AUS DER KLINIK FÜR PSYCHIATRIE UND PSYCHOTHERAPIE

GESCHÄFTSFÜHRENDER DIREKTOR: TILO KIRCHER

DES FACHBEREICHS MEDIZIN DER PHILIPPS-UNIVERSITÄT MARBURG

The neurobiology of cortical music representations

Inaugural-Dissertation zur Erlangung des Doktorgrades der gesamten Humanmedizin
bzw. Zahnmedizin, Humanbiologie oder Naturwissenschaften, Medizinwissenschaften
dem Fachbereich Medizin der Philipps-Universität Marburg
vorgelegt von

Peer Herholz aus Friedrichroda

Marburg, 2019

Angenommen vom Fachbereich Medizin der Philipps-Universität Marburg am 21.02.2019.

Gedruckt mit Genehmigung des Fachbereichs

Dekan: Univ.-Prof. Dr. Helmut Schäfer
Referent: Univ.-Prof. Dr. Andreas Jansen
1. Korreferent: Univ.-Prof. Dr. Arne Nagels

This final print version of this thesis was adapted from the version accepted by the examination committee by means of a layout change. The content, results, etc. of this thesis were not changed during this process. The examination committee as well as the examination office gave their consent to these changes. The layout used is an adapted version of the LaTeX dissertation template of the TU Delft. The author was granted permission to use this template for the here presented thesis.

“If you’re gonna be dumb, you gotta be tough.”

Roger Alan Wade

for Verena & Rosa

CONTENTS

SUMMARY

Music is undeniable one of humanity's defining traits, as it has been documented since the earliest days of mankind, is present in all know cultures and perceivable by all humans nearly alike. Intrigued by its omnipresence, researchers of all disciplines started the investigation of music's mystical relationship and tremendous significance to humankind already several hundred years ago. Since comparably recently, the immense advancement of neuroscientific methods also enabled the examination of cognitive processes related to the processing of music. Within this neuroscience of music, the vast majority of research work focused on how music, as an auditory stimulus, reaches the brain and how it is initially processed, as well as on the tremendous effects it has on and can evoke through the human brain. However, intermediate steps, that is how the human brain achieves a transformation of incoming signals to a seemingly specialized and abstract representation of music have received less attention. Aiming to address this gap, the here presented thesis targeted these transformations, their possibly underlying processes and how both could potentially be explained through computational models. To this end, four projects were conducted. The first two comprised the creation and implementation of two open source toolboxes to first, tackle problems inherent to auditory neuroscience, thus also affecting neuroscientific music research and second, provide the basis for further advancements through standardization and automation. More precisely, this entailed deteriorated hearing thresholds and abilities in MRI settings and the aggravated localization and parcellation of the human auditory cortex as the core structure involved in auditory processing. The third project focused on the human's brain apparent tuning to music by investigating functional and organizational principles of the auditory cortex and network with regard to the processing of different auditory categories of comparable social importance, more precisely if the perception of music evokes a is distinct and specialized pattern. In order to provide an in depth characterization of the respective patterns, both the segregation and integration of auditory cortex regions was examined. In the fourth and final project, a highly multimodal approach that included fMRI, EEG, behavior and models of varying complexity was utilized to evaluate how the aforementioned music representations are generated along the cortical hierarchy of auditory processing and how they are influenced by bottom-up and top-down processes. The results of project 1 and 2 demonstrated the necessity for the further advancement of MRI settings and definition of working models of the auditory cortex, as hearing thresholds and abilities seem to vary as a function of the used data acquisition protocol and the localization and parcellation of the human auditory cortex diverges drastically based on the approach it is based one. Project 3 revealed that the human brain apparently is indeed tuned for music by means of a specialized representation, as it evoked a bilateral network with a right hemispheric weight that was not observed for the other included categories. The result of this specialized and hierarchical recruitment of anterior and posterior auditory cortex regions was an abstract music component that is situated in anterior regions of the superior temporal gyrus and preferably encodes music, regardless of sung or instrumental. The outcomes of project 4 indicated that even though the entire auditory cortex, again with a right hemispheric weight, is involved in the complex processing of music in particular, anterior regions yielded an abstract representation that varied excessively over time and could not sufficiently explained by any of the tested models. The specialized and abstract properties of this representation was furthermore underlined by the predictive ability of the tested models, as models that were either based on high level features such as behavioral representations and concepts or complex acoustic features always outperformed models based on single or simpler acoustic features. Additionally, factors know to influence auditory and thus music processing, like musical training apparently did not alter the observed representations. Together, the results of the projects suggest that the specialized and stable cortical representation of music is the outcome of sophisticated transformations of incoming sound signals along the cortical hierarchy of auditory processing that generate a music component in anterior regions of the superior temporal gyrus by means of top-down processes that interact with acoustic features, guiding their processing.

ZUSAMMENFASSUNG

Musik ist unbestreitbarer Weise eine der definierenden Eigenschaften des Menschen. Dokumentiert seit den frühesten Tagen der Menschheit und in allen bekannten Kulturen vorhanden, ist sie von allen Menschen nahezu gleich wahrnehmbar. Fasziniert von ihrer Omnipräsenz haben Wissenschaftler aller Disziplinen vor einigen hundert Jahren begonnen die mystische Beziehung zwischen Musik und Mensch, sowie ihre enorme Bedeutung für selbigen zu untersuchen. Seit einem vergleichsweise kurzem Zeitraum ist es durch den immensen Fortschritt neurowissenschafticher Methoden auch möglich die kognitiven Prozesse, welche an der Verarbeitung von Musik beteiligt, sind zu untersuchen. Innerhalb dieser Neurowissenschaft der Musik hat sich ein Großteil der Forschungsarbeit darauf konzentriert wie Musik, als auditorischer Stimulus, das menschliche Gehirn erreicht und wie sie initial verarbeitet wird, als auch welche kolossallen Effekte sie auf selbiges hat und auch dadurch bewirken kann. Jedoch haben die Zwischenschritte, also wie das menschliche Gehirn eintreffende Signale in eine scheinbar spezialisierte und abstrakte Repräsentation von Musik umwandelt, vergleichsweise wenig Aufmerksamkeit erhalten. Um die dadurch entstandene Lücke zu adressieren, hat die hier vorliegende Dissertation diese Prozesse und wie selbige durch Modelle erklärt werden können in vier Projekten untersucht. Die ersten beiden Projekte beinhalteten die Herstellung und Implementierung von zwei Toolboxen um erstens, inhärente Probleme der auditorischen Neurowissenschaft, daher auch neurowissenschaftlicher Untersuchungen von Musik, zu verbessern und zweitens, eine Basis für weitere Fortschritte durch Standardisierung und Automatisierung zu schaffen. Im genaueren umfasste dies die stark beeinträchtigten Hörschwellen und –fähigkeiten in MRT-Untersuchungen und die erschwerte Lokalisation und Parzellierung des menschlichen auditorischen Kortex als Kernstruktur auditiver Verarbeitung. Das dritte Projekt befasste sich mit der augenscheinlichen Spezialisierung von Musik im menschlichen Gehirn durch die Untersuchung funktionaler und organisatorischer Prinzipien des auditorischen Kortex und Netzwerks bezüglich der Verarbeitung verschiedener auditorischer Kategorien vergleichbarer sozialer Bedeutung, im genaueren ob die Wahrnehmung von Musik ein distinktes und spezialisiertes neuronalen Muster hervorruft. Um eine ausführliche Charakterisierung der entsprechenden neuronalen Muster zu ermöglichen wurde die Segregation und Integration der Regionen des auditorischen Kortex untersucht. Im vierten und letzten Projekt wurde ein hoch multimodaler Ansatz, welcher fMRT, EEG, Verhalten und Modelle verschiedener Komplexität beinhaltete, genutzt, um zu evaluieren, wie die zuvor genannten Repräsentationen von Musik entlang der kortikalen Hierarchie der auditorischen Verarbeitung generiert und wie sie möglicherweise durch Bottom-up- und Top-down-Ansätze beeinflusst werden. Die Ergebnisse von Projekt 1 und 2 demonstrierten die Notwendigkeit für weitere Verbesserungen von MRT-Untersuchungen und die Definition eines Funktionsmodells des auditorischen Kortex, da Hör-

schwellen und –fähigkeiten stark in Abhängigkeit der verwendeten Datenerwerbsprotokolle variierten und die Lokalisation, sowie Parzellierung des menschlichen auditorischen Kortex basierend auf den zugrundeliegenden Ansätzen drastisch divergiert. Projekt 3 zeigte, dass das menschliche Gehirn tatsächlich eine spezialisierte Repräsentation von Musik enthält, da selbige als einzige auditorische Kategorie ein bilaterales Netzwerk mit rechtshemisphärischer Gewichtung evozierte. Aus diesem Netzwerk, welches die Rekrutierung anteriorer und posteriorer Teile des auditorischen Kortex beinhaltete, resultierte eine scheinbar abstrakte Repräsentation von Musik in anterioren Regionen des Gyrus temporalis superior, welche präferiert Musik enkodiert, ungeachtet ob gesungen oder instrumental. Die Resultate von Projekt 4 deuten darauf hin, dass der gesamte auditorische Kortex, erneut mit rechtshemisphärischer Gewichtung, an der komplexen Verarbeitung von Musik beteiligt ist, besonders aber anteriore Regionen, die bereits genannten abstrakte Repräsentation hervorrufen, welche sich exzessiv über die Zeitdauer der Wahrnehmung verändert und nicht hinreichend durch eines der getesteten Modelle erklärt werden kann. Die spezialisierten und abstrakten Eigenschaften dieser Repräsentationen wurden weiterhin durch die prädiktiven Fähigkeiten der getesteten Modelle unterstrichen, da Modelle, welche entweder auf höheren Eigenschaften wie Verhaltensrepräsentationen und mentalen Konzepten oder komplexen akustischen Eigenschaften basierten, stets Modelle, welche auf niederen Attributen wie simplen akustischen Eigenschaften basierten, übertrafen. Zusätzlich konnte kein Effekt von Faktoren, wie z.B. musikalisches Training, welche bekanntermaßen auditorische und daher Musikverarbeitung beeinflussen, nachgewiesen werden.

Zusammengefasst deuten die Ergebnisse der Projekte darauf, hin dass die spezialisierte und stabile kortikale Repräsentation von Musik ein Resultat komplexer Prozesse ist, welche eintreffende Signale entlang der kortikalen Hierarchie auditorischer Verarbeitung in eine abstrakte Repräsentation von Musik innerhalb anteriorer Regionen des Gyrus temporalis superior durch Top-Down-Prozesse, welche mit akustischen Eigenschaften interagieren und deren Verarbeitung steuern, umwandeln.

ABBREVIATIONS

Acoustic scanner noise – ASN

Anterior superior temporal gyrus – STGa

A-weighted decibels – dBA

Brodmann area - BA

Data acquisition protocol - DAP

Echo-planar imaging – EPI

Electroencephalogram - EEG

Fast Fourier Transformation – FFT

Field of view - FoV

Frequency spectrum - FS

Hearing threshold – HT

Hertz - Hz

Heschl's gyrus – HG

Kilohertz - kHz

Magnetization-prepared rapid acquisition gradient echo – MPRAGE

Magnetic resonance imaging - MRI

Planum polare – PP

Planum temporal – PT

Permission exposure limit - PEL

Posterior superior temporal gyrus – STGp

Population receptive field - pRF

Primary auditory cortex – PAC

Pure tone audiometry - PTA

Radio frequency – RF

Region of interest - ROI

Representational dissimilarity matrix – RDM

Representational similarity analysis – RSA

Stimulus onset asynchrony - SOA

Superior temporal gyrus – STG

Superior temporal plane – STP

Time of echo - TE

Time of repetition - TR

LIST OF FIGURES

LIST OF TABLES

1

GENERAL INTRODUCTION

The first chapter of this dissertation acts as an introduction to the diverse topic of cortical music representations, beginning with how music is defined, understood and utilized in the thesis at hand. Subsequently, an overview of the neuronal processing of music as an auditory percept will be provided, followed by a broad summary of previous research on music and the (human) brain. Lastly, the driving research question of this dissertation is outlined by incorporating the insights of the above mentioned respective parts.

A wave. This is how it all starts. A wave one cannot see but hear. A sound wave. A wave that through an infinite number of neurons, evolution and development is transformed to something that cannot really be explained. Something that is tremendously powerful and can evoke everything from ultimate fear to utmost joy. Something that can heal. Something that shapes the world. That something is music. Dating back at least to the earliest documented days of humankind (Mehr, Singh, York, Glowacki, & Krasnow, 2018; Zatorre & Salimpoor, 2013), music has been crucial to humanity since several thousands of years. Notably, to this day, there is no documented society or culture that does not have music, independent of the continent, country or level of industrialization (Mehr et al., 2018). Despite all these underlying differences, the way music is produced, conserved and utilized is remarkably similar across the entire world (Mehr et al., 2018). Intrigued by its prominence, researchers of nearly all scientific disciplines pursued the quest to unravel, at least in parts, some of its mystery in order to comprehend the significant role it exhibits. Even though an eminent shift from research that was mainly shaped by historical, ethnological and musicological aspects towards the inclusion of biological facets already started with Charles Darwin's "The Descent of Man" (1871) and Hermann von Helmholtz' "On the Sensations of Tone as a Physiological Basis for the Theory of Music" (1954), the neurobiological basis of music processing has gained immense interest especially during the last twenty years due to the tremendous advancement of neurobiological, that is neuroscientific methods. Measurements of brain physiology like Functional magnetic resonance imaging (fMRI) and Electroencephalography (EEG) enabled scientist to capture a glimpse of the extensive cognitive computations that are involved in music processing. In roughly 1,480,000 [1] articles scientists investigated the relationship between music and the human brain. While this is indeed a colossal number of research work, it appears rather small given the high dimensionality and complexity of both music and the human brain. As music itself, its processing by the human brain is characterized by spatial and temporal properties within which small particles are combined to something greater that is more than just the sum of its parts. Like notes that lead to melodies and subsequently musical pieces, neurons lead to brain regions and networks. Thus depicting a processing spectrum, the investigation of neuronal correlates of music as a perceivable phenomena so far vastly centered around both ends of this spectrum: the encoding of musical features and higher order cognition related to music (structure, emotion, etc.). The missing link, how music is represented on an intermediate processing stage between feature driven bottom-up and higher order top-down cognitive processes is however less understood and in accordance with its name giving analogy from evolutionary biology most likely non-linear (Krause, 2012). The thereout arising question how the perception of music is achieved through the aforementioned transformation of sound waves that in turn lead to a seemingly specialized and stable percept that subsequently extends purely auditory domains was at the very core of the here presented thesis. In order to provide the reader with the information necessary to evaluate and situate the particularized projects of this thesis, the latter are framed by an overall introduction that establishes core theories and principles of music, its neuronal

---

[1] Number of articles listed on google scholar yielded through the search term "music brain" on December 19, 2018.

processing, as well as an overall discussion that will connect the different projects with respect to previous and future research. Serving as a rather broad background, detailed introductions and discussions will be provided in the respective chapters. Everything aiming to shed some light on how waves are transformed into one of humanity's most defining characteristic: music.

1.1. MUSIC - AN ATTEMPT OF AN IMPOSSIBLE DEFINITION

The first problem that is at the very beginning of any scientific investigation that focuses on music, is the shear impossibility to define it. As definitions and descriptions of music are at least as diverse and complex as music itself, an operational definition for a given research work is inevitable. With the thesis at hand being no exception to that, this chapter will provide a brief overview of music's properties and how it is characterized within the here presented studies. Ironically, the omnipresence and diversity of music combined with the ability of the majority of humans of all ages and cultures to perceive and process it, is the prevailing reason for its almost impossible definition. However, these factors also capture the dimensions that were identified to be necessary for a sufficient characterization and include the following, highly interacting components: elements of music, culture and subjective experience ("Definition of music", n.d.). Concerning the first, music is, most basically, the result of sound, where sound is both an audible wave that propagates by means of oscillations through substances and the perception of said waves through neuronal mechanisms ("ANSI/ASA S1.1-2013", 2018). It is important to note, that already this circumstance is highly debated, as humans have the ability to imagine music and therefore seemingly do not require a physiological sensation based on a physical stimulus (Halpern, 2001; Zatorre, Halpern, Perry, Meyer, & Evans, 1996). One common trait of sound assumed to be defining for music is its structure and organization, while some concepts of music certainly characterize unstructured sound as noise. However, what "structured" constitutes is not precisely defined. The aforementioned perception is often described through a varying number of characteristics, that are, specifically for music, termed elements of music or music features and can include pitch, melody, harmony, meter, tempo, rhythm, texture and timbre among many others (L. B. Meyer, 1973, 1989; L. B. Meyer, Narmour, & Solie, 1988) . Pitch refers to the audible difference between sounds, often tones, based on their frequencies, in the sense that they are higher or lower compared to one another on a certain scale (Klapuri & Davy, 2007). How such tones are concatenated over time is described and defined through melody, which follows certain rules with regard to the frequency scale tones can originate from, as well as how they are proceeding over time and structured into segments, called motifs. These governing rules are often referred to as musical syntax (Koelsch, 2009a; Patel, 2008) and furthermore extent to principles of how tones can be combined at the same time to create a chord, that is their harmony. Therefore, harmony and melody constitute vertical and horizontal facets of music, especially pitch (Jamini, 2005). These facets are additionally distinguished by the musical element rhythm comprising also e.g., meter and tempo. Here, meter indicates the successive alternation of tones and pauses on the basis of beats with tempo illustrating the number of beats per minute, thus the speed of a musical piece. The combination of the aforementioned elements of music leads to the perception of texture and timbre. While texture refers to the complexity and richness as provided by the range and amount of parts found in a given musical piece (e.g., the number of instruments and singers) and how they their combined (e.g., mono-, hetero- or polyphony), timbre describes the “color” of a sound (Harnsberger, 1996) and is defined as the audible difference between tones of the same frequency, loudness and duration produced by different sources (e.g., a guitar or a piano). Notably, two factors regarding these elements of music are necessary to actually illustrate and characterize a musical piece: their respective expression and their change over time (McClellan, 2000) or termed differently, their spectrotemporal modulations. These properties were furthermore used to derive categorical principles that would group music pieces based on their respective expression of spectrotemporal modulations into categories and subsequently genres. While certain theories and research work attested the above mentioned elements of music an universal character, the majority of contemporary investigations stresses the immense influence of culture (Harwood, 1976), as instead of a concept, music refers to a definition that varies tremendously between cultures and cannot sufficiently be captured by researchers, as well as their respective participants from mainly Western, educated, industrialized, rich, and democratic (WEIRD) countries (Henrich, Heine, & Norenzayan, 2010). Indeed, the scientific field of Ethnomusicology could provide a vast amount of evidence, pointing not only to the variation of elements of music and their spectrotemporal modulations, but also completely diverging concepts of music across societies, leading to the transition of music and its definitions becoming highly culture specific. As a matter of fact, the majority of the above addressed principles vary intensively as a function of the culture they are utilized in. Especially, tonal and rhythmic characteristics seem to be affected by this. Respective research work could hereby hint to effects that entail nearly the entirety of human music processing. For example, differences could be observed with regard to the perception of harmony and rhythm (Jacoby & McDermott, 2017; J. H. McDermott, Schultz, Undurraga, & Godoy, 2016), memory (Demorest et al., 2009), emotion (Balkwill, Thompson, & Matsunaga, 2004; Kwoun, 2009) and preference (Darrow, Haack, & Kuribayashi, 1987; Teo, Hargreaves, & Lee, 2008) that are already present in early stages of human development (Soley & Hannon, 2010). Their unifying and underlying driving factor is referred to as music enculturation, thus the familiarization of adolescent individuals with their surrounding, here music, culture during their process of socialization. In other words, the music culture humans are born into shapes their livelong processing of music given its specificities (e.g., Western or Indian tonal system). This furthermore leads to the third factor, that is subjective experience. Given its incredibly diversity, music enculturation can result in highly diverging percepts and concepts of music, even within the same culture. This is grounded in the individual experience of music and reflected in e.g., music consumption behavior and music preference, thus adding another complex layer to the cultural concept of music, ultimately blurring the lines between music and non-music, as well as different genres. Taken the aforementioned points together the following operational definition of music will be used throughout the here presented thesis: as a category of sound, music is an auditory sensation that is distinguished by the expressions of the spectrotemporal modulations of said sound that are governed by rules typical to western tonal music, i.e. a twelve note chromatic scale and a meter of 2/4, 3/4 or 4/4. A restriction to western tonal music was made in order to account for the music enculturation of the to be included participants and limited resources with regard to both time and finances. Therefore, every background information, theoretical concept, as well as the results of this thesis are exclusively referring to western tonal music and should only rarely, if at all, be transferred to other music systems.

1.2. THERE AND BACK AGAIN - NEURONAL PROCESSING OF MUSIC

As most sensory phenomena, the perception and subsequent processing of music does not start in the cerebral cortex, but in sensory organs (SO) that are connected to the central nervous system (CNS) through a complex cascade of signal transduction and transformation. With regard to music, as a part of sound perception and thus hearing, this cascade describes how sound waves are transformed to percepts through the auditory system. The auditory system achieves this goal via the three parts it is commonly divided into, the ear or peripheral part, the non-neocortex or pathway part and the cortical structures or central part (Gage & Baars, 2018). While the ear and its components are responsible for the transformation from mechanical to electrical signals, the subcortical structures mark the pathways and therefore connections between SO and the CNS, as well as conduct initial neuronal processing that then reaches its full complexity in the cortical parts of the auditory system which processes distinct features of the incoming sound. A simplified graphical depiction of the human auditory system can be found in Figure 1.1. The respective parts and their inherent functions are described hereinafter.

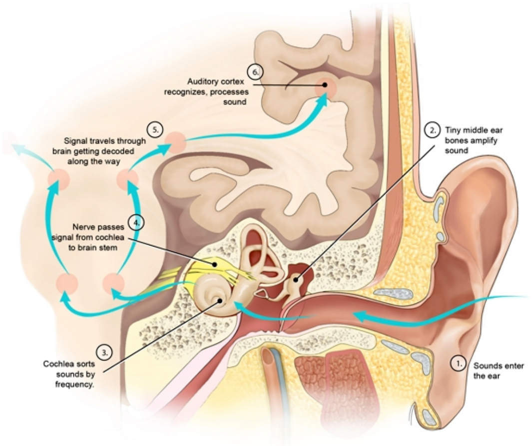


Figure 1.1: *Graphical depiction of the human auditory system, indicating important processing steps in a hierarchical ascending manner (blue arrows) from the ear towards the auditory cortex. Source: Public domain.*

The human ear

Being at the very beginning of hearing, the human ear is already tremendously complex and commissioned with a multitude of tasks. To this end, it is furthermore divided into three components, the outer or external ear, the middle ear and the inner ear as illustrated in Figure 1.2. The outer ear consists of the pinna or auricle and the ear canal or external acoustic meatus.

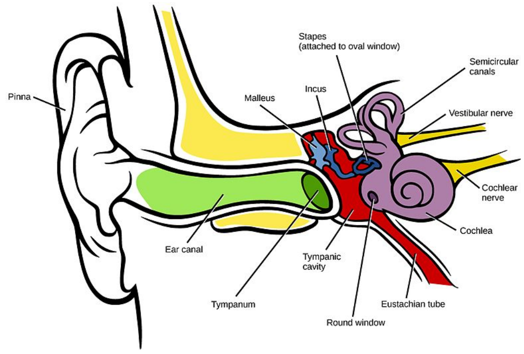


Figure 1.2: *The human ear and its components: the outer or external ear (green), the middle ear (red) and the inner ear (purple). Important structures within each are indicated and labeled through black lines. Source: By CNX OpenStax – https://cnx.org/contents/GFy_h8cu@10.53:RflAUPE2@6/Hearing-and-Vestibular-Sensation, CC BY 4.0, https://commons.wikimedia.org/w/index.php?curid=49935036.*

The first is the visible part laterally attached to the head most individuals refer to as "the ear". It acts as a sort of funnel that gathers sound by means of sound waves and enhances their localization, additionally amplifying incoming sounds of certain frequencies, namely between 1.5 kHz and 7 kHz (Purves et al., 2001)and therefore frequencies important to humans in everyday hearing situations like e.g., speech. The sounds are then directed towards the ear canal that transportes the incoming waves to the middle ear, while further amplifying the sound. The ear canal therefore marks the connection from the pinna to the eardrum or tympanic membrane, with the latter constituting the border between the outer and the middle ear. Within the air-filled middle ear the three tiniest bones of humans are located, the ossicles. They consist of the malleus, incus, and stapes and connect the eardrum to the oval window, that this the opening to the vestibule of the inner ear. More precisely, they enable an impedance matching through the principle of "mechanical advantage", "hydraulic principle" or "lever principle" (Bronzino & Peterson, 2006) which is necessary for the sound transfer from air-filled (outer and middle) to liquid-filled (inner) parts of the ear. This process is inevitable to tackle the tremendous decline of sound energy that would normally occur when sound transfers from air to liquid (Hill, Wyse, & Anderson, 2016) and which would be given in the absence of the ossicles, thus the direct connection from the outer ear and the oval window. In more detail, the surface area of the eardrum is larger by an order of magnitude than that of the malleus, which results in an increase in pressure when the vibrations of the eardrum, that are created through the sounds waves, are transmitted to the malleus. As the first ossicle, it then poses as the long arm of a lever, that moves the stapes as the short arm through a fulcrum, that is the incus. As the stapes are connected to the oval window, this in turn leads to an adequate transmission of sound energy that is only slightly diminished. Being transformed from sound waves to mechanical waves, the vibrations enter the liquid-filled inner ear, which houses the bony labyrinth. Within it, the central organ of hearing in humans is located, the cochlea, including the Organ of Corti. The cochlea is a hollow spiralled structure that contains three scalae: scala vestibuli, scala tympani and scala media. While the first two scalae contain perilymph, the scala media contains endolymph (Nicholls et al., 2012). The scala media is created through Reissner's membrane that separates it from the scala vestibuli and the basilar membrane that separates it from the scala tympani. On the basilar membrane, the Organ of Corti is situated. A respective illustration of the structure and organization of the cochlea is depicted in Figure 1.3. The aforementioned mechanical waves are created through the vibration of the stapes which in turn transmits the vibration to the oval window which is connected to the base of the cochlea and the scala vestibuli (Nicholls et al., 2012). These perilymph waves travel along the cochlea from its base to its apex which houses the helicotrema, the part of the cochlea where scala vestibuli and scala tympani adjoin. The resulting movements of the basilar membrane and thus the movement of the endolymph in the scala media, which is furthermore amplified by the Organ of Corti, lead to a displacement of the inner hair cells of the Organ of Corti. On top of the inner hair cells, cytoplasm structures with embedded bundles of cross-linked actin filaments, termed stereocilia and the tectorial membrane are connected to one another with so called tip links and furthermore arranged by height (Alberts et al., 2002).

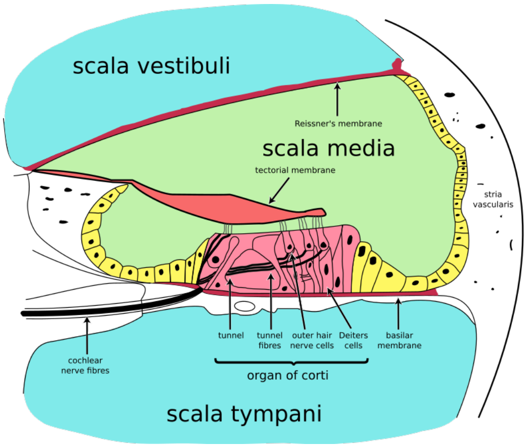


Figure 1.3: *Graphical depiction of the cochlea and its structure, including the three scalae and their respective parts. Source: By Original:OarihVector:Fred the Oyster - Own work based on: Cochlea-crosssection.png, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=9851471.*

Movements towards the taller stereocilia increase the tension of the tip-links that at a certain extent mechanically open the ion channels of the inner hair cells (Alberts et al., 2002). As the endolymph found in the scala media is rich on potassium, with the opposite being true for the perilymph, this opening of ion channels results in an influx of potassium and calcium ions into the inner hair cells which in turn deplorazise. This leads to the opening of voltage-gated calcium channels that are located at the base of the inner hair cells through which glutamate is released into the ribbon synapse based connection to the auditory nerve cells of the spiral ganglion, ultimately resulting in an spatiotemporal firing rates of these neurons that subsequently are transmitted as an electrical signal to auditory structures outside the neocortex. Based on this cascade the signal transduction found in the perception of sound is defined as mechanoelectrical. Notably, certain properties, that is thickness and width, thus the stiffness of the basilar membrane vary as a function of distance from the base to the apex of the cochlear. More precisely the stiffness decreases towards the helicotrema which results in a tuning of the hair cells of the Organ of Corti, as sound waves with a high frequency induce a displacement and thus mechanoelectrical signal transduction near the base of the cochlear and vice versa for sound waves with a low frequency. This encoding of frequencies along the cochlea and therefore also auditory nerve fibers is referred to as tonotopy and schematically depicted in Figure 1.4.

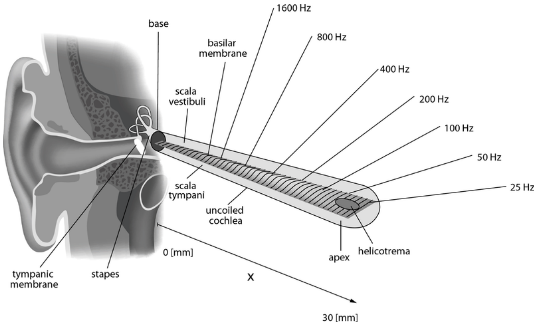


Figure 1.4: *The tonotopic organization of the human cochlear that exhibits a preferential encoding of high frequencies at the base and low frequencies at the apex. Source: Kern A, Heid C, Steeb W-H, Stoop N, Stoop R - Biophysical Parameters Modification Could Overcome Essential Hearing Gaps ([1]), CC BY 2.5, https://commons.wikimedia.org/w/index.php?curid=5929157.*

Non-neocortex structures

After being transformed from a mechanical to an electrical signal, a complex further processing is achieved through a multitude of highly bidirectional pathways that transport information from the spiral ganglion in an ascending manner to the neocortex and in a descending manner from the neocortex back to the spiral ganglion. Comprising the non-neocortex part of the auditory system, this furthermore includes a parallel stream that combines information from both peripheral parts. Overall the pathways entail structures in the brainstem, midbrain and subcortical parts of the cerebrum. Initially, the signal is projected from the spiral ganglion to the cochlear nuclei within the brainstem that receive only input from the corresponding ipsilateral cochlear (Middlebrooks & Bremen, 2013). In general, two nuclei are distinguished, the ventral and the dorsal cochlear nucleus. Exhibiting a tonotopic organization comparable to that of the cochlear (Oline, Ashida, & Burger, 2016; Oliver, Cant, Fay, & Popper, 2018), the nuclei initiate distinct pathways that compute complementary functions. While the ventral cochlear nucleus is concerned with population codes of the auditory nerve, the dorsal cochlear nucleus conducts different spectral analysis aiming to separate different sound sources by integrating information from higher auditory structures through descending pathways (K. A. Davis, 2005). From the ventral cochlear nucleus, the so called ventral acoustic stria originates and projects information to the ipsi- and contralateral superior olivary complex. The latter is enabled through a structure of the pons within the medulla oblongata, that is the trapezoid body (Mendoza, 2011). The superior olivary complex marks the first structure in which input from both peripheral structures are combined, integrated and jointly processed (Kulesza, 2007; Oliver, Beckius, & Shneiderman, 1995). Also located in the pons, it entails primary and secondary nuclei, with the primary being the lateral superior olive and the medial superior olive (Moore, 2000). The first is concerned with evaluating interaural signal differences due to intensity, the latter focuses on temporal aspects, resulting in the superior olivary complex assessing the azimuth, that is the localization of sounds (Kulesza, 2007; Moore, 2000). The ventral acoustic stria subsequently continues to the ipsilateral lateral lemniscus where it is joined by the contralateral dorsal acoustic stria that projects directly from the contralateral dorsal cochlear nucleus. Through its three nuclei (intermediate, ventral, dorsal) it transmits information from the superior olivary complex to the inferior colliculus, while also assessing timing and amplitude features. Here, the ventral nuclei receive information from the ipsilateral superior olivary complex and are not tonotopically organized, whereas the dorsal nuclei receive binaural information through the commissural fibers of Probst (Gil-Loyzaga, Pujol, & Irving, 2016). The inferior colliculus constitutes three parts, the central nucleus and the dorsal and lateral cortex and receives input from ipsi- and contralateral fibers of the lateral lemniscus which all terminate there. It is therefore assumed that the inferior colliculus, especially the central nucleus, integrates complex and distinct information of the ascending sound signal through auditory receptive fields (Lindeberg & Friberg, 2015). Additionally, high-resolution fMRI provided evidence for its tonotopic organization (De Martino et al., 2013). The inferior colliculus then projects information to the last subcortical structure, the medial geniculate nucleus which is "part of the auditory thalamus and poses as a thalamic relay between the inferior colliculus and the auditory cortex" ("Medial geniculate nucleus", 2019). However, recent studies could also show, that the functionality of the medial geniculate nucleus and its multiple subdivisions exceeds purely relay driven computations, to also include complex transformations of sounds comprising for example the extraction and integration of spectral and temporal features (Bartlett, 2013). As the aforementioned points described the ascending pathway of auditory information from the cochlea to the auditory cortex, it is worth mentioning that the descending pathway is comparably complex and involves the same structures in a reversed manner, that is from the auditory cortex to the cochlea. Through its top-down influence it can not only guide sound related attention but also trigger a re- or more fine-grained evaluation of the transmitted signal (Bajo, Nodal, Moore, & King, 2010; Rinne et al., 2008; Slee & David, 2015; Suga & Ma, 2003; Winer & Lee, 2007).

Cortical structures

The ascending pathway concludes in the auditory cortex, the first cortical structure to process auditory signals. In general, the auditory cortex is an assemblage of regions that span most of the superior temporal plane (STP) and superior temporal gyrus (STG) in both hemispheres and is divided into primary and non primary regions (Gage & Baars, 2018). However, contemporary research work suggests a more fine-grained differentiation of the non primary regions into belt and parabelt regions, while primary regions are termed primary auditory cortex (PAC) or core regions (Baumann, Petkov, & Griffiths, 2013; Fullerton & Pandya, 2007; Moerel, De Martino, & Formisano, 2014; Rivier & Clarke, 1997; Saenz & Langers, 2014; Sweet, Dorph-Petersen, & Lewis, 2005). The PAC is approximately located in parts of the Heschl's gyrus (HG) and exhibits a tonotopic organization that is comparable to those of the cochlea and non neocortex structures (Humphries, Liebenthal, & Binder, 2010; Moerel et al., 2014; Saenz & Langers, 2014), which is also present, although to a diminished extent, in the surrounding areas of the belt and parabelt regions (Humphries et al., 2010; Kajikawa et al., 2015). Compared to PAC, belt and parabelt regions extend along the STP, anteriorly towards the planum polare (PP) and posteriorly towards the planum temporale (PT), as well as the STG. While it was long assumed that the auditory cortex is solely concerned with the analysis of low level auditory stimulus features like pitch, timbre and tempo, recent studies could provide evidence that it is a highly connected network that conducts complex computations related to auditory processing and beyond (Gage & Baars, 2018; Winer & Schreiner, 2011). A respective model of cortical music processing was proposed by Koelsch (2011b) and is depicted in Figure 1.5.

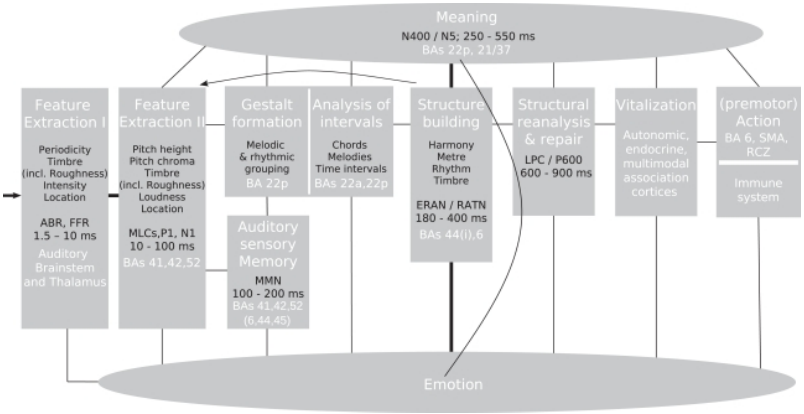


Figure 1.5: *Model of cortical music processing by Koelsch (2011b). It entails important components of music related analysis, assumed to be conducted by the human brain and ranges from rather early (feature extraction) to comparatively late (structure) processing stages that interact with one another. Source: Koelsch (2011b).*

Regarding these computations a two-stream model is presumed (Ahveninen et al., 2013; Hickok & Poeppel, 2007; Tian, Ma, & Wang, 2016), which describes a ventral and dorsal stream that originate from the PAC. In more detail, the ventral stream proceeds in the anterior direction to include PP and anterior parts of the STG (STGa) and is concerned with analyzing “what” was heard. In contrast, the dorsal stream evaluates “where” something was heard and extends towards the posterior direction covering PT and posterior parts of the STG (STGp). Both streams ultimately leave the auditory cortex to adjoin in the inferior frontal cortex, further connecting superior temporal (ventral stream), as well as parietal (dorsal stream) regions with the auditory cortex in an eminently bidirectional manner (Rauschecker & Scott, 2009). However, recent studies challenge both the hierarchical and distinct nature of the pathways (Ahveninen et al., 2013; Woods & Alain, 2009). Additionally, despite extensive investigations the exact localization and parcellation of the auditory cortex itself is still subject to ongoing debates. Throughout these investigations, multimodal descriptions that combine information from structural and functional approaches have been proven to provide the most accurate characterization (Glasser et al., 2016). This furthermore lead to an insufficient understanding of the computations and transformations that are applied to sound within and across the auditory cortex, including its specificity and generation of subsequent stable percepts. Taken together with the enhancing, yet limited possibility of in vivo investigations of the subcortical structures and those outside the neocortex, the here presented thesis focused exclusively on the cortical part of the auditory system.

1.3. A BRIEF OVERVIEW OF MUSIC AND THE BRAIN

After reaching the cortical portions of the auditory system, incoming sounds are further processed in a highly distributed bilateral network that however might be altered depending on the type of sound that needs to be processed (Andoh, Matsushita, & Zatorre, 2018; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Giordano, McAdams, Zatorre, Kriegeskorte, & Belin, 2013; Moerel, Martino, & Formisano, 2012; Peretz, Vuvan, Lagrois, & Armony, 2015; Rauschecker & Scott, 2009; Santoro et al., 2014; Zatorre & Zarate, 2012).Based on that, the outcomes of several studies point to a certain modularity of the auditory system with regard to different sound categories and their specific requirements and the subsequent effects they evoke, with especially language and music being vigorously pronounced (Fedorenko, Patel, Casasanto, Winawer, & Gibson, 2009; Joanisse & DeSouza, 2014, SEP; Peretz & Coltheart, 2003; Zatorre & Baum, 2012; Zatorre, Belin, & Penhune, 2002). Even though exhibiting a comparable social importance, music seemingly takes a special role, which is stressed by its limited direct use for communication (compared to speech), but tremendous influence on cognition, emotion and health (Koelsch, 2011a, 2014; Peretz et al., 2015; Whitehead & Armony, 2018). While the processing of music is an almost universal human trait (Zatorre & Salimpoor, 2013) its complexity and the therewith connected necessary analyses, that in turn can have widespread effects is unbeknownst to the vast majority of individuals (Koelsch, 2011a, 2014). For example, music has inherent syntactic and semantic principles that are automatically and effortless evaluated in real time, allowing individuals to perceive and judge music as irregular, non-fitting or simply wrong (Asano & Boeckx, 2015; Koelsch, 2009b, 2011a, 2011b; Koelsch, Rohrmeier, Torrecuso, & Jentschke, 2013; Maess, Koelsch, Gunter, & Friederici, 2001; Patel, 2003). These principles apply to fine-grained and larger scales of divers acoustic features and other properties of music, as for example both chords and melody sequences can introduce errors that are automatically detected by most individuals (Koelsch, 2009b, 2011a; Maess et al., 2001; Minati et al., 2008; Zatorre & Salimpoor, 2013). This “intramusical meaning” interacts with “extramusical meaning”, that among other aspects might include certain memories about the perceived music excerpts, to generate a context or semantic about a given musical piece (Koelsch, 2011a, 2011b). To a great amount, these effects are also at the core of perceiving and inducing emotions through music (Koelsch, 2014; Zatorre & Salimpoor, 2013). Not only can different emotions that are reflected in music be distinguished by listeners, but also be respectively evoked (Koelsch, 2014; Zatorre & Salimpoor, 2013). While being most prominent for joy, happy, sad and fearful music, the same accounts for the remaining myriad of emotions individuals can perceive (Aube, Angulo-Perkins, Peretz, Concha, & Armony, 2015; Koelsch, 2014; Zatorre & Salimpoor, 2013). The results of previous research work further stress the distribution of these effects, as music apparently affects the majority of brain regions involved in the processing of emotions, as well as their connections (Koelsch, 2014). Notably, the processing of both syntactic/semantic principles and emotion is seemingly highly comparable between music and speech, with their underlying guiding principles and acoustic features being assumed to drive this similarity (Aube et al., 2015; Koelsch, 2011a, 2011b, 2014; Kumar, Kriegstein, Friston, & Griffiths, 2012; Zatorre & Baum, 2012; Zatorre et al., 2002; Zatorre & Zarate, 2012). Through these aspects, one of music’s most prominent use cases is generated, the feeling of pleasure (Juslin & Sloboda, 2011; Zatorre & Salimpoor, 2013). Interestingly, the sensation of pleasure is introduced through an interaction of the aforementioned aspects, as well as other top-down influences that encompass for example memory and culture related factors (Gebauer, Kringelbach, & Vuust, 2012; Reybrouck, Eerola, & Podlipniak, 2018; Salimpoor & Zatorre, 2013; Zatorre, 2015; Zatorre & Salimpoor, 2013), as e.g., recent studies could provide evidence that also sad music, hence thought to evoke negatively connoted emotions, can induce pleasure (Brattico et al., 2016; Sachs, Damasio, & Habibi, 2015). Music induced pleasure furthermore offers two observations that provide additional insights into the special role of music. The first concerns its potential role in evolution as even though non-human primates and other animals have the ability to perceive and produce music (as defined by humans) to a certain extent, none of them seem to utilize it for pleasurable experiences (J. H. McDermott & Hauser, 2004, 2007; Zatorre & Salimpoor, 2013). Second, as individuals exhibit different sensitivities to pleasure evoked by music, especially individuals that display music specific anhedonia (Martínez-Molina, Mas-Herrero, Rodríguez-Fornells, Zatorre, & Marco-Pallarés, 2016; Mas-Herrero, Dagher, & Zatorre, 2018; Mas-Herrero, Zatorre, Rodriguez-Fornells, & Marco-Pallarés, 2014), a certain modularity of cortical music processing is further supported. As the feeling of music evoked pleasure is accompanied by pronounced (neuro-)endocrine responses, that is the release of dopamine in the ventral striatum based on an increased connectivity of the auditory cortex towards the ventral striatum (Martínez-Molina et al., 2016; Mas-Herrero et al., 2014; Zatorre, 2015; Zatorre & Salimpoor, 2013), it additionally provides the possibility of health related applications. Here, several studies have provided evidence that music can be utilized as a powerful tool to release or even prevent stress (Chanda & Levitin, 2013; Koelsch, 2014; Linnemann, Ditzen, Strahler, Doerr, & Nater, 2015; Linnemann, Strahler, & Nater, 2016; Pelletier, 2004; Thoma et al., 2013). Besides from that, music therapy is steadily growing especially in the fields of developmental and acquired disorders. This is based on its complex cortical processing and the subsequent effects as described above. Hence, it is used to to improve or restore certain cognitive functions within which it is heavily involved (Thaut & Hoemberg, 2014). This comprises e.g., memory (Koelsch, 2015; H.-C. Li, Wang, Chou, & Chen, 2015; Prickett & Moore, 1991; Thaut & Hoemberg, 2014) and in particular language where its aforementioned similarities to music being utilized (Krauss & Galloway, 1982; Norton, Zipse, Marchina, & Schlaug, 2009; Schlaug, Marchina, & Norton, 2008; Thaut & Hoemberg, 2014). Furthermore, music has also been proven useful to enhance social abilities and general wellbeing in neuropsychiatric disorders (Geretsegger et al., 2017; Maratos, Gold, Wang, & Crawford, 2008; Sharda et al., 2018).

1.4. THE MISSING LINK BETWEEN WAVES AND PATTERNS

As noted in the introduction of this chapter, the vast majority of studies that investigated neuronal correlates of music processing focused on either ends of the spectrum music processing entails: rather low level representations of features or comparably higher level functionality like syntax, semantics or emotion and pleasure. The same holds true for both the presumably specialized cortical processing and representation of music. Thus the intermediate stage how incoming sound waves are processed and transformed to yield stable and specialized representations of music that are then further utilized is less understood. One key factor proposed to drive the respective computations are the spectrotemporal modulations of acoustic features found in and defining for music (Norman-Haignere, Kanwisher, & McDermott, 2015; Santoro et al., 2014; Santoro et al., 2017; Whitehead & Armony, 2018). However, a growing body of research works points to the insufficiency of a purely bottom-up feature based approaches and hint to the importance of more complex features and their interaction with top-down processes (Casey, 2017; Whitehead & Armony, 2018; Zatorre & Salimpoor, 2013).

2

AIMS OF THE THESIS

The second chapter of this dissertation introduces the aims of the here presented thesis and how they address the gaps outlined in the previous chapter. It furthermore provides an overview of the chapters included in this thesis and their content.

The aim of the here presented thesis therefore was to address the aforementioned missing link, especially how the transformation of incoming sound to specialized and stable representations of music is achieved through the interaction of low and high level features along the cortical auditory system. To this end, the thesis at hand was divided into three consecutive parts and four projects that are also reflected in the different chapters of this research work. The first one was split in two complementary projects and concerned with methodological challenges inherent to neuroscientific research on music as a part of auditory neuroscience. In more detail, this refers to the aggravated measurement settings which are characterized through (heavily) distorted sound presentations due to e.g., the noise individuals are exposed to when laying inside a magnetic resonance imaging (MRI) scanner, as well as the robust definition by means of localization and parcellation of the auditory cortex as the brain structure central to cortical auditory, thus music processing. Targeting these challenges, two open source toolboxes were created in order to provide a methodological and unifying fundament that allowed the standardization and consecutive character of the subsequent projects. One intended to enable the assessment of individual hearing thresholds and abilities given the acoustic scanner noise (ASN) and acoustic properties of auditory stimuli. The other one focused on the automated localization and parcellation of the human auditory cortex by means of multiple approaches and respective metadata gathering. The third project utilized the corresponding outcomes of the latter to investigate if the auditory system is tuned for music or put differently if music evokes a unique cortical processing within the realm of auditory categories. To this end, an fMRI study during which participants were presented with multiple auditory categories was conducted and the resulting data analyzed with distinct, yet complementary approaches. Using categories of equally social importance, that were additionally controlled and matched for basic acoustic features, a musical gradient was created that spanned speech, singing and instrumental music. Via multivariate searchlights and dynamic causal modelling two essential functional and organizational principles of the human brain were probed respectively: segregation and integration. More precisely, the segregation and integration of areas belonging to the auditory cortex. Here the focus was to identify patterns of processing within the auditory system that are able to distinguish music from other categories thus giving rise to its defining representation, while additionally allowing to generalize to an overall "musical component" that is independent of music with (singing) or without speech (instrumental) and cannot be attributed to acoustic features alone. The fourth and last project build upon these insights by further investigating how music is categorized in itself through the auditory system. It especially intended to examine how the incredible diversity and complexity of music is addressed throughout cortical auditory processing and via the combination of low and high level features leads to rather precise and stable percepts that are categorized on different levels, i.e. main and subgenres. In order to tackle the high dimensionality, a versatile set of music genres was utilized within a multimodal approach that entailed fMRI for spatial aspects, EEG for temporal properties, behavior for high level representations, music information retrieval for acoustic features and annotations for semantic factors. Integrated through a unifying analysis approach, that is representational models, these modalities focused on how music genre categorization is achieved across space and time, as well as how this computation can be explained of models of both bottom-up (low level acoustic features) and top-down (high level features such as behavior and concepts) processes.

3

EXPERIMENTAL BASIS OF COGNITIVE & COMPUTATIONAL NEUROSCIENCE RESEARCH ON MUSIC

Within the third chapter of the here presented thesis the first project, which was split in two subprojects, is presented. Starting with a general introduction and outline of two major methodological challenges in auditory neuroscience, the subsequent parts will be dealing with two toolboxes created to target these challenges. Each contains a Methods & Materials section introducing the respective toolbox and its functionality, as well as corresponding results based on example data. The final part will discuss the suitability and limitations of the created toolboxes given the underlying challenges. Additionally, the factor open science and its principles will be an important part, with respect to both the challenges and toolboxes.

Explanation and identification of personal contribution and data usage

Hereinafter the contribution of the thesis' author (PH) with regard to each project part will be outlined and additional information be provided.

General remarks

The project and its parts described in this chapter targeted the creation of respective toolboxes. Therefore, and based on the lack of financial resources no data, except two pilot participants in the first project part "Ain't no sound loud enough - audiometry in MRI settings", was acquired, but publicly available datasets used.

Conceptualization

PH solely conceptualized the herein described project and its parts.

Data usage

All data used was already made publicly and freely available online as part of repositories or other comparable resources. The two pilot participants that are included in the first project part "Ain't no sound loud enough - audiometry in MRI settings" gave their informed consent allowing public data sharing after anonymization (through the "Open Brain Consent Form" (APPENDIX A)). Both toolboxes are publicly available:

Project 1 - "Ain't no sound loud enough - audiometry-like measurements in MRI settings" (https://github.com/C0C0AN/audiometry_mri)

Project 2 – "ALPACA – Automated Localization and Parcellation of Auditory Cortex Areas" (https://github.com/C0C0AN/ALPACA).

