statistic per voxel. Thus a map was produced that revealed significantly higher or lower activations for the original maps compared. However, this map includes significant differences below the threshold of interest (p=.0005) that need to be filtered out.

To do this, we took the pooled z-value map for the condition of interest thresholded at a significant level p=.0005 with a corrected CS threshold of 8 (as estimated using APPENDIX B). As a result, we obtained a binary map indicating voxels for which the pooled z-value is significant only for the condition of interest, which can be used as a mask to extract the significant values from the resulting t-statistic map.

In short, the idea behind the whole map subtraction procedure is to leave the significant z-scores in the map of interest while removing the non-significant overlap between them.