3.1. INTRODUCTION

Spending a closer look on the research field of auditory neuroscience, the impression that a striking difference by means of advances and insights as compared to other cognitive functions and processes exists, emerges. In fact, especially other "core functions" like visual perception and motor abilities seem to have a more profound and reliable knowledge basis, as well as precise computational models that account for the majority of their functionality (Felleman & Van Essen, 1991), which is further indicated by the tremendous success of these skills in artificial intelligence and robotics (Pagallo, Corrales, Fenwick, & Forgo, 2018; Torresen, 2018; Yosinski, 2018). Considering the dissimilarities between these functions and auditory neuroscience, e.g., music, multiple, more precisely three prominent causes become evident. Two of them are highly specific for auditory neuroscience, while the third is based on a general structural problem across all scientific disciplines. Being grounded specifically in neuroimaging, that is MRI research, the first two entail aggravated measurement settings that lead to heavy signal distortions, as well as the localization and parcellation of the human auditory cortex as the main brain region responsible for the processing of auditory signals have to be called. The third refers to the significant absence of data and analysis sharing that result in low reproducibility and research islands instead of collaboration which in turn slows down further advancement tremendously. These problems that affect auditory neuroscience are presented in more detail hereinafter.

Acoustic distortions in neuroimaging settings

The perception of sound is only in rare cases undistorted and free of any influence. However, the influence of e.g., environment noise and transmission quality (i.e. headphones, signal, etc.) can in large parts be ignored and might at worst create an unpleasant or challenging experience for the listeners. Considering the research on sound, these influence on the contrary pose immense problems and confounding factors. That being said, the extensive amount of auditory neuroscience studies that were and are conducted within an (f)MRI setting appear not only ironic but also counterintuitive. In other words, acquiring neurophysiological data that focuses on auditory perception through placing participants inside a machine that produces extremely loud and unpleasant noises will introduce unwanted pronounced distortions in both the auditory stimulus the participant has to perceive and the data that is recorded. These noises are an inherent effect of MRI scanners and are termed acoustic scanner noise (ASN) (Peelle, 2014). In order to enable detailed insights and analyses of ASN two different sources have to be distinguished: gradient switches and radiofrequency (RF) pulses. The first refers to the gradient switches between phase encoding and readout currents that are necessary within certain data acquisition protocols (DAP) like the commonly used echo-planar imaging (EPI) to e.g., capture the blood oxygen level dependency (BOLD) (Peelle, 2014). A "side product" of these switches is the creation of Lorentz forces that subsequently oscillate through the transmissive material of the MRI scanner and thus becomes sound once reaching air (Peelle, 2014). As the Lorentz force is proportional to the magnetic field strength (Bo) and the current of the gradient, an increase of either or both will results in more prominent ASN (Moelker & Pattynama, 2003; Peelle, 2014). This is especially relevant as MRI scanners with comparably high field strengths, that is 7 or 9 Tesla, become more common, as a machine with 3 Tesla already generates ASN that ranges from 85 to 130 decibel of sound pressure level (dB SPL) (Foster, Hall, Summerfield, Palmer, & Bowtell, 2000; Peelle, 2014; Price, De Wilde, Papadaki, Curran, & Kitney, 2001; Ravicz, Melcher, & Kiang, 2000). The second is concerned with the ASN generated by the RF pulses that are usually fairly ignored due to the prominence and overlap of gradient switch based ASN, as RF pulse ASN peak at 57 to 76 dB SPL, which however arise as problematic dependent on the DAP (Peelle, 2014; Ravicz et al., 2000). Even though ASN primarily exhibit frequencies at the rather low end of the human audible spectrum (below 1 kHz) and are with regard to this aspect not as harmful as higher frequencies (Moelker & Pattynama, 2003), the indicated sound level have to be considered as dangerous with respect to hearing impairments or even loss (Sheppard, Chen, & Salvi, 2018). This is motivated by the maximum permission exposure limit (PEL) as formulated by the U.S. Occupational Safety and Health Administration (OSHA) regulations and is defined as 105 A-weighted decibels (dBA) for one hour, a duration typical for (f)MRI experiments (Sheppard et al., 2018). In order to provide the thus necessary hearing protection, participants are usually equipped with earplugs that however result in another type of auditory distortion, as their attenuation and filtering effect alters the incoming auditory signal to a large extent (Peelle, 2014). ASN mainly affect neurophysiological data through three pathways as depicted in Figure 3.1 (Peelle, 2014).

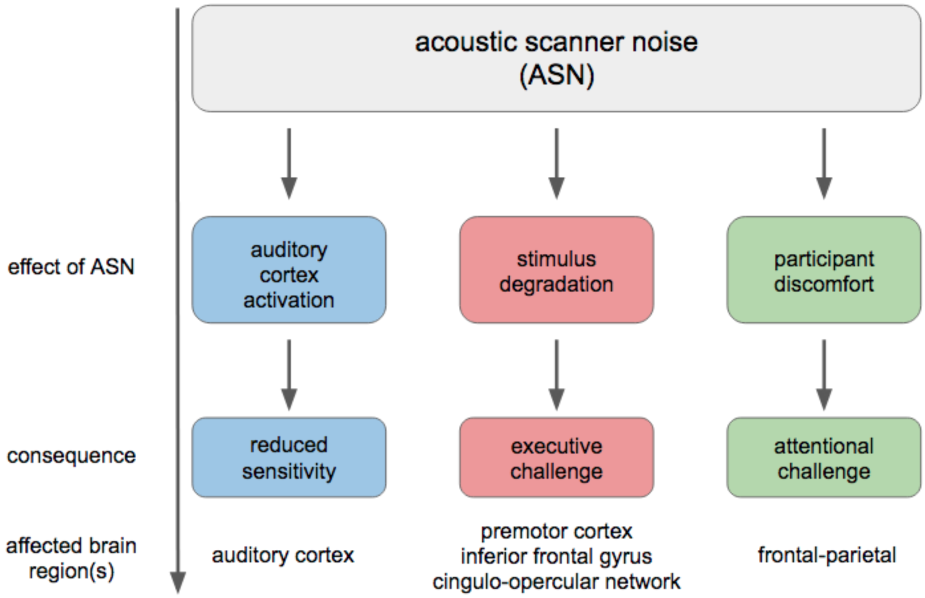


Figure 3.1: *Possible pathways of ASN introduced distortion effects. Colors mark the different pathways: blue - first pathway, red - second pathway, green - third pathway. Modified from Peelle (2014).*

The first pathway is based on the exemplary work of Bandettini, Jesmanowicz, Van Kylen, Birn, & Hyde (1998) & Talavage, Edmister, Ledden, & Weisskoff (1999) that could respectively provide evidence, that ASN alone is sufficient to evoke auditory cortex activation, that exhibits a time course comparable to those of auditory stimuli and is present in both primary and secondary regions. In turn activations evoked by simultaneously presented auditory stimuli are distorted and diminished in their overall extent and spatial distribution (Gaab, Gabrieli, & Glover, 2007; Langers, Van Dijk, & Backes, 2005; Scarff, Dort, Eggermont, & Goodyear, 2004; Shah, Jäncke, Grosse-Ruyken, & Müller-Gärtner, 1999; Talavage & Edmister, 2004; Talavage, Edmister, Ledden, & Weisskoff, 1999). The degradation of auditory stimuli through ASN and its resulting increased auditory challenge are at the core of the second pathway. Despite mainly investigated for speech (Peelle, 2014), a challenging perception of auditory stimuli originating from other auditory categories most likely also leads to a higher cognitive demand and therefore altered activations and the recruitment of additional brain regions (Dosenbach, Fair, Cohen, Schlaggar, & Petersen, 2008; Peelle, 2018; Rönnberg et al., 2013; Scott & McGettigan, 2013; Wingfield, Tun, & McCoy, 2005). Even though these extended networks included regions outside the auditory cortex and are not specifically or exclusively involved in cortical auditory processing, their resulting accompanying top-down influences will affect the processing and representation within early auditory regions (Obleser & Kotz, 2011; Obleser, Wise, Dresner, & Scott, 2007; Sohoglu, Peelle, Carlyon, & Davis, 2012). The last pathway refers to the discomfort induced through ASN that might prevent participant from attending to the auditory stimuli, unsuccessful task performance or even anxiety based alterations of activations (Peelle, 2014). Here, especially task related effects that are know to modulated auditory processing (Alho, Rinne, Herron, & Woods, 2014; J. B. Fritz, Elhilali, & Shamma, 2005; Petkov et al., 2004; Scheich, Brechmann, Brosch, Budinger, & Ohl, 2007) need to be considered, while the factor anxiety might introduce confounds when comparing populations with diverging sensitivities, i.e., infants vs. adults and healthy vs. clinical groups (Peelle, 2014). In order to address the different sources and resulting effects of ASN, a range of methods have been proposed that are mainly concerned with the adaptation and modification of DAPs. Their unifying approach is grounded in a discontinuous DAP in which auditory stimuli are presented exclusively between volume acquisitions and therefore gradient switches and RF pulses. The application of such DAPs is possible to the time delayed response of the BOLD signal that allows to capture its peak in response to an auditory stimulus roughly 6 sec. after the presentation (Liao et al., 2002). Thus, no overlap between ASN and auditory stimuli exists, however at the expense of data quantity. Two famous and heavily used versions of such a DAP exist: sparse imaging (Bilecen, Scheffler, Schmid, Tschopp, & Seelig, 1998; Eden, Joseph, Brown, Brown, & Zeffiro, 1999; Hall et al., 2000; Talavage et al., 1999; Talavage & Hall, 2012) and interleaved silent steady state (ISSS) (Bekinschtein, Davis, Rodd, & Owen, 2011; M. H. Davis, Ford, Kherif, & Johnsrude, 2011; Doehrmann, Naumer, Volz, Kaiser, & Altmann, 2008; Doehrmann, Weigelt, Altmann, Kaiser, & Naumer, 2010; A. Engel & Keller, 2011; Mueller et al., 2011; Rodd, Johnsrude, & Davis, 2012; Schwarzbauer, Davis, Rodd, & Johnsrude, 2006; Yoo et al., 2012). While sparse imaging is characterized through the aforementioned description of discontinuous DAPs, ISSS addressed the reduced amount of stimuli corresponding data points by implementing a DAP during which multiple volumes are acquired instead of one. A comparable DAP would theoretical and technically also be possible in classic sparse imaging but introduce confounds in the time series due to a "lack of steady-state longitudinal magnetization" (Peelle, 2014, p. 7). ISSS accounts for this by combining continuous RF pulses with variable read out times, thus achieving a temporal coverage that is comparable to continuous DAPs (Peelle, 2014, p. 7). While these recent developments (Norman-Haignere & McDermott, 2016) have undeniably increased the quality and validity of auditory neuroscience in imaging settings, they do not account for other prominent distortion factors. Among these are other sources of ASN, frequency spectrum (FS) of ASN, earplugs, headphones and specificities of DAPs. Components of the MRI scanner such as pipes and the helium pump are also creating ASN that are in contrast to gradient switches and RF not modulated by discontiunous DAPs, at least to a certain extent. They are therefore also audible in silent periods (e.g., presentation of an auditory stimulus). Furthermore, ASN cannot only be described by means of their volume, as they also exhibit a certain FS, that is the frequencies that the ASN is composed of and their respective markedness. This might lead to certain frequencies that are audible by humans being more affected by ASN than others, due to an overlap with frequencies prominently included in ASN. As already addressed, the important and necessary attenuation through earplugs also introduces distortions of the incoming auditory signal based on its filter properties that vary the FS of the stimulus (Peelle, 2014). Comparably, headphones that are typically used in auditory fMRI settings to present auditory stimuli vary greatly between brand and type with respect to frequency range, loudness and noise reduction. Furthermore, they can introduce distortion products that can alter cortical responses to incoming auditory signals, especially with regard to pitch (Norman-Haignere & McDermott, 2016), adding yet another source of variance. Additionally, ASN are highly DAP dependent as their underlying sources are varied in their markedness by means of DAP parameters like the time of repetition (TR). Previous research could provide evidence for the extensive effect of DAP on ASN that even lead to diverging lateralization patterns (Angenstein, Stadler, & Brechmann, 2016; Ranaweera et al., 2016). As all the aforementioned factors furthermore heavily interact with the stimulus material of a given study, a high dimensional and complex factor space is created that leads to listening situation and percepts which are highly individualized to a single participant. So far, the majority of research work did not account for this problematic circumstance and if they did, only in a rather undetailed, unspecific and undocumented manner by e.g., asking the participants if they could hear the auditory stimuli during the experiment, that is DAP. Furthermore, only a limited amount of studies included the possibility to adjust the volume of the stimuli and so far no results exist on the interplay between the frequency spectra of the different sound sources, i.e. ASN and stimuli. Taken together this demonstrates the need for the assessment of individual hearing thresholds (HT) in neuroimaging settings dependent on different DAPs to not only account for the aforementioned problems, but also gaining more insights into hearing abilities in neuroimaging settings and how they might be influenced which in turn poses as a necessity to further enhance DAPs in the context of auditory neuroscience.

Localization and parcellation of the human auditory cortex

One of the core objectives of neuroscientific investigations that utilize neuroimaging techniques is the identification and characterization of brain regions that are involved in certain cognitive processes. Albeit a recent paradigm shift that changed the perspective from a rather purely segregation to more integration based approach, the robust localization and parcellation of brain regions remain an important and challenging quest. The auditory cortex, as the main and most crucial brain region responsible for cortical auditory processing, is by no means an exception to that. On the contrary, its arguably the most difficult primary sensory brain region with regard to its localization and parcellation, as nicely outlined by Glasser et al. (2016, S3, p. 35): “In contrast to early visual and somatomotor cortex, parcellation of the early auditory cortex has proven much more challenging in both macaques and humans...” . As mentioned in 1.2, the here presented thesis refers to the auditory cortex not by means of the primary auditory cortex, but an aggregation of regions that span core, belt and parabelt within the STP and STG, thus the cortical auditory network or system, in other words the primary auditory cortex and regions of both the dorsal and ventral pathway within the temporal lobe (Gage & Baars, 2018). Even though the previous scientific literature with regard to this topic mostly agreed on a rough subdivision in core, belt and parabelt ROIs, thus primary and non-primary or secondary auditory ROIs, that in turn contain multiple regions themselves and follow the model of the monkey auditory cortex (Hackett, Preuss, & Kaas, 2001; Hackett, Stepniewska, & Kaas, 1998; Kaas & Hackett, 2000; Moerel et al., 2014; Morosan et al., 2001), to this day a tremendous lack of consensus with regard to the exact localization and broad, as well as fine-grained parcellation of the human auditory cortex exists, which rests in large parts upon the immense between-individual variability of these brain structures (Moerel et al., 2014). Comparable to other brain regions, the investigation of the auditory cortex can be described by a function of time period and the therein present methodological possibilities. In general, two approaches have to be distinguished: structural and functional, with both comprising a set of frequently utilized methods. These approaches as well as their respective mostly used methods are briefly characterized hereinafter and are furthermore illustrated in Figure 3.2.

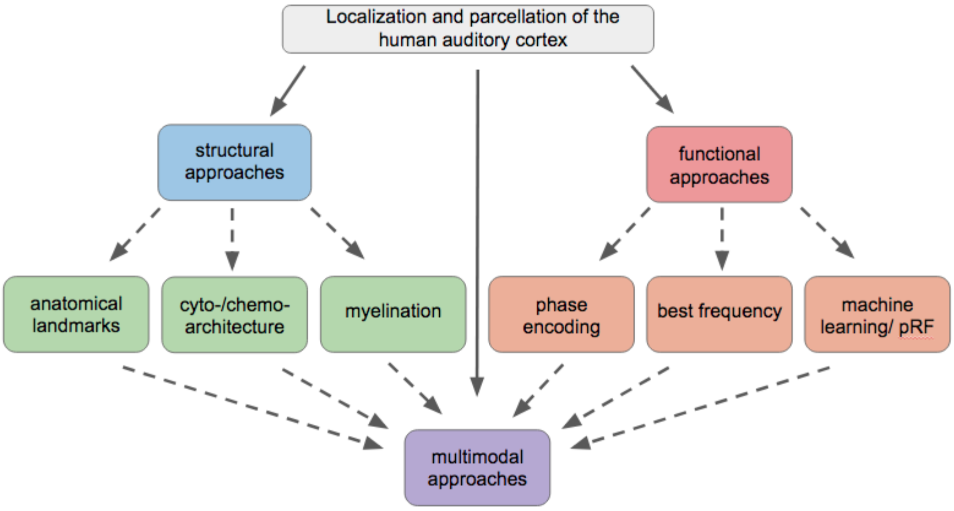


Figure 3.2: *Common approaches to localize and parcellate the human auditory cortex. Colors indicate different approaches,including structural (green), functional (red) and multimodal methods (brown).*

In line with the aforementioned interaction of time period and methodological possibilities, the first endeavours to locate and parcellate the human auditory cortex relied on post-mortem investigations that were driven by diverse features of the neocortex. Here, three procedures have to be discerned: anatomical landmarks, cyto- and chemoarchitecture and myelination. Concerning the first, patterns of gyri and sulci have been used to delineate the borders of the auditory cortex and its regions. Broadly, the auditory cortex comprises the upper parts of the STG, that is the STP and most superior bank of the STG (Celesia, 1976; Gage & Baars, 2018; Galaburda & Sanides, 1980; Moerel et al., 2014; Rivier & Clarke, 1997). Hidden within the lateral sulcus or sylvian fissure, a convolution known as transverse temporal gyrus or HG marks the core or primary auditory cortex (PAC). First discovered by the austrian anatomist Richard L. Heschl, it is circumscribed medially and laterally by the borders of the STP, i.e. medially by the insular cortex and laterally by the upper bank of the STG, with the first transverse sulcus and Heschl's sulcus bounding the HG in the anterior and posterior direction (Celesia, 1976; Gage & Baars, 2018; Galaburda & Sanides, 1980; Moerel et al., 2014; Rivier & Clarke, 1997). While HG, PAC and core of the core ROIs are often used in a rather interchangeable manner, a vast variety of precise localizations and relationships of these terms exist. Most commonly and based on contemporary studies, PAC and core ROIs (e.g., A1 and R) are assumed to lay within HG, comprising only a certain portion of it (Emmorey, Allen, Bruss, Schenker, & Damasio, 2003; Gage & Baars, 2018; Gage et al., 2009; Hubl et al., 2010; Leonard et al., 2001; Penhune, Zatorre, MacDonald, & Evans, 1996; Rademacher, Galaburda, Kennedy, Filipek, & Caviness, 1992; Schneider et al., 2009; Schneider et al., 2002; Warrier et al., 2009; Wong et al., 2008). Notably, the number of convolutions that constitute the HG can very extensively between individuals, ranging from 1 to 3 gyri (Campain & Minckler, 1976) and are therefore on key factor that drives the above mentioned inter-individual variability (Campain & Minckler, 1976; Gage & Baars, 2018; Moerel et al., 2014; Penhune et al., 1996). These duplications of Heschl's gyri can furthermore more be partial or complete (D. Marie et al., 2015; D. Marie, Maingault, Crivello, Mazoyer, & Tzourio-Mazoyer, 2016). Within the STP, the HG is surrounded by the PP (anterior direction) (Kim et al., 2000) and the PT (posterior direction) (D. Marie et al., 2015; Moerel et al., 2014). Notably, in case of a complete HG duplication, the respective posterior part is allocated to the PT (Dorsaint-Pierre et al., 2006; D. Marie et al., 2015). Even though investigated to a much smaller extent, laterally, i.e. within the upper bank of the STG, two regions are differentiated through the rostrolateral end of the first transverse sulcus (Rademacher et al., 1992), the anterior and posterior STG. Using information with regard to the cellular composition of brain tissues, as well as their histochemical properties, the second set of procedures, that is cyto- and histoarchitectonic methods, integrated cell types and their distribution into the labeling of the auditory cortex based on anatomical landmarks (Galaburda & Sanides, 1980; Hackett et al., 2001; Moerel et al., 2014; Morosan et al., 2001; Rivier & Clarke, 1997). However, the precise localization and amount of thereby created auditory ROIs diverges as a function of the increasing number of studies that are concerned with this topic. As this is especially marked for ROIs outside HG, the vast majority of research work focused on the fine-grained subdivision of core ROIs within HG (Moerel et al., 2014). Beginning with Korbinian Brodmann's pioneering work, two of his marked regions correspond roughly to HG: Brodmann Areas (BA) 41 and 42 (Brodmann, 1909). More precisely, they reflect the anterior and posterior part of the HG respectively, with the first, BA 41 most likely constituting PAC and core regions (Brodmann, 1909; Moerel et al., 2014). BA 42 might comprise belt and parabelt regions with the latter extending into BA 22 which entails large portions of the STG (middle to posterior portions of the anterior parts, as well as anterior portions of the posterior parts). Morosan et al. (2001) further divided Brodmann's rather broad auditory cortex into five regions along the STP, including HG, PP and PT. Notably, they introduced three subdivisions of the HG: anterolateral HG, middle HG and posteromedial HG. Here, the middle HG refers to BA 41, thus PAC and core regions and relies on a set of cyto- and histoarchitectural properties are in line with the parcellation schemes of further studies of the same subject (Clarke & Morosan, 2012; Rivier & Clarke, 1997; von Economo & Horn, 1930). Namely, these are a highly granular koniocortex that has small to medium sized pyramidal cells in the III layer and a well-developed IV layer that is characterized through an immense density of Acetylcholinesterase, cytochrome oxidase and parvalbumin in its neuropil (Moerel et al., 2014). In contrast, the anterolateral and posteromedial HG constitute subdivisions of the HG and the belt regions that approximately correspond to homologue regions in the auditory cortex model of the monkey. Whereas the posteromedial HG is considered to correspond to the PAC / core region rostrotemporal subfield and is indicated by a broad III layer with collections of medium sized pyramid cells, the anterolateral HG might resemble the caudolateral or caudomedial field of the PAC / core region surrounding belt areas and is defined through sole medium-sized pyramidal cells in layer III that exhibits less distinct cortical layering (Moerel et al., 2014; Morosan et al., 2001). BA 42 and BA 22 are identified in comparison to BA 41 / PAC / core regions via a less granular parakoniocortex and large pyramidal cells in the III layer (Moerel et al., 2014). The aforementioned distinction between core, belt and parabelt regions could further been replicated by research work that focused on myelination, the third procedure. While initially restricted to post-mortem investigations (Beck, 1928; Galaburda & Sanides, 1980; Hackett et al., 2001; Hopf, 1954; Nieuwenhuys, 2013), recent advances in neuroscientific methods, especially MRI, enabled the examination of myelination in vivo (Dick et al., 2012; Glasser & Van Essen, 2011; Wasserthal, Brechmann, Stadler, Fischl, & Engel, 2014) through DAPs that are capable of capturing myelin contrasts like quantitative T1 (Dick et al., 2012; Sereno, Lutti, Weiskopf, & Dick, 2013; Sigalovsky, Hawley, Harms, & Melchert, 2001) and T2 (Cohen-Adad et al., 2012) or T2/T2* weighted contrasts (De Martino, Moerel, Xu, et al., 2015; Glasser & Van Essen, 2011) as outlined by Moerel et al. (2014). Both procedures yield comparable results in that the PAC / core region exhibits the most dense myelination that decreases towards all surrounding regions, however to a different extent. The decline is less pronounced into posterior direction and more into anterior direction, leading to PT and PP respectively (Glasser et al., 2016; Glasser & Van Essen, 2011; Sigalovsky, Fischl, & Melcher, 2006). Within PAC / core region the above mentioned subdivisions of A1 and RT could be distinguished based on a medial-lateral contrast (Moerel et al., 2014). Unfortunately, little is know about the myelination of auditory cortex ROIs outside the STP, but first results indicate a prominent drop of myelin density at the border of STP and STG (Glasser et al., 2016). It is worth mentioning that the last two procedures confirm that the PAC / core regions are indeed only entail the medial two-thirds of the HG in case of non-duplicated HGs and the medial two-thirds of the anterior

HG in case of duplicated HGs (Braak, 1978; Clarke & Rivier, 1998; Emmorey et al., 2003; Gage & Baars, 2018; Gage et al., 2009; Galaburda & Sanides, 1980; Hackett et al., 2001; Hubl et al., 2010; Leonard et al., 2001; Penhune et al., 1996; Pfeifer, 1920; Rademacher et al., 1992; Rivier & Clarke, 1997; Schneider et al., 2009; Schneider et al., 2002; Von Economo & Koskinas, 1925; von Economo & Horn, 1930; Warrier et al., 2009; Wong et al., 2008). The recent advancements in neuroscientific, that is neuroimaging, methods enabled researchers to shift from almost exclusively post-mortem and limited structural methods not only to in vivo in depth structural approaches using e.g., ultra-highfield DAPs with submillimeter resolution, but also the investigation of functional characteristics of the auditory cortex. While exhibiting a comparable variability, functional approaches are, in contrast to structural ones, limited to two certain properties of the auditory cortex: its responsiveness to sound, as well as certain acoustic features and the tonotopic organization of its primary regions. However, as the first mainly allows the localization of regions and included voxels that are more responsive to sound than silence or percepts of other domains, e.g., visual, the research field converged on the functional localization and parcellation of the auditory cortex through its tonotopic organization that has been exhaustively investigated via a myriad of analysis methods. Concerning the first approach, especially earlier fMRI studies used a broad range of auditory stimuli to identify brain regions that correspond to sound, covering everything from pure tones to complex speech. The respective results indicated the prominent involvement of multiple regions within the bilateral STG, thus rather broad and unspecific activation patterns that did not allow a precise localization or parcellation (Binder et al., 1994; Wise et al., 1991). Nevertheless, also comparably early studies were able to provide evidence that hinted at hierarchical processing within the auditory cortex based on distinct preferred processing of certain stimuli types within diverging regions of the STG. In particular, one region that was localized within HG through anatomical landmarks responded more to pure tones and other comparably simple sounds, whereas regions that surrounded it seemingly preferred the processing of more complex sounds such as speech (Kaas, Hackett, & Tramo, 1999; Rauschecker, 1998; Saenz & Langers, 2014). Together with the simultaneously found evidence for a tonotopic organization of the same region (Bilecen et al., 1998) that was additionally proposed to constitute PAC, these results allowed the functional distinction between primary and non-primary auditory cortex, or core and belt / parabelt regions. Throughout the last two decades this distinction was supported by the results of numerous studies (Y. E. Cohen, Bennur, Christison-Lagay, Gifford, & Tsunada, 2016; de Heer, Huth, Griffiths, Gallant, & Theunissen, 2017; Langers, Backes, & van Dijk, 2007; Norman-Haignere et al., 2015; Overath, McDermott, Zarate, & Poeppel, 2015) with some of them even being capable of creating computational models that could probe and reflect this hierarchical organization (Kell, Yamins, Shook, Norman-Haignere, & McDermott, 2018; Norman-Haignere & McDermott, 2018). Even though widely debated (Moerel et al., 2014; Saenz & Langers, 2014; Striem-Amit, Hertz, & Amedi, 2011), nonprimary regions of the auditory cortex lack the specificity of their primary counterparts with regard to tonotopy along other functional properties which is why their localization and parcellation is even more challenging and less reliable (Di Salle et al., 2003; Schreiner & Winer, 2007). It is worth mentioning that there is evidence for a tonotopic gradient in belt as compared to parabelt regions, with their border however being only unreliably drawn (Belin et al., 2000; Bendor & Wang, 2005; Timothy D. Griffiths & Hall, 2012; Hackett et al., 1998; Kajikawa et al., 2015; Moerel et al., 2014; Moerel et al., 2012). Within the light of the preferential encoding of complex sounds, recent research work focused on a variety of acoustic features and their respective mappings within the auditory cortex in order to evaluate if these functional properties allow a differentiation and thus parcellation between regions (Santoro et al., 2014; Walker, Bizley, King, & Schnupp, 2011). The respective results exhibited a large overlap between features that were yet decodable and through the investigation of spectrotemporal modulations yielded a distinction between anterior and posterior regions of the auditory cortex (Allen et al., 2018; Alluri et al., 2017; Alluri et al., 2012; Alluri et al., 2013; De Angelis et al., 2018; Timothy D. Griffiths & Hall, 2012; Santoro et al., 2014; Toiviainen, Alluri, Brattico, Wallentin, & Vuust, 2014), "...that is posterior/dorsal auditory regions preferably encode coarse spectral information with high temporal precision." and, "...neuronal populations in anterior/ventral auditory regions preferably encode fine-grained spectral information with low temporal precision." (Santoro et al., 2014, p. 1). However, the vast majority of studies investigating the functional properties of the auditory cortex and its regions utilized and integrated anatomical definitions when concerned with approximate localizations and parcellations. For example, Norman-Haignere, Kanwisher, & McDermott (2013) used structural and functional properties to parcellate the auditory cortex in five different regions that run along a posterior-anterior axis within the STP and displayed an equal response to sounds, while at the same time an increasing response to pitch in the anterior direction. Due to the aforementioned problems, a substantial amount of research work focused on primary auditory cortex regions and its tonotopic organization. As noted in 1.2., the primary auditory cortex displays a tonotopic organization which is comparable to the cochlea and thus exhibits a frequency-place code. The precise expression of this organization is subject of a lively debate (Baumann et al., 2013; Moerel et al., 2014; Saenz & Langers, 2014) and changed throughout recent years from a single low-to-high gradient running along an anterior to posterior axis within the HG, to a high-to-low-to-high gradient in which medial parts of the HG that preferably encode low frequency are surrounded by regions that prefer high frequencies and finally to versions of the former in which the high frequency regions converge at the posterior end of the low frequency region, thus creating a "V"-shape instead of a posterior-anterior axis (Baumann et al., 2013; Langers, 2014; Moerel et al., 2014; Saenz & Langers, 2014). Traditionally, this gradient and the reversal at its roughly middle part are used to parcellate the PAC into the core regions A1 and R, serving as human homologue to the subdivisions found in the monkey model of the auditory cortex (Da Costa et al., 2011; Formisano et al., 2003; Hertz & Amedi, 2010; Humphries et al., 2010; Kaas & Hackett, 2000; Langers & Van Dijk, 2011; Merzenich & Brugge, 1973; Reale & Imig, 1980; Riecke, Opstal, Goebel, & Formisano, 2007; Schönwiesner, Dechent, Voit, Petkov, & Krumbholz, 2015; Striem-Amit et al., 2011; Talavage et al., 2004; Upadhyay et al., 2008; Woods & Alain, 2009; Woods et al., 2010; Woods et al., 2009). Although the respective research community mostly agreed on the "V"-shape organization, its extent and orientation remain open for discussion as diverging evidence exists. Focusing the first, the described gradient might reflect PAC / core regions and extents to anterior and posterior belt regions within PP and PT respectively or actually includes not only PAC / core regions, but also belt regions (Da Costa et al., 2011; Humphries et al., 2010; Moerel et al., 2014; Saenz & Langers, 2014; Striem-Amit et al., 2011; Talavage et al., 2004; Woods & Alain, 2009; Woods et al., 2010). Addressing the second, two deviating interpretations with regard to the orientation of the gradient exist. While the classic or parallel interpretation assumes a gradient that runs along the HG, the orthogonal or perpendicular interpretation states a gradient that runs across the HG (Baumann et al., 2013; Moerel et al., 2014; Saenz & Langers, 2014). A modification of the latter is the oblique interpretation from Baumann et al. 2013 that is in line with the perpendicular one, but restricts the extension of the PAC / core regions to the HG with only small portions extending it. Despite empirical evidence for both exists (Baumann et al., 2013; Da Costa et al., 2011; Dick et al., 2012; Formisano et al., 2003; Humphries et al., 2010; Langers, de Kleine, & van Dijk, 2012; Lutti, Dick, Sereno, & Weiskopf, 2014; Striem-Amit et al., 2011), recent studies and developments seem to be in favour of the perpendicular / oblique interpretation (Baumann et al., 2013; Saenz & Langers, 2014). As mentioned above, the extent of tonotopic organization within regions outside PAC / core, that is belt and parabelt, remains questionable, yet a limited amount of studies that point to several identifiable regions that are localized through their frequency specific response and placed around PAC / core regions in anterior and posterior direction (Herdener et al., 2013; Humphries et al., 2010; Langers & Van Dijk, 2011; Moerel et al., 2014; Moerel et al., 2013; Striem-Amit et al., 2011; Talavage et al., 2004). The tremendous differences found in functional approaches to localize and parcellate the human auditory cortex are grounded in two prominent sources of variability: stimuli and analysis methods. Wheres most studies rely on pure-tone stimuli, the specific range and characteristics deployed so far vary tremendously. Overall, mid range frequencies are used based on their importance in everyday hearing situations and in order to avoid confounds due to slight deteriorated HTs in extreme low and/or high frequencies that are common in the general population (Cha, Zatorre, & Schönwiesner, 2016; Humphries et al., 2010; Norman-Haignere, Kanwisher, & McDermott, 2013; Saenz & Langers, 2014; Schönwiesner et al., 2015). More precisely, the frequency range covers 8-10 logarithmically spaced frequencies from 200 Hz to 8 kHz in average. Continuing with the paradigm those frequencies were implemented in, two distinct types of to be distinguished that are based and dependent on the aforementioned gradient and planned analysis technique: ascending/descending or random sequences. While the first presents pure-tones in an ascending or descending order, that is from low to high or high to low frequencies (Barton, Venezia, Saberi, Hickok, & Brewer, 2012; Da Costa et al., 2011; Striem-Amit et al., 2011; Talavage et al., 2004), the second applies a pseudorandom order of pure-tones. Leading to possible variations due to the applied analysis approach, the first is intended to be within the most often applied methods, best-frequency maps and phase-encoded or traveling waves. With all being voxel-wise approaches, best-frequency maps are based on which frequency yields the highest amplitude for a given voxel (as compared to a baseline) (Cha et al., 2016; Formisano et al., 2003; Saenz & Langers, 2014; Schönwiesner et al., 2015), while on the contrary, phase-encoded or traveling waves indicate the preferred frequency for a given voxel through the phase as indicated through fourier transformations or cross-correlation analysis

(Da Costa et al., 2011; Da Costa, van der Zwaag, Miller, Clarke, & Saenz, 2013; Saenz & Langers, 2014). With regard to the latter time-lagged functions based on the ascending or descending pure-tone cycles are created and correlated with each voxel's time course, with the highest correlation determining a given voxel's prefered frequency, therefore inducing a traveling wave through the primary regions of the auditory cortex, resembling the above described gradient (Da Costa et al., 2011; Da Costa et al., 2013; Saenz & Langers, 2014). Going one step further, analyses that target the complete response profile of a given voxel to certain frequencies, thus their tuning width, provide more detailed insights as compared to the former that purely focus on frequency preferences. Tuning width approaches comprise two further advantages, which are the independence of stimuli type and their order. For example, a growing number of studies utilized a broad range of natural sounds in order to compute tonotopic maps that were highly comparable to those obtained through more standard pure-tones (De Martino et al., 2013; Moerel et al., 2012; Saenz & Langers, 2014). The ascending/descending order of stimuli demanded by the so far prominently applied traveling wave design does not pose a necessity and therefore lead to the second distinguished paradigm, random orders. Here, especially population receptive field (pRF) models as an extension of tuning widths approaches seem promising. Originating from research focusing the retinotopy of the visual cortex (Bridge, 2011; DeYoe et al., 1996; Dumoulin & Wandell, 2008; S. Engel, 1997; D. F. Smith & Gedayt, 2001), pRF methods apply a linear temporal model to the timeseries of a given voxel (Thomas et al., 2015). Another method that does not depend on a specific stimuli order was proposed by Schoenwiesner et al. (2015) and entailed a machine learning method, more precisely a searchlight analysis, that relied on the frequency place coding that defines tonotopy. As a searchlight travels through the cortex in spheres of certain voxels, revealing voxels and regions of the brain that are able to distinguish between conditions (stimuli types, cognitive processes, etc.), an application to e.g., pure-tones should yield high prediction accuracies in the (primary) auditory cortex which was supported by the outcome of their study. An addition to the aforementioned analysis methods, a limited amount of studies were also able to provide evidence for different integration of regions within, from and towards the auditory cortex through connectivity measurements. Here, Cha et al. (2016) could show that voxels with a similar tuning width profile exhibited a high functional connectivity with one another, while this was diminished for voxels with a diverging profile. It is furthermore assumed that the cyto-/chemoarchitectural, as well as myelin properties of the primary auditory cortex reflect a high density of thalamocortical connections originating from the MGB (Glasser et al., 2016; Moerel et al., 2014). Besides the paradigms and its stimuli, the planned analysis method furthermore influences the choice of experimental settings and DAPs, as for example a large number of volumes and hence amount of data is prefered with regard to stability and statistical power, this is especially important for e.g., machine learning approaches which furthermore demand a certain number of runs and/or sessions. However, the need of sufficient data is in contrast to the noncontinuous DAPs typically used in auditory neuroimaging and their aim to enable a presentation and perception of sound that is as distortion-free as possible and thus in turn low on confounds, which was addressed in the previous section on individual HTs. As part of an ongoing discussion, prior research work suggests that continuous and noncontinuous DAPs perform equally well to a certain extend (Saenz & Langers, 2014; Scarff et al., 2004; Woods & Alain, 2009) while others showed evidence for the superiority of noncontinuous DAPs when investigating functional properties of the auditory cortex (De Martino, Moerel, Ugurbil, Formisano, & Yacoub, 2015). Comparably, tonotopic maps remain stable across different field strength and voxel size, even though higher field strength and smaller voxel sizes entail more fine-grained information that is otherwise lost (Saenz & Langers, 2014). As noted above, a fair amount of research work, especially those that focus functional approaches, integrate insights from anatomical properties and therefore apply a multimodal approach. Most often, anatomical information is used to gather an approximate outline of the auditory cortex and its regions within which functional properties like tonotopy can then reveal subparcellations (De Martino et al., 2012; Formisano et al., 2003; Norman-Haignere et al., 2013). Furthermore, different functional approaches capturing distinguished functional properties can be combined (Cha et al., 2016; Schönwiesner et al., 2015). The tremendous advantage of multimodal modal approaches was demonstrated by recent study from Glasser et al. (2016) that combined a multitude of structural and functional properties, resulting in the yet most extensive and fine-grained parcellation of the human neocortex, including the auditory cortex. As part of the Human Connectome Project (their localization and parcellation of the auditory cortex is based on a diverse set of structural (e.g., folding, myelin, thickness) and functional (e.g., speech comprehension, functional connectivity) features and yielded the above described and widely accepted distinction in primary and nonprimary auditory cortex regions, with the latter being termed auditory association cortex and both further parcellated in multiple subregions. However, while entailing all benefits of the diverse approaches, multimodal approaches also comprise their problems, as all underlying properties fluctuate immensely across participants, including size and structure of the regions (e.g., duplications of HG (Marie et al., 2015)), myelination, tonotopic maps, etc., thus effectively not allowing any evaluation of the stability of their respective results. This interindividual variability results in the need for individual assessments with localizations and parcellations on the group level only being feasible for displaying main and general properties (Moerel et al., 2014). Summarizing the aforementioned points, the localization and parcellation of the auditory cortex is influenced by a myriad of factors, that together with the interindividual variability so far did not allow a consensus, except on a rather broad level. This encompasses a distinction between primary and nonprimary regios, that is PAC / core, belt and parabelt regions that are distributed across the STP and STG, with the PAC / core regions being situated in parts of the HG and exhibiting a "V"-shaped tonotopic organization that extends, while diminishing, towards belt and parabelt regions. As the further and detailed discussion and focus on more fine-grained subdivisions of the auditory cortex and its regions is unfortunately outside the scope of the here presented thesis, the richness of detail will not be extended and the reader is advised to consult respective overview articles (Baumann et al., 2013; Moerel et al., 2014; Saenz & Langers, 2014).

Open science as a basis to further advance methods of auditory neuroscience?

While both of the above discussed problems are based on different circumstances specific to auditory neuroscience they are unified by a third factor that additionally extends to most if not all scientific fields: the absence of open science and its included principles. With regard to the complication at hand, this especially refers to resources, more precisely analyses and data. Not only leading to low reproducibility and replicability rates (Gilmore, Diaz, Wyble, & Yarkoni, 2017; Millman, Brett, Barnowski, & Poline, 2018; Nichols et al., 2017; Picciotto, 2018; Turner, Paul, Miller, & Barbey, 2018) but also slowing down scientific advancement, a closed scientific environment is furthermore at the very core of our currently limited understanding of auditory processing. Given the two discussed problems auditory neuroscience faces, the second, that is the localization and parcellation of auditory cortex, is in particular affected, as the first, i.e. individual HTs in MRI settings, were not investigated so far. Within that two further distinctions arise, one between anatomical and functional approaches and one between analyses and outcomes. In more detail, the outcomes of anatomical approaches by means of available ROIs are shared more often, both within atlases and for themselves as compared to the results of functional approaches. They however unfortunately display two drawbacks. The ROIs are restricted to group maps and are therefore in a common reference space like the MNI-space which does not allow to address or consider the described colossal interindividual differences. Furthermore, the individual ROIs along with the raw data and analysis code are not shared, with the same holding true for the functional approaches. Independent of the approach and outcome, if the analysis code is provided, the vast majority relies on closed proprietary software that together with a lack of documentation allows little to no reproducibility or further application. This extends to the description of utilized methods and deficiency of available metadata. The thus tremendously needed standardization and automatization that can serve as a basis for further investigations and advancements are hence impossible to achieve. However, like any other research field, auditory neuroscience would highly benefit from publicly available rich and large datasets, including analysis methods that are possible in the upcoming age of big data. This especially holds true for the limited complexity and quality of datasets and the included experiments (Hamilton & Huth, 2018), usually small sample sizes found in auditory neuroscience and the therewith associated confounds (Turner et al., 2018) and the increasing complicatedness of analysis methods.

3.2. THE PRESENT PROJECTS

The initial projects of the here presented thesis therefore aimed to address the discussed problems present in auditory neuroscience by means of two toolboxes that completely rely on open source software and build upon open science principles that involve current best practices with regard to data sharing, management and reproducibility in order to create a basis for investigations within the realm of auditory neuroscience, not only for the thesis at hand, but the entire research world. To this end, the first project was concerned with developing a toolbox that allows the assessment and evaluation of individual HTs in MRI settings by means of measurements that follow the guidelines of a pure tone audiometry. To enable a precise investigation of ASN influence on hearing abilities, the toolbox furthermore supports the analysis of acoustic features of both DAP corresponding ASN and auditory stimuli intended to use in experiments. Through a combination of the above mentioned information informative visualizations by means of audiograms are created automatically within the supplementary functionality of the toolbox. This part additionally contains an open online repository where the obtained data, as well as metadata are stored and gathered across researchers and subsequently can be used for meta analyses and queries in order to apply an open by default investigation of individual HTs and hearing abilities in MRI settings from the beginning on. Within the second project a toolbox targeting the automated localization and parcellation of the human auditory cortex was developed. Comprising a variety of structural and functional approaches, the goal of the toolbox was twofold: enabling an in depth investigation of the auditory cortex and its structure and function, as well as the creation of auditory ROIs that can be utilized in subsequent analyses. While the structural approaches include automated parcellations, the extraction of ROIs from brain atlases and fine-grained registrations and mappings, the functional approaches contain two different paradigms and a broad range of analysis methods. Focusing a multimodal approach, the toolbox also allows the integration of differently obtained parcellations and the assessment of their specificity. Another part of this supplementary component is involved with the curation of an open online repository where researchers can upload, download and evaluate parcellations. Through an implementation based on complete open source code and usage of container technologies, along with a high amount of documentation and usage examples both toolboxes aim to provide a reliable and robust framework for the further standardization and automation of important basics of auditory neuroscience that are open to the whole research community.

3.3. PROJECT 1 - AIN'T NO SOUND LOUD ENOUGH - AUDIOMETRY-LIKE MEASUREMENTS IN MRI SETTINGS

Hereinafter details regarding the first project, that is the toolbox “Ain't no sound loud enough? (ANSL) - audiometry-like measurements in MRI setting”, will be provided. After a description of the structure and included functionality of the toolbox, its application and utility will be depicted on a sample dataset.

3.3.1. MATERIALS & METHODS

TOOLBOX OVERVIEW AND STRUCTURE

ANSL is divided into three major components through which combination and integration the toolbox' functionality is achieved (please see Figure 3.3). These components entail audiometry experiments, analysis tools and supplementary functionality. While the first two are necessary to obtain individual hearing thresholds, some of their subfunctions as well as the third are optional and can be deployed if the user desires so. The typical application of ANSL follows the subsequent procedure. After the participant has been placed inside the scanner and initially necessary DAPs like localizer measurements have been conducted, the experimenter starts the toolbox before the main experiment that is intended to run and indicates the participant identifier, as well as participant demographics (e.g., age, sex, know hearing disorders, etc.), the broad type of DAP (e.g., no DAP, structural, functional, etc.) and metadata regarding the custom setup (e.g., operating system, headphones, etc.). Afterwards the audiometry measurement is started by choosing one of three respective implementations (further information is provided below in 3.3.1.2). Upon completion the obtained data is automatically analyzed to compute individual HT. Within this step the user has the possibility to add further information that allow the assessment of specific hearing abilities with regard to the used DAP and auditory stimuli of the main experiment (a detailed description can be found in 3.3.1.4). Finally, a participant specific folder is created that contains all acquired data in different file formats, which all contain the participant identifier in order to avoid confusion and overlap. These files furthermore follow the Brain Imaging Data Structure (BIDS) (Gorgolewski et al., 2016) to facilitate standardization, reproducibility and data sharing (the section 3.3.1.4 contains more information on this). Participant demographics are saved in simple text format, that is tab-separated values, while information with respect to the audiometry (version, type of DAP, etc.) and the setup are captured within JSON files. The results of the HT analyses are stored as comma-separated values in simple test format and additionally are visualized by means of a line plot to allow an easy and fast visual assessment. In case the user added further information, multiple graphics are created (please see 3.3.1.4). Based on the respective outcomes, the experimenter can then choose to conduct the main experiment or refrain from it. Notably, to address the aforementioned absence of open science in the realm of auditory neuroscience, the entirety of the toolbox is based on open-source software including python and audacity (https://www.audacityteam.org) and exclusively deploys open and basic file formats. More precisely, the audiometry measurements are implemented in PsychoPy (Peirce, 2007), analysis in python and audacity and the supplementary functionality completely in python. Through the BIDS structure and additions that are currently being developed, experimenters that use the toolbox are furthermore encouraged to automatically upload the acquired data in an anonymized way to an online repository in order to provide other researchers access to the data and enable large scale analyses, along with searchable datasets. In the following chapters further information on the respective components will be given.

STIMULI AND PARADIGM

Traditionally, HTs are assessed by means of an audiometry, which in its core refers to the assessment of hearing abilities of a given individual. However, its specific implementations was subject to different adaptations and types since it has been first described in 1802 (Feldmann, 1992). Nowadays, a broad range of audiometry measurements are distinguished with the three most prominent, that is electrophysiologic audiometry, otoacoustic audiometry and pure-tone audiometry being divided into two groups: objective and subjective audiometry. The first group includes measurements that do not require any form of active engagement from the tested individual as the respective measurements are purely based on physical, acoustic or electrophysiologic properties. As such, electrophysiologic and otoacoustic audiometry are heavily used to assess HTs in clinical settings, but less in research. While the first uses evoked po-

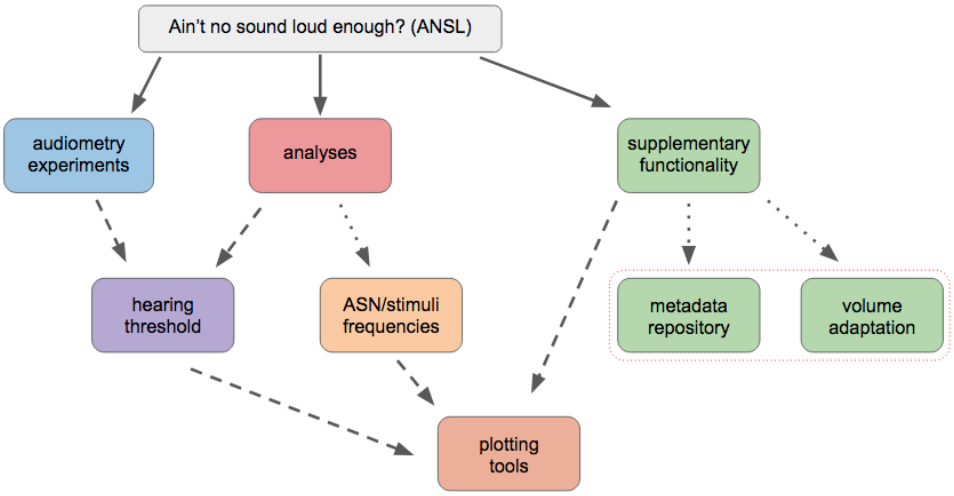


Figure 3.3: *Graphical depiction of the ANSL toolbox and its three major components (solid arrows) including the corresponding functionality. ANSL comprises audiometry experiments (blue), divers analyses (red) and supplementary functionality(green). Dashed arrows indicate core functionality, while dotted arrows imply auxiliary functions. Dotted red contours display parts of the toolbox that are currently not fully supported but in development.*

tentials (e.g., auditory brainstem response (Jewett & Williston, 1971), auditory steady state response (Wilding, McKay, Baker, & Kluk, 2012) or vestibular evoked myogenic potential (Cal & Bahmad Jr, 2009) and is conducted mostly within neurological contexts, the latter utilizes otoacoustic emissions (Kemp, 2002) and is found predominantly in newborn hearing screenings as it can differentiate between sensory and neurological components of hearing abilities. The second group of audiometry measurements demands an active participation from the participant. Here, the pure-tone audiometry (PTA) is the most common approach and involves the presentation of pure tone frequencies (PTF) that range from 250 Hz to 8 kHz through air and bone conduction to which individuals must response once they become audible. Independent of the type of audiometry, an audiogram is the most common outcome and indicates HTs based on intensity (volume in dB) and frequency (in Hz). As audiometry type for ANSL a version of PTA that closely follows the guidelines for PTA from the American Speech-Language-Hearing Association (Association (ASHA), 2005) was chosen. This was motivated by mainly two reasons. First, the equipment necessary for conducting objective audiometry measurements is mostly not compatible with MRI scanners due to the interaction of the electronic components of the first and the strong magnetic field of the latter, that would most likely result in both injuries and financial loss. Furthermore, even if compatible equipment exist, e.g., MRI compatible EEG, the vast amount of researchers do not have access to such devices and the additional preparation time and instruments might lead to other confounds that do not justify the application. Second, PTA offers an advantage over other audiometry measurements as it provides PTF specific and therefore detailed HT profiles. The PTA version included in ANSL differs in multiple core concepts from traditional PTAs in order to allow an adaption for MRI settings. While conventional PTAs assess HTs for each ear independently therefore being monaural, ANSL uses a binaural PTA. Even though ear specific HTs are desirable as they enable an in depth profile, the thus increased time in the MRI setting would lead to an increase in costs and fatigue, with the latter potentially affecting the subsequent main experiment. Additionally, the majority of neuroimaging experiments that involve auditory stimuli employ a binaural presentation. Furthermore, the adaptive design of PTA trials is altered in two ways, as participants do not have to indicate when a tone becomes audible and inaudible within the same, but different trials in order to avoid carry-over effects due to the ASN and PTF are terminated after a response and not repeated within the increase 5 dB decrease 10 dB approach. The PTF HT is hence not based on the 50% criterium of conventional PTAs that are based on individuals responding to a PTF at a certain volume in 50% of the trials or more. Thus, ANSL traded precision for standardization to a certain extent as individual HT might not be as exact as in conventional PTA, but more standardized with regard to effects on the population level aiming to unravel possible interaction effects between ASN and HTs. Additionally, the number of tested PTFs was increased from 7 to 13 to obtain a more fine-grained profile of HTs. The aforementioned adaptions lead to a PTA that is described within the following. 13 PTFs ranging from 150 Hz to 8 kHz in 10 volumes settings ranging from -90dBFS to -10dBFS (in steps of 10dBFS) were included and each presented four times within a paradigm based on a 2x2 design with the factors order of frequencies and volume change. The range of PTFs as a subset of the frequency range audible by humans (20 Hz to 20 kHz) is motivated by the prominence and immense rate of these frequencies in everyday and important listening situations, like e.g., speech in conversations. The volume range was set to avoid exceedance of the PEL in decreasing trials, while at the same provide a high chance of perceivableness in increasing trials. A graphical depiction of the different trial types can be found in Figure 3.4. The two order types are based on conventional PTAs (1 kHz to 8 kHz, followed by 500 Hz to 150 Hz) (order 1) and its reversed form (order 2). Together with either increasing or decreasing in volume, the following four trial types are presented: PTF order 1 - increasing, PTF order 1 - decreasing, PTF order 2 - increasing and PTF order 2 - decreasing. Participants are instructed to either press as soon as they are sure that a PTF is audible (increasing trials) or as soon as they are sure that a PTF became inaudible (decreasing trials). Each PTF is presented for 2 sec. at each volume, with the volume being increased (increasing trials) or decreased (decreasing trials) if participants do not respond at a given volume. In case of a response, the next PTF is presented at the trial corresponding volume start level (-90 dBFS in increasing trials, -10dBFS in decreasing trials) after an interstimulus interval of 2 sec. . The trial order is kept identical across participants with a back to back presentation. Notably, as the duration of the PTA is dependent on a given participant's response, a pronounced difference between participants is possible. In case participants always respond to the first volume setting in increasing trials and the last volume setting in decreasing trials (i.e. "unaffected" HT) the duration would roughly be 5-6 minutes. The same accounts for the contrary (i.e. "high affected" HT) as in this case participant would respond to the last volume setting in increasing trials and the first volume setting in decreasing trials. Even though a significant amount of data is necessary to be certain, both extremes seem rather unlikely as participants are expected to show some extent of affected HT dependent on DAP and PTF as motivated by the points mentioned in this chapter's introduction. However, given that the acquired neuroimaging data is in most cases not of interest, this circumstance does not pose a severe problem, as the experi-

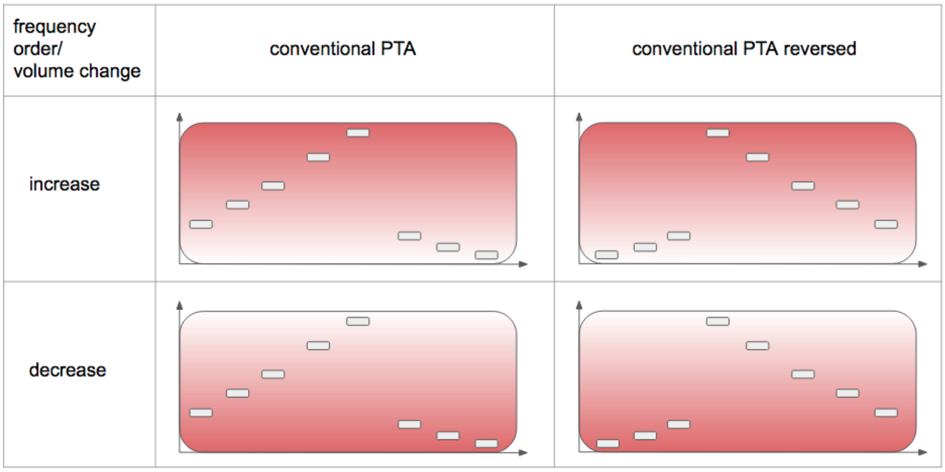


Figure 3.4: *Trial types of the audiometry measurements included in ANSL. Trial types are varied by means of two factors: frequency order (conventional PTA or reversed) and volume change (increase or decrease). X axes indicate time and y axes PTFs. Grey blocks reflect PTFs and shades direction of volume change.*

menter or MRI technician can abort the DAP once the PTA is completed.

ANALYSIS

The analysis components of the ANSL toolbox comprises two distinctive parts that can however be combined to provide further insights into individual HT and abilities. The first part is concerned with the analysis of the obtain participant responses and analyzed automatically, while the second part focuses on characteristics of DAPs, as well as auditory stimuli and is only conducted if the experimenter desires so. The respective parts are further described hereinafter.

Part I - HT

Upon completion of the PTA, HTs are automatically computed. To this end the volume setting during which the participant responded to a given PTF, as well as the trial type is extracted from experiment logfile created through PsychoPy and saved as a CSV file using custom code written in python. Subsequently, responses from both order trial types (conventional and reversed) are averaged within (increase or decrease) and across volume change trial types using NumPy's (Oliphant, 2006) mean function and saved in a separate CSV file. The thus created file therefore contains individual PTF specific HTs that can furthermore be evaluated with respect to different settings: an overall HT and HTs specific for increasing and decreasing trials. This option was included in order to enable an assessment of both stability of HTs and possible effects of volume change direction that might lead to a response or cognitive strategy bias, i.e. varying level of response certainty.

Part II - DAPs & auditory stimuli

Aiming to address the aforementioned possible interaction between DAPs, auditory stimuli and HTs, functionality to investigate the first two was included in ANSL. However, as these analyses require additional preparation steps and might not be of interest to all potential toolbox users, these functions are optional and not conducted automatically. Considering the introduced importance of frequencies and their spectrum as well as markedness, an investigation of these properties as found in DAP corresponding ASN and auditory stimuli of a given main experiment appears significantly relevant. To this end, ANSL supports the extraction of the FS from both through functionality from audacity. In more detail, the experimenter can provide sound files (e.g., WAV) from which the FS is extracted by means of a Fast Fourier Transform (FFT) which indicates the markedness of frequencies the sound is made off in Hz, such that prominent frequencies will obtain a higher Hz as compared to frequencies that are rarely present. Notably, auditory stimuli intended to use during the main experiment can just be employed, whereas DAP corresponding ASN need to be recorded through an MRI compatible microphone and saved as a respective file. The outcomes indicate frequencies against Hz and are saved as CSV files. As this might pose as an extra effort, it is strongly recommended in order to enable a precise evaluation of the HT, abilities and hearing situation with respect to a certain experiment. Assuming that all participants will be presented with the same set of auditory stimuli under the same DAP, these steps need to be conducted only once, preferably before the start of the experiment's data acquisition phase, as this might already indicate potential problems for example in cases where there is a large overlap between ASN and stimuli frequencies. The application of the thus gathered information is outlined in the next section.

SUPPLEMENTARY FUNCTIONALITY

ANSL entails supplementary functionality that is organized in two different sections: visualization and dataset curation. They are located within supplementary functionality as they are not core functions of the toolbox, yet pose as very valuable extension. While the visualization section provides automated and informative graphical depictions of the obtained results across multiple levels, the dataset curation section integrates ANSL into the greater auditory neuroscience community through aspects of open and reproducible science. Both are further described in the following.

Visualization

In order to provide a high level of comparability to conventional audiometry measurements that are conducted outside neuroimaging settings, the outcomes of ANSL are also visualized by means of audiograms. Thus they illustrate the assessed HT through frequencies on the x axis against volume of response (threshold in dBFS) on the y axis. To this end a line plot as implemented in seaborn (Waskom et al., 2018) is used, providing the CSV files with the computed HT as input. By default, multiple graphics are created corresponding to the different HTs, one graphic based on the HTs averaged across both trial types, as well as two graphics based on trial type specific HTs. In cases were ANSL was run in multiple DAPs, the aforementioned graphics are created per DAP along an additional set of these graphics that display multiple DAP corresponding HTs respectively. These graphics aim to provide a quick and comprehensive assessment of individual HTs through visual inspection. If DAPs and/or stimuli were analyzed, the experimenter can indicate the generation of a complementary set of graphics in which the FS related information of DAPs and stimuli are integrated into the above mentioned plots. Thereto, the respective frequency spectra are tailored to the range of PTFs in order to have a congruent x axis. By scaling these values into a range from 0 to 1 using scikit-learn's (Pedregosa et al., 2011) MinMaxScaler, the markedness (in Hz) of each frequency is then indicated via drawing one line per frequency which thickness is set through its corresponding scaled markedness value. Thus, areas of the frequency range with darker shades depict that these frequencies were highly represented and vice versa for lighter shades. These overlay plots can be generated per additionally FS, that is DAP or stimuli, or by combining both in one via placing them on top of each other. In addition to the HTs, these plots furthermore reveal possible interactions between HTs, DAPs and stimuli.

Metadata repository

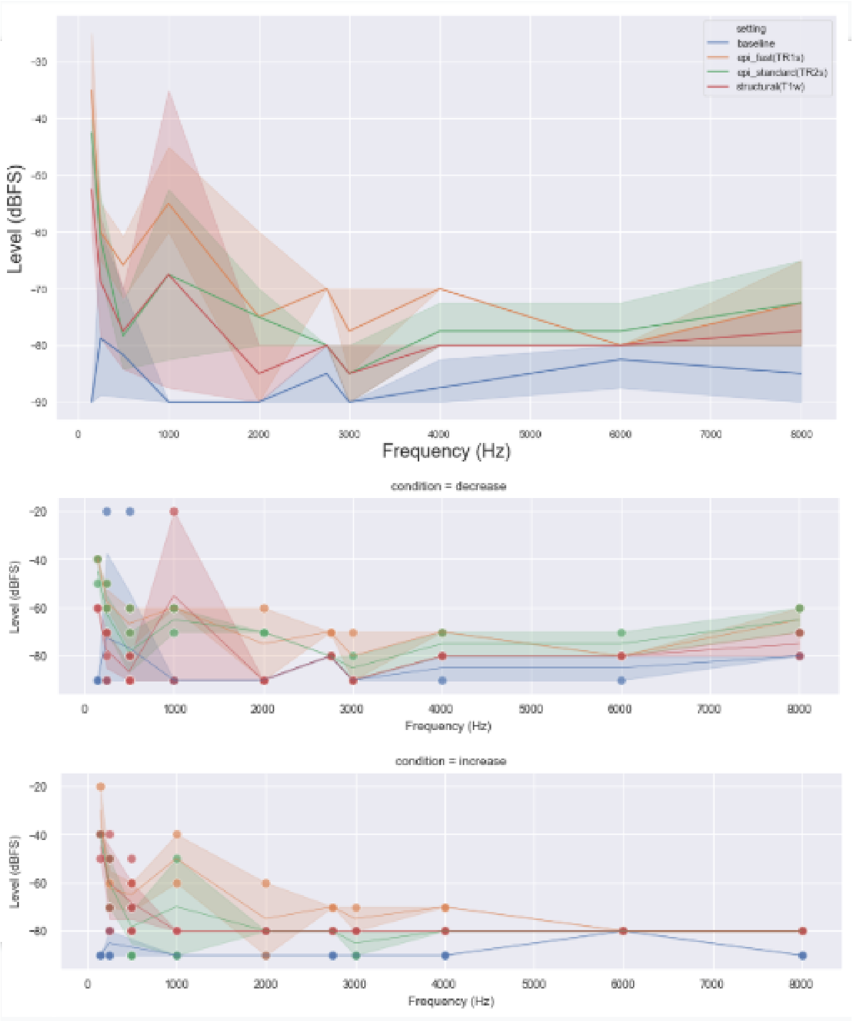
As so far no extensive and profound data with regard to individual HTs exist, the possibly tremendous influence of DAP related HT, especially in relation to deployed stimuli, can barely be estimated. Targeting the creation of a comprehensive dataset that is furthermore enriched with detailed metadata, ANSL is currently (by the time this thesis is written) extended to overcome this gap. Experimenters will have the option to upload the obtained data, that is HTs, participant and setting information, as well as ASN and stimuli frequency spectra in an anonymized manner to an online repository. The ultimate goal of this repository is to provide a comprehensive overview, as well as enable large scale meta-analyses focusing the distribution and characteristics of DAP evoked ASN and their influence on individual HTs in interaction with diverse stimuli sets. This will allow experimenters to estimate these effects for their intended experiment before actually running it, based on their planned DAP and stimulus properties and in turn to adapt certain parameters if necessary. Through an integration with Datalad (https://www.datalad.org), the dataset will be version controlled and searchable, for example to obtain the best suited experimental settings.

3.3.2. EXAMPLE DATA

To demonstrate the usability and necessity of ANSL, an example data set was acquired, measuring the HT of the same participant through PTA across different DAPs. The participant was a male individual with no diagnosed or subjective hearing disorders, neither past or present. The DAPs consisted of three DAPs that are heavily used in neuroimaging settings, hence also auditory neuroscience. In more detail, one structural and two functional DAPs were deployed. The "structural" DAP entailed a whole brain T1 weighted MPRAGE with the following parameters: TR = 1900ms, TE = 2ms, FoV = 256 x 256 mm, flip angle = 9°, voxel size = 1 x 1 x 1 mm and 176 slices. Such DAPs are often utilized to acquire a high resolution structural image that is employed for coregistration or structural analysis approaches. The first functional DAP was specified as follows: EPI, TR = 2000ms, TE = 25 ms, FoV = 192 x 192 mm, flip angle = 70°, voxel size = 3 x 3 x 3 mm and 20 slices. These parameters reflect typical settings used within fMRI, that is BOLD sensitive DAPs. The second functional DAP was completely identical with the exception of a TR of 1100 ms. Such DAPs aim to provide a better temporal resolution through the higher sampling rate which is especially of interest in auditory neuroscience, given the dynamic structure of the stimuli that evolves over time. As the functional DAPs are distinguished by means of their temporal resolution, the first one will be termed the "standard EPI" and the second "fast EPI". Besides the already mentioned prototypical character, these DAPs reflect different types and aspects of ASN, as the structural and the slow EPI had nearly identical TR's but highly distinct TE's and the slow EPI and the fast EPI identical TE's but diverging TR's. Thus the ASN introducing factors gradient switches and RF pulses were varied in a systematic manner. To further stress the influence of DAPs on individual HTs, a baseline HT was assessed through a PTA that was conducted in the absence of a DAP but just laying inside the MRI scanner. All data was acquired on a 3 Tesla Siemens Magnetom Trio, using a 12-channel head matrix Rx-coil and MRI compatible MR Confon HP VS03 headphones. The participant completed the PTA via an MRI compatible keyboard, pressing a button once a PTF become audible or inaudible. The thus acquired data was analyzed through the corresponding component of ANSL and also included the analysis of DAP and stimuli frequency spectra. To this end the ASN was recorded using a MRI compatible microphone (Optoacoustics, http://www.optoacoustics.com) and exported as WAV.

3.3.3. RESULTS

In the following, the results of the above described example data are described. The HT of one individual was assessed within four settings: one baseline without DAP, one structural DAP and two functional DAPs (standard and fast EPI). HTs were computed within setting, as well as within setting and trial type. The respective outcomes are depicted in Figure 3.5 and reveal a prominent influence of ASN on individual HTs that are furthermore highly dependent on the DAP. The HTs deteriorate as a function of ASN introducing factors, as the baseline without any DAP exhibits the best HT and vice versa for the fast EPI which includes the most gradient switches and RF pulses. This is furthermore supported by the observation that the structural DAP and standard EPI DAP, while having nearly an identical TR, resulted in fairly different HTs due to the change in RF pulses. However, this effect was most pronounced for PTFs that were below 4 kHz. Hence, 4 kHz as the middle of the PTF spectrum marked the boundary of a pronounced interaction between PTFs, DAPs and HTs, as lower PTFs showed a high variability across DAP and generally aggravated HTs, while higher PTFs converged towards better HTs across DAPs. This was present for both increasing and decreasing trials. Focusing the relationship between HTs on the hand and ASN, as well as stimuli frequencies on the other, the above mentioned interaction has to be extended. Even though both ASN and stimuli display a broad FS, their lower frequencies are marked the most. While the FS of the example stimuli was also pronounced in higher frequencies, the ASN FS diminished around 4 kHz. Hence, the visual inspection furthermore unveiled a prominent overlap between ASN/stimuli FS and HT variability/values. A respective illustration can be found in Figure 3.6, which displays the HTs and FS of the structural DAP as an example.


condition = decrease

condition = increase

Figure 3.5: *Individual HTs across different settings visualized as audiograms displaying PTFs (in Hz) on the x axis against volume (in dBFS) on the y axis: baseline (orange), structural (red), standard EPI (green) and fast EPI (blue). Top: HTs within setting, across trial types. Middle: HTs within setting within increase trials. Bottom: HTs within setting within decrease trials. Shades display HT variance across trials and dots in the lower two graphics single responses.*

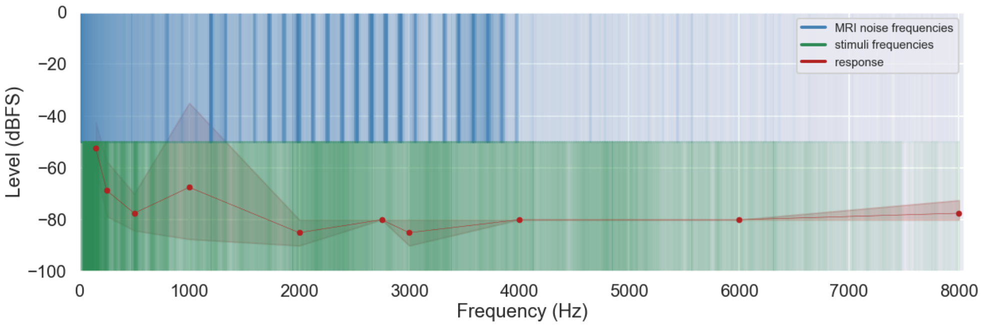


Figure 3.6: *HTs (red dots) as assessed within the structural DAP visualized as audiograms displaying PTFs (in Hz) on the x axis against volume (in dBFS) on the y axis. Shades display HT variance across trials. Blue and green lines depict the FS of the ASN and stimuli respectively, with their thickness indicating the markedness of a given frequency.Thus darker lines and shades reflect parts of the FS that were more pronounced.*

3.4. PROJECT 2 - ALPACA - AUTOMATED LOCALIZATION & PARCELLATION OF AUDITORY CORTEX AREAS

The subsequent chapter outlines the structure and functionality of the seconds project "ALPACA - Automated Localization & Parcellation of Auditory Cortex Areas". Furthermore, its usability is demonstrated on multiple examples.

3.4.1. MATERIALS & METHODS

TOOLBOX OVERVIEW AND STRUCTURE

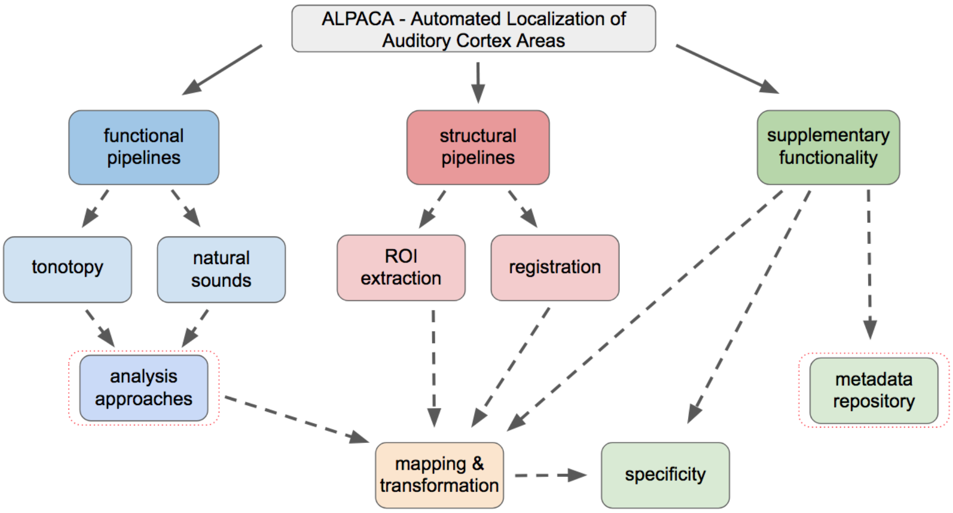


Figure 3.7: *Graphical depiction of the ALPACA toolbox and its three major components (solid arrows) including the corresponding functionality. ALPACA comprises functional pipelines (blue), structural pipelines (red) and supplementary functionality (green). Dashed arrows indicate core functionality. Dotted red contours display parts of the toolbox that are currently not fully supported but in development.*

Overall, the ALPACA toolbox is divided into three components (please see Figure 3.7) which can be run serially or in parallel to obtain optimized results. In more detail, these components consist of functional pipelines, structural pipelines and supplementary functionality. As the application of the toolbox is highly user dependent, different workflows are necessary for the respective components and no prototypical procedure exists. In general researchers have the choice to either run only one of the pipelines or both. While the structural pipeline only requires a structural T1 weighted image in case participant specific analyses in native space are planned or no further data if analyses will be conducted in a reference space. It contains two main parts, the extraction of auditory cortex ROIs from publicly available brain atlases and the registration between participants' native and a reference space. If the experimenter decide to utilize the structural pipeline within native space, the toolbox expects that a participant specific structural image was acquired, processed using the FreeSurfer recon-all pipeline (Dale, Fischl, & Sereno, 1999; Dale & Sereno, 1993; Fischl & Dale, 2000; Fischl, Liu, & Dale, 2001; Fischl et al., 2002; Fischl, Salat, van der Kouwe, et al., 2004; Fischl, Sereno, & Dale, 1999; Fischl, Sereno, Tootell, & Dale, 1999; Fischl, van der Kouwe, et al., 2004; Han et al., 2006; Jovicich et al., 2006; Reuter, Rosas, & Fischl, 2010; Reuter, Schmansky, Rosas, & Fischl, 2012; Ségonne et al., 2004) and the respective outcomes being organized according to BIDS (Gorgolewski et al., 2016). It then makes use of the resulting data to compute a highly precise registration between participants' native and a reference space. With regard to the latter, ALPACA relies on the MNI space due to its prominent usage and data availability within the neuroimaging research world (Brett, Johnsrude, & Owen, 2002). Furthermore, the experimenters can indicate which auditory cortex ROIs from which brain atlas they want to use within their analyses.

FUNCTIONAL PIPELINES

As noted in the former section, as well as in the introduction (3.4.1), two types of paradigms are predominantly utilized in the functional localization and parcellation of the auditory cortex: classic tonotopy, that is pure tones and natural sounds. Targeting an in depth and extensive investigation to enhance standardization, ALPACA includes both while following recent advancements and insights. The same accounts for the analysis methods that have been shown to result in diverging maps and ROIs. To address this part of the investigation, ALPACA includes a range of analysis methods, reflecting different aspects and properties: best frequency maps, pRF and searchlights. While both experiments are conducted through PsychoPy (Peirce, 2007), the analysis methods are implemented in Nipype (Gorgolewski et al., 2011) workflows. Notably, phase-encoding / traveling-waves designs were not included, as they impose restrictions to the paradigms and analysis methods (Saenz & Langers, 2014). Focusing the first, ascending or descending pure-tone sequences are necessary which can introduce "...biases due to habituation, expectation, and estimation artifacts, as well as reducing the effects of spatio-temporal BOLD nonlinearities." (Thomas et al., 2015, p.428) that can subsequently affect analyses. Concerning the seconds, the respective outcomes capture less information as compared to other approaches (Saenz & Langers, 2014; Thomas et al., 2015). Therefore, both paradigms included in ALPACA are based on (pseudo-) random sequences that allow the application of diverse analyses. As both paradigms exhibit a fairly long runtime, no task was given, but a passive listening design implemented in order to reduce fatigue over the time course of the experiments. Hence, participants are only instructed to stay awake while watching a nature documentation of their choice. This was further motivated by research work that provided evidence of a prominent task influence on auditory cortex activation (J. B. Fritz, Elhilali, David, & Shamma, 2007; Häkkinen, Ovaska, & Rinne, 2015; Otazu, Tai, Yang, & Zador, 2009), while at the same the tonotopic organization is only marginally affected by attention (Petkov et al., 2004; Woods & Alain, 2009). A simplified graphical depiction of the paradigms can be found in Figure 3.8.

Classic tonotopy (pure tones)

Based on previous research work, ALPACA includes a version of a classic tonotopy measurement using pure tones that follows the paradigms of Schoenwiesner et al. (2015) and Cha et al. (2016). While the stimuli followed classic tonotopy measurements, the paradigm was adapted to yield robust and reliable tonotopic maps. Pure-tone bursts that were centered around eight nominal and logarithmically scaled frequencies (200, 338.8, 573.8, 971.9, 1646.2, 2788.4, 4723.1 and 8000 Hz) are used as stimuli, with all of them having a 24.4 kHz sampling rate and 24-bit amplitude resolution in order to cover most important fields of normal human hearing abilities and reflect the bandwidths of the cochlear and auditory cortex (Glasberg & Moore, 1990; Schönwiesner et al., 2015). The frequencies are presented for 187.5 ms, plus a 20 ms squared-cosine onset and offset ramp, every 250 ms for 4 seconds to increase SNR and prevent onset distortions (Schönwiesner et al., 2015). Each tone-burst varies in a range of one semitone around the nominal frequency to avoid response adaptation through audible difference that are yet processed in the same cochlear channel (Glasberg & Moore, 1990; Schönwiesner et al., 2015). Furthermore, all stimuli are presented on background noise that exhibit an equivalent rectangular bandwidth to minimize variations in individual hearing thresholds between both, frequencies and participants, as well as inhomogeneities in the transfer functions of the headphones (Glasberg & Moore, 1990; Schönwiesner et al., 2015). A pseudorandom trial order that ensured a balanced transition probability between frequencies is used to present the stimuli, with each frequency being repeated 20 times that together with 160 baseline and 4 initial trials yields a total of 324 trials (Glasberg & Moore, 1990; Schönwiesner et al., 2015). Focusing multivariate and model based analyses, the paradigm from (Schönwiesner et al., 2015) was altered to include 6 instead of 2 runs of equal length to enable a sufficient within participant cross-validation of the parameters. In total the pure-tone paradigm lasts approximately 60 min..

Natural sounds

The natural sounds paradigm was adopted from Moerel et al. (2012) & (2013) using the stimuli from the latter and paradigm of both. A myriad of natural sounds, totalling 160, of various categories ranging from human and animal vocalizations over music to nature scenes and tools are included, with all of them having a sampling rate of 16 kHz and a duration of 1000 ms, adding a 10 ms linear slope to the onset and offset. Furthermore, the energy levels by means of root mean square were equalized. Each sound is presented three times over the course of eight runs in a pseudorandomized event-related design, incorporating a jittered interstimulus interval and baseline trials. In contrast to Moerel et al. (2012) & (2013), no task is given due to the aforementioned reasons and the same passive listening setting as in the pure-tone paradigm is provided. With each run lasting 10 min., the total runtime of the natural sounds paradigm yields roughly 80 - 90 min.

Data acquisition protocol

Technically, the paradigms could be run with any DAP, independent of continuous or non-continuous schemes, which is furthermore more supported by research work that both DAP types perform adequate (Schönwiesner et al., 2015b; Thomas et al., 2015). However, considering the prominent distortions and confounds introduced through ASN as outlined in the chapters concerning the ANSL toolbox and studies showing the advantage of non-continuous DAPs (De Martino, Moerel, Ugurbil, et al., 2015), the paradigms that are included in ALPACA are intended to run within this type of DAP, that is sparse, or preferably ISSS.

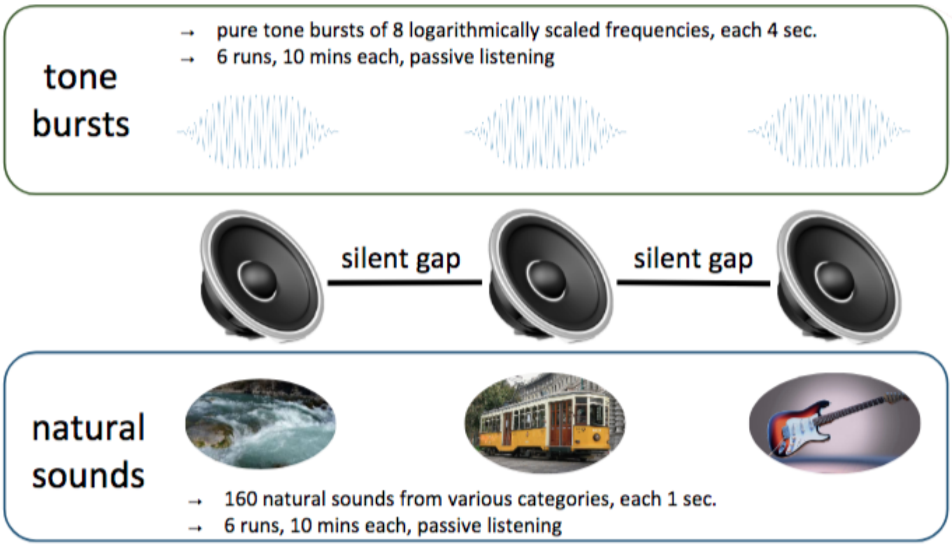


Figure 3.8: *Graphical description of the tonotopy / pure-tone (top) and natural sounds (bottom) paradigms that are included in ALPACA and are utilized to obtain tonotopic maps.*

Analysis

Due to prominent resource restrictions and the near absence of data sharing from previous research work, no functional data could be acquired or existed to complete the functional analysis pipelines. Hence, the hereinafter described analysis approaches are based on development versions. Aiming at a high standardization, reproducibility and quality control across all processing stages, the preprocessing of the functional data is conducted via the preprocessing workflow of fmriflows (https://github.com/miykael/fmriflows), a toolbox that includes a consortium of automated and extensive pipelines for mass-univariate and multivariate processing within which development the author of the here presented thesis is involved. It includes the detection of non-steady-state volumes via Nipype (Gorgolewski et al., 2011), brain extraction through nilearn (Abraham et al., 2014), lowpass filtering via AFNI (Cox, 1996), realignment through FSL, two-stage coregistration with the corresponding structural image implemented in FSL (if desired and the necessary files exist), slice-time correction via SPM (not utilized within ALPACA due to the non-continuous DAP) and temporal filtering through AFNI (Cox, 1996). Additionally, an extensive set of confounds are computed that entails Friston's 24-parameter model for motion parameters (Friston, Williams, Howard, Frackowiak, & Turner, 1996), framewise displacements and DVARS, average signal in the total volume, gray matter, white matter and cerebrospinal fluid, anatomical and temporal CompCor components and independent components (Behzadi, Restom, Liau, & Liu, 2007). As ALPACA recommends precise analysis in participant specific native space (Moerel et al., 2014), smoothing and image transformations (e.g., to a reference space) are not conducted by default, but can however be included if the experimenter desires so. The analysis approaches intended to be included in ALPACA cover standard and advanced methods, including best frequency maps, searchlights and pRF. The first applies a general linear model (GLM) comprising the different conditions (frequencies or natural sounds) on each voxel's time series across all runs using nistats (https://nistats.github.io) integrating all the computed confounds, exclusive the independent components, to diminish distortions as good as possible. Subsequently, the best frequency maps are computed by calculating the centroid of each voxel's response profile, that is the response to the presented stimuli as indicated by the beta weight resulting from the GLM, using functionality from NumPy (Oliphant, 2006) and scikit-learn (Pedregosa et al., 2011). The centroid is used instead of the pure amplitude to adjust for multipeaked response profiles. The second method utilizes the beta weights to run a searchlight analysis that can either be whole-brain or restricted to a mask of the temporal lobe in order to decrease computation time (the latter requires outputs of the structural pipeline). Two searchlight options exist: a classic and a model based one. The classic searchlight aims to identify voxels and regions that distinguish between the presented stimuli by means of a sphere that travels across the volume and applies a multiclass support-vector-machine (SVM), yielding high accuracies in regions where stimuli representations strikingly diverge. The model searchlight approaches the problem from a complementary point of view by using a representational model that constitutes in representational similarity matrix (RDM) displaying unique representations of each stimulus as compared to the remaining ones via assumed a high distances between them. The searchlight then computes an RDM for each sphere position and correlates this RDM based on functional data with the model, where a high correlation indicates voxels and regions that exhibit a representation similar to the model. Both searchlights are implemented in PyMVPA (Hanke et al., 2009), a python package for multivariate analysis of neuroimaging data, and should reveal high accuracies or correlation values near and within the primary auditory cortex, given its frequency place coding. pRF as the third method, creates a predicted time series based on a one-dimensional Gaussian that is applied to the stimuli and a given standard deviation, that is subsequently fit to a given voxel's time series. Here, stimuli pose as best frequency, while the standard deviation defines the tuning width. Using an iterative approach, these values are varied until a combination is found that maximizes the correlation between the predicted and the observed time series for a given voxel, with the respective values being assigned to this voxel. This approach is implemented in popeye (DeSimone, Viviano, & Schneider, 2015), a python package for pRF analysis.

STRUCTURAL PIPELINES

Consisting of two parts, the structural pipelines support the extraction of auditory cortex ROIs from publicly available atlases and the computation of highly precise registrations between participants' native and a reference space. As the individual localization and parcellation of the auditory cortex is still very challenging, a lot of researchers rely on brain atlases that include respective ROIs. Depending on their basis, e.g., structural, functional, multimodal, etc., their number, extent and specificity very tremendously. Addressing this circumstance and enabling one of ALPACA's intended use cases, the gathering of ROIs for subsequent analyses, ALPACA includes a multitude of auditory cortex ROIs originating from a broad range of atlases of varying foundations. Among those are structural, functional, resting state and multimodal ones. While some ROIs were provided as is (e.g., Norman-Haignere et al. (2013)), the majority of ROIs was extracted from atlas files through a combination of nilearn's datasets and Nipype's

PickAtlas functionality. To this end, the atlases were obtained using nilearn and visually inspected for ROIs either that fell within the anatomical boundaries of the STP and STG or were already named accordingly (e.g., "HG"). Using PickAtlas, those ROIs were then extracted and saved as volume files in .nii.gz format. Furthermore, there were special cases in which different approaches were conducted. One example is the multimodal parcellation from Glasser et al. (2016) which required a separation into the different included ROIs. However, as for the remaining ROIs, all are available in ALPACA as volumetric versions within the MNI space. If FreeSurfer's recon-all was run on the structural images, ALPACA furthermore includes functionality to extract and save the ROIs that fell within the temporal lobe of the aparc+aseg.mgz volume, including "transverse temporal", "superior temporal", "middle temporal" or "banks sts" using FreeSurfer's mri_binarize function as implemented in Nipype. As a registration between participants' native and a reference space is necessary to enable across participant and group level analysis, as well as a usage of ROIs that are present in the reference space, ALPACA's structural pipelines entail such a registration. More precisely, a non-linear symmetric diffeomorphic mapping between a given participant's anatomical space and the MNI space through ANT's antsRegistration (Avants, Tustison, & Song, 2009) function as implemented in Nipype. To this end, the skull-stripped structural image as obtained from FreeSurfer's recon-all output is registered to the ICBM 2009c nonlinear asymmetric version of the MNI template space, generating forward (participant to reference space) and inverse (reference space to participant) transformation matrices. In the unlikely case that no structural image was acquired, the mean functional image as computed in the functional pipeline can be utilized, although this is not recommended.

FURTHER FUNCTIONALITIES

The final component of ALPACA, the supplementary functionality, includes three parts: mapping & transformation, specificity and a metadata repository.

Mapping & transformations

The first allows the previously addressed necessary transformations between participant and reference space by utilizing the computed transformation matrices. For example, a certain set of ROIs can be transformed to a given participant's native space or vice versa. If the functional pipeline was also run, an additional transformation between a participant's functional and anatomical space is possible through a concatenation of the transformation matrices obtained within the functional and structural pipeline respectively. This would for example allow either a transformation from ROIs in reference to functional space, or the transformation of tonotopic maps from functional to reference space. Additionally, ROIs can be mapped from volume to surface format or vice versa. To this end ALPACA builds upon the surface reconstructions yielded through FreeSurfer's recon-all and applies FreeSurfer's mri_vol2surf or mri_surf2vol respectively. This functionality accepts ROIs from both, the functional and structural pipeline and can be before or after a transformation. In the former case, ALPACA furthermore applies FreeSurfer's mri_surf2surf, which as all mapping procedures, is implemented in a Nipype workflow.

Specificity

The second part includes functionality to integrate and combine different ROIs, as well as to evaluate their overlay. Through Nipype's version of FSL's ImageMaths the intersection or difference of a set of ROIs can be computed. For example, within the structural pipeline larger ROIs can be parcellated to more fine-grained regions or small ROIs can be combined to create masks. Considering functional and structural pipelines, tonotopic masks can be restricted to certain parts of the HG or used to obtain subdivisions of the latter. As an estimate of concordance between ROIs of different atlases or analysis approaches, ALPACA supports the computation of dice and jaccard indices through Nipype's FuzzyOverlap and stores the results in a matrix that contains the respective pairwise comparisons. This also targets the reported brain atlas concordance problem (Bohland, Bokil, Allen, & Mitra, 2009).

Metadata repository

Aiming to address the prominent differences between localizations and parcellations of the auditory cortex and the resulting absent convergence and standardization that lead to a deceleration of advancements, ALPACA includes an open online metadata repository that integrates with its other components. As this functionality is currently being development (by the time this thesis is written) an outline of the planned capabilities will be provided. Foremost, the metadata repository is concerned with the generation of an extensive dataset that includes detailed metadata and is curated through Datalad based version control and the research community. Following the example of other BIDS-apps like MRIQC (Esteban et al., 2017) or fMRIPrep (Esteban et al., 2019), ALPACA enables the automatic upload and download of anonymized data, while at the same time enabling an interactive exploration that can build the basis for comprehensive and large scale meta-analysis which might be used to unravel the influence of stimulus material and analysis approaches. In order to engage the entire research community and create an open collaboration, the metadata repository will also adapt recent advantages of open science based citizen science frameworks that integrate with big data. Especially, pioneer projects like Braindr (https://braindr.us) and mindcontrol (Keshavan et al., 2018) act as role models for the respective ALPACA functionality which will allow researchers to evaluate auditory cortex ROIs and their localization and parcellation through an interactive web app.

3.4.2. RESULTS

Due to the absence of functional data, ALPACA's functionality and feasibility is depicted through a set of atlas based ROIs of varying complexity in both, volume and surface based format, as well as the exemplary evaluation of correspondence between HGs of different atlases. Focusing the surface examples, Figure 3.9 displays auditory cortex ROIs from multimodal (Glasser et al., 2016), anatomical landmark (Desikan et al., 2006), cytoarchitecture (Morosan et al., 2001) and anatomical-functional (Norman-Haignere et al., 2013) based approaches. All parcellations are displayed on the inflated left hemisphere of the FreeSurfer fsaverage brain.

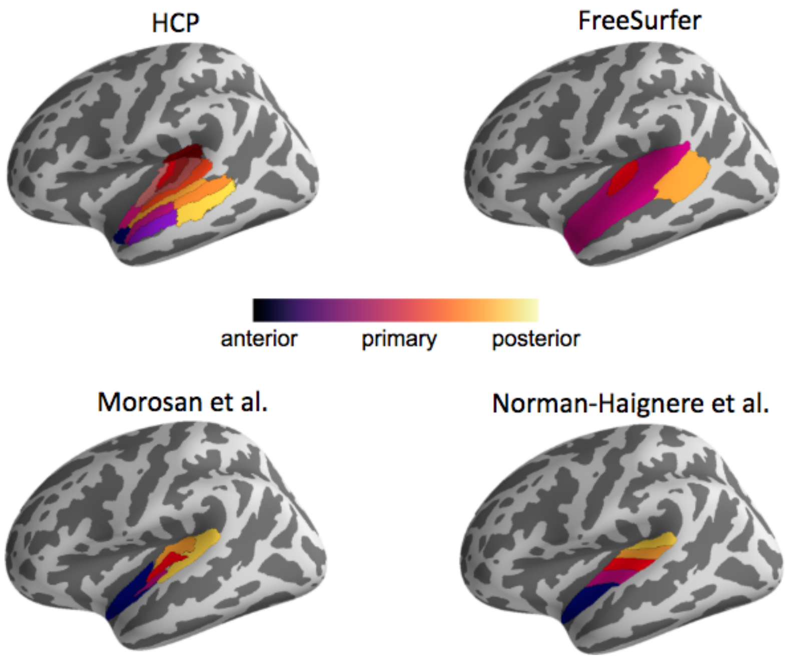


Figure 3.9: *Examples of auditory cortex parcellations in surface format. Parcellations include multimodal (top left, Glasser et al., (2016)), anatomical landmark (top right, Desikan et al., 2006), cytoarchitecture (bottom left, Morosan et al., (2001)) and anatomical-functional (bottom right, Norman-Haignere et al., (2013)) approaches and are displayed on the inflated version of the FreeSurfer fsaverage brain's left hemisphere. Colors indicate a gradient from primary to non primary regions towards anterior and posterior direction.*

The results stress the already noted variability between parcellations, especially those that are based on diverging underlying approaches. While all ROIs are restricted to the STP and STG, their complexity varies to an extreme extent, as for example the parcellation obtained through FreeSurfer's recon-all (Desikan et al., 2006) includes only three ROIs, whereas in contrast the HCP parcellation by Glasser et al. (2016) comprises 12 ROIs in nearly the same coverage of the STP and STG. The same accounts for the distinction between primary and non primary areas, with the FreeSurfer parcellation including one ROI resembling the HG, thus PAC / core regions and one covering nearly the entire STG and the remaining parcellations entailing three (Morosan et al., 2001; Norman-Haignere et al., 2013) or more (Glasser et al., 2016) subdivisions of the PAC / core regions and two (Morosan et al., 2001; Norman-Haignere et al., 2013) or more (Glasser et al., 2016) of non primary regions. The latter is furthermore the only enabling a precise partitioning between core, belt and parabelt regions. Another difference is visible through the borders between ROIs. More precisely, some parcellation display rather fuzzy borders (Glasser et al., 2016; Morosan et al., 2001) while others show more clear-cut margins (Desikan et al., 2006; Norman-Haignere et al., 2013). All of the aforementioned observations are also evident in the volume examples. They included resting state (Varoquaux, Gramfort, Pedregosa, Michel, & Thirion, 2011), macroscopic anatomical (Tzourio-Mazoyer et al., 2002), connectional architecture (Fan et al., 2016) and probabilistic anatomical (Desikan et al., 2006; Frazier et al., 2005; Goldstein et al., 2007; Makris et al., 2006) based parcellations and are displayed on the right hemisphere of the MNI template brain as distributed with nilearn in Figure

3.10. Here, also contrasts between approaches that roughly rely on the same properties, that is functional connectivity became apparent, as the MSDL atlas (Varoquaux et al., 2011) , which is based on resting state measures displays a very broad and unspecific auditory cortex ROI that covers large parts of the STP and STG while not being further parcellated and the Brainnetome atlas (Fan et al., 2016), which is based on functional connectivity, includes a parcellation within a comparably larger extent, that however is divided into six fairly large ROIs. On the contrary the AAL (Tzourio-Mazoyer et al., 2002) and Harvard-Oxford atlas (Desikan et al., 2006; Frazier et al., 2005; Goldstein et al., 2007; Makris et al., 2006) follow the primary / non primary region distinction more precisely and include an approximate version of PAC / core regions within the HG and several non primary regions within the STP and STG. Notably, the AAL atlas comprises two non primary ROIs along a posterior-anterior axis that cover STP and STG, while the Harvard-Oxford atlas (Desikan et al., 2006; Frazier et al., 2005; Goldstein et al., 2007; Makris et al., 2006) contains four non primary ROIs along the same axis that furthermore are divided into STP and STG portions.

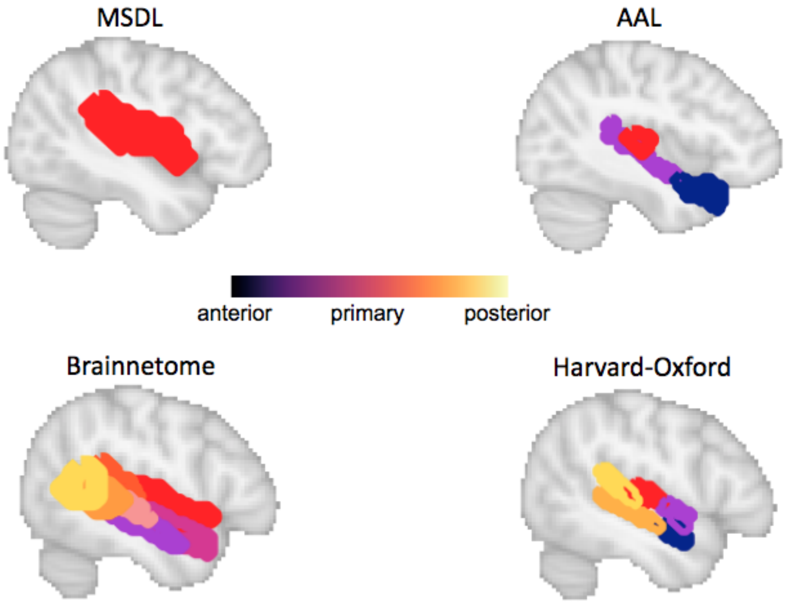


Figure 3.10: *Examples of auditory cortex parcellations in volume format. Parcellations include resting state (top left, Varoquaux et al., 2011, macroscopic anatomical (top right, Tzourio-Mazoyer et al., 2002), connectional architecture (bottom left, Fan, L. et al., 2016) and probabilistic anatomical (bottom right, Frazier et al., 2005; Desikan et al., 2006; Makris et al., 2006; Goldstein et al., 2007) approaches and are displayed on the right hemisphere of the MNI template brain as distributed with nilearn. Colors indicate a gradient from primary to non primary regions towards anterior and posterior direction.*

The ROIs of the volume examples that approximately resemble HG and thus PAC / core regions were additionally evaluated according to their concordance which was assessed by their pairwise dice index. The outcome is illustrated in Figure 3.11 as a matrix containing all respective values and reflects the already mentioned unspecificity of the MSDL atlas as it exhibits a prominent overlap with all other atlas ROIs due to its large size. All remaining atlas ROI yielded an intermediate concordance pointing to a roughly identical location with varying extent.

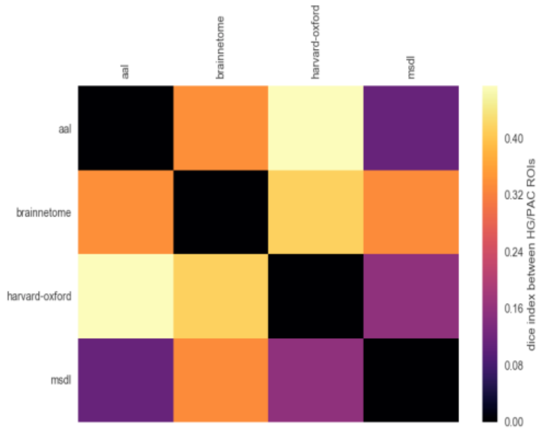


Figure 3.11: *Matrix containing all pairwise dice indices of the HG / PAC / core regions found in the volume examples. The matrix is indexed by the amount of ROIs with their atlas labels being displayed along the x and y axis. Colors indicate dice indices as reflected in the colorbar. Darker colors reflect large overlap and concordance and vice versa for bright colors.*

3.5. DISCUSSION

In the following, the feasibility and necessity of the created toolboxes are discussed based on the exemplary results. Subsequent to a summary of the obtained results, the toolboxes as well as their possible combination will be focused. Finally, an outlook to further functionality will be given.

Targeting the aggravated measurement settings present in auditory neuroscience, that is a highly distorted stimuli perception and deteriorated HTs through ASN, as well as the unreliable and tedious localization and parcellation of auditory cortex areas, the first part of the here presented thesis was concerned with the generation of toolboxes that address these circumstances via a systematic investigation and robust straightforward application. Furthermore, both aim to enable the creation and curation of respective comprehensive databases through an integration with corresponding metadata repositories. To this end both toolboxes are additionally completely based on open-source software and already freely available to enable the application for every interested researcher, independent of financial background and other possibilities, hence following the principles of open and inclusive (neuro)science.

The interaction between DAPs, ASN, stimuli & HTs

The results of the obtained example dataset elucidate that ANSL is capable of assessing individual HTs, as well as unravel the effect of different DAPs and the interaction between the thus generated ASN, stimuli and HTs in a straightforward and easily usable manner. Especially the audiogram visualizations provide a distinct support for experimenters as they allow an immediate evaluation of individual HTs that can additionally be used to decide if a certain participant should be further tested or included at all. In line with previous research, the results furthermore indicate the pronounced effect ASN has on the perception of auditory stimuli in neuroimaging settings (Gaab et al., 2007; Langers et al., 2005; Peelle, 2014, 2018; Ranaweera et al., 2016; Scarff et al., 2004; Talavage & Edmister, 2004; Talavage & Hall, 2012). More precisely, HTs extend the ASN effect on pathways by a new level. Regarding the first pathway, individual HTs can hint to the markedness of the ASN and how severe hearing abilities are affected. Albeit most likely not capable to disentangle the variance in auditory cortex activation introduced through ASN and stimuli of interest, the outcomes of ANSL could be utilized within correlation based approaches to shed some light on the question how strongly hearing abilities relate to the decreased activation proposed by several previous studies (Bandettini, Jesmanowicz, Van Kylen, Birn, & Hyde, 1998; Gaab et al., 2007; Harms & Melcher, 2002; Langers et al., 2005; Peelle, 2014; Peelle, Eason, Schmitter, Schwarzbauer, & Davis, 2010; Scarff et al., 2004; Schmitter et al., 2008; Shah et al., 1999; Talavage & Edmister, 2004; Talavage et al., 1999). However, with respect to the structure and function of the auditory cortex that heavily relies on pitch perception, especially the HG and its tonotopic organization (Moerel et al., 2014; Saenz & Langers, 2014), a new factor, that till now was only marginally discussed and considered, has to be introduced and inherently leads to the second pathway: the distribution of frequency spectra. As the output of the example analysis showed, all important components that create an individual listening experience, i.e. ASN, stimuli and HTs, exhibit highly specific frequency spectra, yet display a large overlap between components. Notably, across all tested DAPs, 4 kHz acted as a threshold between extensively varying and deteriorated HTs on the one hand and stable and acceptable HTs on the other, while at the same time HTs increased as a function of ASN introducing factors, that is gradient switches and RF pulses. In more detail, the higher the amount and prominence of these two factors, the more deteriorated the HT, but almost exclusively for frequencies below 4 kHz. Becoming evident through the audiogramm, this frequency also marks the border between salient and less pronounced parts of the frequency spectra of both ASN and stimuli. This is line with the outcomes of previous research pointing at ASN volume increase due to fast parameters within the corresponding DAP (Ahmed & Shellock, 2001; Price et al., 2001) and that the vast amount of DAPs exhibit a frequency spectrum that is particularly expressed in lower components, i.e. below 4 kHz (Hattori, Fukatsu, & Ishigaki, 2007; Moelker, Maas, Lethimonnier, & Pattynama, 2002). Taken together, this points to confounds that mainly affect the lower end of the frequency spectra, independent of the DAP. This large overlap between ASN, stimuli and HT furthermore results in a heavily changed percept of a given stimulus as for most auditory stimuli used in auditory neuroscience settings only certain portions of their frequency spectrum will be masked and hence altered (Scarff et al., 2004). Technically constituting a degradation of the stimulus, this outcome could moreover be used to create a new sub-pathway within the second pathway that describes the aforementioned ASN based confounds. With regard to HTs, the interaction of frequency spectra poses as a valuable extension to describe individual hearing abilities, especially in MRI settings. It is worth mentioning that particularly research work that is interested in spectral components such as pitch (Oxenham, 2012; Stern, 2017) and timbre (Allen, Burton, Olman, & Oxenham, 2017; Tużnik, Augustynowicz, & Francuz, 2018) and its respective outcomes will suffer from this confound. The same holds true for the extensive body of studies that indicated the highly important role of these acoustic features within the perception and categorization of music (Alluri et al., 2012;

Alluri et al., 2013; Casey, 2017; Casey, Thompson, Kang, Raizada, & Wheatley, 2012; Güçlü, Thielen, Hanke, & van Gerven, 2016; Hoefle et al., 2018; Levitin & Grafton, 2016; Toiviainen et al., 2014). Effectively adding a bottom-up to the already described top-down influence (Dosenbach et al., 2008; Peelle, 2018; Rönnberg et al., 2013; Scott & McGettigan, 2013; Wingfield et al., 2005), this newly introduced ASN confound has to be exhaustively investigated by further studies. However, as HTs heavily depend on ASN and thus DAPs, the volume adaptation approach conducted by the majority of the few studies that actually performed it (Hanke et al., 2015), that is adjusting the presentation volume within a DAP with highly different parameters (e.g., the structural DAP before the main experiment, i.e. functional ones) or the absence of any DAP appears to be not favourable. The above mentioned altered percept of a given auditory stimulus can furthermore also influence the third pathway, not in the sense of physiological and/or psychologically discomfort, but through participant specific stress and annoyance due to the masking of certain stimulus properties. Focusing individual hearing abilities while at the same time expanding automatization and standardization, the outcomes of the example data combined with the characteristics of ASN discussed in this chapter's introduction, important limitations and their improvements in terms of ANSL's foundation and its further functionality become evident. First and most important, a large amount of additional and diverse data has to be collected and curated within ANSL's metadata repository part. This especially concerns DAPs and setup settings. Focusing the first, the example data aimed to provide further evidence of the stringent effects of ASN of heavily utilized continuous DAPs by means of individual HTs, hence prominent DAPs that already address the ASN like sparse imaging and ISSS (please see 3.1) were not tested so far. Therefore, no HT related data for this type of DAP, that was created and is well suited for auditory experiments in MRI settings, exists. Considering the ASN introducing factors, at least the influence of the most distinguished one, that is gradient switches, should be heavily diminished, as auditory stimuli are presented between the gradient switches. Nevertheless, RF pulses as the other main factor are still present and might indeed be more important within this type of DAPs (Peelle, 2014). Another non-ignorable source of HT deterioration that is also present in non-continuous DAPs, although to a smaller extent, is the possibility of short-term (or even long-term) hearing loss and/or tinnitus due to the short-term exposure of ASN that has been reported by numerous previous studies (Brummett, Talbot, & Charuhas, 1988; Dewilde, Grainger, Price, & Renaud, 2007; Mollasadeghi et al., 2013). Regarding this, ANSL could furthermore be utilized to assess and monitor the extent as well as recovery. Another type of DAP that should be considered is a continuous DAP with reduced acoustic noise in which parameters are modified to either change the characteristics of the auditory stimulation (Harms & Melcher, 2002; Seifritz et al., 2006) or its overall volume (Peelle, 2014; Peelle et al., 2010; Schmitter et al., 2008). As for the non-continuous DAPs, ANSL could gather further evidence that stresses the importance and feasibility of such DAPs for auditory neuroscience in MRT settings. Concerning the second necessary data extension, additional information centering around effects based on specific setup settings has to be acquired. Not only the MRI machine itself can have a tremendous impact on ASN and thus individual HTs (Dewilde et al., 2007), but also every other piece of technique that is involved in presenting sound to participants in MRI environments, namely OS and stimulation software, headphones and hearing protection. Already the used OS and stimulation software can introduce large difference in the perception of auditory stimuli even within the identical experiment, as e.g., their underlying libraries might have different sound encodings and/or volume levels differ to a certain amount. Outlined by previous research work, the type of headphone can have a large influence on the perception of auditory stimuli and can additionally initiate new distortions in the signal (Norman-Haignere & McDermott, 2016), not to mention the variation of the headphones themselves which range from in-ear to over-ear and pneumonic to electric forms. Once presented, the percept can additionally be altered through the type of hearing protection or noise control at hand. The majority of studies provided participating individuals with earplugs which pose as a form of passive noise control and hearing protection. Yet, as human ears, earplugs come in a vast variety of shapes and therefore introduce another source of variability as the resulting percept might diverge systematically as a function of the earplug type (hard, soft, etc.) and unsystematically through the combination with participant specific needs. On the contrary, if in-ear headphones are used, no ear-plugs, but a different form of hearing protection has to be provided, e.g. noise diminishing muffs or foam padding. With respect to the aforementioned points, ANSL's metadata repository, once extensive enough, could provide a diverse database that in turn enables meta-analyses to identify combinations of setup parameters that grant the best possible auditory percept, as well as an evaluation of these settings. ANSL's ASN frequency spectrum analysis functionality could be furthermore expanded and integrated with its included audiometry experiments to implement active noise control within which the ASN is used to create "antinoise" that is presented simultaneously with the auditory stimuli in order to cancel parts of the ASN out (Chambers, Akeroyd, Summerfield, & Palmer, 2001; Chambers, Bullock, Kahana, Kots, & Palmer, 2007; Hall et al., 2009; M. Li, Rudd, Lim, & Lee, 2011; Peelle, 2014). Outside the need for further and diverse data, measurements that target other substantial dimensions of auditory perception which were ignored so far should be added to ANSL's functionality, with the most obvious being temporal properties. As outlined in the music theory related chapter of this thesis' introduction (, temporal properties of auditory stimuli, especially music, like meter and rhythm are at least as important during the creation of an auditory percept as their spectral counterparts (i.e. pitch, timbre, etc.). This is especially motivated by meter being identified through both pauses and tones. Comparable to the discussed frequency spectrum interaction, respective effects in the temporal domain are conceivable, e.g., certain types of auditory stimuli, in particular different music genres, are more affected than others. Unfortunately, so far no insights on the potentially altered perception of meter and rhythm in MRI settings exist, which is why the introduction of a meter audiogram in ANSL poses an extremely valuable extension. The results of the obtained example dataset elucidate that ANSL is capable of assessing individual HTs, as well as unravel the effect of different DAPs and the interaction between the thus generated ASN, stimuli and HTs in a straightforward and easily usable manner. Especially the audiogram visualizations provide a distinct support for experimenters as they allow an immediate evaluation of individual HTs that can additionally be used to decide if a certain participant should be further tested or included at all. In line with previous research, the results furthermore indicate the pronounced effect ASN has on the perception of auditory stimuli in neuroimaging settings (Gaab et al., 2007; Langers et al., 2005; Peelle, 2014, 2018; Ranaweera et al., 2016; Scarff et al., 2004; Talavage & Edmister, 2004; Talavage & Hall, 2012). More precisely, HTs extend the ASN effect on pathways by a new level. Regarding the first pathway, individual HTs can hint to the markedness of the ASN and how severe hearing abilities are affected. Albeit most likely not capable to disentangle the variance in auditory cortex activation introduced through ASN and stimuli of interest, the outcomes of ANSL could be utilized within correlation based approaches to shed some light on the question how strongly hearing abilities relate to the decreased activation proposed by several previous studies (Bandettini et al., 1998; Gaab et al., 2007; Harms & Melcher, 2002; Langers et al., 2005; Peelle, 2014; Peelle et al., 2010; Scarff et al., 2004; Schmitter et al., 2008; Shah et al., 1999; Talavage & Edmister, 2004; Talavage et al., 1999). However, with respect to the structure and function of the auditory cortex that heavily relies on pitch perception, especially the HG and its tonotopic organization (Moerel et al., 2014; Saenz & Langers, 2014), a new factor, that till now was only marginally discussed and considered, has to be introduced and inherently leads to the second pathway: the distribution of frequency spectra. As the output of the example analysis showed, all important components that create an individual listening experience, i.e. ASN, stimuli and HTs, exhibit highly specific frequency spectra, yet display a large overlap between components. Notably, across all tested DAPs, 4 kHz acted as a threshold between extensively varying and deteriorated HTs on the one hand and stable and acceptable HTs on the other, while at the same time HTs increased as a function of ASN introducing factors, that is gradient switches and RF pulses. In more detail, the higher the amount and prominence of these two factors, the more deteriorated the HT, but almost exclusively for frequencies below 4 kHz. Becoming evident through the audiogramm, this frequency also marks the border between salient and less pronounced parts of the frequency spectra of both ASN and stimuli. This is line with the outcomes of previous research pointing at ASN volume increase due to fast parameters within the corresponding DAP (Ahmed & Shellock, 2001; Price et al., 2001) and that the vast amount of DAPs exhibit a frequency spectrum that is particularly expressed in lower components, i.e. below 4 kHz (Hattori et al., 2007; Moelker et al., 2002). Taken together, this points to confounds that mainly affect the lower end of the frequency spectra, independent of the DAP. This large overlap between ASN, stimuli and HT furthermore results in a heavily changed percept of a given stimulus as for most auditory stimuli used in auditory neuroscience settings only certain portions of their frequency spectrum will be masked and hence altered (Scarff et al., 2004). Technically constituting a degradation of the stimulus, this outcome could moreover be used to create a new sub-pathway within the second pathway that describes the aforementioned ASN based confounds. With regard to HTs, the interaction of frequency spectra poses as a valuable extension to describe individual hearing abilities, especially in MRI settings. It is worth mentioning that particularly research work that is interested in spectral components such as pitch (Oxenham, 2012; Stern, 2017) and timbre (Allen et al., 2017; Tużnik et al., 2018) and its respective outcomes will suffer from this confound. The same holds true for the extensive body of studies that indicated the highly important role of these acoustic features within the perception and categorization of music (Alluri et al., 2012; Alluri et al., 2013; Casey, 2017; Casey et al., 2012; Güçlü et al., 2016; Hoefle et al., 2018; Levitin & Grafton, 2016; Toiviainen et al., 2014). Effectively adding a bottom-up to the already described top-down influence (Dosenbach et al., 2008; Peelle, 2018; Rönnberg et al., 2013; Scott & McGettigan, 2013; Wingfield et al., 2005), this newly introduced ASN confound has to be exhaustively investigated by further studies. However, as HTs heavily depend on ASN and thus DAPs, the volume adaptation approach conducted by the majority of the few studies that actually performed it (Hanke et al., 2015), that is adjusting the presentation volume within a DAP with highly different parameters (e.g., the structural DAP before the main experiment, i.e. functional ones) or the absence of any DAP appears to be not favourable. The above mentioned altered percept of a given auditory stimulus can furthermore also influence the third pathway, not in the sense of physiological and/or psychologically discomfort, but through participant specific stress and annoyance due to the masking of certain stimulus properties. Focusing individual hearing abilities while at the same time expanding automatization and standardization, the outcomes of the example data combined with the characteristics of ASN discussed in this chapter's introduction, important limitations and their improvements in terms of ANSL's foundation and its further functionality become evident. First and most important, a large amount of additional and diverse data has to be collected and curated within ANSL's metadata repository part. This especially concerns DAPs and setup settings. Focusing the first, the example data aimed to provide further evidence of the stringent effects of ASN of heavily utilized continuous DAPs by means of individual HTs, hence prominent DAPs that already address the ASN like sparse imaging and ISSS (please see 3.1) were not tested so far. Therefore, no HT related data for this type of DAP, that was created and is well suited for auditory experiments in MRI settings, exists. Considering the ASN introducing factors, at least the influence of the most distinguished one, that is gradient switches, should be heavily diminished, as auditory stimuli are presented between the gradient switches. Nevertheless, RF pulses as the other main factor are still present and might indeed be more important within this type of DAPs (Peelle, 2014). Another non-ignorable source of HT deterioration that is also present in non-continuous DAPs, although to a smaller extent, is the possibility of short-term (or even long-term) hearing loss and/or tinnitus due to the short-term exposure of ASN that has been reported by numerous previous studies (Brummett et al., 1988; Dewilde et al., 2007; Mollasadeghi et al., 2013) . Regarding this, ANSL could furthermore be utilized to assess and monitor the extent as well as recovery. Another type of DAP that should be considered is a continuous DAP with reduced acoustic noise in which parameters are modified to either change the characteristics of the auditory stimulation (Harms & Melcher, 2002; Seifritz et al., 2006) or its overall volume (Peelle, 2014; Peelle et al., 2010; Schmitter et al., 2008). As for the non-continuous DAPs, ANSL could gather further evidence that stresses the importance and feasibility of such DAPs for auditory neuroscience in MRT settings. Concerning the second necessary data extension, additional information centering around effects based on specific setup settings has to be acquired. Not only the MRI machine itself can have a tremendous impact on ASN and thus individual HTs (Dewilde et al., 2007), but also every other piece of technique that is involved in presenting sound to participants in MRI environments, namely

OS and stimulation software, headphones and hearing protection. Already the used OS and stimulation software can introduce large difference in the perception of auditory stimuli even within the identical experiment, as e.g., their underlying libraries might have different sound encodings and/or volume levels differ to a certain amount. Outlined by previous research work, the type of headphone can have a large influence on the perception of auditory stimuli and can additionally initiate new distortions in the signal (Norman-Haignere & McDermott, 2016), not to mention the variation of the headphones themselves which range from in-ear to over-ear and pneumonic to electric forms. Once presented, the percept can additionally be altered through the type of hearing protection or noise control at hand. The majority of studies provided participating individuals with earplugs which pose as a form of passive noise control and hearing protection. Yet, as human ears, earplugs come in a vast variety of shapes and therefore introduce another source of variability as the resulting percept might diverge systematically as a function of the earplug type (hard, soft, etc.) and unsystematically through the combination with participant specific needs. On the contrary, if in-ear headphones are used, no ear-plugs, but a different form of hearing protection has to be provided, e.g. noise diminishing muffs or foam padding. With respect to the aforementioned points, ANSL's metadata repository, once extensive enough, could provide a diverse database that in turn enables meta-analyses to identify combinations of setup parameters that grant the best possible auditory percept, as well as an evaluation of these settings. ANSL's ASN frequency spectrum analysis functionality could be furthermore expanded and integrated with its included audiometry experiments to implement active noise control within which the ASN is used to create "antinoise" that is presented simultaneously with the auditory stimuli in order to cancel parts of the ASN out (Chambers et al., 2001; Chambers et al., 2007; Hall et al., 2009; M. Li et al., 2011; Peelle, 2014). Outside the need for further and diverse data, measurements that target other substantial dimensions of auditory perception which were ignored so far should be added to ANSL's functionality, with the most obvious being temporal properties. As outlined in the music theory related chapter of this thesis' introduction (1.1), temporal properties of auditory stimuli, especially music, like meter and rhythm are at least as important during the creation of an auditory percept as their spectral counterparts (i.e. pitch, timbre, etc.). This is especially motivated by meter being identified through both pauses and tones. Comparable to the discussed frequency spectrum interaction, respective effects in the temporal domain are conceivable, e.g., certain types of auditory stimuli, in particular different music genres, are more affected than others. Unfortunately, so far no insights on the potentially altered perception of meter and rhythm in MRI settings exist, which is why the introduction of a meter audiogram in ANSL poses an extremely valuable extension.

The curious case of the auditory cortex

As previously noted, the assessment of ALPACA's functional pipelines was not possible due to prominent restrictions with regard to available resources. Hence, the hereinafter discussion of the toolbox within the light of its functionality will be mostly restricted to the remaining components, the structural pipelines and supplementary functionality. However, already the exemplary visualized and evaluated atlas based ROIs underline the aforementioned variability of auditory cortex localization and parcellation (Moerel et al., 2014). Parcellations range from one ROI based on one feature to 12 ROIs that are based on a myriad of features. This is especially drastic, as all differences originate from diverging approaches but not include the tremendous interindividual variability inherent to functional properties that defines the outlined problems (Baumann et al., 2013; Moerel et al., 2014; Saenz & Langers, 2014). Additionally, it demonstrates the need for further standardization and automation to create a basis for further investigations. In particular, most structural and functional approaches so far rely on a manual definition at some point that increase variability and decrease reliability, as well as reproducibility even further. Hence, an automated localization and parcellation based on standardized stimuli and analysis methods provide the possibility to address this gap. This is in line with the discussed feasibility of multimodal approaches, as they integrate distinct properties covering diverse structural (anatomical landmarks, cytoarchitecture, myelination, etc.) and functional features (tonotopy, spectrotemporal modulations, etc.), with all evidence of usability and importance existing for all of them (Baumann et al., 2013; Moerel et al., 2014; Saenz & Langers, 2014). Given the long lasting and currently emerging problems, the combination of multimodal investigations with automated analyses appear to be highly promising. The same holds true for the application of further analysis methods that are going to be added to ALPACA. Here, encoding models have already been shown to be suited for the examination of the auditory cortex and its properties (Moerel et al., 2013; Moerel et al., 2012; Santoro et al., 2014; Santoro et al., 2017), as they provide two main advantages. The first is the testing of multiple features at the same time which allows to incorporate spectral and temporal aspects, thus capturing both crucial properties any sound is defined through (Santoro et al., 2017), while most methods so far purely focused on the spectral domain. This would allow to extent the assessment of tuning widths to additionally include their response profile over time. The second advantage is based on the application of computational models to probe neuronal responses which enable the testing of both hypothesis and data driven assumptions that can furthermore be used to inform one another. Through their capability of out-of-sample testing they would also allow to validate and adjust the same set of models to different participants and paradigms (Moerel, Martino, Uğurbil, Formisano, & Yacoub, 2018; Norman-Haignere & McDermott, 2018). Another feasible approach, especially considering the automated parcellations, is the computation of gradient-sign maps from the best frequency maps, which are assumed to reveal subdivisions of the human PAC / core regions (Formisano et al., 2003; Petkov et al., 2004). A respective approach was introduced by Schoenwiesner et al. (2015) who used an edge detection filter to mark gradient reversals. Additionally, the multimodality of the investigation itself can be extended through the inclusion of time sensitive methods like MEG and EEG, as even the fastest ISSS DAP will not be able to capture the already stressed temporal properties of the auditory cortex sufficiently which in turn limits analysis approaches, like encoding models. For example, Su, Zulfiqar, Jamshed, Fonteneau, & Marslen-Wilson (2014) showed that spatiotemporal searchlights based on representational similarity analysis using both recorded electrophysiological data from MEG and EEG, as well as models obtained through stimulus features can yield tonotopic maps that are comparable to those obtained via MRI in location and extent.

Finally, a growing body of research work points to the importance of subcortical structures that are involved in auditory neuroscience in order to advance the understanding of their cortical counterparts. Exemplary studies by De Martino et al. (2013) and Moerel et al. (2015) revealed that also subcortical structures like the IC and MGB exhibit a tonotopic organization which, together with their ascending pathways towards cortical structures and the dense corticothalamic connections of the auditory cortex, can be used to achieve more detailed and precise characterizations. Even though still in rather early stages of development ALPACA is able to tackle the demanding task of localizing and parcellating the auditory cortex to a certain limited extent by providing an open framework that includes and combines a myriad of resources and information. Together with the addressed need for standardization, it was therefore also used to define a respective working model of the auditory cortex that was used within all subsequent projects of the here presented thesis. In more detail, this model consisted of the parcellation found in the Harvard-Oxford atlas (Desikan et al., 2006; Frazier et al., 2005; Goldstein et al., 2007; Makris et al., 2006) and motivated by previous studies that used ROIs comparable in extent and localization (Allen et al., 2018; Regev, Honey, Simony, & Hasson, 2013). Notably, its terminology was altered by means of more broadly defined primary and non primary regions that expand the concept of core, belt and parabelt regions, but yet followed past working model definitions (Moerel et al., 2014). Here, the auditory cortex is assumed to consists of five ROIs that are divided into core, belt and parabelt and cover the STP and STG, running along a posterior-anterior axis. The core and belt regions are situated on the STP and include the HG as the core region, as well as the PP and PT as belt regions that surround the HG in anterior and posterior direction respectively. This is in contrast with the majority of other working models that situate both, core and belt regions within the HG and too a very limited extent, within PP and PT, with the latter usually assumed to be part of the parabelt (Javitt & Sweet, 2015; Moerel et al., 2014). The parabelt regions are located on the STG and include its anterior and posterior portion. It reflects a rather broad and unusual definition that was driven to highlight and investigate proposed functional and computational differences of these regions important to the processing of sound, therefore also music, which is furthermore referred to in the respective chapters of this thesis. It is thus also a tradeoff between specificity and complexity that is grounded in the current limitations with respect to special resolution and costs of MRI measurements, as well as absence of publicly available data. It is furthermore intended to provide a basis for subsequent more fine-grained investigations that entail more precise parcellations and are not subject to potential problems regarding too small ROIs and spatial correlation between neighboring voxels. A graphical depiction of the working model in surface format can be found below in Figure 3.12.

Figure 3.12: *Graphical depiction of the auditory cortex working model utilized within the here presented thesis. The model is displayed in surface format on the inflated version of FreeSurfer's fsaverage left hemisphere (left) and right hemisphere (right) respectively. The white outlines indicate the borders of the auditory cortex, while the remaining colors reflect the included ROIs: HG - yellow, PP - green, PT - blue, STGa - violet and STGp - red. Note: the visible overlay of ROIs is based on the mapping from volume to surface format.*

Open auditory neuroscience

Both toolboxes indicate the necessity of data sharing and automatization to capture, understand and tackle the immense individual variability that affect auditory neuroscience most likely more prominent than other research fields, as both the presentation and perception of stimuli and localization and parcellation of important brain regions is tremendously aggravated. This extends to further technical development that is needed to increase measurement and analysis quality. Aiming to create a framework of open auditory neuroscience, both toolboxes are built to enable standardized and reliable applications themselves with their output being shared and evaluated in an open and inclusive manner through open source code and publicly accessible metadata repositories. By defining the investigations of individual HTs and hearing thresholds as open by default from the beginning on and shifting the investigation of the auditory cortex from a rather closed to a completely open manner, the goal of the toolboxes is furthermore the creation of a sustainable basis that allows the rapid advancement of our understanding of auditory processing.

4

STRUCTURE & FUNCTION OF THE CORTICAL AUDITORY SYSTEM: INSIGHTS FROM CONNECTIVITY & MULTIVARIATE PATTERN ANALYSIS

The third project of this dissertation is presented in this fourth chapter. After providing a general introduction into the topic, this chapter will proceed with the depiction of the utlized dataset, as well as a detailed description of the applied analysis methods. Subsquently, the corresponding outcomes will be presented in the results section of this chapter. Finally, these outcomes will be integrated with one another and critically discussed in relation to the hypotheses and working models outlined in this chapter's introduction.

Explanation and identification of personal contribution and data usage

Hereinafter the contribution of the thesis' author (PH) with regard to each project part will be outlined and additional information be provided.

General remarks

The here described project and included analyses were part of and realized within a collaboration between the author of this thesis, Peer Herholz (PH) and members of the International Laboratory for Brain, Music and Sound Research (BRAMS), more precisely, Jorge Armony (JA) and Jocelyne Whitehead (JW).

Conceptualization and implementation

The overall conceptualization of the herein described project was conducted by JA and JW, while the presented approaches and analyses were conceptualized solely by the author (PH).

Stimulus set creation and data acquisition

JA and JW created the stimulus set and acquired the data, as well as conducted the auditory feature analysis.

Analysis approaches and implementation

Conceptualization of the approaches and subsequent implementations, as well as the execution of the resulting analysis, as well as their interpretation were done solely by PH.

Data usage

The usage of the data was allowed within the collaborative nature of this project.

4.1. INTRODUCTION

Even though tremendous advances regarding the amount and localization, as well as structural and functional properties of brain regions involved in auditory processing have been made, one particular question still drives a large amount of research work: how do these brain regions process different auditory categories, what is shared and what is different in terms of resource allocation, usage and integration? A question that was extensively applied primarily to music and speech processing, yielding to numerous studies investigating this query at different levels such as behavior and neuronal (Koelsch, 2009b; Koelsch & Friederici, 2003; Koelsch et al., 2013). Together with their striking similarities concerning structure or syntax, that is the rules they follow to chunk and connect information to create larger structures (Koelsch, 2009b; Koelsch & Friederici, 2003; Koelsch et al., 2013), the results of these studies led to the development of hypotheses and theories focusing on the possibly shared origin of music and speech that to this day are the subject of intense debates (Besson & Schön, 2001). Arguably, the most prominent theory proposing shared origins was formulated by (S. Brown, 2000) and was termed the 'musilanguage' hypothesis. In it's core it states that, given their similarities, that also extend to the domains of emotion, culture, development and social interaction, music and speech might have evolved from the same origin, having a common root and diverging slowly over the course of thousands of years, becoming distinguishable entities in terms of cognitive functions, therefore also drawing an analogy to the evolution of humankind and apes (Wilson & Sarich, 1969). It is important to note that the mentioned syntactic similarities were the focus of the majority of previous studies (Koelsch, 2009b). Starting from a theoretic point of view, a fast variety of attempts have been made to describe the governing principles of linguistic and musical syntax (Koelsch, 2009b). Common to all is the organization of sequential events into hierarchical structures (Koelsch et al., 2013), which includes the progression of words and phrases in speech and notes and chords in music (Koelsch et al., 2013). Studies introducing errors in these structures while applying behavioral and neuroimaging research techniques therefore enabled the investigation of cognitive processes involved in the respective computations. Initially studied in isolation, syntactic errors in music and speech yielded comparable results that exhibited a different weight with regard to their lateralization. Focusing EEG, an "early left anterior negativity" or ELAN was observed for speech (Friederici, Rüschemeyer, Hahne, & Fiebach, 2003) and an "early right anterior negativity" or ERAN for music (Koelsch, 2011a; Koelsch & Friederici, 2003), both describing a strong negatively poled event-related potential (ERP) 100-200 ms after the onset of a stimulus, here the onset of a harsh syntactic violation (missing word or tone), most prominent in frontal electrodes (Koelsch, 2011a; Koelsch & Friederici, 2003). The respective effects were localized in a range of brain areas, including parts of the IFG and STGa using fMRI for both speech (Friederici et al., 2003) and music (Koelsch & Friederici, 2003; Koelsch & Siebel, 2005). A second cognitive process, related to a different type of syntactic violation, that is violations with regard to the integration of entities (wrong word or tone), is represented in the so called P600, a strong positively poled ERP around 500-600 ms after the onset of the respective syntactic violation (Kaan, Harris, Gibson, & Holcomb, 2000). In contrast to the early potentials, the P600 does however not exhibit a lateralized pattern, but a bilateral one which is expressed most prominently in centro-parietal electrodes (Kaan & Swaab, 2003) within speech (Gouvea, Phillips, Kazanina, & Poeppel, 2010; Sassenhagen, Schlesewsky, & Bornkessel-Schlesewsky, 2014) and music (T. D. Griffiths, Johnsrude, Dean, & Green, 1999; Peretz & Coltheart, 2003). Using fMRI, the respective process was localized in IFG (Brauer & Friederici, 2007; Friederici, 2002), STGa (Friederici et al., 2003) and STGp/superior temporal sulcus (STS) (Friederici, Kotz, Scott, & Obleser, 2009) for speech and in IFG and STGp for music (Koelsch & Siebel, 2005; Sinex, Guzik, Li, & Henderson Sabes, 2003). Further work, including studies that utilized paradigms presenting music and speech simultaneously (Koelsch et al., 2002; Koelsch, Gunter, Wittfoth, & Sammler, 2005; Tillmann, 2012), complemented by a fast amount of fMRI studies providing evidence for overlapping networks (Fedorenko et al., 2009; Koelsch, Fritz, Schulze, Alsop, & Schlaug, 2005; Koelsch et al., 2002; Koelsch, Gunter, et al., 2005; Levitin & Menon, 2003; Patel, Gibson, Ratner, Besson, & Holcomb, 1998) paved the way for theories such as the „Shared Syntactic Integration Resource Hypothesis" (SSIRH) postulated by Patel (2003), proposing shared cortical resources and cognitive processes for music and speech, together with modality-specific elements. Comparably, but to a way lesser extent, the same hypotheses in terms of overlap have been proposed for semantic processing, that is cognitive processes related to the processing of meaning (Koelsch et al., 2004). Despite their pioneering character, the mentioned studies and their unifying approach face limitations that lead to several problems and challenges. More precisely this includes the pure focus on cognitive processes that are comparably late in terms of hierarchical processing, a dichotomous perspective ignoring other categories and analysis methods not sensitive enough and therefore not suited to investigate potential fine grained characteristics, especially given the limits of the applied data acquisition techniques. Focusing the first point, assuming a hierarchical model (please see 1.2), the processing of hierarchical structures and meaning are suited in comparably high and late processing stages involving abstract concepts and in a vast amount of studies a task. Following established models of general auditory (Kaas et al., 1999), speech (Friederici, 2002) and music (Koelsch, 2011a) processing, large portions, especially early components, of the processing stream are therefore completely ignored and potential insights missed. This furthermore results in the recruitment of cortical resources involved in domain general abstract processes, leaving out the structures primarily involved in auditory processing, which is indicated through the prominent role of the IFG in the mentioned previous studies. Addressing this possible pitfall, a growing number of research work focused on basic acoustic features, hence lower level, processing within the auditory cortex and network. This is especially interesting given the tremendous differences between the acoustic features underlying music and speech, such as pitch, timbre, rhythm, as well as their spectrotemporal modulations and how they are represented and processed within the auditory cortex (Allen et al., 2017; Santoro et al., 2014; Santoro et al., 2017). Briefly, high temporal and low spectral resolution are needed for speech processing, with the opposite, low temporal and high spectral resolution, being true for music (Allen et al., 2017; Santoro et al., 2014; Santoro et al., 2017; Zatorre & Zarate, 2012). This is based on results providing evidence for a preferred encoding of the mentioned features which differs with regard to specific cortical regions and networks, as well as hemispheres (Alluri et al., 2012;

Alluri et al., 2013; Tervaniemi & Hugdahl, 2003; Toiviainen et al., 2014; Zatorre & Belin, 2001). The correspondent studies used diverse stimuli sets ranging from words and tones over sentences and melodies to almost naturalistic paradigms presenting longer text passages and musical pieces or even a multitude of varying sound categories (Dehaene-Lambertz et al., 2010; Santoro et al., 2014). Especially the belt and parabelt areas (PP/PT and STGa/STGp) exhibited the above addressed differentiated encoding preferences, that is anterior regions (PP and STGa) preferably encoded information with low temporal and high spectral resolution (music) and vice versa posterior regions (PT and STGp) information with high temporal and low spectral resolution (speech) (Santoro et al., 2014; Santoro et al., 2017). Indicating first differences of the auditory cortex and network's specificity and sensitivity with regard to the processing of music and speech in terms of an intrahemispheric distinction, other research work also suggests interhemispheric effects related to a diverging lateralization. Complementary to the ELAN and ERAN, these studies provide evidence for a modality-specific lateralization of the mentioned overlapping bilateral network, including core (HG), belt (PP/PT) and parabelt regions (STGa/STGp) (Schirmer, Fox, & Grandjean, 2012), with speech being more lateralized to the left hemisphere and music more to the right (Angulo-Perkins et al., 2014; Aube et al., 2015; Dehaene-Lambertz et al., 2010; Fedorenko, McDermott, Norman-Haignere, & Kanwisher, 2012; Zatorre et al., 2002). This mirrored lateralization is hypothesized to be grounded in a lateralized feature encoding, for example indicated through a higher frequency selectivity in intrinsic FC of the RH HG, as compared to the LH HG (Cha et al., 2016). In more detail, the encoding and processing of temporal features is proposed to be more prominent in the LH, while the contrary is indicated for spectral features in the RH (Zatorre & Belin, 2001). Thus, not only does the processing of acoustic features, hence early cortical processing stages, point to marked differences between music and speech in terms of the auditory cortex and network's functionality, but also to the mentioned potential second problem of previous studies: a dichotomous perspective, focusing rather entirely on music and speech. Given that underlying acoustic features alone, while otherwise matched stimuli (familiarity, attention, valence, etc.), could be held responsible for any observed dissimilarity (Abrams et al., 2011), this poses an extensive confound based on the already addressed deviating characteristics of music and speech, preventing the investigating of certain subtle effects focusing on only crucial differentiating features. In order to overcome this gap to a certain degree, the incorporation of singing into the stimulus set, more precisely a capella singing, constitutes a valuable possibility. As "it is impossible, and indeed undesirable (Leaver & Rauschecker, 2016), to remove all possible acoustic differences between music and speech (the same way it is not possible to do so for vocal vs. non-vocal sounds, or face to nonface visual stimuli), it is important to minimize them," (Whitehead & Armony, 2018, p.2), which is why singing can act as an intermediate stage, filling the addressed gap between music and speech as dichotomous categories. This is based on the acoustic features as singing is characterized through expressing properties of both music and speech. A "more regular rhythm and pitch stability" (Peretz et al., 2015, p.4) as found in singing compared to speech leads to a spectrotemporal pattern located between those of music and speech, therefore taken all categories together creating a "musical gradient" from speech with high temporal and harsh spectral modulations over singing with intermediate modulations and acoustic features similar to speech (Whitehead & Armony, 2018) to music with low temporal and fine-grained spectral modulations. Additionally, the investigation of cognitive processes involved in the processing of singing offers the possibility to test the sensitivity and specificity of the auditory cortex and network proposed by previous studies, as it enables to evaluate if the regions that seemingly preferably encode music or speech are robust. More precisely, if they encode music independent of instrumental or including vocals and speech independent of spoken or sung, respectively. In fact, a growing number of studies investigated the cortical processing of singing and the potential overlap of cortical resources it has with speech. They yielded results comparable to those comparing music and speech, that is very similar networks along the STG and STS, as well as intrahemispheric differences such as a preferential encoding of singing compared to speech in anterior regions (PP/STGa) (Callan et al., 2006; Schön et al., 2010). However, these studies also share the potential problems, as they didn't include instrumental music, creating another different dichotomous perspective and the majority of them furthermore utilizing methods not suited for the investigation of the apparently subtle modality differences, leading to the third problem. Considering the structure and function of the auditory cortex and network, especially how auditory input is processed in early stages, an overlap of music and speech, as well as singing and speech is, to some degree not surprising (Patel, 2014). Additionally, fMRI even though providing a very high spatial precision, contains very limited information about the underlying neuronal populations with common processing steps like smoothing and analysis approaches such as mass-univariate statistics further diminishing, or even eliminating important information. In more detail, these voxelwise approaches focus purely on spatially averaged responses in relation to certain experimental perturbations, leading to potential problems in subsequent analysis steps, especially if cognitive processes which tend to have highly similar response profiles are investigated, as they will in turn cancel each other out to a prominent extent when compared applying the standard contrasting approach (which condition activates more where?). Taken these points together, using rather classical analysis that are not sensitive to the fine grained differences present in the activity profiles evoked by music, singing and speech it is worth noting that "neural overlap does not necessarily mean neural sharing." (Peretz et al., 2015, p.3). Indeed, studies applying methods more sensitive to fine-grained difference, such as multivoxel pattern analysis (MVPA) were able to provide evidence that, even though their spatially averaged response were highly similar, their underlying (not spatially averaged) patterns could distinguish between the respective categories, as proposed by Norman, Polyn, Detre, & Haxby 2006 & Peelen & Downing 2007. More precisely, a growing body of research work indicated that either the networks and regions overlapping during the processing of music and speech (Abrams et al., 2011; Rogalsky, Rong, Saberi, & Hickok, 2011), as well as singing and speech (Merrill et al., 2012) could actually distinguish between categories if investigated with MVPA, that is there underlying pattern could be used to predict a certain modality or that in fact diverging neuronal populations are involved during the respective processing (Norman-Haignere et al., 2015; Whitehead & Armony, 2018). What unifies those results with those from univariate analysis, is that they targeted spatial patterns of brain responses and therefore, at least to a certain extent the segregation of the respective regions. Focusing on segregation and therefore one of the proposed two major principles of brain structure and function however, more or less totally ignores the other, which is integration (J. R. Cohen & D'Esposito, 2016; Deco, Tononi, Boly, & Kringelbach, 2015). Rather targeting a temporal domain, it describes the collaboration of different regions with respect to cognitive processes (J. R. Cohen & D'Esposito, 2016; Deco et al., 2015), something that, despite a recent paradigm shift highlighting the crucial view of cognitive processes being organized in networks rather than single regions, was not investigated so far for potentially shared resources between auditory categories. As auditory categories are not consistent over time (e.g., as compared to pictures in the visual domain), but evolving and changing, the application of methodological approaches that can possibly target theses changes, such as connectivity analyses, probes a very powerful opportunity, especially given the already addressed importance of the spectrotemporal patterns found in auditory categories with regard to potentially shared and distinct cortical resources, as well as the structure and function of the auditory cortex, here how its regions are connected and interact with one another (Upadhyay et al., 2008). Summarizing the aforementioned points, the results of previous studies indicate overlapping, yet specialized regions and networks involved in the processing of different auditory categories such as music, singing and speech which are, at least partially, grounded in comparable cognitive processes that vary based on distinct spectrotemporal modulations. However, due to either too restricted (dichotomous: music and speech or singing and speech) or too broad (natural sounds) stimuli sets utilized in paradigms evoking or already including a certain task (therefore most likely higher processing levels) together with the application of methods not sensitive enough to detect fine-grained spatial differences, as well as the near absence of network approaches, still only little is known about similarities and differences of the auditory cortex and network with regard to the basic and early processing of different auditory categories. This furthermore extends to the proposed modality specificity of certain regions such as a preferential encoding of music in anterior (PP/STGa) and speech in posterior (PT/STGp) regions, as well as their integration within the cortical auditory network. Given the focus of this thesis, the initially marked question can therefore ultimately be adapted to: how does the auditory cortex and network distinguish between music and other auditory categories, what is shared and what is different in terms of resource allocation, usage and integration?

4.2. THE PRESENT STUDY

The study presented here thus aimed to address the mentioned and described existing gaps by means of a three step procedure: a precisely defined and controlled stimuli set, functional segregation and integration. In more detail, a stimuli set containing diverse music (instrumental), singing and speech excerpts (a vast variety of instruments and languages), controlled with regard to their acoustic features, was presented without an active task in a passive listening paradigm to participants while they underwent an fMRI session. Using complementary analysis approaches, the processing of these categories within the auditory cortex and network, as defined by the amalgamation of formerly identified regions, was investigated subsequently. Following the results of previous studies, the regions covered anterior and posterior regions of the STP and STG respectively (HG, PP, PT, STGa & STGp). In order to examine a potential segregation between those regions multiple searchlight analysis implementing multiclass and binary classification tasks between the categories, targeting fine-grained differences of their respective spatial responses, were conducted. Providing potential insights into where in the auditory network voxel patterns are able to distinguish between certain categories instead of purely focusing on their respective spatially averaged response, this approach was extended by an ROI-modality-specific decoding in which the ability of each region's pattern to predict a certain modality was assessed, enabling a more detailed review of the specificity and sensitivity of those regions and therefore the auditory cortex. Focusing the functional integration among those regions, connectivity analyses, more precisely dynamic causal modelling (DCM), were applied to test different models of hypothesized connectivity patterns, resembling proposed organizational functional principles of the auditory network with regard to a hierarchical and lateralized processing that varies as a result of the characteristics of the different categories. Ultimately, the goal of the study at hand was to provide a comprehensive characterization of the auditory cortex and network with respect to similarities and differences within the early processing of music as compared to other categories. Especially if certain regions and their connections are tuned for music, independent of instrumental or vocal and if the proposed musical gradient has a neuronal correlate, indicated through feature processing that is more similar between music and singing than for music and speech and vice versa.

4.3. MATERIALS & METHODS

4.3.1. PARTICIPANTS

All together, twenty-four participants were included in this part of the project. More precisely, they had a mean age of 25.5 years (SD = 3.021) and eleven of them were females. All participants reported right-handedness, the absence of hearing impairments and were able to fluently communicate in English, nine of them being native speakers. The remaining participants had a broad range of mother tongues, including French (n=11). The number of languages participants were able to communicate in ranged from one to three. Comparably, participants showed a vast variety of musical expertise (M = 4.2 years, SD = 4.72) covering a diverse set of instruments.

4.3.2. STIMULI AND PARADIGM

To address the aforementioned problem of the dichotomous stimulus space, the stimuli included in this part of the project spanned three categories: instrumental music, a capella singing and speech, each containing 60 different examples. The music stimuli consisted of different pieces produced by a variety of musical instruments. Singing excerpts were sung in 19 different languages by individuals of varying age, balanced between females (32) and males (28). Spoken phrases in 45 different languages were used as speech stimuli. All stimuli had a sampling rate of 44,100 Hz and were normalized with regard to their short-term loudness through the Moore and Glasberg Loudness Model (Glasberg & Moore, 1990). To further validate the stimulus set, divers acoustic features were extracted using several open source toolboxes (Boersma, 2002; Ewender, Hoffmann, & Pfister, 2009; Lartillot, Toiviainen, & Eerola, 2008) and compared between the three categories. The respective outcomes are depicted in Table 4.1.

Hence, the resulting stimulus set was very rich in its nature and structure, while at the same time highly matched with regard to basic acoustic features and created the stated important and necessary gradient of musicality from music to speech, using singing to fill the gap between these categories (Schön et al., 2010). The experiment was divided into three runs of equal length (8 min). Within each run 30 stimuli from each category (therefore, 90 in total) were presented in a pseudo-randomized order having an identical number of first-order transitions to avoid carry-over effects and a jittered pause between them (M = 2.49 sec., SD = 0.20 sec). Each stimulus was presented once and the run specific subsets were counterbalanced across participants. While participants passively listened to the stimuli in a continuous design, natural scenes were presented to keep them awake and attentive. This passive listening paradigm was motivated by studies that provided evidence for the task dependency of auditory processing, which however seemingly not activates different areas in the auditory cortex (Alho et al., 2014; Hall et al., 2000). An example run structure is depicted in Figure 4.1. E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA) and MRI-compatible Sensimetrics headphones (Model S14) were used to deliver the stimuli. Notably, two of the three runs were using a continuous multi-band sequence, while one implemented an ISSS sequence. This was done in order to enable a comparison between sequence types, more precisely the evoked responses and their possibly alterations in both (Bandettini et al., 1998; De Martino, Moerel, Ugurbil, et al., 2015; Gaab et al., 2007; Schwarzbauer et al., 2006). However, as this question is unfortunately outside the scope of the here presented work, only the two multi-band sequence runs are described and were used hereinafter. Results based on this respective research question and therefore applying different analysis techniques, are published in Whitehead & Armony (2018).

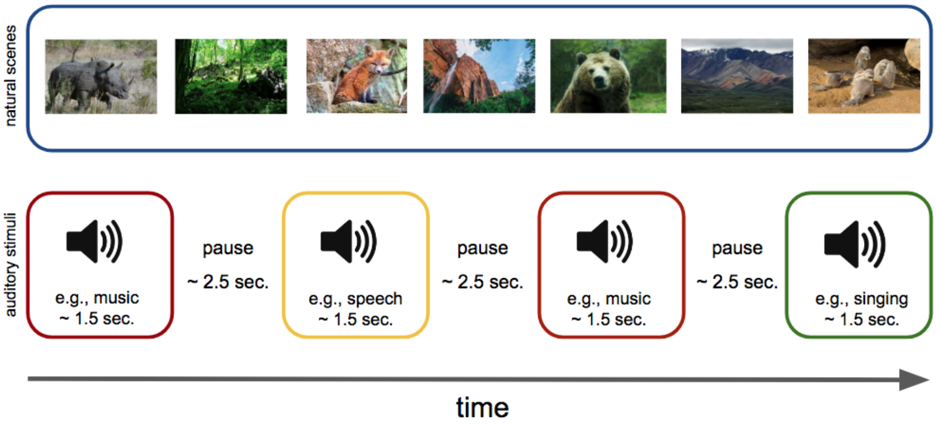


Figure 4.1: *Example of run structure with natural scenes segments (blue rectangle, top) and altering auditory stimuli (categories) (bottom) over time.*

	Music	**Singing**	**Speech**
Duration			
Length (seconds)	1.49 (0.13)	1.51 (0.22)	1.51 (0.23)
Audio features			
Articulation (a.u.)	.32 (.21)a	.27 (.15)a	.44 (.09)a
Root mean square (dB)	.13 (.05)	.16 (.04)b	.13 (.04)
Tempo (bpm)	125 (30)	137 (29)	126 (30)
Spectral centroid (kHz)	2.3 (1.5)	2.4 (1.0)	2.1 (1.0)
Spectral brightness (>1.5 kHz)	.44 (.26)	.42 (.16)	.37 (.15)
Spectral spread (Hz)	5.8 (3.3)	6.0 (2.6)	5.1 (1.6)
Spectral Skewness (a.u.)	.21 (.20)*	.37 (.36)*	.30 (.28)
Spectral kurtosis (a.u)	.73 (1.09)	1.5 (3.0)	.86 (1.43)
Spectral roll off 95th percentile (kHz)	4.2 (2.7)	5.0 (2.3)	4.3 (2.0)
Spectral Spectentropy (bits)	.76 (.08)b	.80 (.05)	.81 (.04)
Spectral flatness	.05 (.08)	.06 (.05)*	.04 (.03)*
Spectral irregularity	.78 (.32)	.95 (.38)*	.67 (.37)*
Zerocross (s−1)	1,335 (1,206)	97 (517)	1,137 (548)
Low energy ratio	.54 (.10)	.48 (.08)b	.52 (.07)
Key clarity (a.u.)	6.8 (3.3)	6.0 (3.3)	7.1 (3.4)
Tonal mode (minor-major, a.u.)	−.02 (.12)	−.02 (.10)	−.02 (.08)
Pulse clarity (a.u.)	.28 (.17)b	.18 (.09)	.23 (.08)
Mean fundamental frequency (F0)	275 (138)	273 (90)	185 (56)b
Std. Dev. Fundamental frequency (F0)	47.2 (37.9)b	31.6 (24.9)	29.6 (15.4)
Minimum fundamental frequency (F0)	204 (109)	217 (79)	134 (45)b
Maximum fundamental frequency (F0)	353 (169)	327 (111)	246 (78)b
Fraction of locally unvoiced frames (%)	10.6 (13.3)	7.6 (9.1)	23.7 (13.1)b
Jitter (local) (%)	2.24 (2.57)	1.43 (1.19)b	2.25 (.71)
Shimmer (local) (%)	12.7 (6.8)b	10.1 (5.1)	10.2 (3.2)
Mean HNR	11.4 (8.0)	13.7 (5.4)	11.4 (3.1)

Table 4.1: Outcome of the acoustic feature comparison between modalities. a – all significantly different, b – significantly different from the other two, * - significantly different from each other. Reproduced and modified from Whitehead & Armony (2018).

4.3.3. PROCEDURE

After the content and objectives were explained to the participants, they gave their informed consent to participate in the experiment and were prepared for the subsequent imaging part of the study. To ensure the audibility during the continuous sequences and avoid a possibly too loud and thus painful presentation during the ISSS sequence, sound tests were conducted accordingly before each scanning session. After the acquisition of the structural image which took roughly 5 min., two functional runs of approximately 7 min. each were conducted during which participants were presented with the above described paradigm (4.3.2). Together with the ISSS sequence not included here, the imaging part of the experiment spanned a time of roughly 30 min. .

4.3.4. MRI ACQUISITION

As mentioned in the previous sections, one anatomical and two functional runs were acquired, both on a Siemens TIM Trio scanner with a field strength of 3T and a 32-channel head coil. The high-resolution anatomical image (1 mm isotropic) was acquired using a T1-weighted magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence with a TR of 2300 ms and a TE of 3 ms, covering 192 slices. Following Setsompop et al. (Setsompop et al., 2012), 800 functional images with 72 slices (interleaved acquisition) covering the whole brain were acquired within each run using a multiband accelerated pulse sequence with a factor of 12 and the following specificities: TR = 529 ms, TE = 35 ms, FoV = 208 x 208 mm, matrix = 104 x 104, voxel size = 2 x 2 x 2 mm.

4.3.5. DATA PREPROCESSING AND STATISTICAL ANALYSIS

Data conversion and preparation

All images were converted from DICOM to NIfTI and the functional images were merged into one 4D image per run using FSL's Merge functionality, discarding the first ten images in each due to T1 saturation, leaving a total of 790 images per run. To facilitate further processing, as well as to enable open & reproducible neuroscience, the data was adapted to the Brain Imaging Data Structure (BIDS) and the anatomical image additionally de-identified using pydeface (https://github.com/poldracklab/pydeface). To control the dataset's quality and for possibly artifact related outliers that would introduce problems and errors in later processing steps, MRIQC (Esteban et al., 2017) in its docker version was run on the complete dataset. It assesses the quality of anatomical and functional images by computing a vast amount of so called Image Quality Metrics (IQMs), for example signal-to-noise-ratio (SNR), entropy focus criterion (EFC) and framewise displacements (FD). IQMs are computed in modality specific Nipype workflows utilizing FSL (S. M. Smith et al., 2004), AFNI (Cox, 1996) and ANTs (Avants et al., 2009), extracted and saved as CSV files, along with report pages in HTML format. Allowing for a comparably easy and straightforward assessment of image quality, MRIQCs superiority furthermore lies within the automated and reproducible application which drastically reduces time consumption, while highly increasing inter-/intra-rater reliability. After visual inspection of the whole dataset and participant specific reports, the MRIQC classifier for structural images was applied to the IQMs, using a (on publicly available data) pre-trained random forests and leave-one-subject-out cross-validation (CV), predicting the quality of images to either accept or reject the data for further analyses. Based on the respective outputs, no participant had to be excluded.

Preprocessing

Data preprocessing was completely conducted using docker (Merkel, 2014), bash (https://www.gnu.org/software/bash/) and python (Python Software Foundation. Python Language Reference, version 3.6. Available at http://www.python.org) on a desktop machine running Ubuntu 16.04 (https://www.ubuntu.com), divided into structural and functional preprocessing. Structural images were preprocessed using the docker version of mindboggle (A. Klein et al., 2017), implementing a comprehensive Nipype processing pipeline (Gorgolewski et al., 2011), including FreeSurfer's recon-all (Dale et al., 1999; Dale & Sereno, 1993; Fischl & Dale, 2000; Fischl et al., 2001; Fischl et al., 2002; Fischl, Salat, van der Kouwe, et al., 2004; Fischl, Sereno, & Dale, 1999; Fischl, Sereno, Tootell, & Dale, 1999; Fischl, van der Kouwe, et al., 2004; Han et al., 2006; Jovicich et al., 2006; Reuter et al., 2010; Reuter et al., 2012; Ségonne et al., 2004), ANTs' antsCorticalThickness (Avants et al., 2009) and mindboggle itself. In more detail, after running the first two processing pipelines, mindboggle effectively combines the output of both to compute a variety of shape measures that complete the already extensive outputs, which includes segmentations into different tissue types, cortical surface reconstructions and parcellations. In the following, the outputs of the respective pipelines are depicted. FreeSurfer's recon-all creates a skull-stripped version of the anatomical image, along with segmentations into different brain tissues (gray (GM) & white matter (WM), cerebral spinal fluid (CSF)) and reconstructs surfaces from the gray and white matter. Additionally, different parcellation schemes are applied to the segmented data, in both, the volume and surface format, obtaining brain regions of varying amount and specificity. Through ANTs' volume based antsCorticalThickness a cortical brain mask, six tissue segmentations (GM, WM, CSF, deep gray matter (DGM), brainstem (BS) and cerebellum (CE)) and a cortical thickness image are computed. Subsequently, mindboggle combines the segmentation from FreeSurfer and ANTs to create a hybrid segmentation to adjust for over-, respectively underestimation of gray and white matter (Klauschen, Goldman, Barra, Meyer-Lindenberg, & Lundervold, 2009), additionally parcellating the newly created segmented file. A variety of shape measures is then extracted within all regions included in the parcellation. Among them are single unit, based on volume/voxels (volume & thickness) and surface/vertices (surface area, mean curvature, geodesic depth, travel depth, convexity (FreeSurfer), thickness (FreeSurfer)), as well as multi unit ones, based on collections of vertices (surface area, Laplace-Beltrami spectra, Zernike moments). Ultimately, summary statistics (mean, standard deviation, kurtosis quartiles, etc.) for all measures within each labeled region are computed. The further usage of the data generated by the anatomical preprocessing pipeline is referenced throughout the description of the functional preprocessing and statistical analysis pipelines. Functional images were preprocessed by means of a Nipype pipeline spanning SPM12 (standalone version) (www.fil.ion.ucl.ac.uk/spm), FSL, ANTS, FreeSurfer and ARTDETECT (https://www.nitrc.org/projects/artifact_detect/). Initially, both 4D functional images were motion corrected via SPM's realignment function using a two step procedure: first, all images from both runs were realigned to first image of the first run to create a mean image; second, the original images were then realigned to this mean image. The resulting mean image, motion parameters and motion corrected functional images were subsequently submitted to Nipype's ArtifactDetect in order to evaluate the time series with regard to possible outliers based on motion parameters and global intensity of the signal. While the first used a euclidean combination of rotations and translations of 1 mm as a threshold, the later marked images with variations higher than three standard deviations as outliers. Afterwards, FreeSurfer's bbregister was used to compute a coregistration matrix between participant's functional and high-resolution structural space, setting the mean image from the realignment step as input and the skull stripped structural image (from FreeSurfer's recon-all within the structural preprocessing pipeline) as reference. The coregistration was further specified with an SPM affine initiation and an optimization for T2-weighted images. Within the next step, functional images were smoothed using SPM's smooth function, applying a 6 mm FWHM 3D Gaussian kernel. In order to restrict the planned statistical analysis to cortical gray matter voxels, a mask was created within 3 steps. First, an image containing voxels belonging to the gray matter in the left or RH as indicated by the labels in the aparc+aseg.mgz volume (multi-tissue-class image computed by FreeSurfer's recon-all within the structural preprocessing pipelines) was created and transformed to participant's functional space using the coregistration approach described above, but setting the mean image from SPM's realignment as reference. Second, using ImageMaths from FSL, the transformed and mean image were subtracted from one another, resulting an image that only includes voxels in the functional space, covered by the sequence's field of view and restricted to anatomical boundaries of the gray matter, hence excluding voxels not originally covered during the acquisition and voxels belonging to areas of no interest (potentially producing artifacts), like ventricles, WM or subcortical regions. This approach increases specificity and sensitivity, while at the same time limiting the influence of possible confounds and number of multiple comparisons to control for during later statistical analyses. Third, the image was dilated by one voxel and binarized with FreeSurfer's mri_binarize. The last part of the preprocessing implemented a coregistration between participant's skull-stripped structural image and FSL's MNI152 template through a non-linear symmetric diffeomorphic mapping in ANTs' antsRegistration function.

Statistical analysis

The statistical analysis of the preprocessed data was conducted in the participant's native space, using Nipype and SPM functionalities implemented in a Nipype workflow. In order to estimate spatiotemporal voxel responses to the three different categories, a general linear model (GLM) was set up through Nipype's SpecifySPMModel and SPM's Level1Design. The preprocessed functional images along with the modality specific onsets and durations (recorded and generated by E-Prime) were included as separate sessions, defining the respective categories as regressors of interest. Both sessions furthermore contained realignment parameters (computed during realignment within the preprocessing workflow) and intercepts as regressors of no interest, resulting in 10 regressors per run. Participant specific GLMs were subsequently estimated by convolving the regressors with the canonical hemodynamic response function (HRF), applying a high-pass filter of 128 seconds and modelling the serial autocorrelation with SPM's FAST option in order to account for the comparably fast TR. The estimation was done via SPM's EstimateModel function and restricted to voxels that fell within the mask created during preprocessing for the above mentioned reasons. Following this, multiple contrasts were defined by comparing the respective regressors against the implicit baseline per run and across runs, together with one effect of interest contrast per run, resulting in a total of 11 contrasts. Contrast image estimation was conducted through SPM's EstimateContrast.

4.3.6. MULTI VOXEL PATTERN ANALYSES

To investigate the spatial pattern of the auditory cortex and its subregions, targeting fine grained differences that are usually not detectable in mass-univariate statistics and the included potential pitfalls, especially with regard to categories and stimuli that elicit a highly similar voxel response profile (Kriegeskorte, Goebel, & Bandettini, 2006), a set of multi voxel pattern analyses (MVPA) was conducted. Throughout all MVPAs the same model parameters were used, as described below. Notably, the necessary CV of the model parameters and permutation tests to assess the significance of the results were implemented in an across participants, more precisely, k-Fold CV. This was motivated and required by several characteristics of the data set at hand. As stimuli and volume onsets weren't temporally aligned spatio-temporal searchlights would only be possible by applying stimulus-volume onset specific HRFs on a trial by trial basis or using other basis functions during to model voxel responses. Even though the multiband acquisition scheme is increasing the signal-to-noise-ratio (SNR) tremendously, the comparably high temporal resolution of potential voxel responses (Boubela, Kalcher, Nasel, & Moser, 2014) cannot be investigated due to the jittering of stimulus onset asynchronies (SOAs) and non-alignment to the volume onsets. Another problem was introduced via the paradigm structure. In more detail, the event related design across two runs did not allow for a sufficient within participant CV that would typically carried out across runs and the respective results analysed on the group level through permutation tests as suggested by e.g., Stelzer, Chen, & Turner (2013). Implementing a "classical" leave-one-run-out CV based on three beta or contrast images per run (representing averaged responses to the three categories) would only allow a very simple and limited testing of the model parameters, as each run would serve as training and test run only once, effectively resulting in restrictive and rather uninformative scoring metrics (e.g. accuracy minus chance ranging from -.50 to .50 in steps of .25). A common work around applied in various studies (Correia, Jansma, & Bonte, 2015; Douglas et al., 2013) is the modelling of single trial instead of averaged run estimates. However, this potentially drastically decreases SNR. To address this dilemma, the aforementioned CV approach was implemented in three steps. First, one contrast image per modality across both runs was estimated to further increase SNR, resulting in three contrast images per participants. Second, contrast images were transformed to a common template (ICBM MNI 152) space, using the participant specific transformation matrices computed within the coregistration step during preprocessing. To avoid multiple transformations of the same (contrast) image, the transformation matrix of the functional to structural coregistration (conducted through BBRegister) was converted to the ITK format (Yoo et al., 2012), allowing it to be merged with the coregistration matrix between the structural and template space computed via ANTs using Nipype's Merge function. The resulting (participant specific) matrix was then applied to the contrast images. Third, NiBabel's concatenate function was utilized to concatenate the respective contrast images from all participants into one 4D image that subsequently had the structure $n_{\text{contrasts}}$ X $n_{\text{participants}}$. Assuming anatomical feature correspondence, this enabled a CV in which participants can be set as chunks, applying the CV across them. This furthermore permitted a more powerful permutation test to assess the significance of the results. As pointed out by previous studies (Haynes, 2015; Kriegeskorte et al., 2006) every choice and variation of model parameters, no matter which and how small, can potentially affect performance and results in a substantial manner. CV is no exception to that, especially the most prominent and used form "leave-one-out" seemingly "yields more variable results" (Varoquaux, 2017, p.172) while at the same time producing a negative bias, both based on variance in the test set (especially in studies having a small number of samples). As shown by e.g., strategies such as k-fold which divide the data in a number of folds, leaving out a certain percentage as test data while training the classifier on the remaining data provide a sufficient cross-validation. Therefore, a stratified k-fold CV, as implemented in scikit-learn (Pedregosa et al., 2011), with 5 splits was used to create 5 folds of varying training and test set compositions. Within each fold the classifier was trained on ~ 80% (~ 18 participants) of the data and tested on the remaining previously hold out ~ 20% (~ 5 participants). The stratified version of the k-fold CV was applied in order to avoid a class imbalance in the training set (Bruns et al., 2015; Raamana, Weiner, Wang, & Beg, 2015; Varoquaux, 2017). Notably, however due to the number of participants and contrasts, as well as the structure of the data set a small imbalance of 1 couldn't be avoided between training and test sets of different folds. The scikit-learn implementation of a Support-Vector-Machine (SVM) was used as a classifier, utilizing a linear kernel and a hyperparameter of C = 1. Prediction accuracy was chosen as a scoring metric. Theses choices were motivated by previous studies implementing similar approaches to comparable research questions (Haxby, 2012; Haxby, Connolly, & Guntupalli, 2014; Kriegeskorte, 2011), additionally providing a default setting if no strong assumptions e.g., about the distribution of the data can be made. The scoring metrics across each fold were averaged to obtain one prediction accuracy. Assessing the significance of cross-validated machine learning scoring metrics, e.g. prediction accuracy as used here, using parametric tests are inappropriate given their characteristics (distribution, dependency, etc.) and was therefore computed using a non-parametric permutation test as suggested by (Ojala & Garriga, 2010). More precisely, the above described CV was repeated 1000 times while randomly shuffling labels to create a distribution of prediction accuracy scores. Following the assumption that labels and features are dependent, the null hypothesis in this setting would be, that they are in fact independent and there's no differences between the to be predicted classes (Ojala & Garriga, 2010). On a more conceptual level, computing a distribution of prediction accuracies based randomly shuffled labels against which the initial, correct accuracy is tested, allows for an assessment of how likely it is that the classifier produced a certain prediction accuracy by chance. A simplified graphical representation of the above described approach is depicted in Figure 4.2. As previously noted, different MVPA approaches were applied in order to facilitate a versatile investigation. Namely, searchlights and decoding. Details on both procedures are given hereinafter respectively.

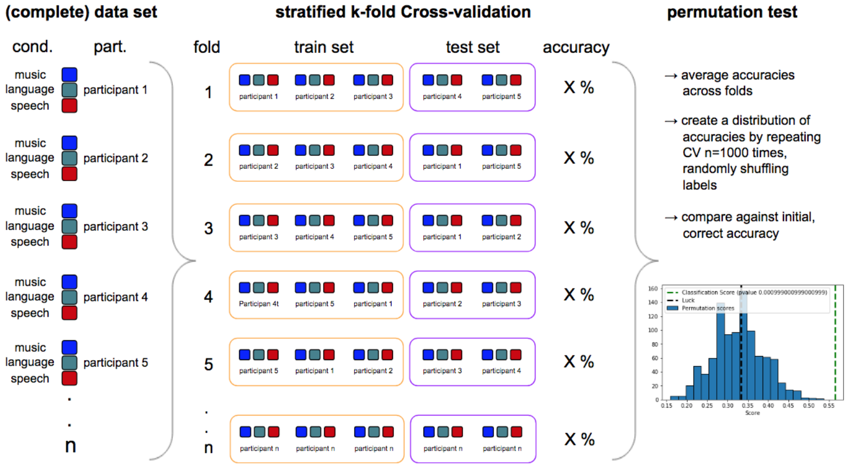


Figure 4.2: *Graphic depicting the applied CV scheme and permutation test. Cond. = condition, part. = participant.*

Searchlight analyses

Initially proposed by Kriegeskorte et al. (2006), so called searchlight analyses are an MVPA approach that perform a classification task at a multitude of locations, e.g. at each given voxel (or vertex) or a subset of them, across the brain as captured by a 3D (or 4D) volume. In contrast to “classic” mass-univariate approaches like the combination of GLMs and t-/F-tests which focus on the spatially averaged responses evoked by certain varying experimental manipulations/conditions, searchlights are “searching” for changes in local activity profiles throughout the brain (Kriegeskorte et al., 2006). Hence, the first is usually termed “activation based mapping”, while the latter is often referred to as “information based mapping”. This approach is especially suited for research questions that involve processes, like the perception of music, singing and speech, which are assumed to evoke highly similar spatially averaged activity profiles and thus potential differences between these processes can not be observed through the limitations of standard voxel-wise analyses. This is because their similar spatially averaged activity will cancel each other out when compared via the “typical” contrast approach applied to voxelwise mass univariate statistics. However, rather focusing on the underlying patterns of voxels enables an investigation of the fine grained spatial structure that carries important information, possibly allowing to distinguish between these processes. Within a searchlight analysis a 3D sphere of a given radius travels along all voxels included in a given volume, with one voxel being the center of the sphere once. The voxel pattern of each sphere location is extracted and a classifier trained and tested to distinguish between given processes or conditions. The accuracy within each sphere is subsequently written/mapped to the respective voxel that lies in the center of that sphere. Hence, the output of a searchlight analysis is not an activation, but accuracy map with each voxel value representing the prediction accuracy of the sphere centered around this voxel. This approach was used in the here described study to investigate which voxels and regions are capable of predicting if participants either heard music, singing or speech and therefore carrying information about the perception of these categories. As previous studies using comparable stimuli (Peretz et al., 2015; Whitehead & Armony, 2018) provided evidence for the vast importance of regions covering the bilateral superior temporal gyrus in the differential processing of these categories (please refer to 4.1), a mask was created to restrict the searchlight analysis to this previously identified regions, aiming to provide complementary insights. To this end the following regions (left and RH respectively) were extracted from the Harvard-Oxford-Atlas (Desikan et al., 2006; Frazier et al., 2005; Goldstein et al., 2007; Makris et al., 2006) as available within the nilearn software package using functionality from the ALPACA toolbox 3.4 and afterwards resampled to the resolution of the functional images via nilearn's resample function: PAC, PP, PT, STGa and STGp. These regions were subsequently combined to create one image containing all regions using nilearn's math_img function. The resulting image was binarized and dilated by 1 voxel through FreeSurfer's mri_binarize. A graphical depiction of this searchlight mask and an its included regions is illustrated in Figure 4.3. Nilearn's Searchlight function was used to set up and conduct multiple searchlight analyses, defining the model parameters as described above: one multiclass and three binary searchlights. While the first applied a one vs. rest strategy identifying regions that contain information to discriminate between all conditions, the latter used a one vs. one strategy to reveal regions that discriminate between respective modalities pairs (music vs. singing, music vs. speech, singing vs. speech). Motivated by disentangling discriminating voxels and regions with respect to categories, this procedure yielded four searchlight accuracy maps, enabling a more detailed coverage. As a radius for the searchlight sphere a value of 5.6 mm was chosen based on investigation using simulated data in (Kriegeskorte et al., 2006). The necessity of multiple (or at least one) searchlight accuracy map per participant demanded by typically applied approaches (parametric or non-parametric/permutation) to assess the significance of searchlight accuracy maps regardless of the specific approach, prohibited the application of the respective approaches in the analysis described here, as the values of interest (classification accuracies) cannot be assessed independently for each participant, given their contribution to all folds of the CV (Kaplan & Meyer, 2012). Therefore, a thematically related proposal by Etzel, Zacks, & Braver (2013) was adapted to the needs and possibilities of this study. The original approach suggests to apply post-searchlight tests to clusters or regions providing discriminative information in order to test if the included voxels "..themselves are informative." (Etzel, Zacks, & Braver, 2013, p.268). In order to prevent a certain degree of circularity which was also addressed and described by Etzel et al. (2013), post-searchlight tests by means of a decoding with permutation test approach were conducted in the searchlight mask and the anatomically defined regions it consisted of. Notably, the respective results therefore do not directly relate to the SMs, but rather allow insights into the discriminative ability of the complete searchlight mask's voxel pattern. Additionally, these tests provide further and more detailed data about the modality related information that is contained in each ROI. The obtained searchlight maps were then mapped and displayed on FreeSurfer's fsaverage using pysurfer. Permutation test results were visualized using Matplotlib.

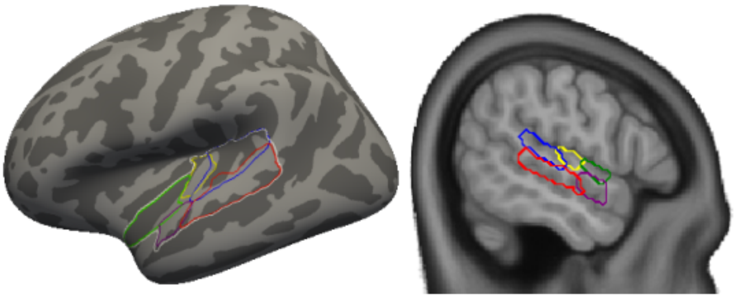


Figure 4.3: *Searchlight mask (white outlines) and its included ROIs: PAC (yellow outlines), PP (green outlines), PT (blue outlines), STGa (purple outlines) and STGp (red outlines) displayed in the LH, respectively on the inflated version of FreeSurfer's fsaverage surface (left) and on the ICBM152 template in MNI volume space (right). Note: Overlap of ROIs are not present in the data, but only in the graphical depiction due to sampling on the surface (left) and limited viewability of 3D data in 2D (right).*

ROI-based decoding

The approach, in terms of parameters, applied for ROI based decoding was identical across all ROIs and conducted as described above. Initially, voxel patterns for each condition in each ROI for all participants were extracted and their values standardized using nilearn's NiftiMasker function, resulting in 1 x nvoxels arrays. As the utilization of masks already restricts the amount of voxels, no further form of feature selection (e.g. anova) was applied. Besides the five anatomically defined ROIs (HG, PP, PT, STGa, STGp) per hemisphere, the searchlight mask was deployed as an additional bilateral ROI. Analyses in the latter consisted of the same classification tasks as in the searchlight analyses. Namely, how well the voxel pattern in the searchlight mask ROI could distinguish between all categories (multiclass) and respective modality pairs (binary). Targeting sensitivity and specificity of the included regions, analyses were adapted to modality specific classification tasks (how well each modality could be predicted by the voxel pattern) within the respective ROIs. The corresponding results were visualized using Matplotlib.

4.3.7. DYNAMIC CAUSAL MODELLING

As mentioned in this chapter's introduction (4.1), recent neuroscientific work pointed to the importance of analysis approaches focusing on cognitive functions as network systems that integrate different brain regions, rather than assuming a more isolated localization (J. R. Cohen & D'Esposito, 2016; Deco et al., 2015). To investigate the possible influence of the different categories on the network structure of the auditory cortex and connectivity patterns between its regions, Dynamic Causal Modelling (DCM) (Friston, Harrison, & Penny, 2003), as implemented in SPM12 and Matlab (Mathworks, Natick, MA, USA) was used. DCM is a Bayesian framework, that aims to describe neuronal states and their dynamics to investigate how experimental conditions influence neural networks and their perturbation, by applying a linear differential equation (Frässle et al., 2016; Frässle et al., 2015). In more detail, this neuronal state equation consists of three important parameters: the A, B and C matrix. The A matrix describes the endogenous connection strengths between regions that compose a network in the absence of any possible manipulation through experimental conditions. The latter, referred to as the modulatory connectivity, is characterized through the B matrix. Direct effects on neuronal states are defined via the C matrix. All together, neuronal states (z) are defined as follows:

$$\frac{dz}{dt} = F(z,u,\theta) = \left(A + \sum_{n=1}^{m} u_j\right)z + Cu \tag{4.1}$$

Bilinear differential equation. Source: Frässle et al. (2016, p. 979).

Convolving the neuronal signal predicted by this equation with the canonical hemodynamic model generates predicted BOLD time series (Frässle et al., 2016) for each given ROI included in the specified network with respect to certain experimental conditions. Hereinafter, a set of models can be tests with regard to their ability to describe the activity pattern of the observed real data. To do so, the Laplace approximation is used to estimate the posterior densities of the model parameters via maximizing the negative free energy (Frässle et al., 2016; Frässle et al., 2015; Friston et al., 2003), while at the same time applying a lower bound approximation to the log model evidence.

Model space

In order to test the possible influence of the different categories on the structure and parameters of the auditory cortex network, a set of models was defined. Aiming to provide a trade-off between accuracy and complexity, especially considering the Bayesian methodological approach, the definition of the models that would subsequently create the model space was based on hypothesis and results of prior research work. The basic structure of the network, including the number of ROIs followed the outcome of previous work aiming to localize and parcellate the auditory cortex, as outlined in 3.4, as well as suggested functional properties of certain ROIs. The network included the following regions: HG, PP and PT, respectively in both hemispheres. Compared to the MVPA described in 4.3.6, STGa and STGp were not included for several reasons. Even though PP/PT and STGa/STGp are sometimes referred to as interchangeable and used to describe subparts of corresponding bigger areas, within the working model of the auditory cortex utilized in this thesis, PP and PT are considered to be part of the auditory cortex' belt region, while STGa and STGp are not. In more detail, PP and PT are localized along the STP whereas STGa and STGp are situated within the STG. Both, therefore being lateral but bordering and not overlapping regions to the HG. This is crucial as a recent study investigating connectivity related properties of the auditory cortex provided evidence for an effective connectivity from HG to the border region of STP and STG (Upadhyay et al., 2008). However, Cammoun et al. (2015) pointed to a hierarchical connectivity pattern in which parabelt regions such as STGa and STGp are mainly connected to the HG via traveling connections through the belt areas (e.g., PP and PT). Furthermore, the results of a recent study analyzing the same dataset with a series of uni- and multivariate analysis methods, indicate the important role of PP and PT regarding a differentiation between the three categories in the absence of STGa and STGp. Together with the maxime (based on the Bayesian nature of DCM) to define possible models as precise as possible, restricting their number and complexity to a certain amount, STGa and STGp were therefore not included in the network architecture. The connections between the remaining regions and therefore the A matrix was defined as follows. Within each hemisphere, the HG had reciprocal connections with the PP and PT (Andoh, Matsushita, & Zatorre, 2015; Cammoun et al., 2015; Upadhyay et al., 2008). Intrahemispheric direct connections between the PP and PT were not included, as prior work (Cammoun et al., 2015) indicated no such pathways, but only between adjacent regions of the belt area. Based on structural connectivity, interhemispheric connections between homotopic regions were limited to HG and PT (Andoh et al., 2015; Elmer, Hänggi, & Jäncke, 2016), both being connected reciprocally. As there is little to no evidence of interhemispheric connections between heterotopic regions (Catani & Thiebaut de Schotten, 2008), respective connections were not included in the network. All regions had an intrinsic connection (Cha et al., 2016; Häkkinen & Rinne, 2018). The A matrix was kept identical across all models. The same accounts for the C matrix, as the input to the network was always set via the bilateral HG. Hence, the varying modulatory parameters of the B matrix set and defined the model space. A graphical depiction of the A and C matrix can be found in Figure 4.4.

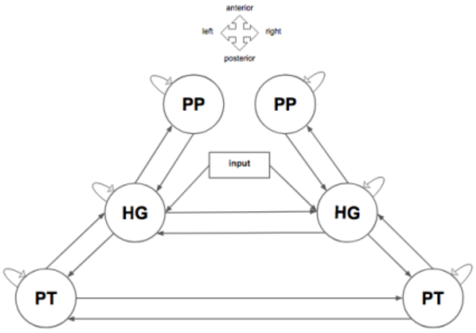


Figure 4.4: *Graphical depiction of the A and C matrix, including bilateral HG, PP and PT. The C matrix is indicated through input entering the network via the bilateral HG.*

Different models were defined based on the results of previous studies investigating the auditory system and are characterized by three components: interhemispheric connections between HG, intrahemispheric connections from HG to PP/PT, as well as site of intrahemispheric connection from HG to PP/PT. Together with the A and C matrix, more precisely auditory input always entering the network via the bilateral HG, the definition of the model space followed the hypothesized hierarchical processing order among regions involved in auditory processing, especially perception (Andoh et al., 2018; Bidelman & Grall, 2014; Chevillet, Riesenhuber, & Rauschecker, 2011; Kell et al., 2018; Zatorre et al., 2002; Zatorre & Salimpoor, 2013; Zatorre & Zarate, 2012). The first component describes the direction of modulated connections between both HG and could manifest as either from left to right HG, right to left HG or both. It therefore formulates one of the hypothesized core principles of the auditory system: the preferred processing of music in the RH and the predominantly processing of speech and speech in the LH at early stages of cortical processing of auditory stimuli, which is grounded in results indicating an underlining lateralized feature processing, especially proposing a distinction between temporal and spectral features (please see 4.1 for a more detailed depiction). Hence, this component was used to define three different respective model families as depicted in Figure 4.5, which together with an identical parametric variation across the remaining two components furthermore allowed a collective comparison between families as advised in terms of inference methods within the DCM framework. Concerning the second component, modulated intrahemispheric connections between HG and parabelt areas could either be present in one (either to PP or PT) or both (to PP and PT) directions.

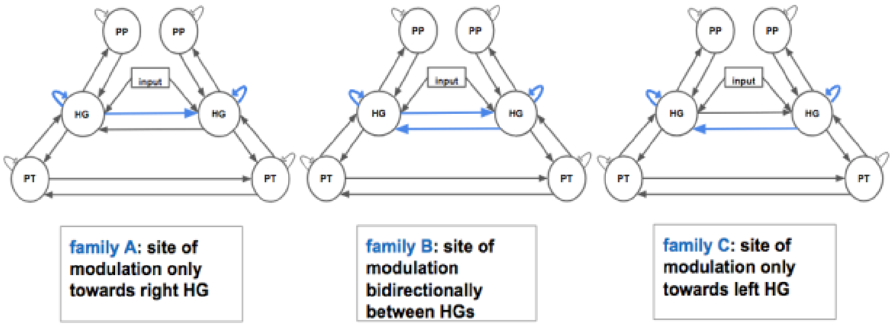


Figure 4.5: *Deployed model families as defined by the modulatory connection between the HGs.*

These possible perturbations enclose evidence that PP and PT preferably encode different temporal features, respectively high temporal rigor in PT and low temporal rigor in PP (Santoro et al., 2014). Notably, both feedforward and feedback connections between the respective regions were always modulated.The hemisphere of these modulations (left, right or both) was varied within the third component, therefore combining assumptions about lateralization and direction of modulated connections. In case modulations towards PT were present, respective models furthermore included a modulated connection from this PT to its homotop counterpart. The modulation of interhemispheric PT connections was not included as a separate component given that were no strong prior assumptions and in order to restrict the number of models. Experimental perturbation could modulate the intrinsic HG connection in all models. This parametric approach therefore yielded 27 different models, grouped in three families of 9 models each. As an example, the model space of family A is illustrated in Figure 4.6. Given the aim of this analysis, the assessment and comparisons of auditory system connectivity patterns in relation to the different categories, the following procedure was applied. A new GLM (differing from the one described in 4.3.5) was set up, using the same functionality as noted in 4.3.5, to include four regressors: three modality specific regressors (music (M), singing (Si), speech (Sp)) and one regressor containing all categories (AAM). These regressors served as input for the DCM analyses, while the aforementioned GLM (4.3.5) and the included regressors were used to extract time series from the ROIs (please see 4.3.6). The A and C matrix were kept identical across models and also categories, as the main question was how the different categories would possibly change certain connections between regions. Thus AAM was set as driving input (C matrix) and the experimental perturbations (M, Si and Sp) were set as input for the B matrix, respectively.

<u>Extraction of time series</u>

To extract BOLD time series from the network constituting ROIs, the same ROIs as within the MVPA (4.3.6) were used. As the functional images were not spatially normalized during prepro-

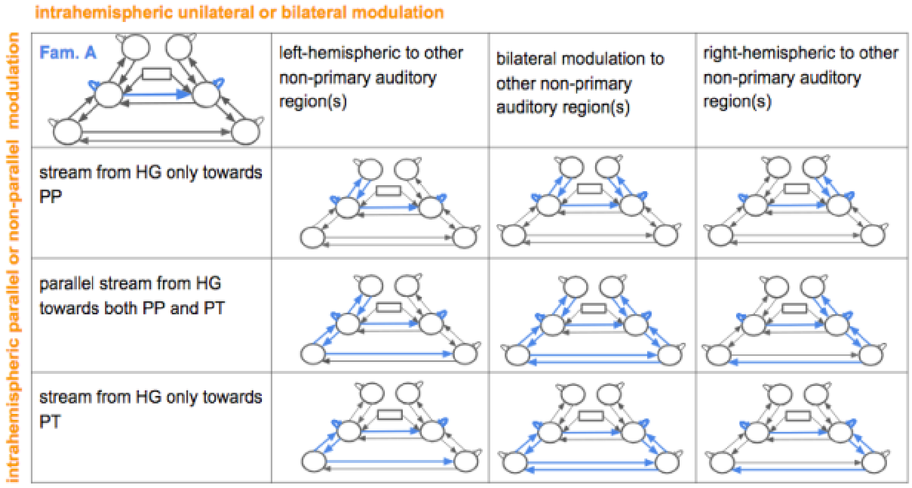


Figure 4.6: *Model space for model family A. Displayed are all 9 models of this family, sorted by their parametric variation.*

cessing or the statistical analysis and this was also not necessary for the DCM analysis (please see 4.3.6 for the respective motivation within the MVPA), BOLD time series were extracted in native space. Therefore, each ROI was transformed to a given participant's native space using ANTS' ApplyTransform, setting the inverse composite transform computed during the coregistration steps within preprocessing as transformation matrix. BOLD time series per ROI and run were extracted using the following approach: within the initial GLM (please see 4.3.5) a contrast of all regressors of interest (M, Si, Sp) compared to the implicit baseline, thresholded at $p < 0.05$ (uncorrected) and inclusively masked within a given ROI was used to define a set of voxels from which subsequently the first eigenvariate was computed. The run-specific F-contrast was used to mean-center the BOLD time series, additionally removing movement related artifacts. This culminated in the BOLD time series of voxels that showed activation within a certain ROI in reaction to the experimental design. Therefore, 6 BOLD time series (one for each ROI) per participant and run were included in the DCM analysis.

Inference and statistical analysis

After all models from each participant and run were estimated, the inference and statistical analysis of connectivity parameters was conducted in a three step procedure: Bayesian Model Selection (BMS) and Bayesian Model Averaging (BMA), connectivity parameter-wise t-tests and connectivity pattern prediction using Random Forests. Initially, all models from both runs were compared on the family level by means of a random effects BMS (Penny et al., 2010; Stephan, Penny, Daunizeau, Moran, & Friston, 2009). The negative free energy, being "an approximation to the log model evidence." (Frässle et al., 2016, p.982) is used to assess the goodness of a given model, yielding posterior evidence, indicated through exceedance probabilities, for each model and family. Probabilities for both, model families and models were visualized via seaborn (Waskom et al., 2018) and Matplotlib (Hunter, 2007). Subsequently, participant and connection specific parameters were computed via random effects BMA (Penny et al., 2010) using only models from the winning family and an Occam's window of $p < 0.05$, weighting each model by its probability. The thus yielded parameters for the A and different B matrices were submitted to t-tests on the group level to assess their statistical significance. In more detail, each parameter was entered into a one-sided one-sample t-test, thresholded at $p < 0.05$ and corrected for multiple comparisons (FDR) within each parameter class via pingouin's ttest function (Vallat, 2018). The same function was used to additionally compute cohen's D, the power and Bayes factor of a given connection. Focusing differences in connectivity patterns, that is the entire network's connectivity parameters, rather than only certain connections, a Random Forest classifier (RFC) as implemented in scikit-learn was used to investigate potential differences between the categories. The number of estimators within the RFC was set to 100 and CV, as well as significance testing was applied in the same manner as in the MVPA. The choice of applying an RFC was furthermore motivated by its ability to assess the importance of the provided features for a given classification task, thus providing additional potentially important and useful information, especially in cases where no above chance and/or significant accuracy could have been reached. To compute feature importance scikit-learn employs gini importance or also called mean decrease impurity (Breiman, Friedman, Stone, & Olshen, 1984) which "is defined as the total decrease in node impurity (weighted by the probability of reaching that node (which is approximated by the proportion of samples reaching that node)) averaged over all trees of the ensemble" ("scikit learn - How are feature importances in RandomForestClassifier determined?", 2013). Feature importance thus allows the assessment of each feature's contribution to the overall predictive performance of the specified model and therefore, in the case of this study, provides details on the respective properties of network connections. Comparable to the MVPA, four different analysis with identical model parameters were conducted to enable a detailed investigation of connectivity patterns related to the three different categories: a one vs. rest classification task (multiclass) assessing to ability to discriminate between all categories and three binary classification tasks focusing on pairs of categories (music vs. speech, music vs. singing, singing vs. speech). The importance of each connection across all classification tasks was visualized through a lineplot using seaborn and Matplotlib.

4.4. RESULTS

In this chapter the results of the previously described analyses will be presented. Beginning with the outcomes of the searchlight analysis, this part proceeds with details regarding the ROI-based decoding. Subsequently, details on the DCM and follow up analyses will be given.

4.4.1. MULTIVOXEL PATTERN ANALYSES

Searchlight analyses

Four different searchlight analyses restricted to a searchlight mask, covering the STP and STG in both hemispheres, were conducted in order to identify voxels and regions that carry distinctive information about the three categories. More precisely, one searchlight applying a multiclass classification task aiming at discriminating between all categories and three binary classification tasks amining at distinguishing between respective pairs of categories. Given the limitations regarding CV and significance testing, grounded in the experimental design, the significance of the classification performance (accuracy) was assessed via a decoding approach that used all voxels within the searchlight mask and permutation tests. The results are graphically summarized and illustrated in Figure 4.7. Overall, all searchlight analyses yielded comparable SMs, showing high similarities with regard to the spatial distribution of informative voxels and clusters. Bilaterally, large portions of both, STP and STG were able to distinguish between the categories above chance level with their accuracies varying between classification tasks. To a certain spatial extent, two clusters within the LH, showing the highest accuracies, were observed regardless of the classification task. These clusters covered bordering regions of STP and STG, respectively within their anterior and superior parts, representing PP/STGa and PT/STGp, whereas a stripe running orthogonal between them displayed low accuracy values. In general, SMs in the RH yielded lower accuracies as compared to the LH and predominantly showed a cluster bordering the anterior parts of STP and STG. All post-searchlight tests yielded significant accuracies of varying degree. The SMs of the respective classification tasks are described in more detail hereinafter.

Multiclass searchlight

The SM associated with the multiclass classification task showed high accuracies in the already mentioned two left and one right hemispheric clusters, resembling anterior and posterior parts of the STP and STG respectively. Those clusters reached accuracy scores of 0.56 % given a chance level of 0.33 %, while the majority of the remaining voxel's values were distributed around 0.44 %. Notably, low accuracy scores were observed in posterior parts of the right STP and STG, as well as the HG region, bilaterally, with a more prominent markedness in the LH. The two left hemispheric clusters showed a seemingly segregation between STP and STG regions, therefore PP/PT and STGa/STGp, with a stronger and wider manifestation in the STP part. In contrast to its right hemispheric counterpart, both clusters were clearly delimited and the anterior cluster mainly observed in the posterior part of the PP. In particular, the anterior cluster in the RH was more distributed and also involved anterior parts of the PP. The post-searchlight test yielded an accuracy of 0.59 % that was significant at $p < 0.001$.

Music vs speech

A very precise cluster of high accuracy values reaching 0.8 % (chance level: 0.5 %) that spannend anterior parts of the RH STP and STG was observed in the SM that distinguished between music and speech. While bordering both STP and STG, the majority of the cluster was localized in STGa. Posterior parts of the PP, in adjacency to the HG, showed a second smaller cluster. Originating from these regions, accuray values diminished into inferior and posterior direction, with most parts of the searchlight mask showing accuracy scores around 0.6 %. PT and

STGp displayed the lowest values, with caudal portions not reaching chance level. The LH demonstrated a disparate pattern, as the lowest values were observed along a stripe covering the superior end of the searchlight mask and most of the remaining areas showing an evenly spread distribution of accuracy scores around 0.65 %. Additionally, two prominent clusters of high accuracy scores could be localized in anterior and posterior parts of the STP and STG. The latter were nearly equally distributed in caudal components of the PT and STGp. In contrast, the anterior cluster spannend a wide range within middle and caudal portions of the PP, with only little involvement of STGa. An accuracy of 0.86 % at $p < 0.001$ was obtained via the post-searchlight test.

Music vs singing

Accuracy scores in the SM distinguishing music and singing reached peak values of 0.7 % at a chance level of 0.5 % that spanned multiple clusters in PP and PT bilaterally. Furthermore, small extensions to the corresponding counterparts in the STG could be observed. LH PT showed the most prominent and precise cluster, while the homotopic PP exhibited a more scattered pattern. The lowest accuracy scores (around or below chance level) in the LH were obtained in regions corresponding to HG and anterior PP. Around the peak clusters, accuracy scores were roughly 0.6 %. The RH displayed a comparable pattern. However, clusters were smaller in their extent, more restricted and in case of the anterior part, shifted towards STG. In this hemisphere, superior and posterior parts of the searchlight masks showed the lowest accuracy scores. In both hemispheres, clusters were in caudal portions of the PP and middle portions of the PT. At $p < 0.001$, the post-searchlight test yielded an accuracy of 0.69 % (chance level: 0.5 %).

Singing vs speech

Bilateral anterior parts of STP and STG yielded prominent contribution, in terms of high accuracy scores around 0.72 % (chance level 0.5 %), within the singing and speech distinguishing SM. While the LH cluster was roughly delimited and primary in caudal portions of the PP, a nearly opposite pattern was observed in the RH. Remarkably, a reversed pattern between LH and RH was furthermore expressed in posterior parts of STP and STG. More precisely, in the LH showed a second small cluster surrounded by medicore accuracy scores bordering PT and STGp whereas accuracy scores drastically decreased towards the corresponding region in the RH. The post-searchlight test resulted in an accuracy of 0.72 % (chance level: 0.5 %), being significant at $p < 0.0001$.

Differences between searchlights

Comparing the SMs, striking similarities between the multiclass and music vs speech SM became obvious, namely two clear distinct and precise patterns within anterior and posterior parts of the LH STP and STG, as well as one pattern in anterior parts of the RH STP and STG. However, patterns were mainly in the STP in the multiclass SM, whereas music vs speech elicited broader patterns in both STP and STG, that additionally were more anterior and in case of the RH more precise and localized in the STGa. Notably, the SMs regarding music vs. singing and singing vs. speech showed variations of the aforementioned patterns that were additionally to some extent reversed between these two SMs. Focusing the LH, two clusters, comparably localized as the corresponding cluster in the multiclass and music vs. speech SMs, were observed. Their overall distribution and extent was however drastically diminished while additionally being more prominent exclusively in either anterior or posterior regions, but not both as in the other SMs. While the PT showed a more strongly pronounced and precise cluster in the music vs. singing SM, its PP counterpart displayed reduced accuracy scores in a rather scattered distribution. The almost reverse pattern was observed in the singing vs. speech SM with the PP showing high accuracy scores in a scattered pattern and PT lower values in a defined region. As indicated through the regions, patterns were localized similar to those of the multiclass SM, that is within the STP. On the contrary, clusters in the RH were substantially different, as the music vs. singing SM showed very few and small clusters of high accuracy scores of the same anterior - posterior distinction as in the LH. Notably, this was the only SM showing this pattern in the RH, especially with regard to the cluster in the posterior part bordering PT/STGp. Highly scattered and small clusters in anterior parts of the STP and STG were observed in the singing vs. speech SM. Comparing the results of the binary searchlight and post-searchlight analyses, accuracies in both dropped from 0.8 % in the music vs. speech SM to around 0.7 % in the music vs. singing and singing vs. speech SM respectively.

ROI-based decoding

In order to enable a more detailed investigation of the information encoding and functional profiles of the ROIs included in the deployed searchlight with regard to the different categories, an ROI-based and modality specific decoding approach was conducted. In particular, it was tested how well each ROI could predict each modality. The respective results are depicted in Figure 4.8 and show an interaction between ROIs and categories, that enfolds along the hierarchy of the cortical auditory pathway and manifested in three main points. First, classification accuracies increased from early (HG) over belt (PP/PT) to parabelt (STGa/STGp) regions. Second, this holds only true for music and speech, with singing always ranging around 0.5 % and therefore showing the lowest classification accuracy (except for the left and right HG). Third, till parabelt regions LH ROIs reached higher classification accuracies for speech and RH ROIs for music. In more detail, classification accuracies reached a maximum of 0.61 % in early regions (speech in LH HG), while the majority of values fluctuated across categories, reaching accuracies around 0.55 % and displaying no prominent differences between categories. Beginning in the PP, accuracies started to increase, but only for music and speech, whereas it decreased for singing, creating a pattern that additionally was hemisphere specific and mirrored between LH and RH PP, that is music reached higher scores in the RH as compared to speech and vice versa. This pattern intensified in the PT, as the classification accuracy for singing decreased even further (falling below the permutation based chance level in the LH PT), whereas values for music

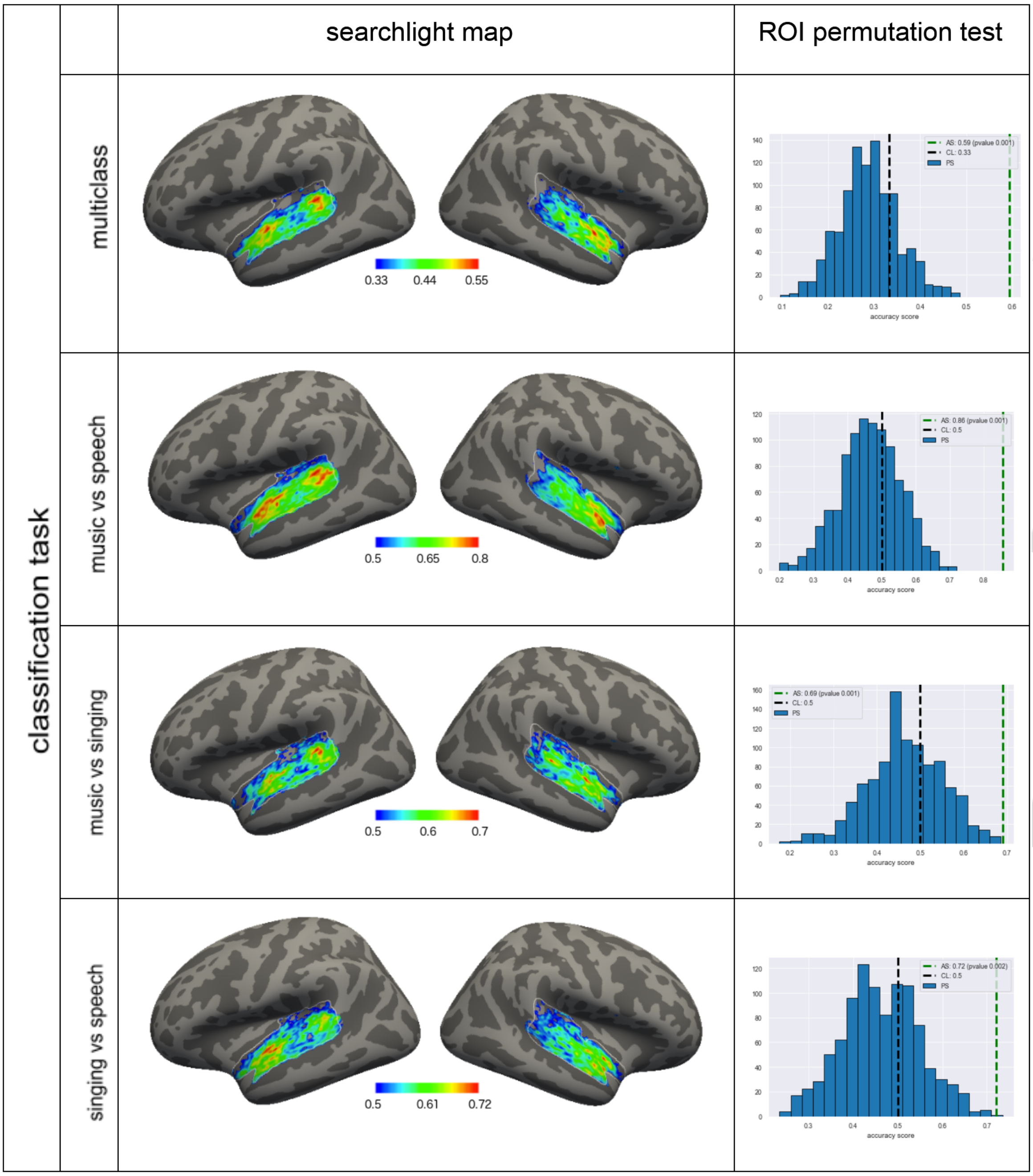


Figure 4.7: *Results of the four-folded searchlight analyses. Displayed are classification task corresponding SMs (left column) and respective permutation test results using all voxels within the searchlight mask (right column). SMs were mapped on the surface using FreeSurfer's mri_vol2surf function (smoothing of 4 FWHM) and are displayed on the inflated surface of FreeSurfer's fsaverage. White outlines depict the mask the searchlights were restricted to. All SMs are thresholded from chance level to maximal reached accuracy as indicated by the colorbars. Hence, the more reddish the higher the accuracy and vice versa for blue voxels and regions. The multiclass searchlight had chance level (cl) of 0.33 (based on three categories), while chance level in the remaining pairwise searchlights was 0.5 (based on a binary classification task). These values also hold true for the permutation test and are visualized by a black dotted line in the respective graphics. The initial, non-shuffled accuracy (as) is marked by a green dotted line, with blue bars represent the distribution of accuracy values yielded by 1000 permutations with randomly shuffled labels.*

(0.64 % in LH and 0.63 % RH) and speech (LH: 0.67 %, RH: 0.58 %) slightly increased. Following this trend, classification accuracies reached maximum values in STGa while at the same time discontinuing the modality-hemisphere interaction, as LH STGa yielded comparable values for music and speech (around 0.74 % respectively) and RH STGa reversing the pattern with a higher classification accuracy for speech (0.8 %) than for music (0.74 %). Furthermore, singing was below chance level (0.5 %) in LH (0.49 %) and RH (0.43 %). STGp also exhibited the highest classification accuracies for speech, independent of the hemisphere (LH 0.67 %, RH 0.72 %), followed by music (LH 0.63 %, RH 0.62 %) and singing (LH 0.54 %, RH 0.58 %), with the values for music and speech decreasing and for singing increasing, as compared to STGa.

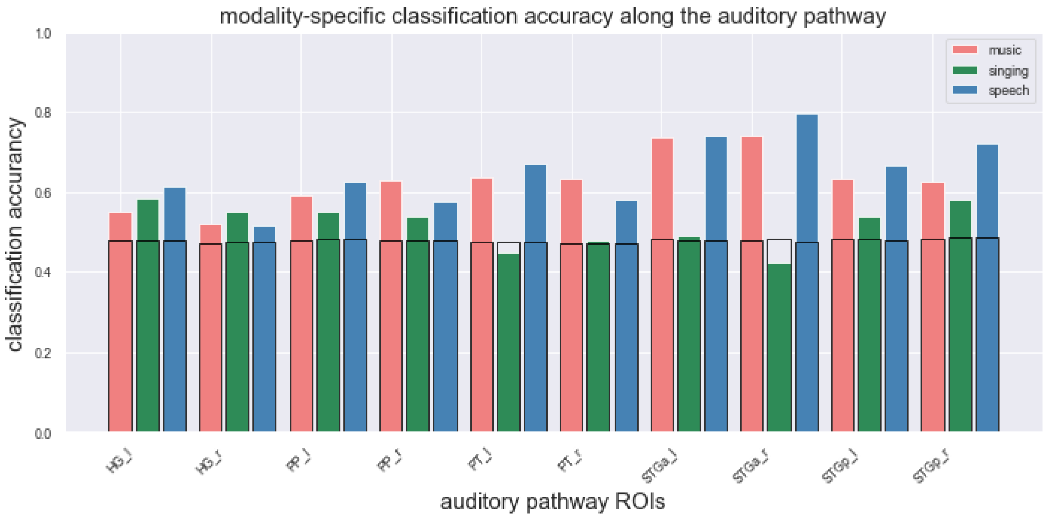


Figure 4.8: *Modality-specific classification accuracies with all ROIs of the utilized auditory network. Colors represent categories: music - red, singing - green, speech - blue. Black outlines indicate the chance level as assessed via a permutation test (n = 1000).*

4.4.2. DYNAMIC CAUSAL MODELLING

A set of 27 models of varying B matrix complexity and organized in three families, which were based on the interhemispheric connections between the HGs, were included in a DCM analyses in order to analyze the connectivity pattern of an auditory cortex network consisting of three regions per hemisphere: HG, PP and PT and its potential underlying processing mechanism related to the different categories.

Bayesian model selection & Bayesian model averaging

Figure 4.9 displays the results of the random effects BMS that was applied to compare the models on the family level. Strong posterior evidence (exceedance probability: 0.89) was found for model family A, grouping models which included a modulated connection from LH HG to RH HG. The remaining model families, either grouping models with a modulated connection from RH HG to LH HG (model family B) or a bidirectionally modulated connection between both HGs (model family C), yielded strikingly low posterior evidence of 0.07 (model family B) and 0.04 (model family C). On the model level, highest posterior evidence (0.52) was found for model 5 within model family A, that is the fully connected model (all connections are modulated). Followed by model 2 (0.13), the remaining models of family A yielded little (0.01) (model 1, 3, 6, 8) to no exceedance probability (model 4, 7, 9). Except model 2, 5 and 6, which resulted in low exceedance probabilities of 0.04, 0.13 and 0.01 respectively, no model of family B yielded an exceedance probability. The same also holds true for model family C and models 4 (0.01), 5 (0.02) and 6 (0.10). Results of the model level BMS are depicted in Figure 4.10. BMA was used to estimate individual connectivity parameters only across the 9 models from the winning family A, weighting them based on their posterior probability.

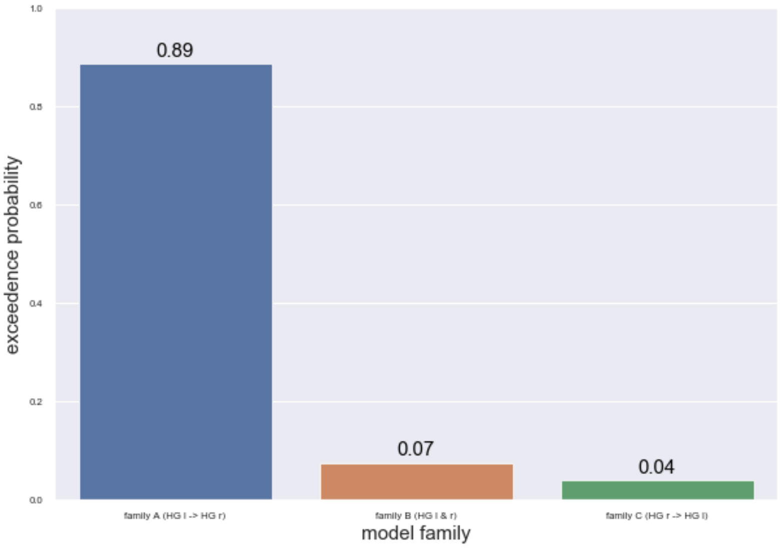


Figure 4.9: *Exceedance probabilities for the different model families. Exact values for each model family are displayed above their respective bar.*

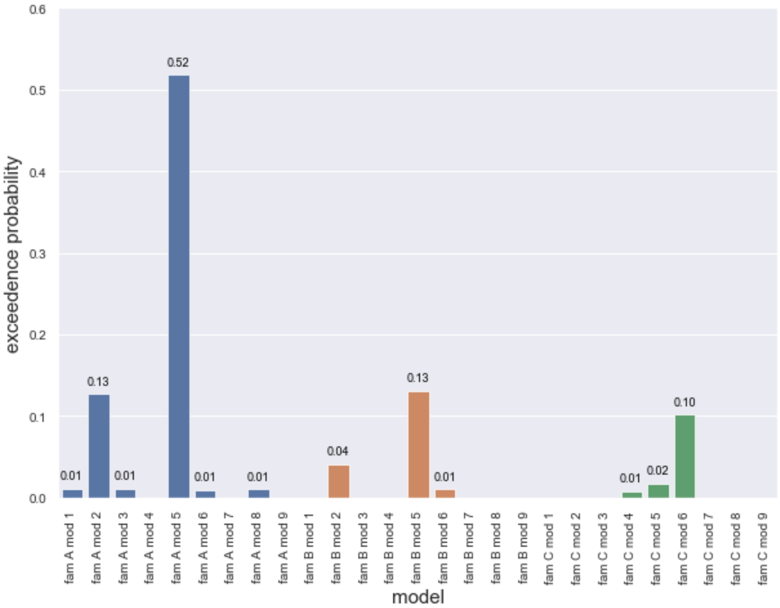
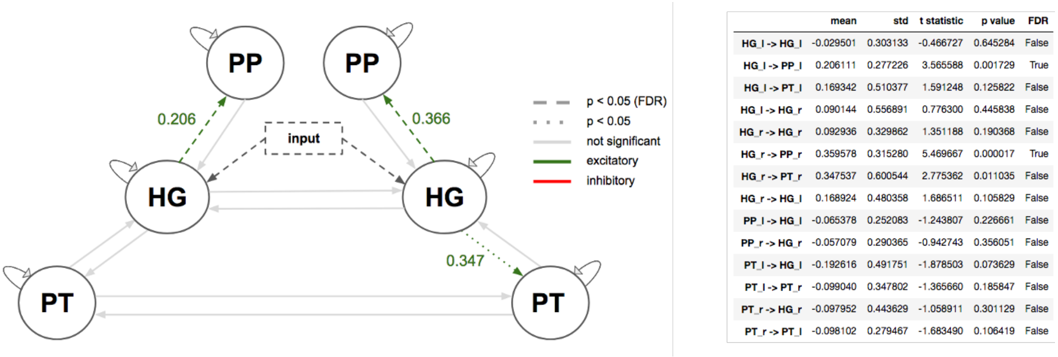


Figure 4.10: *Exceedance probabilities for the different models. Exact values for each model are displayed above their respective bar.*

Endogenous connectivity - A matrix



	mean	std	t statistic	p value	FDR
HG_l -> HG_l	-0.029501	0.303133	-0.466727	0.645284	False
HG_l -> PP_l	0.206111	0.277226	3.565588	0.001729	True
HG_l -> PT_l	0.169342	0.510377	1.591248	0.125822	False
HG_l -> HG_r	0.090144	0.556891	0.776300	0.445838	False
HG_r -> HG_r	0.092936	0.329862	1.351188	0.190368	False
HG_r -> PP_r	0.359578	0.315280	5.469667	0.000017	True
HG_r -> PT_r	0.347537	0.600544	2.775362	0.011035	False
HG_r -> HG_l	0.168924	0.480358	1.686511	0.105829	False
PP_l -> HG_l	-0.065378	0.252083	-1.243807	0.226661	False
PP_r -> HG_r	-0.057079	0.290365	-0.942743	0.356051	False
PT_l -> HG_l	-0.192616	0.491751	-1.878503	0.073629	False
PT_l -> PT_r	-0.099040	0.347802	-1.365660	0.185847	False
PT_r -> HG_r	-0.097952	0.443629	-1.058911	0.301129	False
PT_r -> PT_l	-0.098102	0.279467	-1.683490	0.106419	False

Figure 4.11: *Group level endogenous connectivity (A matrix) parameters and their respective statistics. Left: Group level endogenous connectivity (A matrix) parameters. Connections that were significant at $p < 0.05$ uncorrected are indicated via a dotted arrow, while those that survived a correction for multiple comparisons are display as dashed arrows. Green arrows represent excitatory connections. Non significant connections are depicted as gray solid arrows. Numerical values represent the strength of a given connection indicated through the mean coupling parameter in Hz. Right: Group level results for the endogenous connectivity parameter estimates after BMA. For each parameter the group based mean and std are displayed, along with their t statistic and p value yielded through a one-sample t test. FDR indicates if parameters survived the correction for multiple comparisons.*

The endogenous forward connections from HG to PP were excitatory in both hemispheres at $p < 0.05$ (FDR corrected). Furthermore, the intrahemispheric connection from HG to PT was excitatory, but only in RH and at $p < 0.05$ (uncorrected). All remaining connections did not reach significance. All connections, their parameter estimates and p-values, as well as graphical representation of the endogenous connectivity pattern are displayed in Figure 4.11.

Modulatory connections - B matrix

Significant parameter estimates for the modulatory connections varied in extent and manifestation with regard to the different categories, while at the same time displaying similarities resampling the structure of the endogenous connectivity pattern. Modality-specific graphical network representations, as well as statistics for each parameter are depicted in Figure 4.12. The patterns of the modality specific parameter estimates are described in more detail hereinafter.

Music

Only connections within or originating from the RH were positively (excitatory) modulated by passive listening to music, supporting the hypothesis of music being predominantly processed in the RH. In more detail, this included the intrahemispheric connections from HG to PP (0.564 Hz, ± 0.842) and PT (0.845 Hz, ± 1.076), as well as the interhemispheric connection from RH PT to LH PT (0.596 Hz, ± 1.203). The intrinsic HG connection exhibited negative (inhibitory) modulations in both hemispheres. The same holds true for the connection from the PP to HG in RH (-0.218 Hz, ± 0.401). The music modulated version of the network along with statistics for all connections can be found in Figure 4.12 (top row).

Singing

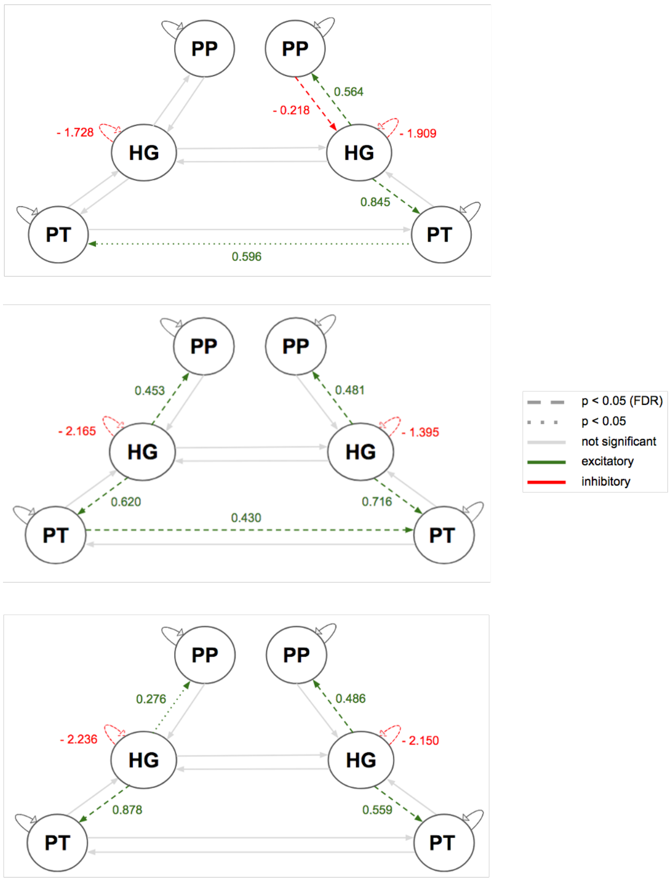
Focusing singing, a pattern including bilateral positively modulated connections from HG to PP, as well as PT was observed. Together, with stronger modulated connections from HG to PT (LH: 0.620 Hz, ± 0.872; RH: 0.716 Hz, ± 0.764) than from HG to PP (LH: 0.453 Hz, ± 0.604; RH: 0.481 Hz, ± 0.515) this emphasizes the important role of the LH and PTs in the perception of singing. An excitatory modulated interhemispheric connection from LH PT to RH PT (0.430 Hz, ± 0.869) provides further evidence for this assumption. However, post-hoc tests aiming to compare the respective connections between hemispheres revealed no significant differences, neither for HG to PP (t(21) = 111, p = 0.876 (wilcoxon)) nor HG to PT (t(21) = 0.394, p = 0.697 (paired t-test)). Negatively modulated connections were observed for the bilateral intrinsic HG connections (LH: -2.165 Hz, ± 1.711; RH: -1.395 Hz, ± 1.578). Figure 4.12 (middle row) displays the network as modulated by singing, as well as statistics for all connections.

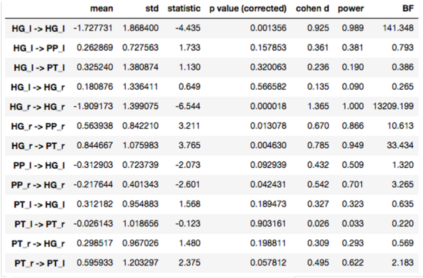
Speech

In contrast to previous studies (Poeppel, 2003; Robert J Zatorre et al., 2002), proposing a rather strong left hemispheric lateralization, a bilateral pattern of positively modulated connections was found related to the passive processing of speech. The connection from HG to PT was more pronounced in the LH (0.878 Hz, ± 0.962) as compared to its RH counterpart (0.559 Hz, ± 0.853). The reverse was observed for the HG to PP connection, as the RH yielded a higher coupling rate (0.486 Hz, ± 0.498) than the LH (0.276 Hz, ± 0.618) with the LH additionally not surviving the correction for multiple comparisons. Post-hoc tests comparing the respective connections between hemispheres yielded no significant difference. Both HG exhibited a negatively modulated intrinsic connection (LH: -2.236, ± 1.844; RH: -2.150, ± 1.555). The statistics of all connections, together with a graphical depiction of the speech induced network modulations can be found in Figure 4.12 (bottom row).

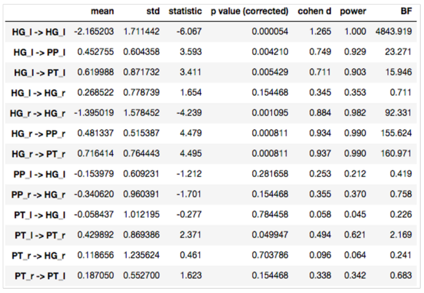
Comparison of connectivity patterns between categories

As mentioned above, all modality specific connectivity patterns resemble the endogenous connectivity to a certain extent. Notably, all categories exhibit positively modulated connections from HG to PP and HG to PT in the RH, as well as, except for music, HG to PP in the LH. This pattern might therefore point to a general importance of these connections during auditory perception. Following the already addressed diverging connectivity pattern of music, two characteristics should be emphasized. First, it is the only modality that yielded a lateralized pattern with positively modulated connections all being within or originating from the RH. Second, the observed negatively modulated connection from PP to HG is the only modulated feedback connection and besides the bilateral HG, the only negatively modulated connection across all categories. In terms of intrahemispheric PT connections, music mirrors the connectivity pattern found for singing. Focusing the later, a bilateral pattern, including HG to PP and HG to PT is extended by an excitatory connection from the LH PT to the RH PT. Therefore, the connectivity pattern for singing represents the most complex one and furthermore the only one containing a modulated connection from LH to RH. Vice versa the speech perception related connectivity pattern was special in that sense that no interhemispheric modulated connections were observed. When tested for differences between hemispheres, both singing and speech yielded no significant characteristics, neither for HG to PP nor HG to PT. However, the prominent involvement of the LH, which is only absent in the RH, indicates its relevance for the processing of auditory stimuli that contain speech, independent of being sung or spoken. It is worth mentioning that all categories yielded negatively modulated intrinsic connections for both HG.

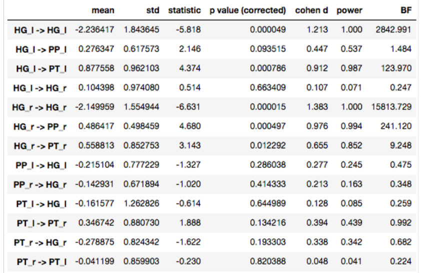




	mean	std	statistic	p value (corrected)	cohen d	power	BF
HG_l -> HG_l	-1.727731	1.868400	-4.435	0.001356	0.925	0.989	141.348
HG_l -> PP_l	0.262869	0.727563	1.733	0.157853	0.361	0.381	0.793
HG_l -> PT_l	0.325240	1.380874	1.130	0.320063	0.236	0.190	0.386
HG_l -> HG_r	0.180876	1.336411	0.649	0.566582	0.135	0.090	0.265
HG_r -> HG_r	-1.909173	1.399075	-6.544	0.000018	1.365	1.000	13209.199
HG_r -> PP_r	0.563938	0.842210	3.211	0.013078	0.670	0.866	10.613
HG_r -> PT_r	0.844667	1.075983	3.765	0.004630	0.785	0.949	33.434
PP_l -> HG_l	-0.312903	0.723739	-2.073	0.092939	0.432	0.509	1.320
PP_r -> HG_r	-0.217644	0.401343	-2.601	0.042431	0.542	0.701	3.265
PT_l -> HG_l	0.312182	0.954883	1.568	0.189473	0.327	0.323	0.635
PT_l -> PT_r	-0.026143	1.018656	-0.123	0.903161	0.026	0.033	0.220
PT_r -> HG_r	0.298517	0.967026	1.480	0.198811	0.309	0.293	0.569
PT_r -> PT_l	0.595933	1.203297	2.375	0.057812	0.495	0.622	2.183



	mean	std	statistic	p value (corrected)	cohen d	power	BF
HG_l -> HG_l	-2.165203	1.711442	-6.067	0.000054	1.265	1.000	4843.919
HG_l -> PP_l	0.452755	0.604358	3.593	0.004210	0.749	0.929	23.271
HG_l -> PT_l	0.619988	0.871732	3.411	0.005429	0.711	0.903	15.946
HG_l -> HG_r	0.268522	0.778739	1.654	0.154468	0.345	0.353	0.711
HG_r -> HG_r	-1.395019	1.578452	-4.239	0.001095	0.884	0.982	92.331
HG_r -> PP_r	0.481337	0.515387	4.479	0.000811	0.934	0.990	155.624
HG_r -> PT_r	0.716414	0.764443	4.495	0.000811	0.937	0.990	160.971
PP_l -> HG_l	-0.153979	0.609231	-1.212	0.281658	0.253	0.212	0.419
PP_r -> HG_r	-0.340620	0.960391	-1.701	0.154468	0.355	0.370	0.758
PT_l -> HG_l	-0.058437	1.012195	-0.277	0.784458	0.058	0.045	0.226
PT_l -> PT_r	0.429892	0.869386	2.371	0.049947	0.494	0.621	2.169
PT_r -> HG_r	0.118656	1.235624	0.461	0.703786	0.096	0.064	0.241
PT_r -> PT_l	0.187050	0.552700	1.623	0.154468	0.338	0.342	0.683



	mean	std	statistic	p value (corrected)	cohen d	power	BF
HG_l -> HG_l	-2.236417	1.843645	-5.818	0.000049	1.213	1.000	2842.991
HG_l -> PP_l	0.276347	0.617573	2.146	0.093515	0.447	0.537	1.484
HG_l -> PT_l	0.877558	0.962103	4.374	0.000786	0.912	0.987	123.970
HG_l -> HG_r	0.104398	0.974080	0.514	0.663409	0.107	0.071	0.247
HG_r -> HG_r	-2.149959	1.554944	-6.631	0.000015	1.383	1.000	15813.729
HG_r -> PP_r	0.486417	0.498459	4.680	0.000497	0.976	0.994	241.120
HG_r -> PT_r	0.558813	0.852753	3.143	0.012292	0.655	0.852	9.248
PP_l -> HG_l	-0.215104	0.777229	-1.327	0.286038	0.277	0.245	0.475
PP_r -> HG_r	-0.142931	0.671894	-1.020	0.414333	0.213	0.163	0.348
PT_l -> HG_l	-0.161577	1.262826	-0.614	0.644989	0.128	0.085	0.259
PT_l -> PT_r	0.346742	0.880730	1.888	0.134216	0.394	0.439	0.992
PT_r -> HG_r	-0.278875	0.824342	-1.622	0.193303	0.338	0.342	0.682
PT_r -> PT_l	-0.041199	0.859903	-0.230	0.820388	0.048	0.041	0.224

Figure 4.12: *Modality specific connectivity patterns and their respective parameter statistics for music (top row), singing (middle row) and speech (bottom row). Left: Connections that were significant at $p < 0.05$ uncorrected are indicated via a dotted arrow, while those that survived a correction for multiple comparisons are display as dashed arrows. Green arrows represent excitatory and red arrows inhibitory connections. Non significant connections are depicted as gray solid arrows. Numerical values represent the strength of a given connection indicated through the mean coupling parameter in Hz. Right: Group level results for the modulatory connectivity parameter estimates after BMA. For each modality and parameter, the group based mean and std are displayed, along with their t statistic and p value obtained through an one-sample t test. FDR indicates if parameters survived the correction for multiple comparisons.*

Random forest classifier

Targeting differences between the connectivity patterns, indicated via their potential ability to predict a given modality, multiple classification tasks following the architecture of those applied in the searchlight and ROI decoding analyses were conducted. In more detail, an RFC was utilized and its significance assessed by means of a permutation test, additionally allowing the assessment of each connection's importance for a given classification task. The corresponding outcomes are displayed in Figure 4.13. All classification tasks neither reached prominent above chance accuracy scores, nor statistical significance. With an accuracy score of 0.31 %, at a p-value of 0.516, the multiclass task was below chance level (0.33 %). The binary tasks also centered around, but above chance level, with singing vs. speech yielded the highest accuracy score (0.57 %, p=0.177) as compared to music vs. speech (0.52 %, p=0.357) and music vs. singing (0.51 %, p=0.397). Focusing the computed importance of each connection, an overall highly comparable pattern with punctual prominent contrasts emerged across the different classification tasks. Notably, four connections show striking similarities. The intrinsic connection of the RH HG yielded a high importance across all classification tasks. The same holds true for the LH feedback connection from PT to HG, pointing to a general importance in terms of diverging functionality of these connections during the processing of the investigated categories.

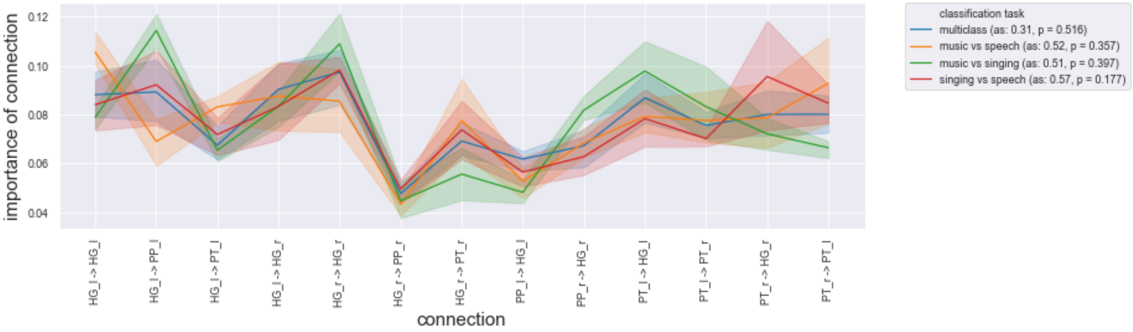


Figure 4.13: *Feature importance of all connections across all classification tasks. Feature importance is displayed on the y-axis and connections on the x-axis. Colors indicate classification task: blue – multiclass, orange – music vs. speech, green – music vs. singing, red – singing vs. speech. Shades indicate variance across CV folds.*

On the contrary, the forward connection from HG to PP in the RH exhibited the lowest importance for all classification tasks, along with the connection from PP to HG in the LH. This is turn indicates comparable characteristics of these connections, independent of the modality. A converse pattern of importance for the classification tasks music vs. speech and music vs. singing was found for the connections intrinsic HG, HG to PP (both LH) and RH PT to LH PT. In more detail, the LH intrinsic HG connection yielded the highest importance for music vs. speech and the lowest importance for music vs. singing, with the same being observed for the connection RH PT to LH PT and the opposite for the LH HG to PP connection. Combining the aforementioned points, another prominent results emerges: while the RH connections intrinsic HG and HG to PP show similar importance values across all classification tasks (high for intrinsic HG and low for HG to PP, respectively), their LH counterparts show the greatest differences between classification tasks, hinting to lateralized different network properties. Notably, the feedback connection from PT to HG in the RH showed a high importance but only for the singing vs. speech classification task.

4.5. DISCUSSION

Hereinafter the obtained results will be critically discussed according to the aims of the here presented study, as well as previous research work. After providing a brief summary of the results, the seemingly hierarchical and lateralized processing patterns will be focused on. Subsequent to addressing points related to acoustic features and the special cortical processing of music, this chapter will conclude with limitations of the current and an outlook to possible further work.

This part of the thesis aimed to describe properties related to the structure and function of the auditory cortex and network associated with the processing of different categories by using a set of complementary approaches. The focus was to identify resources and functional principles in which the auditory system differs for the processing of music as compared to other categories that are comparable with regard to spectrotemporal modulations of acoustic features and social importance. It extends previous research work by three factors: the inclusion of singing as a modality to prevent the often found dichotomous approaches, additionally creating a musical gradient that allows a more in-depth investigation of acoustic feature related differences; the utilization of multiple spatially sensitive analyses that focus on underlying patterns rather than spatially averaged responses enabling to disentangle the specificity of regions involved in the processing; as well as connectivity analyses intended to unravel the functional integration among the auditory network constituting ROIs, facilitating the assessment of the hypothesized hierarchical and lateralized processing. Notably, the same set of ROIs was used in all analysis approaches to provide a detailed description of similarities and differences between music and other categories across multiple levels and properties.

The interaction between hierarchical processing and lateralization

One of the most prominent patterns observed was the hierarchical and lateralized processing that was found across all categories and that varied with respect to them. Focusing the first, the MVPA approaches revealed an increasing specificity and sensitivity along the auditory pathway, more precisely from core to belt and parabelt regions, which is in line with previous research work (Kell et al., 2018; Norman-Haignere et al., 2015; Whitehead & Armony, 2018). In both, the searchlight and ROI-modality-specific decoding analyses, accuracy scores reached their maximum in anterior and/or posterior parts of the auditory cortex, near the border of the STP and STG, respectively PP/STGa and PT/STGp, while accuracy scores in parts around the core region HG were low throughout. As this was observed in all analyses and independent of the classification task it furthermore points to a general organizational and functional principle of the auditory cortex that is not special to the processing of music, but seems to be shared among the majority, if not all, auditory categories. This is in agreement with previous research work and postulated models of cortical auditory processing (Koelsch, 2011a; Zatorre

& Salimpoor, 2013) and furthermore stresses the proposed distinction between primary and non-primary auditory cortex regions (Norman-Haignere et al., 2015; Norman-Haignere & McDermott, 2018). The ROI-modality-specific decoding also illustrates this and additionally indicates a lateralized pattern, as RH ROIs showed the highest accuracy scores for music and LH ROIs for speech, following the hypothesized contrary lateralization of music and speech, despite largely overlapping networks (Peretz et al., 2015; Zatorre et al., 2002; Zatorre & Zarate, 2012). Together with the expressed increase towards non-primary regions, the ROI-modality-specific accuracy scores allow an even more fine-grained distinction between belt and parabelt regions: even though values increase till STG regions, they did not exhibit the mentioned lateralized pattern, as it does not exceed STP regions. This supports the assumption of lateralized specificities in earlier stages that lead to a more prominent categorization and thus distinction of sounds in later stages (Santoro et al., 2017). Furthermore, it points to the importance of applying a network perspective, as the HG seemingly recruits other regions depending on the network necessary for a given processing. This is motivated by two observations. First, throughout nearly all categories, the feedforward connections from HG towards PP and PT are positively modulated, whereas feedback connections were only modulated (negatively) from PP to HG in the RH during the processing of music. Second, its intrinsic connectivity was highly important for the distinction between the category corresponding network patterns, as revealed by the RFC. In combination with the absence of modulated interhemispheric connections between the HGs, it appears that especially the RH HG plays a very important role in the distribution of information and thus generation of networks. Further support is added by the results of the BMS, with modal family A being the winning family and the only family that does not include a modulated connection from RH to LH HG. Together with a bilateral input to HG, the aforementioned remarks point to some sort of evaluative process within the right HG that is heavily involved in the decision of the further processing of sound in diverging auditory categories. This is line with results of a recent study, that could show that integration of the right HG is further advanced that those of other regions which might lead to a more prominent role within the network (Mišic et al., 2018). While both, anterior and posterior regions are recruited, the RFC provides evidence for the significance of the LH HG - PP connection, with the reverse being true for its RH counterpart. Moreover, it exhibits a diverse pattern as its especially important for the distinction between music and singing and least important for the distinction between music and speech, whereas the corresponding RH connection displays a low importance for all distinctions. The A-matrix underlines the importance of these connections, as there they are the only connections that are modulated, additionally in a positive manner. This could hint to a disposition state that enables a fast response to and processing of incoming sound. Considering the high accuracies found in the PP for all classification tasks with a predominant LH weight, it becomes evident, that the DCM and searchlight approach provide distinctive, yet complementary insights with regard to the timescale of processing. In more detail, the searchlight maps can be understood as a result of the connectivity patterns obtained through DCM. As the HG recruits other regions based on an altered intrinsic connectivity, that is more prominent in the RH in order to enable the processing of different sound categories, it yields comparably low accuracies, while the recruited areas, that is PP and PT display high accuracies, especially in the LH. Moreover, accuracies in the HG were slightly higher in the RH than in the LH, which is in line with their importance as revealed by the RFC. Notably, only music exhibited a clearly lateralized pattern, here to the RH, while the remaining categories displayed bilateral patterns, as well as the only modulated feedback connection, that is an inhibition from PP to HG. Besides providing another evidence for the preferred RH processing of music (Koelsch, 2011a; Peretz & Zatorre, 2005; Warrier et al., 2009; Zatorre et al., 2002), it also stresses the importance of the PP (Norman-Haignere et al., 2015; Whitehead & Armony, 2018). On the contrary, the remaining categories, i.e. singing and speech always included modulated connections in the LH, which is in line with the results of previous studies, suggesting a LH dominance for the processing of auditory signals that contain voice (Poeppel, 2003; Zatorre et al., 2002).

Auditory cortex and network characteristics as a function of stimulus properties

The aforementioned properties of the auditory cortex and system, as well as its difference with regard to the processing of diverging auditory categories can be described as a function of the category constituting stimuli and thus are pointing to the importance of their spectrotemporal modulations and resulting feature encoding which supports the majority of previous research work and furthermore builds upon it. This is for example reflected in the high similarity of searchlight maps for the multiclass and music vs. speech classification task. Upon a closer look that focuses on stimuli accompanying features, both the multiclass, as being task with differentiating between all categories and the music vs. speech classifier include the highest possible dissimilarity possible given the stimuli, or in other words the broadest and most distinct disparity. This includes large scale contrasts like the presence and absence of voice, as well as resulting small scale differences in pitch, timbre, tempo and rhythm (Hanke et al., 2015). However, as most of these acoustic features not differed significantly, it appears unlikely that the observed patterns can be attributed to one specific, a subset or a simple linear combination of them, which is in line with research work that used the same data set but different analysis methods and thus also points to the robustness of category distinctions and overlap (Whitehead & Armony, 2018). A further and more fine-grained characterization of the patterns and their evoking parts becomes evident through the fine-grained pairwise searchlight maps (music vs. singing, singing vs. speech), the ROI-modality-specific decoding results, as well as the connectivity patterns as their combination allows a corresponding dissection. Overall, the fine-grained pairwise searchlight maps revealed comparably high accuracy values that were more clustered and fewer in number. Additionally, they were able to resolve the posterior - anterior distinction predominantly observed in the multiclass and music vs. speech searchlight maps. The anterior cluster was present in both of these maps and seemingly resembles a music component. This is grounded in a combination of two observations. This cluster was present in all searchlight maps but showed a degradation exclusively in the music vs. singing map. Hence, it is involved in the differentiation between categories if one of them exhibits spectrotemporal modulations that are more or less characteristic for music. However, if music and singing should be differentiated, the cluster is only marginally involved, possibly due to the shared acoustic features between singing and instrumental music, yet showing some sort of engagement as they are not entirely identical for example with regard to timbre and complexity. This is furthermore supported by the distinguished expression of the anterior cluster in the singing vs. speech map. As here both categories also share acoustic features, but more related to speech, their main difference is the shift of spectrotemporal modulations towards music in singing. The results of the DCM analysis support this assumption due to the prominent engagement of the PP. This also holds true for the cluster's observed LH advantage, as the LH connection from HG to PP yielded an importance gradient from high, within the differentiation between music vs. singing to low, for music vs. speech. This possibly reflects the increased demand of this connection and the PP when two comparably musical categories (music and singing) need to be processed and discerned, whereas the opposite being true for the processing and comparison of highly distinct categories (music and speech). Together with the immense variety of the stimulus material, the PP appears to be crucial for the processing of music, independent of the specific type or the existence of vocals or instruments. This is in favour with the results of the majority of previous research work, while adding insights from two complementary and novel (connectivity) approaches (Angulo-Perkins et al., 2014; Aube et al., 2015; Fedorenko et al., 2012; Leaver & Rauschecker, 2010; Norman-Haignere et al., 2015; Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002; Rogalsky et al., 2011; Tierney, Dick, Deutsch, & Sereno, 2013; Whitehead & Armony, 2018). When focusing the posterior cluster a reversed pattern emerged. While also being present in both broad searchlight maps, it was exclusively restricted to the LH and furthermore might reflect a speech, more precisely voice representation. Like the music component, this also becomes evident through a detailed investigation of the fine-grained searchlight maps. Here the singing vs. speech map revealed a reduction of the posterior cluster, which was no the case in the music vs. singing map where it was the cluster with the highest accuracies. Considering the categories' underlying and defining acoustic features, both singing and speech entail comparable respective spectrotemporal modulations that are representative for the human voice, whereas (instrumental) music and singing are distinguished by those features. Hence, this cluster seems to be heavily involved in the processing of voice related acoustic features, which is why it is only marginally present in the singing vs. speech map, as here the respective differences are drastically smaller as within the comparison of music and singing, with the latter completely lacking those features. This distinction between music and singing stresses this even further, as the corresponding searchlight map is the only one to include a posterior cluster of high accuracies in the RH, even though to a small extent. Additionally, this is supported by the obtained connectivity patterns, as all categories yielded a positively modulated connection from HG to PT in both hemispheres and a high importance of the LH feedback connection from PT to HG within the differentiation between music and singing. In line with prior studies, this cluster could therefore resemble the temporal voice area that preferably encodes and processes the fast temporal and coarse spectral modulations of acoustic features found in the human voice (Agus, Paquette, Suied, Pressnitzer, & Belin, 2017; Belin & Zatorre, 2003; Belin, Zatorre, & Ahad, 2002; S. Brown, 2000; Fecteau, Armony, Joanette, & Belin, 2004; Kriegstein & Giraud, 2004; Pernet et al., 2015; Santoro et al., 2017; Whitehead & Armony, 2018). The aforementioned connectivity patterns and searchlight maps, especially the fine-grained ones furthermore show the advantage of avoiding dichotomous stimuli sets and comparisons that were addressed in the introduction, as a broad comparison between music and speech alone would not have allowed the above described detailed differentiation that is enabled through the inclusion of singing as an auditory category that shares acoustic features with both music and speech. Its intermediate role is also visible in the corresponding connectivity pattern and the ROI-modality-specific decoding results. Incorporating acoustic features of music and speech, singing yielded a highly bilaterally connectivity pattern that largely overlaps with those of the former two, that also poses as the only one to include a modulated interhemispheric connection from LH to RH PT. This overlay was also reflected in the ROI-modality-specific decoding results, as singing was the only category not showing an increase an accuracy from core to parabelt regions. Thus singing is not as specific and precisely differentiated and represented as music and speech. Its rather blurred lines between the former might therefore also loosen their proposed lateralization as also suggested by others (Brattico et al., 2011). From another point of view the searchlight maps and connectivity patterns also allow the dissection of music and speech related acoustic features evident in singing. It is worth mentioning that the stimulus features lead to two further outcomes that are described in this and the preceding section and are connected to the observed lateralization. The first concerns the LH weight of the searchlight maps and is based on the inclusion of categories that contain voice (singing and/or speech) in all classification tasks. Given its social importance and prominence, as well as specialized cortical processing that is predominantly represented in the LH due to the preferential encoding of precise temporal and coarse spectral modulations, the most dominant processing differences with regard to other auditory categories could be found in the LH regions of the auditory cortex. The second entails the bilateral connectivity pattern that was obtained for speech which is in contrast to an assumed strong lateralization to the LH. A possible explanation therefore could be the characteristics of the speech stimuli, as they were very short in duration and included examples from a myriad of languages. Thus they contained little to no syntactic and/or semantic information, while additionally being largely incomprehensible and therefore missed two essential components of speech and language, in turn evoking a rather purely acoustic feature driven processing as intended.

Getting in tune for music

Considering the main question of this part of the thesis at hand, that is “How does the auditory cortex and network distinguish between music and other auditory categories, what is shared and what is different in terms of resource allocation, usage and integration?”, several important points have to be addressed to provide a sufficient answer based on the obtained outcomes. It appears that music shares most, if not all, of the resources of the auditory cortex and system with other auditory categories like singing and speech. However, as noted above and addressed through the applied analysis methods: overlapping does not necessarily mean sharing, as the usage and integration of these resources varies tremendously between categories and results in highly distinguished and precise percepts. This computation is implemented in a most likely hierarchical processing that achieves an increased specificity from primary towards non primary areas of the auditory cortex. Within that music, with its fine-grained spectral and coarse temporal modulations predominantly modulates connections in the RH, which subsequently result in a bilateral processing that exhibits a specialized component in the anterior parts of the STG and STP, that is the STGa and PP respectively.

Limitations and outlook

Given the high dimensionality of auditory processing that is already reflected in the here included categories and their respective stimuli, some inherent limitations and problems need to be addressed. Overall these are situated in three classes: stimuli, methods and different populations. Focusing the first, it should be noted that even if the stimulus set was highly diverse, it does not reflect the complexity of hearing situations humans face every day. Hence, the inclusion of further categories is necessary to further investigate the specialized processing of music (Norman-Haignere et al., 2015; Santoro et al., 2017). This especially refers to other types of music itself. For example, so far insights on the difference between instrumental music, acapella singing and music with vocals is missing. In contrast to the here applied categories this would allow the investigation of an even finer musical gradient, as music with vocals integrates acoustic features of both instrumental music and acapella singing, therefore constituting a highly complex percept. Additionally, excerpts of a myriad of music cultures could shed light on the influence of music enculturation on cortical music processing and possibly unravel the extent to which the seemingly specialized components of music processing are universal. With respect to the stimuli and the paradigm they were included in, the absence of a task has to be considered as a potential source of confounds, as it does not allow an assessment of the stability and robustness of the results. The second class is foremost marked by a lack of multimodal investigations, as most research work, like the one presented here, exhibits a bias towards fMRI and thus spatial characteristics. Even though to a lesser extent in this project (due to the multiband acquisition), this neglects the temporal aspects of the processing, which are tremendously important for the investigations of music as an auditory stimulus, given its dynamic, non-static nature. This could be addressed by two approaches, that is spatio temporal searchlights and generalization across time decoding. While the first is based on fMRI response profiles and enables the examination of specificity and sensitivity of certain ROIs or the whole brain over time, the latter utilizes the high temporal resolution of EEG data to indicate which time points of the perception distinguish most between categories. In light of the aforementioned necessity of more complex and diverse stimulus sets, encoding models based on another naturalistic experiments pose as an attractive approach that could tackle both, the specialization for music and its temporal evolution. The need for multimodal investigations furthermore extends to the inclusion of structural properties of auditory cortex regions, as their correlation with functional characteristics has been repeatedly shown (D. Marie et al., 2015). Albeit Whitehead & Armony (2018) compared the results of the same data set as used here between different DAPs, a certain influence cannot be ruled out, as they applied different methods within their analysis. In order to partially resolve this pitfall, a respective comparison of the here utilized multivariate and computational approaches needs to be conducted. The third and final class depicts the inherent need for the investigations of different populations, especially those assumed to display an altered cortical organization. This for example includes left-handed individuals and especially individuals with musical training, as the latter exhibit a potential diverging cortical organization and function of the auditory cortex and beyond due to the tremendous cortical plasticity evoked by musical practice and engagement.

5

INVESTIGATING MUSIC GENRE PROCESSING USING A MULTIMODAL APPROACH: INSIGHTS FROM MUSIC FEATURE BASED, CONCEPTUAL & BEHAVIORAL MODELS

In this chapter the last project of the thesis at hand is presented. As this project was highly multimodal in its nature and used identical stimuli, paradigms, procedures and analysis approaches across it's sub-parts, the chapter's overall introduction is followed by a methods section containing details to the aforementioned points. Afterwards the chapter proceeds by means the sub-parts, starting with the project's fMRI component, ensued by the conducted EEG experiment and concluding with a purely behavioral implementation targeting different populations. Subsequently, the different parts are discussed together and in relation to one another within an overall discussion.

Explanation and identification of personal contribution and data usage

Hereinafter the contribution of the thesis' author (PH) with regard to each project part will be outlined and additional information be provided.

General remarks

Within the project, one master thesis and two bachelor theses were conducted. They are connected to the three parts of the project:

"A Pilot Study of the Neuronal Processing of Different Music Genres in Auditory Brain Areas" by Mirjam Schneider (MS) as a master thesis for the program "Cognitive and Integrative Systems Neuroscience" at the Philipps-University of Marburg.

"Neuronal processing of different music genres investigated with EEG: The role of music Features" by Marie Frerich (MF) as a bachelor thesis for the program "Psychology" at the Philipps-University of Marburg.

"Music genre perception similarities in musicians and non-musicians: a behavioral approach by inverse multidimensional scaling" drafted by Klara Brinkman (KB) as a bachelor thesis for the program "Psychology" at the Philipps-University of Marburg.

Conceptualization and implementation

The herein described project and its parts were conceptualized solely by the author of this thesis (PH). The same accounts for the way the different parts were implemented.

Stimulus set creation and validation

PH and MS created the stimulus set. PH, MS and KB implemented the online validation survey. MS, MF and KB acquired the online validation survey data. PH, MS and KB analyzed the online validation survey data with analysis scripts written by PH.

Nothing compares to you I - music genre representations revealed by fMRI

PH and MS implemented the paradigm and collected the data. Although demonstrating comparable research questions with the thesis of MS, the remaining work related to this part of the project described and presented in this thesis was solely conducted by the author (PH). Notably, different analysis approaches were used.

Nothing compares to you II - time course of music genre representations

PH implemented the paradigm. MF collected and preprocessed the data. Although demonstrating comparable research questions with the thesis of MF, the remaining work related to this part of the project described and presented in this thesis was solely conducted by the author (PH). Notably, different analysis approaches were used.

Nothing compares to you III - the influence of musical training on music genre perception
PH implemented the paradigm. PH and KB collected the data. Although demonstrating comparable research questions with the thesis of KB, the remaining work related to this part of the project described and presented in this thesis was solely conducted by the author (PH). Notably, different analysis approaches were used.

Data usage
As all participants in all three projects signed the "Open Brain Consent Form" (APPENDIX A) the entirety of the acquired data can freely be reused and shared. To aid and facilitate the analysis across all project parts the author of this thesis (PH) developed the open source toolbox "URIAL - Utilities for Representational Similarity Analysis in Python" which is freely and publicly available in an online repository (https://github.com/PeerHerholz/URIAL) and was used in the theses of all parties involved in this project.

5.1. INTRODUCTION

The application of novel and advanced analysis methods within the realm of auditory neuroscience enabled recent studies to investigate the sensitivity and specificity of the auditory cortex with regard to different modalities (Cha et al., 2016; Fedorenko et al., 2009; Whitehead & Armony, 2018; Zatorre & Baum, 2012). The respective results provided evidence for overlapping, yet functionally distinct regions and networks involved in the processing of various types of auditory stimuli and pointed to the imported role of their corresponding underlying acoustic features, as well as their combination, in creating highly specific percepts (Santoro et al., 2017; Whitehead & Armony, 2018). These are in turn most likely responsible for the aforementioned processing as previous research work could further show that regions of the auditory cortex exhibit a differentiated preferential encoding of acoustic features along the hierarchy of cortical auditory processing, which is furthermore influenced by the hemispheres (Alluri et al., 2012; Alluri et al., 2013; Santoro et al., 2014; Santoro et al., 2017; Zatorre & Baum, 2012; Zatorre et al., 2002). Due to their striking similarities with regard to higher cognitive processing and social importance, a vast amount of studies focused on speech and music (Peretz et al., 2015), in some cases also including singing (Aube et al., 2015; Whitehead & Armony, 2018). Depending on the question at hand, as well as the processing level of interest, these modalities can be both too broad and too narrow at the same time. Focusing the first, speech, music and singing can pose as very coarse classes of sound that include a tremendous variety of classes themselves and are defined through a certain expression and combination of acoustic features which are shared among their included subclasses. Any effect related to possible differences and similarities can therefore be confounded by the versatility of the stimulus material included in each class (Livingstone, Peck, & Russo, 2013; Saitou, Goto, Unoki, & Akagi, 2007). Addressing this potential issue might on the other hand lead to a decrease in ecological validity of the obtained results, as heavily controlled stimuli, e.g., using syllables for speech and tone progressions for music (Hasson, Nusbaum, & Small, 2007; Kilian-Hütten, Valente, Vroomen, & Formisano, 2011; Leaver & Rauschecker, 2010), most likely not capture and reflect the natural complexity neither of a given sound class nor of the extremely rich acoustical environment humans are surrounded by (Giordano et al., 2013). Hence, this leads to the second point, as the number of stimuli within a given class is limited and even the most diverse stimulus set will be restricted to a certain class. Addressing this circumstance and gap, a growing number of research focuses on the categorization of sounds (Giordano et al., 2013). In more detail, "Auditory categorization is a computational process in which sounds are classified and grouped based on their acoustic features and other types of information (e.g., semantic knowledge about the sounds)" (Tsunada & Cohen, 2014, p.1). Notably, besides the immense extension to a multitude of sound categories, this line of research inherently includes an additional question that is at the core of auditory neuroscience and therefore also the here presented thesis: how are acoustic features processed and possibly combined with each other, as well as higher order information to yield a stable percept that is abstract in nature, acting as a representation that allows to distinguish and therefore categorize a myriad of sounds? The majority of previous research work on auditory categorization proposes that this computation is achieved along an "...anteroventral axis in temporal cortex..." (Hjortkjær, Kassuba, Madsen, Skov, & Siebner, 2018, p.295) which is in turn grounded in the dual stream model of auditory perception. Originally formulated and observed in the visual domain (Goodale & Milner, 1992) it was later extended to auditory perception (Hickok & Poeppel, 2007; Mithen, Morley, Wray, Tallerman, & Gamble, 2006; Romanski et al., 1999; Zatorre & Salimpoor, 2013) and constitutes that cortical auditory processing is organized in two distinct streams that originate from core towards belt and parabelt regions of the auditory cortex and subsequently to other brain regions in a hierarchical and parallel manner. In more detail, a ventral and a dorsal stream are identified, with the first progressing towards “targets in superior and inferior temporal sulcus and gyrus, eventually terminating in the inferior frontal cortex” (Zatorre & Salimpoor, 2013, p.10431) and focusing on “what” was heard, while the latter progresses towards regions in the “parietal, premotor, and dorsolateral frontal cortices” (Zatorre & Salimpoor, 2013, p.10431) and is involved with assessing “where” it was heard, as well as audiomotor processing (Bizley & Cohen, 2013). Interestingly, the dual stream model apparently already manifests itself in brainstem regions of the auditory nerve (Pickles, 2015) and furthermore constitutes another striking similarity between the auditory system of humans and non-human primates (the other being tonotopy (Baumann et al., 2013; Schönwiesner et al., 2015), as it is found in both (Rauschecker & Scott, 2009) (for a graphical depiction of the dual stream model in humans and non-human primates please see Figure 5.1).

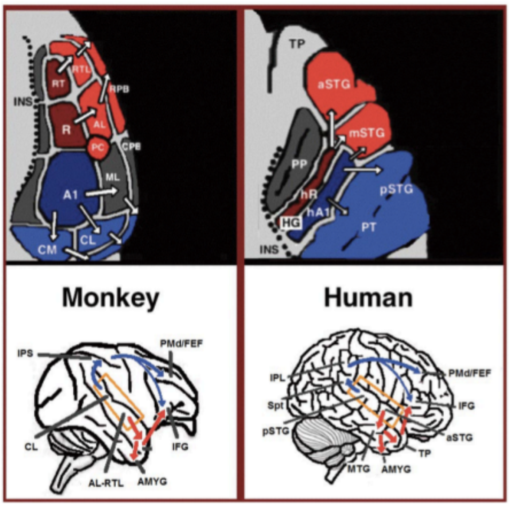


Figure 5.1: *Dual stream model of auditory processing in non-human primates (left) and humans (right). Upper row displays an axial view with outlined and labeled species-specific subfields, as well as white arrows indicating the processing stream, i.e. direction. Colors reflect ventral (red) and dorsal (blue) streams respectively. Lower rows illustrate a schematic visualization of both species' brain from a sagittal view with the orange rectangle indicating auditory cortex and colored arrows processing streams (corresponding to the upper row) towards labeled regions outside of the auditory cortex. Source: Poliva et al. (2017, p. 4).*

Even though studied extensively, the mentioned preferred computation of auditory representations along the ventral stream is however subject to an ongoing lively debate as an increasing number of studies provided evidence that also the dorsal stream (Timothy D. Griffiths & Warren, 2002), core regions (Kilian-Hütten et al., 2011) or even the entire auditory cortex (Formisano, De Martino, & Valente, 2008; Staeren, Renvall, De Martino, Goebel, & Formisano, 2009) are involved in this highly important part of auditory processing. Additionally, studies focusing auditory categorization utilized a broad range of naturalistic sounds ranging from human and non-human vocalizations to diverse nature sounds (Altmann, Doehrmann, & Kaiser, 2007; Belin et al., 2002; Belin et al., 2000; De Lucia, Camen, Clarke, & Murray, 2009; Doehrmann et al., 2008; L. R. Engel, Frum, Puce, Walker, & Lewis, 2009; Fecteau et al., 2004; Galati et al., 2008; Giordano et al., 2013; Leaver & Rauschecker, 2010; Murray, Camen, Andino, Bovet, & Clarke, 2006; Pizzamiglio et al., 2005). While this addresses the aforementioned low ecological validity of studies including only a very small subset, it also exhibits their too-broad-too-narrow problematic confounds as a very broad stimulus set will create a context dependency through that more similar sounds will be perceived as more related and further apart from other sounds. Creating a broad representation and categorization, it however renders fine-scale differences within categories that might lead to the formation of subcategories unresolvable. Indeed, only a limited amount of research work was dedicated to within-category diversity and information and if so, this was mainly restricted to speech (M. H. Davis & Johnsrude, 2003; Goll, Crutch, & Warren, 2010; Hasson et al., 2007; Leaver & Rauschecker, 2010, 2016; Okada et al., 2010; Rauschecker & Scott, 2009; Rauschecker & Tian, 2000; Romanski et al., 1999; Scott, Blank, Rosen, & Wise, 2000). While "Speech is indeed a primary exponent of robust category perception in humans...the ability of the primate brain to extract detailed category information from sound sources presumably generalizes beyond speech processing." (Hjortkjær et al., 2018, p.301). One, if not the ideal, candidate to extend the investigation of auditory categorization and representations capable of providing insights into the presumed generalization is music. Evolutionarily and socially at least as important as speech (S. Brown, 2000), it is also a very frequent sound category and as, or even more, diverse as the former. It is furthermore well suited, because it is highly feature, that is acoustic feature, based which leads to both a precise clustering of music excerpts when compared with other sound categories and a highly distinct organization when compared within its category (Gjerdingen & Perrott, 2008; Istók, Brattico, Jacobsen, Ritter, & Tervaniemi, 2013; Sturm, 2013; Tzanetakis & Cook, 2002) and hence the formation of subcategories. In combination with, yet unknown, high level features, these low level features lead to the formation of subcategories and representations within the sound category music, or in other words music genres. Considering the high dimensionality and complexity of sound categories, especially music genres, humans are remarkably good at recognizing and categorizing them, even within a very short period of time (Gjerdingen & Perrott, 2008; Ogg, Slevc, & Idsardi, 2017). Together with music genre based subcultures and a proposed connection between personality traits and preference for certain music genres (Greenberg et al., 2016) this points to the tremendous social significance and arguably tuning of the human brain to auditory, that is, music categories. A vast amount of previous research work targeting the perception of music and its categories focused on low level, meaning acoustic features. These respective studies deployed a vast variety of decoding and encoding approaches to investigate the mapping between acoustic feature and fMRI-voxel space, i.e. where in the human brain certain acoustic features are encoded (Casey, 2017). Exemplary work of Alluri et al. (2012, 2013) and Toiviainen et al. (2014) used a set of encoding models in combination with a naturalistic music listening paradigm to achieve this goal. Within their studies, participants were presented with a range of music excerpts varying in length and type. While Alluri et al. (2012) used an 8 min. modern tango within a passive listening paradigm, Alluri et al. (2013) and Toiviainen et al. (2014) used the same set of stimuli and data comprising two different medleys of 15 min. each (one being the B-side from The Beatles' Abbey Road (1969) and one entailing four different music pieces from four different genres) within an attention task where participants should respond to a target word that was presented along the music excerpts. Afterwards they extracted a myriad of acoustic features and utilized them within various regression models in order to predict the brain's responses to these features, hence their encoding and mapping. Their results pointed to the extensive importance of the STG and HG in both hemispheres with respect to the processing of music, while also providing evidence for the involvement of other regions outside the auditory cortex (e.g., cerebellum and hippocampus). Being most prominent for timbral features, comparable results were also obtained for tonal and rhythmic features. Toiviainen et al. (2014) furthermore showed that certain segments of the music piece could be classified through the fMRI data. Despite having a high ecological validity, their work deployed a rather restricted number of music genres and thus expression and combination of acoustic features. Conversely, Hoefle et al. (2018) used a more diverse set of 40 music genres with a shorter duration of 46 sec. each in a comparable approach where they combined encoding and decoding analyses, while additionally varying the spatial and temporal coverage of the included data. Based on that, they were able to generalize across the variety of included music genres and identify novel, previously unheard, music pieces. In line with the aforementioned research work, acoustic features were encoded broadly across the auditory cortex, comprising all hierarchical stages. While these approaches allow a precise mapping of acoustic features, that is independent of music genres, they provide only limited insights of how the combination of these features that leads to the perception of music genres are represented and how they could possibly be explained. Even though comparably established in more general and broader contexts (Giordano et al., 2013), research work on auditory categorization and their respective representations, additionally incorporating high level features, solely within music is rather limited. With the majority of the respective studies being conducted in the behavioral (Gjerdingen & Perrott, 2008; Pálmason, Bjoern, Jónsson, Schedl, & Knees, 2017; Siedenburg & Müllensiefen, 2017) and computational domain (Ajoodha, Klein, & Rosman, 2015; Dawson, 2018; Esparza, Bello, & Humphrey, 2015; Leon & Martinez, 2014) a sufficient amount of neuroscientific data is missing. In fact, so far only four studies investigated the neuronal correlates of music categorization (Casey, 2017; Casey et al., 2012; Güçlü et al., 2016; Levitin & Grafton, 2016), with two of them relying on the same data set (Hanke et al., 2015) and thus music genres, as well as paradigm (Casey, 2017; Güçlü et al., 2016). All of them used representational similarity analysis (RSA) (Kriegeskorte, Mur, & Bandettini, 2008) to assess representations of music genres reflecting their categorization through representational dissimilarity matrices (RDMs). However, their respective research subject and implementation varied to a great extent. Levitin & Grafton (2016), as a case study, conducted a set of three experiments within the same participant. Only one of them targeted the categorization of music genres while the others focused on creativity across domains and imagined and heard music respectively. However, the included music genres were highly specific as they were specifically gathered for the participant to furthermore include a distinction with regard to familiar and unfamiliar music. Their computation of RDMs was based on a subset of voxels, entailing a highly distributed pattern covering several parts of the temporal and frontal lobes, yielded through comparing trials during which music was presented to a silent baseline. The obtained results hinted at a categorization of music genres that followed theoretical descriptions and definitions, e.g. popular music was similar to one another and dissimilar to classic. Besides spanning a comparably small and less structured music genre space and unspecific, as well as a broad selection of neuronal features, the exploration of the results remained on a rather descriptive stage, neither including low nor high level feature based models. In contrast, the work of Casey et al. (2012) used a more precise set of music excerpts spanning five genres, including 5 examples per each genre. Additionally, they tested the predictive ability of four acoustic features, i.e. spectrum, pitch, chroma and timbre with regard to the observed categorization in the bilateral STS. In line with previous research (Allen et al., 2018; Halpern, Zatorre, Bouffard, & Johnson, 2004; Leaver & Rauschecker, 2010) they found that the timbral feature accounted best for the neuronal categorization, in both a comparison between neuronal and acoustic feature RDMs, as well as a multivariate multiple regression between these feature vectors. Due to the manually defined mask of the STS the results unfortunately provide no insights with respect to the potentially diverging computation and categorization mechanisms along the processing streams of the cortical auditory system. Furthermore, no high level feature models were included and as in Levitin & Grafton (2016) categories were rather broad and allowed no exploration of subcategorization, that is genres. In more detail, their music genre space entailed ambient, rock ‘n’ roll, heavy metal, symphonic and roots country with their respective five examples being gathered by using a representative seed artist within each genre. Unfortunately, no motivation for this genre selection was provided, which is furthermore, comparably to Levitin & Grafton (2016), restricted and not exhibiting a certain level of structure, as well as theoretical categorization, as e.g., prominent and frequent music genres are not included and for example, rock ‘n’ roll and heavy metal could be perceived as two sub categories to the main category rock. Addressing the points related to spatial specificity and features, Casey (2017) used a newly acquired data set (Hanke et al., 2015) that utilized the identical music genres and procedures as in Casey et al. (2012), but included more brain regions and models. More precisely, they assessed the representation of the music genres in three bilateral ROIs, covering core (HG) and parabelt (anterior and posterior STG) auditory cortex thus increasing the spatial resolution of music categorization along the auditory pathways, yet data from belt regions like PP and PT, found to be heavily involved in acoustic feature encoding and categorization by previous studies (C. Chen, Halpern, Bly, Edelman, & Schlaug, 2000; Timothy D. Griffiths & Warren, 2002; Norman-Haignere et al., 2015; Whitehead & Armony, 2018), was not obtained. The ROI corresponding categorizations were highly similar to one another, as well as to the one from Casey et al. (2012). The model space was extended by means of a behavioral model based on a categorization task that displayed tremendous correlations with the representations of all ROIs, pointing to an assumed reflection of neuronal categorization in the observed behavior (Casey, 2017). Notably, in contrast to their earlier work, they used RSA on rather formal level, that is matrices reflecting the distinguishability between music genres, as the RDMs were based on binary classifications and therefore entailed confusion matrices. While certainly possible, it might not be the best suited approach for the investigation of representations, as recent research suggests (Walther et al., 2016), that RSA based on confusion matrices tends to be less reliable as when based on continuous distances (e.g., like the euclidean distance). Addressing a more fine-grained representation, their analyses was additionally twofold in that confusion matrices were either based on genres or examples. Here example based confusion matrices reflected genre based ones across all ROIs, as ambient and its examples achieved the highest classification accuracy and were the least confused with all other examples and genres, followed by symphonic. However, as all examples within a given genre should be representative and no further characterization between the examples were made, information with regard to potential sub categories within a genre are yet again nonexistent. Using the identical data set (Hanke et al., 2015) Güçlü et al. (2016) focused on complex computational models, i.e. convolutional neural networks (CNN), to evaluate their ability to resemble neuronal processing and categorization. Being their primary research subject, they compared the representation of different CNN layers to neuronal data on the one hand and to acoustic features on the other. In this setting, the latter served as control models. In accordance with Casey et al. (2012) their neuronal target was rather broad as they utilized the entire bilateral STG, yet again missing potentially fine-grained differences along the auditory pathways. However, their results indicate a significant correspondence between deeper CNN layers and more posterior regions of the auditory cortex which is in line with previous research work suggesting an increasingly complex representation towards these very regions (Casey, 2017; Casey et al., 2012; Timothy D. Griffiths & Warren, 2002). Besides the repeatedly addressed lack of spatial sensitivity and restricted music genre space, the above discussed studies focusing neuronal correlates of music categorization share further potential confounds which are outlined in brief hereinafter. The used model spaces are rather small (or even nonexistent) and additionally exhibit an absence of high level features models, thus not enabling the assessment of proposed interaction between bottom-up (low level acoustic features) and top-down (high level features) crucial to the computation of auditory (Giordano et al., 2013; Hjortkjær et al., 2018), here music categories. Additionally, the acoustic features were solely reflecting spectral properties, thus missing temporal characteristics like rhythm and tempo. Thus, characteristics and low level features crucial to the perception and categorization of music into genres (Honing, 2012) and their predictive ability were not tested so far. As all studies exclusively used fMRI, no results concerning the temporal evolution of music categorization exist, even though music and its genres are largely defined by means of their spectrotemporal feature modulation over time. With the exception of Levitin & Grafton (2016), the participants were always presented with a task related to the music excerpts. Albeit not being an explicit categorization but an attention task asking participants about certain characteristics of the previously heard music excerpt (e.g., “Was there a guitar in the song you just heard?”), this results in a modulated processing and therefore most likely modulated the categorization of the perceived music as outlined by studies focusing different sound categories (Hjortkjær et al., 2018), while additionally evoking a fairly unnatural (besides the MRI setting) listening situation. The latter refers to music consumption as a more holistic process that usually does not entail listening or paying attention to certain acoustic features only, therefore not representing everyday music listening, which is furthermore often used for other purposes (Linnemann et al., 2016). Moreover, all studies used sets of music excerpts that included both vocal and non-vocal music. Although reflecting the contemporary landscape of music genres in which most genres included lyrics and are therefore vocal, the mixture of vocal and non-vocal leads to two potentially prominent confounds that can be assigned to low level and high level features. Regarding the first, the inclusion of vocals will alter the acoustic features of a given music excerpt tremendously as their underlying spectrotemporal modulations are shifted towards those of speech, at least to a certain extent (Saitou et al., 2007). In fact, all of the discussed studies indicated that the largest dissimilarities between their music excerpts are grounded in the vocal vs non-vocal distinction of the respective genres (Casey, 2017; Casey et al., 2012; Hanke et al., 2015). Thus, a modulated processing within early stages of the hierarchy, which is assumed to heavily rely on acoustic feature encoding (Koelsch, 2011a), cannot be precluded and might distort potential insights. Concerning the second, high level features reflected in comparably late stages of auditory processing or even outside the auditory domain, like syntax and semantics, as well as emotions and aesthetics, might be triggered through vocals (Koelsch, 2011a, 2011b). In turn, these features might influence top-down processes involved in the categorization of music via e.g., introducing factors that are conveyed through the included speech rather than the music itself. Based on within-category studies focusing on speech (Bonte, Hausfeld, Scharke, Valente, & Formisano, 2014; Formisano, Martino, Bonte, & Goebel, 2008; Kilian-Hütten et al., 2011) or on emotion evoking music and the difference between vocal and non-vocal (Brattico et al., 2011), furthermore pathways and even hemispheres could be recruited differently (Brattico et al., 2011). One additional factor that hasn’t been addressed so far is the potential influence of musical training which is in fact often referred to as a role model for the human mind’s intense plasticity (Schlaug, 2015). A multitude of studies provided evidence that musical training leads to long term alterations of processes involved in auditory perception and beyond, including memory and motor (Habib & Besson, 2009; Herholz, Lappe, Knief, & Pantev, 2008; Herholz & Zatorre, 2012; Jäncke, 2009; Luo et al., 2012; Pantev & Herholz, 2011). As these changes are reflected in brain structure and function, as well as behavior (C. Marie, Kujala, & Besson, 2012; Micheyl, Delhommeau, Perrot, & Oxenham, 2006; Parbery-Clark, Skoe, Lam, & Kraus, 2009; Schellenberg & Moreno, 2010; Strait, Kraus, Parbery-Clark, & Ashley, 2010; Tervaniemi, Just, Koelsch, Widmann, & Schröger, 2005), with their expression and manifestation being dependent on the duration and type of musical training, they pose as potentially vigorous confounds. While Levitin & Grafton (2016) explicitly tested one highly trained musician and the sample of Casey et al. (2012) including participants with varying musical expertise, no information with regard to this characteristic was assessed or provided by Hanke et al. (2015) and Casey (2017). Summarizing all of the aforementioned studies, previous research work suggests that the computation of auditory categories is achieved through a combination of low and high level features along a ventral pathway from core towards parabelt regions of the auditory cortex. However, based on broad sound classes, including a myriad of natural sounds, comparable little is know about the subcategorization, thus potential high complexity, within a given category. Concerning this, speech posed as the primary target of previous studies, while the complex subcategorization of other equally important and complex sound classes like music received fairly insufficient attention, especially in the realm of neuroscientific investigations. The limited amount of research dedicated to this question shares a number of restrictions with regard to specificity and sensitivity which are reflected in three factors: music genre space, modality and coverage and possible models of explanation. The so far included narrow space of music genres prevented insights into the potential subcategorization of music in main and subgenres, as well as the relationship between genres across both of these levels. As exclusively fMRI was used to this point, additionally including a delimited extent of covered brain regions, information on the temporal evolution of music categorization, also along the auditory processing hierarchy, incorporating the dorsal pathway and hemispheric differences, is almost nonexistent. Comparably, the thus far tested models covered mainly spectral aspects of low level, i.e. acoustic features, and therefore did not account for other, especially high level, i.e. semantic features that could possibly provide an explanation for the observed categorization of music genres. Based on inconsistent participant samples that were not controlled for musicianship the results of the previous might additionally be distorted by an altered music categorization present in trained musicians as compared to non musicians.

5.2. The present study

Aiming to provide new insights into the computation of music genre categorization and possible underlying mechanisms, the here presented study addressed all of the aforementioned potential problems in a highly multimodal approach. Beginning with the generation of a more pronounced music genre space, a twofold procedure was conducted. Initially, a diverse set of music genres was defined based on previous studies, as well as insights from musicology and computational signal processing that additionally allowed a more fine-grained categorization in main and respective subgenres. Moreover, supplementary factors like unfamiliarity and frequency were incorporated. Next, the entire set was rated with regard to its representativeness by an independent group of participants that was not included in the main experiments. Intending to foster a precise characterization of music genre categorization, the thus created music genre space was deployed within a highly multimodal approach in order to capture different aspects of the corresponding processing. In more detail, this entailed fMRI and EEG for neuronal correlates (targeting spatial and temporal properties respectively), behavior as a high level feature, as well as computational implementations that focused on both low level (acoustic features) and high level (semantic features) explanation attempts. To this end, the study was divided in three parts: an fMRI project, an EEG project and a purely behavioral version. With respect to the parts that involved imaging, the music genre were used within a passive listening paradigm to achieve an unmodulated processing thus categorization of the excerpts. Addressing a high resolution in the spatial and temporal domain, the categorization was assessed within fMRI through a collection of anatomically defined ROIs that covered multiple processing stages along the ventral and dorsal pathway from core over belt to parabelt regions that were additionally hemisphere-specific and within EEG through a set of predefined time windows reflecting different segments and therefore processing stages of the excerpts. The same behavioral paradigm, capable of gathering the high dimensionality and complexity of music genre spaces, was utilized throughout all projects in order to obtaine and enable a comprehensive linking between neuronal and behavioral categorization. Furthermore, the identical set of models was tested across all parts of the study. This comprised a broad range of low level, i.e. acoustic feature based models which entailed spectral and temporal properties, as well as high level, i.e. models based on hypotheses and semantics related to the music genre categorization. In line with previous research work, the study was settled within the framework of representational models, more precisely RSA, as it offered several advantages which are grounded in the level of abstraction that is inherent to RSA. Here, the categorization of the music genres was described through their dissimilarity to one another thus their representation, yielding to the two terms being exchangeable in the here presented study. Besides providing an intuitive way to assess the amount and structure of the categorization, RDMs, containing pairwise distances for all items, are indexed by the total number of music genres and therefore can be used as a summary statistic which in turn can be compared and integrated between modalities and models. A graphic, illustrating the conducted approach can be found below (Figure 5.2). Taken together, the objective of the study at hand was to provide an extensive assessment of neural correlates related to the categorization thus creation of representations of music genres across space and time, as well as how these processes can be explained through models of varying source and complexity. Probing the potential influence of musical training, the third part furthermore included a comparison between highly trained musicians and non musicians in a pure behavioral setting, but otherwise using the above described approach. Thus, enabling first insights of music categorization within different populations.

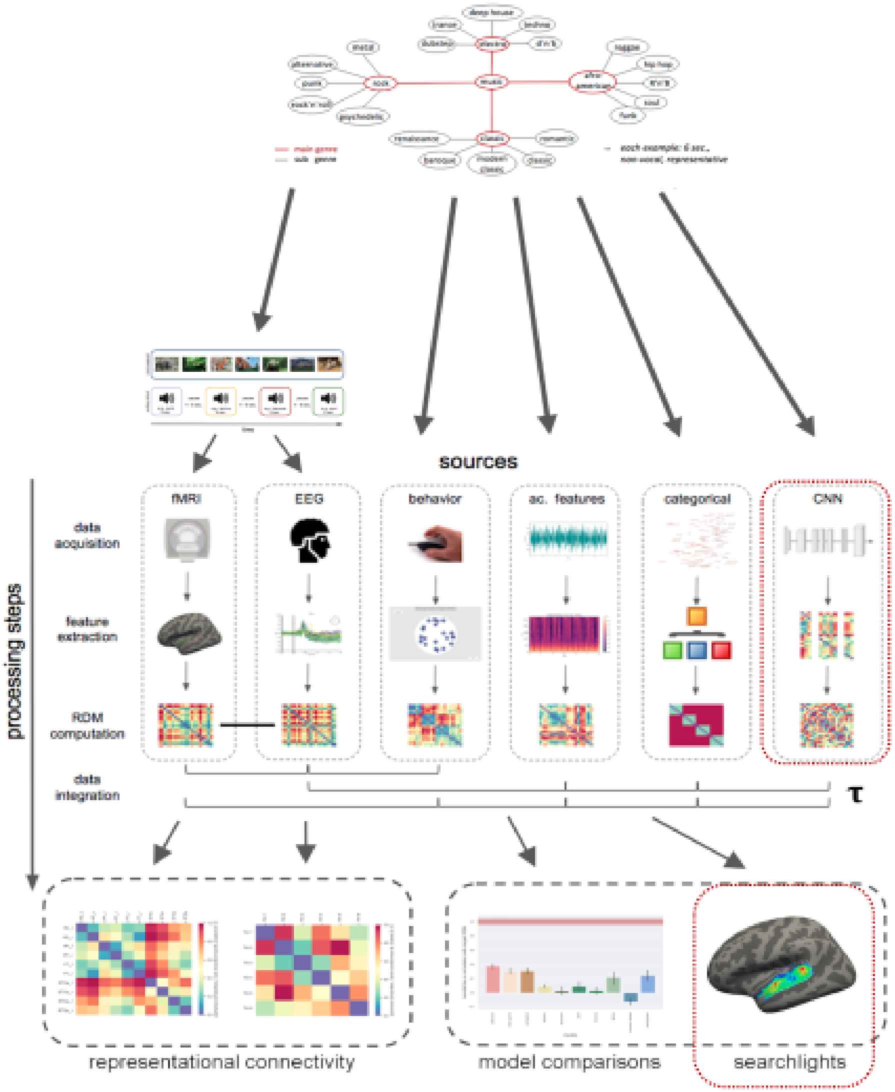


Figure 5.2: *Graphical depiction of the applied multimodal approach. Dotted red rectangles indicate possible extensions.*

5.3. NOTICE ON ANALYSIS SOFTWARE

The analysis within this project of the thesis made heavy use of the following python packages: SciPy (Jones, Oliphant, & Peterson, 2001), NumPy (Oliphant, 2006), Pandas (McKinney, 2010), seaborn (Waskom et al., 2018), Matplotlib (Hunter, 2007), Nipype (Gorgolewski et al., 2011), MNE (Gramfort et al., 2014), scikit-learn (Pedregosa et al., 2011) and URIAL (https://github.com/PeerHerholz/URIAL). The respective functions are referred to within the corresponding parts of this chapter. All other utilized analysis software is particularly named and referenced.

5.4. STIMULI SET CREATION AND VALIDATION

The first part of the project consisted of generating a diverse and validated stimuli set that possibly can overcome the aforementioned potential problems. The respective endeavours are described hereinafter and cover all parts of the conducted stepwise procedure beginning with the initial selection of music genre and examples. Afterwards the stimuli validation process is described, including the utilized design and analysis of the obtained data.

5.4.1. STIMULUS SET PREPARATION

Selection of music genres

To address the mentioned possible problems in previous studies (Casey, 2017; Casey et al., 2012; Hanke et al., 2015; Levitin & Grafton, 2016) with regard to stimuli selection, the first step of the here described project targeted these points and involved the creation of a project specific map of music genres, divided in main and subgenres, as well as a stimulus validation regarding representativeness through an online survey. The initial music genre space was based on the studies of Hanke et al. (2015), Casey et al. (2012)(2012) and Casey (2017) and further extended through an intense review focusing genre definitions, descriptions and clustering from musicology (Holt, 2007; Pachet & Cazaly, 2000; Pulman, 2006) and computer science (signal processing) related literature (Ajoodha et al., 2015; Esparza et al., 2015; Ghosal, Chakraborty, Dhara, & Saha, 2015; Leon & Martinez, 2014), while also prominently using publicly available resources that combined these two realms, such as www.musicmap.info, www.musicgenrelist.com and www.everynoice.com. Overall, the selection of to be included genres followed three specific rules: First, all genres had to part of the western tonal system to ensure that potential participants had a certain familiarity with and exposure to the genres, as well as to rule out possible artifacts related to their development, more precisely music enculturation (Campbell, 2010; Cross, 2001; Morrison, Demorest, & Stambaugh, 2008). Second, (sub) genres had to be distinct in such a way that they would be similar enough to belong to the same main genre (in comparison to all other (sub) genres), while at the same time be also clearly separable within such. Third, both main and subgenres should reflect a broad range of music categories average WEIRD (Henrich et al., 2010) participants are regularly exposed to, e.g., in radio, streaming platforms, TV and movies. Taken all of these points together, a project-specific music genre map was generated, consisting of 20 genres, clustered as 4 main genres, having 5 sub genres each. A graphical representation of this map is depicted below in Figure 5.3. Notably, prominent and well known genres like pop, blues or jazz (among others) were deliberately excluded as they were either or too broad in themselves or could not fulfill the second rule regarding genre selection.

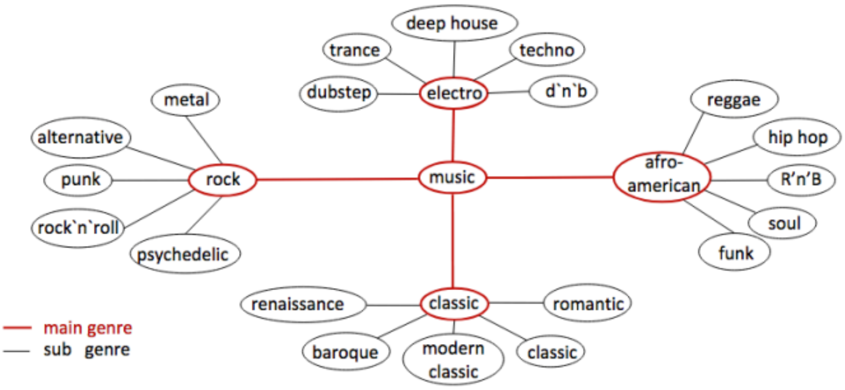


Figure 5.3: *Graphical representation of the project specific music genre map. Four main genres (red outlines) included, each further divided into five subgenres (black outlines), totaling in twenty genres.*

Selection of genre examples/stimuli

To capture the variability and enable the validation of a representative example of a given genre, ten different examples were gathered for each, based on characteristics defined via musicology and music theory work focusing genre definitions and descriptions (Gjerdingen & Perrott, 2008; Holt, 2007; Lena & Peterson, 2008; Pulman, 2006): instrumentalisation, melody, harmony, pitch, tempo, rhythm, etc. . All examples for each given genre were extracted from initial or middle parts of the song, non-vocal and six seconds in length. While the first point concerns representativeness and sound/pattern homogeneity over time (preventing fast and harsh changes with regard to the mentioned music characteristics within a given example), the second is related to major differences that occur when comparing vocal and non-vocal music, especially in terms of spectral sound features (e.g., spectrum, etc.) (Casey, 2017; Casey et al., 2012; Hanke et al., 2015), memory (Alley & Greene, 2008; Weiss, Trehub, Schellenberg, & Habashi, 2016; Weiss, Vanzella, Schellenberg, & Trehub, 2015) and valence (Eerola & Vuoskoski, 2013; Kreutz, Ott, Teichmann, Osawa, & Vaitl, 2008; Weiss et al., 2016). As the here described project was interested in low-level primary perception and the encoding of certain features, this choice was furthermore motivated to prevent certain higher cognitive encoding like semantic and/or syntactic processing due to lyrics (J. Zhang et al., 2018), as well as to compensate for the characteristic of certain genres being completely non-vocal (e.g. most classical and electronic music), while genres usually having vocals (e.g. rock and afro-american) could also be reliably identified in non-vocal versions (Gjerdingen & Perrott, 2008). The duration of six seconds for all examples was driven by two points: the planned analysis and neurophysiology. Focusing the first, a high number of stimuli repetitions and experimental runs are crucial for any machine learning (pattern recognition) approach, as they allow for a high signal-to-noise ratio and a sufficient cross-validation scheme, thus enabling certain learning rates and inferential statistics. While recent studies show that more, high quality data from only a few participants is preferred to the contrary and more classical data situation (Kolossa & Kopp, 2018), it is still important to prevent any (experiment related) fatigue naturally occurring during long experimental settings (Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1990; Haller & Bartsch, 2009). Therefore, the stimulus duration was additionally influenced by the importance of a trade-off between number of stimuli repetitions/experimental runs and total length of the experiment to counterbalance the amount of data and it's possible influence by attention and fatigue. Targeting the second point, recent studies have shown that the assumptions for most statistical tests (e.g., random field theory), that are applied in neuroimaging, are to some extent highly determined by the used stimulus, task and paradigm (Birn, Saad, & Bandettini, 2001; G. Chen, Saad, Adleman, Leibenluft, & Cox, 2015). In more detail, especially the so-called BOLD linearity seems to be actively influenced and furthermore also varies as a function of brain region and therein grounded local changes of neurovascular coupling (Devonshire et al., 2012; Soltysik, Peck, White, Crosson, & Briggs, 2004). This line of research suggests that auditory stimuli of intermediate length (e.g., four to eight seconds) are preferable in terms of BOLD linearity and adaption. After extraction all examples were normalized to have equal root mean square values and saved in WAV, having a sampling rate of 44.1 kHz. Both, extraction and normalization were conducted through the open source software audacity (https://www.audacityteam.org). A professional musician and musiciologist assessed all examples regarding their audio and music theory features deciding if a certain example fitted the respective genre or not. Non fitting examples were exchanged till all examples (10 per genre, 200 in total) were rated as appropriate.

5.4.2. VALIDATION OF STIMULUS SET

Stimulus validation online questionnaire

The final step of stimulus validation was carried out through an online questionnaire in which naive participants rated the theory and study based, pre-validated examples with respect to their representativeness for a given genre. This was done to obtain a representativeness rank order from all examples within a genre and necessary, as the rank order allowed to further narrowing down the stimulus set as only the most representative example per genre was used included in the final stimulus set and therefore experiments. Besides allowing for more repetitions per genre, while keeping the total experiment length as short as possible, this also enabled a more straightforward integration of and comparison between the different modalities (neuroimaging, behavior, etc., 5.1). In order to share the survey in an open and reproducible manner and not violating the stimulus' copyright restrictions, the online survey was implemented via SurveyJS (https://surveyjs.io), an open source and freely available JavaScript library and hosted within the private University of Marburg webspace. The questionnaire itself consisted of three different parts: demographic information, music preference and usage, stimulus validation. The first addressed information important to the validation sample's demographic characteristics to ensure, first, the non-existence of any factors possibly influencing their music perception and consumption/listening behavior (including subjective hearing ability, hearing disorders, as well as neurological and/or psychiatric diseases) and second, demographic similarities to the (to be in the main experiment included) participants (age, education, etc.). Focusing their music consumption and listening behavior, as well as their music preference, the second part of the questionnaire contained digitized versions of the Music Preference Questionnaire (MPQ-R) (Nater, Krebs, & Ehlert, 2005) and Short Test of Music Preferences (STOMP) (Rentfrow & Gosling, 2003). Both assess the preference for a range of music genres on a 5 (MPQ-R) or 7 (STOMP) point Likert scale, respectively. The MPQ-R additionally asks for typical usage purpose (relaxation, etc.) and consumption type of music (concerts, etc.), as well as musical training, overall importance of music and music related chills. The third part of the questionnaire consisted of the stimuli validation implemented as a free sorting task, asking the participants to sort examples within a given subgenre with regard to their representativity for this certain genre, the representativity of the subgenres for their respective main genre and how familiar they were with a given sub and main genre. The order of main and subgenre was always identical and followed the subsequent pattern: all subgenres of one main genre were processed after one another and before each new subgenre, all ten examples for this given subgenre were presented in order to give participants a brief glimpse and scope of the to be arranged stimuli. After that, the free sorting task started and demanded that stimuli should be sorted according to their representativeness, where the most representative example should be at the topmost position and the least representative at the bottommost. An example is depicted in Figure 5.4.

The sorting itself could be achieved in a drag and drop manner and participants could additionally listen to each example, both as often as participants wanted and without a editing and/or time limit.

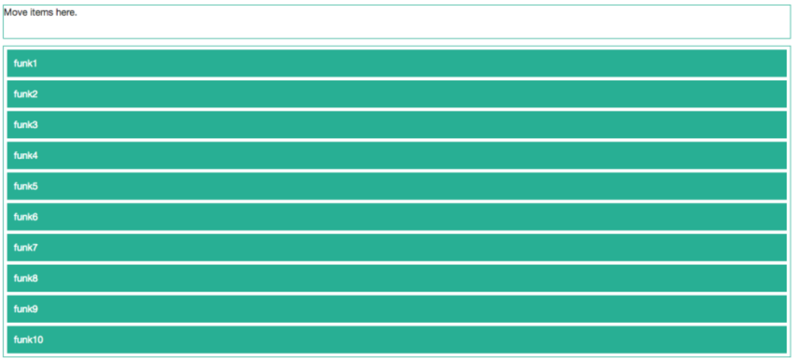


Figure 5.4: *Free sorting task example for the subgenre funk.*

Only, if all examples were sorted, participants could continue to the next section where they were asked how familiar they were with the previously processed genre. After the sortings of all subgenres of a given main genre were conducted, subgenres had to be sorted with regard to their representativeness for the main genre they belonged to. A respective example is depicted in Figure 5.5.

Figure 5.5: *Free sorting task example for the main genre african-american*

Validation survey sample

33 participants completed the online questionnaire (age M: 27.87, SD: 11.28, 20 females), with none of them reporting any neurological and/or psychiatric disease. Four participants reported slightly impaired hearing abilities while the remaining individuals reported that their hearing ability was not impaired. The sample was nearly equally divided into participants that were currently actively making music (n=15) and those who did not (n=18). None of them indicated any problems with regard to the online questionnaire, neither in terms of usability nor technical issues. The participants did not receive any financial or other compensation and could terminate the questionnaire at any given point without consequences or drawbacks.

Procedure

Prospective participants were contacted via e-mail and asked if they were interested and willing to take part in the online questionnaire. Those who replied with a consensual answer received an email response with a private link leading them to the online questionnaire in order to keep the access to the included material restricted, fulfilling the already mentioned copyright related constraints. Each link was generated specifically for one person and manufactured to become inactive after a certain amount of time and/or if the participant had used it once. After conducting the first two parts of the questionnaire (demographic information and music preference/usage) in which participants responded via checkboxes and text fields, they were presented with the stimuli validation, which was conducted as described above. Depending on participant's speed the questionnaire took roughly 60 - 90 minutes. Participants' responses were stored in a SQL-database hosted in the same webspace as the questionnaire and exported as JSON files.

Analysis

The resulting JSON files were analyzed using custom written python code, mainly utilizing the python packages json (https://docs.python.org/3/library/json.html), NumPy and Pandas. Initially, necessary information was extracted from the JSON files and written to participant specific pandas data frames in CSV format. As a next step the data from all participants was concatenated and stacked to create one data frame, which was used for the subsequent analyses. Within each subgenre each example was assigned a weight, respectively points, based on their assorted position in the questionnaire in reverse order. Therefore, examples received 10 points if they were set as the most representative example (at position 1) and 1 point if they were set as the least representative example (at position 10). The thereby assigned points for each example were summed across participants using NumPy's sum function and written to a pandas dataframe in CSV format. Based on this approach and the number of participants (n=33) each subgenre had a range from 330 (most representative) to 33 (least representative) points. Within each subgenre the example that received the highest points was further used for the main experiments. Subgenre-specific results were visualized by means of bar and violin plots using seaborn.

Results

Although a high variability between subgenres could be observed, a winning example could be obtained from each subgenre. This was especially enabled through the violin plots of the ratings, as they provided insights not only to the average rating and its SD, but also the distribution of the ratings. As the detailed respective outcomes, although highly interesting, are outside the scope of this thesis, only the winning examples that constituted the used stimuli set, including 20 music excerpts of 5 subgenres per 4 main genres, are indicated in Table 2, naming the composer and song.

Genre	Track	Composer
Alternative	Don’t listen to the radio	The Vines
Punk	I don’t give a fuck	The Partisans
Heavy Metal	Vengeance is mine	Iced Earth
Rock’n’Roll	Roll over Beethoven	Chuck Berry
Psychedelic	Heartbreaker	Led Zeppelin
Baroque	Sonata a 2	Dietrich Buxtehude
Viennese Classic	4. Symphonie	Carl Philipp Emanuel Bach
Modern Classic	Suite for String Quartett	Alban Berg
Renaissance	Canzoni e sonate	Giovanni Gabrieli
Romantik	6. Symphonie	Antonin Dvorak
Deephouse	Round Two	New Day (feat. Andy Cain)
Drum’n’Bass	Grand Funk Hustle	Die & Break
Dubstep	Invincible Firebeats	Dub Vision ft. Ruby Prophet
Techno	Tranquilizer	Oilver Huntemann
Trance	Amsterdam	Luminary (Smith Pledger Mix)
Funk	Hit & Run	The Bar-Kays
Hip Hop	Criminal	War Game
Reggae	Chase the devial	Max Romeo
R’n’B	Diary	Alica Keys
Soul	Nobody Taught Me	Eartha Kit

Table 5.1: Composer and title of the music excerpts included in the experiment, sorted by genre.

5.5. REPRESENTATIONAL MODELS

Hereinafter, the applied methodological approach is introduced and its utilization within the here described project of the thesis motivated. After a general outline of the theoretical framework is given, details on the analyses method and its parts, as well as steps are provided.

As already stated, the conception of a given cognitive process is, even though the tremendous advancement of data acquisition and analysis methods, rather limited. This holds especially true for the investigation of highly complex stimuli using methods displaying a restricted sensitivity with regard to fine-grained differences. Given the focus of this chapter and complete thesis, music genre perception and categorization and how it could possibly be explained is highly affected by the aforementioned circumstances. In more detail, the nonlinear combination of acoustic features (Whitehead & Armony, 2018) responsible for the definition and processing of music genres (Casey, 2017; Casey et al., 2012) might not be encoded prominent enough to be sufficiently traceable by certain approaches. For example, based on the same set of features, that are however nonlinearly combined, the evoked responses will share an extensive amount of variance and thus become troublesome to distinguish. Therefore, the investigation of how the human brain processes music genres in terms of similarities and differences demands settings and analysis approaches suited to capture responses related to these features and their high dimensionality. The questions of where music is represented in the human brain and what information these regions contain, hence have to be extended via also addressing how this information is represented (Diedrichsen & Kriegeskorte, 2017). According to that, the framework of representational models poses a valuable approach as it can address those questions, as they aim to "...fully characterize the representational geometry, defining all represented features in a region, how strongly each of them is represented (signal to noise ratio), and how the activity patterns associated with different features relate to each other. Representational models therefore fully specify the representational content of an area." (Diedrichsen & Kriegeskorte, 2017, p.3). In it's core it focuses on the so called second moment of activity profiles as reflected by population codes. As the framework itself is targeting computational aspects of brain activity, population codes refer to the activity of populations of neurons which can be assessed via a broad range of techniques like EEG and fMRI, reflecting the summarized population or hemodynamic activity and therefore thousands of neurons respectively (Diedrichsen & Kriegeskorte, 2017). Notably, "...a "representation" constitutes a functional interpretation, which requires not only that the represented variable (such as a perceptual property, some cognitive content, or an action parameter) is encoded in the pattern of activity in a format that can be read out by downstream neurons, but also that the information is actually used by other brain regions and, thus, serves a functional purpose." (Diedrichsen & Kriegeskorte, 2017, p.1). In order to entail "...an explicit representation, another area must be able to read out the represented variable directly using a neurobiologically plausible readout mechanism, such as linear or radial-basis-function decoding..." (Diedrichsen & Kriegeskorte, 2017, p.1). Here, it is important to describe and structure a given population activity, which is indicated by n channels (e.g., electrodes within EEG, voxels within fMRI, etc.) and n conditions (e.g., different music genres) by means of two factors: the activity pattern and the activity profile. While the first describes the response of all channels across one condition, the latter characterizes the response of one channel across all conditions. In turn, these factors can be used to generate spaces that are either spanned by the measurement channels or the conditions and subsequently projecting either activity patterns or activity profiles as points in these spaces (Diedrichsen & Kriegeskorte, 2017). Focusing the latter, representational models aim to define the respective probability distributions, further the second moment of activity profiles distributions and therefore the decodability of any feature from the population activity. Hence, the second moment matrix constitutes a useful "summary statistic for characterizing representations." (Diedrichsen & Kriegeskorte, 2017, p.25). However, the exact type of information depicted in the second moment matrix varies between existing methods of the representational models framework which includes encoding analyses, pattern component modelling (PCM) and representational similarity analysis (RSA). Given the condition-rich nature of the stimulus set, the exploratory character of this project and that no assumptions of linearity between the to be obtained data and constructed models could have been made beforehand, only RSA was applied and will be described in further detail hereinafter. For a very informative, well-written and more precise introduction regarding representational models and the different included methods, the reader is advised to consult Diedrichsen & Kriegeskorte (2017).

Motivation for the application of representational models

Applying a certain abstraction, representational models allow the comparison of representations, not the activity profiles themself, across a vast amount of levels, as the second moment matrix is indexed by the number of conditions and can be computed using a multitude of measurement channels. For example, across participant comparisons bypass the need for between-participant registration and therefore data transformations via focusing on abstract representations rather than activity patterns. Furthermore, representations can be compared within different channels of the same (e.g., multiple ROIs in fMRI and multiple time points in EEG) or different modality (e.g., certain ROIs in fMRI with certain time points in EEG), thus also enabling data integration. This also extends to models that can be constructed in various ways and aim to predict the observed data. This encloses complex and simple computational models (e.g., representations generated by deep and convolutional neural networks), conceptual models (representations formulated based on hypotheses), behavior (representations based on behavioral responses regarding the conditions, like e.g., reaction times and arrangements) and feature models (representation of certain stimuli features across conditions). It therefore poses as a highly suited approach for the project at hand for two reasons: first, the representations of different music genres can be formulated via the second moment matrix and second, by keeping the conditions (music genres) identical throughout varying measurement channels (fMRI, EEG, behavior, acoustic features, conceptual models) a multimodal investigation, additionally allowing the assessment of the predictive ability of a variety of models, was enabled. To this end, the three studies of this thesis part shared not only the same stimuli set (please see 5.4), but also paradigms, models and exact method of the representational model framework. In more detail, studies that involved imaging, that is subproject 1 and subproject 2, used the same paradigm (please see 5.5), while the behavioral paradigm was identical across all subprojects (further information provided in 5.7). This also accounts for the set of tested models which are depicted in 5.8. Further information with regard to the respective parts are given in the subsequent chapters.

Representational similarity analysis

RSA is special among the representational model methods, as it's the only one that spans the space via measurement channels and projects activity patterns as points into this space, with the opposite being true for encoding models and PCM. It then computes the distance between these points to obtain a matrix containing pairwise distances between all activity patterns. This so called representational similarity matrix (RDM) serves as the summary statistic and therefore entails information on how dissimilar activity patterns are represented within a given (set of) measurement channel(s). Framed differently, it characterizes the information that's presented in a measurement channel by defining how it distinguishes "different mental states (i.e., stimulus percepts)" (Kriegeskorte et al., 2008, p.2). Notably, in the light of the introduced representational models framework distances between activity patterns in a space spanned by measurement channels pose as a converse as the second moment matrix is described through distances between activity patterns (Diedrichsen & Kriegeskorte, 2017). Yet it can be considered a useful and informative complementary approach, given that certain requirements are fulfilled. One important necessity is the application of distance measures suited to entail the same information as the second moment matrix. Among those are correlation and Euclidean distance, with the later being focused on hereinafter, as it can serve as "a function the second moment of the activity profiles." and is "independent of the resting baseline and generally easier to interpret." (Diedrichsen & Kriegeskorte, 2017, p.15). As the representational models framework was formalized on imaging data, more precisely fMRI (and in parts M/EEG), two methodological points are of high importance in order to remove potential sources of bias: multivariate noise normalization and CV of distances. Concerning the first, due to the limitations of current data acquisition methods, the to be estimated activity patterns are always corrupted by noise. As they pose as random vectors when projected into a high dimensional space, this noise corruption leads to an orthogonal shift of their respective coordinates and in turn to positively biased distance estimates (Walther et al., 2016). It is therefore advised to apply noise normalization before computing distances, ideally in a multivariate manner to not only downweight measurement channels with a high error variance, but also include the noise covariance matrix between measurement channels to render their noise components "approximately independent" (Diedrichsen & Kriegeskorte, 2017, p.15) to one another (Walther et al., 2016). This is especially motivated by the spatial dependency of neighboring measurement channels (Diedrichsen &

Kriegeskorte, 2017; Friston, Jezzard, & Turner, 1994). It is important to note, that computing the squared Euclidean distance on multivariate noise normalized activity patterns will result in a second moment matrix that is defined by the squared Mahalanobis distance (Diedrichsen & Kriegeskorte, 2017; Walther et al., 2016). Even when noise normalized, the activity patterns still remain noisy to a certain extent, causing yet again a biased distance estimate. To address this potential problem, CV as an approach common and important to multivariate and machine learning analysis can be adapted for distance measures. Comparing the activity pattern from splitted and independent data partitions will yield an unbiased distance estimate which is only zero if the activity patterns between partitions differ only by noise (Diedrichsen & Kriegeskorte, 2017; Walther et al., 2016). In more detail, the activity pattern is estimated within certain parts of the portioned data and the distance computed to an activity pattern estimated within another part. For example, if the data was portioned into 8 parts, the activity pattern will be estimated within a training set, that is through all but one part (7), as well as a testing set, that is the remaining part (1) and the distance computed between these training and test sets. This is repeated till all parts acted as the test set once. The final RDM is then obtained by averaging the CV fold corresponding RDM (Diedrichsen & Kriegeskorte, 2017; Walther et al., 2016). Notably, both operations are crucial to obtain a high reliability of RDMs and its underlying distance measures (Diedrichsen & Kriegeskorte, 2017; Walther et al., 2016). Combining the multivariate noise normalization and CV distance yields the cross-validated squared Mahalanobis distance, which was used to obtain RDMs based on neural data within the parts of the project that utilized imaging methods (subproject 1 and 2). However, as the precise steps varied between subprojects further details on RDM computation will be given in the subproject corresponding chapters. In total, applied to the project at hand, RSA can provide information on how different music genre are perceived, which is operationalized through "..the dissimilarities between the activity patterns elicited across channels by the experimental conditions" (Diedrichsen & Kriegeskorte, 2017, p.15).

Comparison of neuronal and model data

As mentioned above, representational models and therefore also RSA enables a comparison between RDMs of a vast amount of sources, like different imaging modalities, computational models and behavior. As one aims to understand and explain these target RDMs, the second moment matrix as a summary statistic can be used to test how well the observed representation can be predicted by a certain model or multiple of it. Like the target RDM, these model RDMs can originate from a multitude of source, like behavioral responses, stimulus properties and hypotheses with all of them being present in the here described study. The comparison between RDMs therefore constitutes a comparison between observed and predicted dissimilarities via a second-order isomorphism as it attempts "...to establish a correspondence between the relations among the stimuli on the one hand and the relations among their representations on the other..." (Kriegeskorte et al., 2008, p.4). This comparison is furthermore characterized and defined by four circumstances: symmetry of RDMs, scaling/linearity, statistical inference and noise ceiling. Concerning the first, as RDMs are symmetrical about the diagonal, using the complete RDM would introduce a bias, which is addressed by only using the lower or upper diagonal of given RDMs when comparing them (Kriegeskorte et al., 2008; Swaroop Guntupalli, 2013). Focusing the second, the scaling of the second moment matrix across RDMs and their assumed linearity in relation to models predetermine a certain set of possible comparison measurements. As the second moment matrix underlying distances are highly affected by a myriad of factors like "physiological responsiveness, physiological noise, and head movements—in short, by all the factors contributing to signal strength or the noise distribution" (Diedrichsen & Kriegeskorte, 2017, p.16) the observed distances will vary tremendously between participants or even different measurement channels of the same type within participants (e.g., brain regions in fMRI or time points in EEG). Hence, a certain type of scaling between distances is needed. Furthermore, a linearity between target and model RDMs cannot be assumed except it is verifiable that a given "model captures not only the neuronal representational geometry but also its possibly nonlinear reflection in our response channels (e.g. fMRI patterns)" (Nili et al., 2014, p.4). Addressing both points, the application of rank correlations in the comparison of target and model RDMs offers a suitable approach as it accounts for the unknown scaling and enables the testing of rank ordering of dissimilarities which poses as a cautious alternative in the absence of linearity (Diedrichsen & Kriegeskorte, 2017; Kriegeskorte et al., 2008; Nili et al., 2014). Among those are Spearman and Kendall's τ_a. The latter poses as the method of choice when models that predict tied ranks, more precisely identical ranks for more than one condition pair, are tested as it more likely prefers the true over a simplified model (Diedrichsen & Kriegeskorte, 2017; Kriegeskorte et al., 2008; Nili et al., 2014). Notably, when comparing RDMs, the target RDM is referring to multiple RDMs of the same type (e.g., a given brain region in fMRI or time point in EEG) across participants, as one is usually interested in the predictive ability of a given model RDM on a population level. As some sort of statistical inference, the third point, is needed to obtain respective evidence, the aforementioned points are therefore conducted for each included participant, that is the upper triangle of the target RDM of each participant is correlated with the upper triangle of each model RDM. This yields an array of participant specific correlation values for each model RDM (for example, if 15 participants were included, there would be 15 correlation values per tested model RDM). In turn, these correlation values are then submitted to a one-sided signed-rank test (Wilcoxon, 1945) per model RDM, treating participants as a random effect. The predictive ability of a given model RDM as then expressed as the mean correlation across participants. Following up on this, it is important to access the "goodness" of a model, indicated here by a high correlation and why it potentially fails to achieve high correlation values and is captured by the fourth point, the noise ceiling. In RSA, "The noise ceiling is the expected RDM correlation achieved by the (unknown) true model, given the noise in the data." (Nili et al., 2014, p.7). As the ground truth of representations obtained by different measurement channels (e.g., fMRI & EEG) is uncertain given the noise in the underlying data, a suited estimate to the ground truth is necessary. Imagining a space spanned by the participant's target RDMs, the RDM predicted by the ground truth that maximizes the average correlation across the participants should be located at the center of this space, which exact position is dependent on the applied comparison approach. Following Nili et al. (2014), when applying Spearman or Kendall's τ_a, the participant specific target RDMs are rank transformed and then averaged to obtain an average target RDM that can serve as an estimate of the ground truth RDM. As this estimate is based on participant specific target RDMs it constitutes an inherent bias towards them. It is therefore important to compute both an upper and lower bound of the noise ceiling. While the upper noise ceiling is obtained by the addressed bias, as the average correlation between participant specific target RDMs and the average target RDM is overestimated, the lower noise ceiling is obtained by using a leave-one-participant-out CV in which all but one participant are used to compute the average target RDM and correlate this RDM with the left out participant specific target RDM. The average of these correlations is underestimated due to the left out data and hence serves as the lower bound of the noise ceiling. Taken all of the aforementioned points together, the described comparison of target and model RDMs offers a variety of insights aiming at representations based on neuronal data and how they can potentially be explained through model representations. First, the amount of variance in a given representation explained by each model RDM can be statistically assessed and its significance tested. Second, the "goodness" of each model RDM is indicated by its distance to the noise ceiling which acts as the estimated correlation with the estimated ground truth target RDM. Third, the range of the noise ceiling provides insights into the stability of representation across the included group of participants (a broad noise ceiling hints to large variations and vice versa for a narrow noise ceiling). Fourth, the distance between the upper bound of the noise ceiling and a perfect correlation of 1 on the one hand and the distance between the lower bound of the noise ceiling on the other hand, provide insights into the overall capability of the data at hand to enable the investigation of the representations of interest. In more detail, if the data fails to generate a noise ceiling that is close to 1, the applied acquisition technique (e.g., measurement channels, paradigm, etc.) should be revised and enhanced. If even the best model RDM, that is the model RDM with the highest average correlation across participants, does not reach the noise ceiling, the tested model RDMs should be improved. The above outlined procedure of RDM comparison was implemented within URIAL as rdm_compare and applied throughout all parts of the project. In short, a set of model RDMs was created that included RDMs based on behavioral responses (please see 5.7), simple computational models based on stimuli properties (acoustic features, as described in 5.8) and conceptual models based on hypotheses about the representations (more information provided in 5.8). Notably, all parts utilized Kendall's τ_a to compare the same model RDMs to the RDMs obtained through fMRI and EEG (subproject 1 & 2) and Spearman's r when comparing the latter with behavior based RDMs. More details on the comparisons are provided within the respective chapter of the subprojects.

5.6. IMAGING PARADIGMS

As mentioned in the preceding chapter, sub-projects that involved neuroimaging, hence I (fMRI) and II (EEG), used the same experimental paradigm. In more detail, this passive listening task consisted of eight runs during which each stimulus was presented two times. Pauses between stimuli were either 4, 6 or 8 seconds to account for habituation effects and maximize the BOLD response with regard to condition rich event related designs (Diedrichsen & Kriegeskorte, 2017; Walther et al., 2016). To pseudo-randomize the order of stimuli and pauses in order to prevent the aforementioned possible problems, as well as sequence related artifacts, Optseq2 (https://surfer.nmr.mgh.harvard.edu/optseq/) was used, applying a first order counterbalancing between stimuli. 100.000 possible run schemes were generated, the eight best were further used and their respective order in the experiment again pseudo-randomized across participants. Each run had am approximate duration of eight minutes, during which, besides the music genre stimuli, movie segments from the BBS documentary “Planet Earth”, cut to the length of a run, were presented without sound and centrically displayed on a screen (for detailed information on the experiment settings, please see 5.10.1 & 5.11.1 respectively). An example run structure is depicted in Figure 5.6. As participants weren't given a task, except to stay awake and listen to the music, these movie segments were included to help participants to achieve this goal and keep them attentive. To avoid any arousal related effects, the movie segments were controlled for graphical content and cut around potentially problematic scenes. The choice of a passive listening design with simultaneous movie segment presentation was motivated by previous studies that either implemented comparable designs (Cha et al., 2016; Schönwiesner et al., 2015; Whitehead & Armony, 2018) or provided evidence for a prominent influence of certain utilized tasks (e.g., n-back) on neural responses and therefore processing and representation of auditory inputs (Carcea, Insanally, & Froemke, 2017; J. B. Fritz et al., 2007; Häkkinen et al., 2015; Otazu et al., 2009; Petkov et al., 2004; Scheich et al., 2007). The task accompanying attention furthermore, intentionally or not, guides and shapes the processing and therefore percept towards certain features of the stimulus, leading to attention depending activation patterns that disturb the original, more feature based percept (Ahveninen et al., 2011; Ahveninen et al., 2006; Brechmann & Scheich, 2005; Da Costa et al., 2013; Degerman, Rinne, Salmi, Salonen, & Alho, 2006; Hillyard, Hink, Schwent, & Picton, 1973; Jäncke, Buchanan, Lutz, & Shah, 2001; Kauramäki, Jääskeläinen, & Sams, 2007; Paltoglou, Sumner, & Hall, 2009, 2011; Riecke et al., 2017). Additionally, studies provided evidence that no further regions in the auditory cortex are recruited in active compared to passive tasks (Alho et al., 2014).To avoid these probable factors, as well as to reconstruct the most frequent form of music consumption the described design was used. After each run participants were asked about their wellbeing and had the chance to take a pause of variable length, depending on their needs. During these pauses no sounds or movie segments were played. With the beginning of a new run a new movie segment started. All together this part took roughly 70 minutes to complete, independent of the sub-project. The paradigm was implemented and presented using PsychoPy (Peirce, 2007) on a desktop machine running Ubuntu 16.04 (https://www.ubuntu.com).

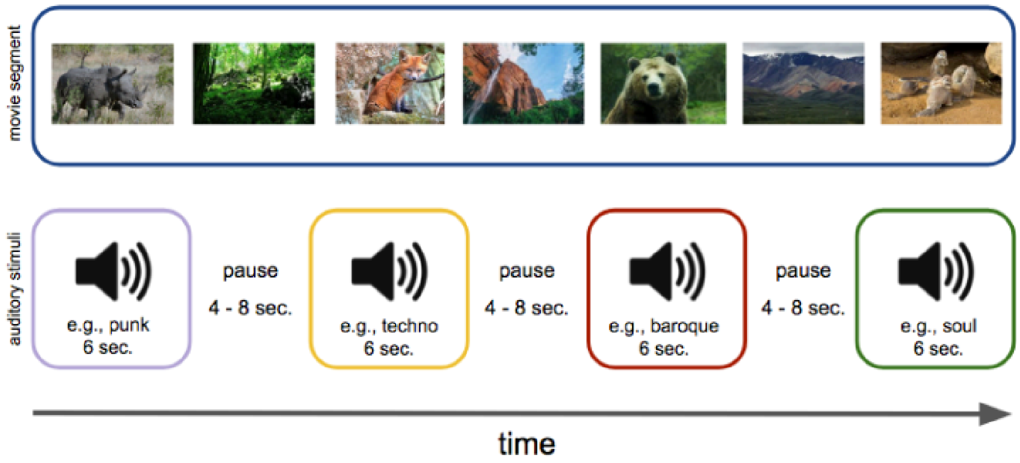
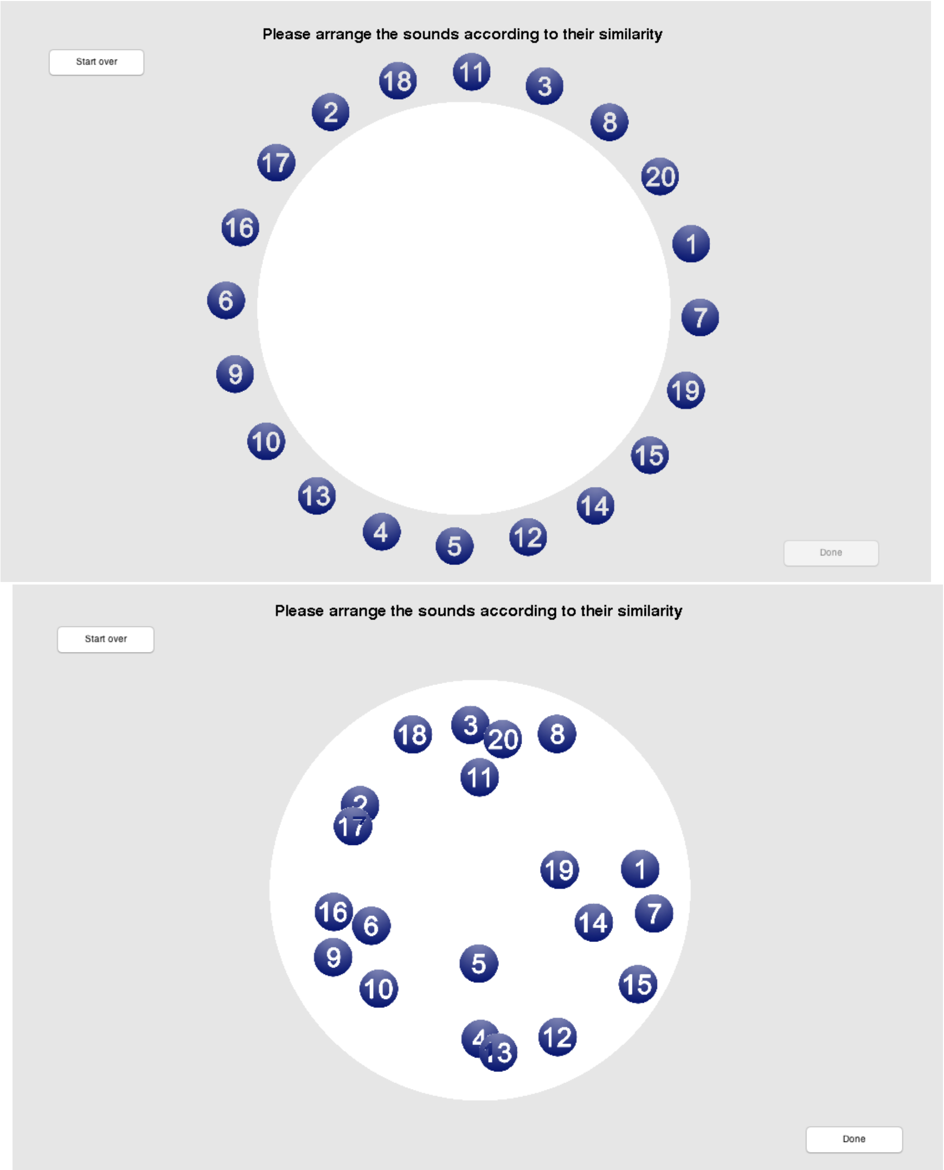


Figure 5.6: *Example of run structure with continuous movie segment (blue rectangle, top) and altering auditory stimuli (subgenres, bottom) over time.*

5.7. BEHAVIORAL PARADIGM

When exploring the representation of certain stimuli or processes, indicated through RDMs based neuronal data like fMRI and EEG, one might also be interested in the relationship between those representations and observed behavior related to the stimuli or processes. In more detail, one wants to excess to what extend, if any, certain cortical representations might, potentially, be reflected in and therefore driving behavior (Kriegeskorte, 2011). As noted before, the representational models framework offers an attractive possibility to address this question by using the second moment matrix as a summary statistic and therefore bridges a long ignored gap within neuroscientific research, the relationship between brain function and behavior (Krakauer, Ghazanfar, Gomez-Marin, MacIver, & Poeppel, 2017). As a matter of fact, the investigation of cortical or mental representations have a long history within the behavioral research domain with many of them relying on indicated dissimilarities between stimuli and processes as a meaningful conceptualization, therefore assuming a geometrical model (Coombs, 1954; Edelman, 1998; Kriegeskorte & Mur, 2012; Rudolf Carnap, 1928; Shepard, 1958; Torgerson, 1958, 1965). As this provides tremendous consensus with representational models, it constitutes an ideal candidate for the behavioral aspect of this study. Even though recent methodological advances (Decock & Douven, 2011) targeted identified problems of the geometrical model, e.g., context dependency, intransitivity and asymmetry (Goldstone, Medin, & Halberstadt, 1997; Goodman, 1972; Tversky, 1977), one crucial uncertainty remains: how should the dissimilarity judgements that form the second moment matrix corresponding summary statistic be obtained (Kriegeskorte & Mur, 2012)? Previous studies applied a multitude of measurements, including reaction times, pairwise judgements, free sorting and forms of arrangements. While they all show individual limitations like the suitability for small item sets, binary dissimilarities, drifts in interpretation and dissimilarities, they share a common problem, which is the constraint to a low dimensional, that is 2 dimensional space (Kriegeskorte & Mur, 2012). Even though still highly interesting and worth applying, these measurement types therefore don't capture higher dimensional dissimilarity structures. This becomes especially problematic for stimuli and processes that are very rich and complex in nature, like the here investigated representation of music genres. In more detail, if to be tasked with indicating the dissimilarity of music excerpts, participants would perceive two examples from different subgenres of the same main genre, e.g., alternative and punk rock as very distinct. However, if a third example of another main genre, e.g., techno would be added, the two initial examples would be perceived as very similar to one another, but at the same time being very dissimilar to the third example. While most approaches would capture the initial dissimilarity structure accordingly, all of them would miss the later one and therefore a crucial amount of information with regard to the representation. In order to address this limitation Goldstone (1997) among others extended the classical single arrangement through a multi-arrangement during which participants arrange multiple randomly chosen subsets of the entire stimuli set. Albeit the arrangements still being in 2 dimensions, a higher dimensional dissimilarity structure can be obtained via the redundant distance information (Kriegeskorte & Mur, 2012). However, creating trials by randomly pulling a subset of stimuli from the entire stimuli set leads to several problems with the two most prominent and severe being the potential missing of certain stimuli and the distribution of stimuli within the stimuli subsets. For example, applied to the stimuli set of this project (please see 5.4), certain subgenres or even entire main genres could potentially be left out, leading to missing dissimilarities for certain stimuli pairs or stimuli, as well as sets with the same main genres, but different subgenres that don't allow the assessment of distances between subgenres of one main genre. Kriegeskorte & Murr (2012) build upon the early work on multi-arrangements and introduced two approaches, implemented in a MATLAB (https://www.mathworks.com) based toolbox, aiming to resolve the aforementioned problems: an adaptive trial design and how representations based on subsets should be averaged. Both are described briefly hereinafter. The adaptive trial design is implemented in a two step procedure where in the first step, participants were encouraged to arrange the entire stimuli set which was displayed around a circle via drag and drop. As their method was introduced for visual stimuli the procedure had to be adapted for the auditory stimuli used within this project. To this end, icons consisting of blue circles with white numbers, ranging from 1 to 20 (one for each of the 20 music excerpts) were assigned and connected to the auditory stimuli in a pseudo-randomized manner to prevent any pattern that would exhibit a concurrence between numbers and music excerpts (e.g., numbers 1 to 5 for the stimuli from the main genre rock). Participants could then click on a given icon to hear the assigned auditory stimulus as often as they want, also after placing the icons in the white circle. An example for the initial arrangement is depicted in Figure 5.7. Notably, the instructions were limited to "Please arrange the sounds according to their similarity" as to not evoke a certain arrangement strategy (e.g., based on certain stimulus properties like timbre or rhythm). After the initial arrangement, the adaptive trial design proceeded with the second step, that is the creation of subsets which is dependent on the arrangement of the previous trial(s). Within that the arrangement of stimuli is prone to errors as they come from the same 2 dimensional Gaussian distribution and therefore stimuli that were placed close together have a low "dissimilarity signal-to-noise ratio" and vice versa for stimuli that were placed further apart (Kriegeskorte & Mur, 2012, p.4). The adaptive trial design keeps track of the evidence weights of each item pair and aims to maximize those weights which leads to the creation of subsets consisting of stimuli which dissimilarities have the weakest evidence weight and hence the underlying algorithm is termed “lift-the-weakest” (Kriegeskorte & Mur, 2012). In more detail the evidence utility, defined as u(w) = 1 - e-wd, uses the function u(w), with u being the usefulness of the current evidence and w the current evidence weight to indicate that stimuli pairs which dissimilarity evidence weight is weak should be preferred and therefore lifted, that is included, in the next trial, as compared to those which already provide a sufficient evidence weight (Kriegeskorte & Mur, 2012, p.11). Importantly, d defines the evidence utility exponent and therefore when “the utility saturates as a function of the evidence weight” (Kriegeskorte & Mur, 2012, p.11). The arrangements after the initial arrangements therefore include stimuli that were placed closed together, with the subsequent trial subsets further adapting based on the outlined approach. An example for an adapted subset of stimuli is provided in Figure 5.8. It is worth mentioning, that this approach leads to relative instead of absolute distances as the stimuli subsets are arranged in the same circle, with regard to size, as the entire stimuli set. As this adaptive trial design which is based on current estimate of a given RDM from the previous trial can lead to stimuli subsets varying drastically in size and composition (depending on the behavior of the participants) it is important to combine the thereby obtained RDMs to a final estimate in such a way that takes the respective distortions into account (Kriegeskorte & Mur, 2012). In other words, the partial RDMs, as they only include dissimilarities for a certain set of stimuli, need to be combined to obtain a RDM including dissimilarities for all stimuli pairs. Therefore, an approach, fulfilling these necessities, while additionally addressing the above noted arrangement errors is needed. To do so, (Kriegeskorte & Mur, 2012) introduced the concept of inverse multidimensional scaling (iMDS) in which arrangements in low dimensional space are used to infer dissimilarities in higher dimensional space, such as a RDM through inferring distances in a RDM via the on screen distances between stimuli. While multidimensional scaling (MDS) has been traditionally used to achieve the opposite, that is a low level representation of high dimensional data, iMDS offers the possibility of estimating the reverse. This however creates an inherent problem, as there is usually more than one optimal high dimensional arrangement representation of a low dimensional arrangement, in other words, a multitude of RDMs could represent a single low dimensional arrangement (Kriegeskorte & Mur, 2012). To this end, iMDS computes an initial estimate of the RDM which is based on the initial arrangement of the entire stimuli set and then uses the distances between subsets of stimuli obtained throughout the multi-arrangement to iteratively adapt the current RDM estimate and reduce arrangement errors. In more detail the current RDM is used to predict an MDS arrangement which is then compared to actual arrangement created by the participants. Afterwards the disparities between the predicted and actual MDS arrangement are computed and used to adjust the estimate of the current RDM accordingly (Kriegeskorte & Mur, 2012). This allows the mentioned combination of subset arrangements as the adjustment can be applied to certain distance estimates of the RDM. Combining iMDS with the adaptive trial design leads to a precise estimate of the participant’s perceived context dependent and continuous dissimilarity between all stimuli within a stimuli set and thus their representation, while at the same time capturing the high dimensionality of the representations. It therefore poses as an ideal candidate for the behavioral assessment of music genre representations as it takes their tremendous complexity into account and provides a solid estimate of their dissimilarities on both a large (between main genres) and small scale (between subgenres of a given main genre). Furthermore arguably capturing mental representations, the resulting behavioral representations pose as a very high and complex feature model comprising abstract information with regard to the categorization of music genres.



Please arrange the sounds according to their similarity

Please arrange the sounds according to their similarity

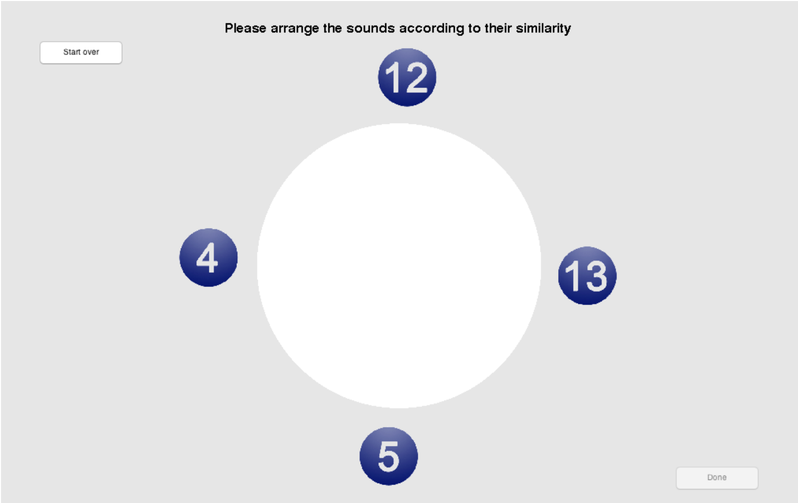
Figure 5.7: Initial trial of the behavioral paradigm before (top) and after arrangement (bottom).

5.8. MODEL RDMS

As outlined in 5.2 a diverse set of model RDMs was created and their predictive ability with regard to the observed representations tested. In more detail, this entailed simple computational (acoustic feature) models and conceptual (hypotheses) models. Following recent advances in the field of auditory categorization research, these types of models reflect low level (acoustic features) and high level (semantic/category) features, thus enabling an assessment of bottom-up and top-down processes (at least to a certain extent).

Acoustic feature models

The first focused, as opposed to deep and convolutional neural networks (D/CNNs), not on models aiming to simulate information processing in the human brain, but on the role of stimuli properties and therefore if the assumed low level processing of features could be explained or characterized by such stimuli driven models to a certain extent. Applied to the stimuli set of the here described study, acoustic features of the music genres were used to compose respective models. Each auditory stimulus, thus also music and its genres, can be defined by



Please arrange the sounds according to their similarity

Figure 5.8: *Example for a trial with an adapted subset of stimuli.*

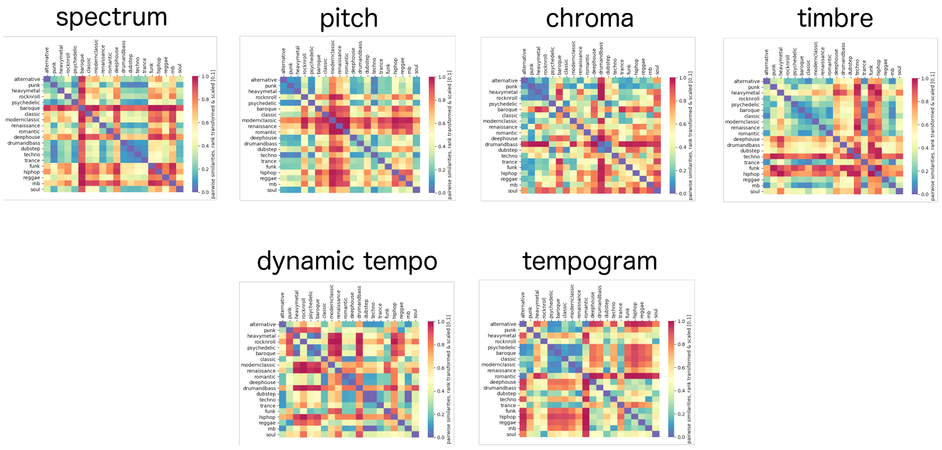
a combination of differently expressed acoustic features (Ajoodha et al., 2015; Esparza et al., 2015) and recent research pointed to a feature encoding during early stages of cortical auditory processing (Alluri et al., 2012; Alluri et al., 2013; Moerel et al., 2012; Santoro et al., 2014; Santoro et al., 2017) (from core to parabelt regions), therefore RDMs based on distances between stimuli with regard to a respective feature could provide potential insights into the representations obtained by neuronal or behavioral data. As mentioned in this project's introduction (5.1), a certain set of acoustic features has been identified as highly important when targeting a description and classification of music genres and their perceived differences on both a formal theory and an empirical level. Following the outcomes of these studies, already the selection of stimuli as outlined in 5.4 was based on these prominent features, which were therefore also used to compose acoustic feature models. While the space of acoustic feature models was comparable to previous research work on music genre perception (Casey, 2017; Casey et al., 2012; Güçlü et al., 2016; Hanke et al., 2015) in terms of models rooted in the spatial domain, that is models related to spectrum, pitch, chroma and timbre, the project at hand extended the model space into the temporal domain by also including dynamic tempo and tempogram models. This extension and thereby yielded model space that covered spatial and temporal features was of high importance in order to address the proposed representations and processes of the auditory cortex with respect to the spectrotemporal modulations of acoustic features which was outlined in chapter 4 of this thesis. A brief explanation of the acoustic features, their respective extraction and RDM computation is provided hereinafter. Notably, the python based open source library librosa (McFee et al., 2015) was used to extract the acoustic features, while URIAL's feat_dist function was used to compute acoustic feature corresponding RDMs.

Acoustic features and extraction

Four spatial and two temporal features were extracted from each music genre stimulus. The spatial features comprised spectrum, pitch, chroma and timbre. Spectrum describes the variation and manifestation of a signal's underlying components or its frequency-domain representation ("Signal Spectra", n.d.), which in terms of music provides insights into how strong different frequencies are expressed over the duration of the music excerpt. In order to extract the spectrum a Fast Fourier Transformation (FFT) with a window size of 2048 and a hop length of 512 was used. More precisely, a mel-frequency spectrum covering 48 dimensions was extracted via librosa's melspectogram function as the logarithmic character of the mel-scale resembles pitch relation judgements and thus organization found by respective psychophysiological studies involving humans (Logan, 2000). Pitch on the other hand reflects the property of frequencies or in turn musical notes being able to be ordered along such a scale as they can be perceived as higher or lower with respect to one another (Klapuri & Davy, 2007). In western tonal music, like the one investigated in this thesis, pitch is organized in 12 classes and ordered on an equal-tempered scale. In order to extract pitch information from a given music excerpt, cepstral coefficients were assessed by computing the logarithm of the obtained mel spectrogram and multiplying it with the discrete cosine transform (Hanke et al., 2015). From the obtained cepstral coefficients, only the higher coefficients were extracted as they most likely encode fine scale spatial feature structure such as pitch (Muller, Ewert, & Kreuzer, 2009). The relation of pitches to one another and over time is captured by the chroma of a music excerpt, that is their composition during progression. Hence, chroma related features describe the melody and harmony of a given music excerpt through the expressed strength of all pitch classes over time and under the rules that govern western tonal music. In contrast to the remaining features, chroma was not extracted by means of an FFT, but a constant q transform, as implemented in librosa as chroma_cqt, as the latter applies a logarithmically scaled axis and thus a more fine-grained resolution that better resembles the human perception of pitch classes (J. C. Brown, 1991). Addressing a higher and more complex feature, timbre is used to describe the tonal color or quality of a musical piece, as it can distinguish between different sources of a given auditory stimulus, that is e.g., the difference between the same tone, equally loud and long, but on different instruments. In its core it depends on the spectrum and envelope of a music excerpt over time, although other and different definitions were proposed (Erickson, 1975; Glorig, 1965; McAdams & Bregman, 1979). Therefore, the low coefficients from the mel spectrogram were extracted as they represent the shape of the coarse spectral envelope. Focusing temporal features, the dynamic tempo and tempogram of each stimulus were computed. The dynamic tempo is involved in the processing of rhythm and characterizes the speed of a music excerpt through beats per minute (bpm), that is how often events (e.g., notes) occur within a certain period of time. The extraction of the respective features was achieved via computing the bpm based on the spectral flux onset strength envelope over the time course of the excerpt (onset_strength and tempo in librosa). In contrast, the short-time autocorrelation of the spectral novelty function was used to compute an autocorrelation tempogram which “encodes local tempo information” (Grosche, Muller, & Kurth, 2010, p.5522). This was achieved using librosa’s tempogram. All of the obtained features, independent of type, were then represented as vectors that constitute the expression of the respective feature over time. The specific dimensions varied across features but were identical for all stimuli with a certain feature.

Acoustic feature RDM computation

Within each feature, the dissimilarity of the resulting music excerpt corresponding feature vectors and thus music genre where assessed by means of the Euclidean distance between the vectors. The resulting distance vector was then transformed into a symmetrical matrix, setting the diagonal to zero. This yielded one RDM for each extracted feature (6 in total) containing pairwise distances for all 20 stimuli included in this project. The respective operations were conducted through feat_dist as implemented in URIAL. A graphical depiction of the acoustic feature models can be found in Figure 5.9.


spectrum pitch chroma timbre

dynamic tempo tempogram

Figure 5.9: *Utilized acoustic feature models. The model space includes four spectral (top row, spectrum, pitch, chroma, timbre) and two temporal features (bottom row, dynamic tempo, tempogram). The RDMs are indexed by the amount and category of music genre ordering subgenres within main genres and main genres along the following order: rock, classic, electronic music and afro-american. The respective subgenres are identified through ticks on both x and y axis. Colors in the RDM indicate dissimilarities where blueish colors reflect very small or no distance and reddish colors extremely large distances. Values range from 0 (no distance) to 1 (extreme distance).*

Conceptual models

Besides obtaining model RDMs from different components of the data and/or stimuli, one might also have hypotheses about the representation in terms of its structure which are based on the results of prior studies or certain theories. While it tends to be difficult to implement and assess such models within most approaches, the framework of representational models, here RSA, allows the generation, integration and testing of such model RDMs which are termed conceptual models. In more detail, the formulation of conceptual models allows the investigation of hypotheses with regard to the representation observed in different brain regions (fMRI), time points (EEG) or behavior. Applied to this project, conceptual models thus can be used to capture assumptions about the representation of music genres based on previous research work and their predictive ability assessed through the representations obtained by the mentioned modalities. To this end three conceptual model RDMs were created: a main genre model, a subgenre model and a random model. The first two are based on the used stimuli allocation described in 5.4 into different genres, divided into a broad and narrow structure. The main genre model assumes that a given representation is based on the four main genres, that is subgenres of a main genre are not distinguished and therefore identically represented, while they are highly distinguishable from all other subgenres of all other main genres and therefore distinctively represented. It therefore constitutes a broad structure which does not differentiate between the subgenres of a main genre, given the whole stimuli set. The respective RDM was created by setting all within-main-genre distances to 0 and all remaining distances to 1, providing the most extreme distinction between main genres. The subgenre model is defined as an alteration of the main genre model in that main genres are highly dissimilar to one another and the subgenres of a main genre being very similar but distinguishable from each other. The narrow structure of this model thus assumes a representation based on main genres like the main genre model, however also a subtle differentiation between subgenres of a main genre, in terms of a RDM small distances (0.25) between subgenres of a main genre and extremely large distances (1) between main genres. Both models rely on a hypothesized allocation of the stimuli and therefore music excerpts into different categories, that is genres, based on their genre specific combination of acoustic features as indicated in chapter 5.4 (which also points to evidence from behavioral and computational studies). Hence, they are termed categorical models. The possibility to include hypothetical models furthermore enables the testing of models which assumed representation can act as a sanity check. The random model serves this function as it assumes that the distances that constitute the representation are completely random and not exhibiting any pattern, therefore not including any type of categorization and structure leading to a representation that contains no meaningful information. In turn, it should yield low correlation values with any target RDM that comprises a representation indicating that the underlying data (fMRI brain region, EEG time point, behavior, etc.) entails information. It is therefore a control model and was created via URIAL's gen_model_rdm. All conceptual models are illustrated in Figure 5.10.

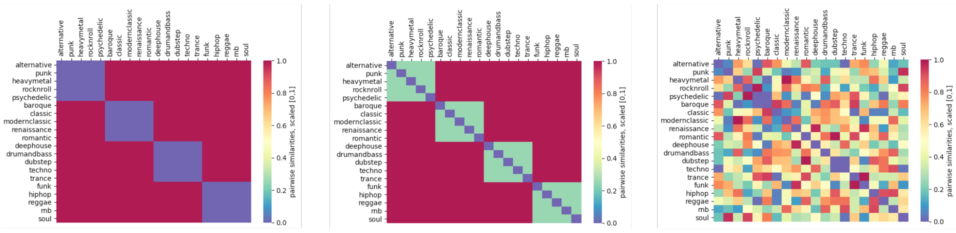


Figure 5.10: *Utilized conceptual models: main genre model (left), subgenre model (middle) and random/control model (right). The RDMs are indexed by the amount and category of music genre ordering subgenres within main genres and main genres along the following order: rock, classic, electronic music and afro-american. The respective subgenres are identified through ticks on both x and y axis. Colors in the RDM indicate dissimilarities where blueish colors reflect very small or no distance and reddish colors extremely large distances. Values range from 0 (no distance) to 1 (extreme distance).*

5.9. BEHAVIORAL TEST BATTERY

In order to describe the included participant population as precise as possible, a behavioral test battery including a multitude of assessments targeting different characteristics was created. The tests included a general demographic questionnaire, three handedness questionnaires, a questionnaire with regard to musical training, the Musical Ear Test (MET) (Wallentin, Nielsen, Friis-Olivarius, Vuust, & Vuust, 2010), the Music Preference Questionnaire (MPQR) (Nater et al., 2005), the Short Test for Music Preference (STOMP) (Rentfrow & Gosling, 2003) and a questionnaire that assessed participants' familiarity with the music excerpts utilized in this project. The entirety of tests was implemented in PsychoPy (Peirce, 2007). All participants, independent of the subproject, processed the tests after the respective main experiment and in case of subproject 1 and 2 additionally after the behavioral paradigm. The order of tests and settings were kept identical across subprojects. Notably, the majority of test results was not used in the here described project, but assessed as either control factors or to enable further investigations. Focusing the first, only the data from the general questionnaire was utilized in all subprojects to evaluate potentially problematic factors such as neurological and/or psychiatric, auditory or developmental disorders. The data from the MET served as a post-hoc confirmation of the self-declared musicianship and therefore group assignment in subproject 3. Regarding the second point, all remaining participant related variables were acquired in order to foster further investigations, especially by other research groups considered that all data will be made publicly available. All questionnaires are briefly described hereinafter.

General questionnaire

Aiming to provide a general, but precise characterization of the included participant groups, the general questionnaire requested information regarding sex, age, education, history of illness (participant and family) and development among others.

Handedness

Even though recent research suggests no influence of handedness on brain structure and function (Groen, Whitehouse, Badcock, & Bishop, 2013) it is still subject of ongoing intense discussions. Hence, and because handedness was comparably less investigated in the light of auditory, especially music processing, three handedness tests were included. Namely, Edinburgh Handedness Inventory (Oldfield, 1971), the Waterloo Handedness Questionnaire (Elias, Bryden, & Bulman-Fleming, 1998) and a custom questionnaire assessing handedness in close relatives. Notably, handedness did not serve as an exclusion criterion in the study at hand.

Musical training

A custom questionnaire was used to access information related to musical training, past or present. It covered years and amount, as well as different types of musical training (formal class, instrument, etc.).

MET

The MET is a short behavioral test that aims to measure musical competence, whilst additionally enabling a division in musicians and non-musicians based on the results of the test. In its core it poses as a classical matching task in that two subsequent musical phrases have to be identified as either identical or non-identical. Within two subtests it focuses on melody and rhythm as core aspects of music, which each take approximately 10 minutes to complete and include 52 trials during which the second musical phrase is either identical or varied with respect to the first one. Variations are introduced via a pitch alteration in the melody subtest and a rhythm variation in the rhythm subtest.

MPQR

Addressing music consumption behavior, the MPQR uses scales to assess the preference towards a broad of range contemporary music, along with the purpose of music listening, occasions of music listening, as well as phenomena like the frequency and intensity of music induced chills.

STOMP

In contrast to the MPQR, the STOMP solely focuses on the preference of certain music genres, assessed through a scale, which can be used to identify five factors of music preference.

Stimulus familiarity

The last test of the behavioral test battery was used to obtain familiarity ratings for each music excerpt included in the stimuli set. To this end all music excerpts were presented one after another while participants rated their subjectively perceived familiarity with a given stimulus on a Likert scale. This assessment was included to control for potentially confounding factors due to a high familiarity with the stimulus material as indicated by previous studies who provided evidence for an altered auditory processing of familiar music in contrast to unfamiliar music, which is most likely grounded in a more fine-grained representation based on frequent perception.

5.10. NOTHING COMPARES TO YOU I - MUSIC GENRE REPRESENTATIONS REVEALED BY FMRI

Hereinafter, the MRI part of this project, respectively the acquired data, conducted analyses and obtained results, will be described.

5.10.1. MATERIALS & METHODS

Beginning with the description of the characteristics of the included sample, this chapter will provide further details on the procedure applied in the MRI part of this project, data acquisition parameters and processing steps, concluding with the utilized representational model approach.

PARTICIPANTS

The sample of this project part consisted of sixteen individuals (three left-handed, six male) with a mean age of 24.3 years (SD=2.7). Participants were native German speakers and indicated neither neurological nor psychiatric disorders. While all of them were used to western tonal and had musical training for at least three years (M=10.4 years, SD=6.18), only seven were still actively making music by the time of their participation in the study. None of them reported normal hearing abilities and no enduring or temporary impairments (based on a self-rating, as no formal hearing test was conducted). All participants were instructed about the nature of the project and signed two consent forms: one for their willingly participation (APPENDIX B) and the “Open Brain Consent Form” (APPENDIX A) which enables the free and open sharing of their anonymized data in public repositories. Participants could choose between 20 € or a CD with their own structural data as compensation. This project part and its contents (which are described in 5.10.1.2) were approved by the ethics committee of the faculty of medicine, Philipps-University Marburg, Germany.

PROCEDURE

Before the experiment itself, participants received information about its content, structure and potential risks of (f)MRI investigations. After they gave their informed consent and were checked for circumstances that would not allow a participation in the (f)MRI part of this study (as indicated by a respective anamnesis form), participants were prepared for the imaging part and subsequently placed inside the scanner, additionally providing them with MR compatible headphones (MR Confon, HP VS03) in order to allow auditory stimulation during the experiment. The imaging part of the experiment was roughly divided in two parts, namely structural and functional scans in the following order: one structural T1 weighted image (~ 5 min.), 8 functional runs (partial coverage) (~ 8 min. each), one functional run (whole brain) (~ 1 min.), one structural T2 weighted image (~ 5 min.). During the first sequence, the structural T1 weighted image, participants had the opportunity and task to adjust the volume of the stimuli for the subsequent functional runs. To this end one example from each main category (please see 5.4) was presented and participants could either increase or decrease the volume using buttons on an MR compatible keyboard. Once they reached their preferred volume they could indicate and logg their volume selection, which then in turn was used for the functional runs. Notably, all participants used the highest possible volume setting. Afterwards, 8 functional runs, during which the experiment (as described in 5.6) was presented, were acquired, followed by a short additional functional run. The movie segments were presented on screen which was placed roughly 2 m behind the MR scanner in another room, with the participants being able to see them via mirror system mounted on the head coil. Lastly, an additional structural image, more precisely a T2 weighted image was acquired. Between each sequence, participants were asked about their wellbeing and could indicate if they needed a short break. All together, this part of the experiment lasted around 75 minutes. After the imaging part and a short break, participants conducted the behavioral paradigm (please see 5.7) on a MacBook Pro laptop running Matlab (R2014a, MathWorks Inc., Natick, MA, USA) and using Sennheiser HD 380 pro headphones which took roughly one hour. Lastly, they provided the information asked for in the behavioral test battery (5.8), culminating in an overall experiment duration of roughly 2 1/2 hours.

MRI ACQUISITION

All imaging data was acquired on a 3T Siemens Tim Trio MRI scanner (Siemens, Erlangen, Germany) at the clinic for psychiatry and psychotherapy of the university hospital Giessen-Marburg, Hesse, Germany. A 12-channel head matrix Rx-coil was used throughout all measurements and participants. A whole brain T1 weighted structural image of 176 slices was acquired using an MPRAGE sequence with the following specifications: TR = 1900 ms, TE = 2 ms, FoV = 256 x 256 mm, flip angle = 9°, voxel size = 1 x 1 x 1 mm. Subsequently, a T2* weighted echo-planar imaging (EPI) sequence with a TR of 1500 ms., TE of 25 ms, FoV of 190 x 190 mm, voxel size of 2.5 x 2.5 x 2.5 mm and a flip angle of 70° was used within 8 runs, containing 320 images each, to acquire functional images while participants underwent the experiment as described in 5.5. As the focus and hypotheses were centered around the auditory cortex, only a partial coverage, covering the temporal lobe in 25 slices, was captured allowing for higher spatial and temporal resolution as compared to a whole brain coverage. Using the participants structural T1 weighted image, this partial coverage was aligned along the sylvian fissure, using the temporal and occipital pole as inferior borders. In order to aid the later on coregistration between structural and functional images during the preprocessing of the data (5.10.1.4), a T2* weighted sequence containing 5 images (TR = 4 ms, TE = 25 ms, flip angle = 70°, voxel size = 2.5 x 2.5 x 2.5 mm) and covering the whole brain was acquired after the 8 experiment runs. The imaging part of this study concluded with a T2 weighted structural image of the following specifications: TR = 3200 ms, TE = 4 ms, flip angle = 70°, voxel size = 1 x 1 x 1 mm. The last sequence aimed to optimize the planned processing of the structural data.

PREPROCESSING AND STATISTICAL ANALYSIS

Data conversion and preparation

For each participant, all sequence corresponding DICOM files were converted to zipped 4D NIfTI files, while at the same time restructuring them. Focusing the latter the dataset was adapted to BIDS in order to facilitate later processing steps, as well as open & reproducible neuroscience through a clear, machine readable structure, including dataset and file corresponding informative meta data. This process was conducted using HeuDiConv (https://github.com/nipy/heudiconv) which in turn utilized dcm2niix (X. Li, Morgan, Ashburner, Smith, & Rorden, 2016) for the conversion of the files. To follow the BIDS specifications and as all data will be made publicly available (allowed through the Open Brain Consent form as described in 5.10.1.2), the anatomical images were furthermore de-identified using pydeface (https://github.com/poldracklab/pydeface) in order to assure participant's privacy. Before any subsequent processing, the dataset was validated using the online version of the BIDS validator (http://bids-standard.github.io/bids-validator/). The first step of data processing focused on quality control, which was assessed using the open source software MRIQC (Esteban et al., 2017) in its docker based BIDS app form (Gorgolewski et al., 2017). Here, the data quality is described by means of so called image specific IQMs, that for example include SNR, tSNR, EFC and FD, among many others. These IQMs are extracted via a Nipype workflow that utilizes FSL (S. M. Smith et al., 2004), AFNI (Cox, 1996) and ANTs (Avants et al., 2009), as well as custom functions, generating informative CSV files, as well as easy to grasp and interactive HTML pages. These outputs were used within a manual visual assessment of the data's quality, as well as the MRIQC automated classifier for structural images. The latter uses a pre-trained random forest classifier within a leave-one-subject-out CV to predict a given image's quality in a binary accept or reject form and therefore provides an automated, reliable and bias reduced approach. Upon inspection of the respective outputs, one participant had to be excluded from further analyses due to prominent motion artefacts, leaving a total of 15 participants.

Preprocessing

The preprocessing of the data was twofold, having a structural and functional processing pipeline which will be described in more detail hereinafter respectively. However, all processing steps were conducted on a desktop machine running Ubuntu 16.04 and using docker, bash and python (Python Software Foundation. Python Language Reference, version 3.6. Available at http://www.python.org). Focusing the structural pipeline first, the docker version of mindboggle (A. Klein et al., 2017) was applied to the respective images. Implemented in three parts, that is FreeSurfer (Dale et al., 1999; Dale & Sereno, 1993; Fischl & Dale, 2000; Fischl et al., 2001; Fischl et al., 2002; Fischl, Salat, van der Kouwe, et al., 2004; Fischl, Sereno, & Dale, 1999; Fischl, Sereno, Tootell, & Dale, 1999; Fischl, van der Kouwe, et al., 2004; Han et al., 2006; Jovicich et al., 2006; Reuter et al., 2010; Reuter et al., 2012; Ségonne et al., 2004), ANTs and mindboggle itself, it constitutes an extensive Nipype workflow that connects and integrates the mentioned parts. Initially, the T1 and T2 weighted structural images were submitted to FreeSurfer's recon-all, which skull-strips and subsequently segments the images into three tissue classes: GM, WM and CSF. Those segmented images are afterwards used to reconstruct the respective surfaces. Both, volumes and surfaces are then additionally parcellated within different schemes of varying complexity. As a second step the T1 weighted structural images were processed using ANTs' antsCorticalThickness function, leading to a brain mask, a six tissue segmentation (GM, WM, CSF, DGM, BS and CE) and an image containing the cortical thickness. Lastly, mindboggle itself combines the previously computed segmentations to a hybrid segmentation in order to account for certain over- and underestimation problems related to different tissue classes in FreeSurfer and ANTs (Klauschen et al., 2009). This hybrid segmentation is then parcellated and a vast amount of volume and surface based shape measures like thickness, travel & geodesic depth, surface area, laplace beltrami spectra, etc. are extracted and subsequently summarized using respective statistical characteristics such as mean, standard deviation, quartiles, etc. on a per label basis. Even though the entire variety of computed results is of potential interest, only a small portion was further used within this thesis, as the remaining parts were unfortunately, outside the scope. The usage of respective outputs is referenced within the depiction of the functional preprocessing and statistical analysis pipelines. Continuing with the first, the functional preprocessing pipeline broadly spanned three sections (that is artifact correction, coregistration and mask creation) and was implemented in Nipype and included functionalities from SPM (standalone version) (www.fil.ion.ucl.ac.uk/spm), FSL, ANTS, ARTDETECT (https://www.nitrc.org/projects/artifact_detect/) and FreeSurfer. The first step of this pipeline targeted motion correction and therefore used SPM's realignment function is depicted in the following. All 8 functional runs (partial coverage) were submitted to an overall two-step procedure: initially, all images from all runs were realigned to the first image of the first run in order to create a mean image and afterwards, the original images were realigned to this mean image. A second, independent node was used to apply the same procedure to the whole brain functional data. Respectively, this yielded one mean image (per node) and six realignment parameters (per node, per run). In order to evaluate the time series with regard to outliers based on motion parameters and global intensity, the outputs of the 8 functional runs node were submitted to Nipype's ArtifactDetect function. The threshold for motion parameter related outliers was set to

1 mm as defined by a euclidean combination of translations and rotations, while the threshold for global intensity related outliers was indicated through three standard deviations. Continuing with the coregistration between functional and structural images, a five step scheme was conducted. First, FSL's BET was used to create skull-stripped versions of the mean images from the partial coverage and whole brain functional scans. Second, using FSL's FLIRT, a coregistration matrix between these two images was computed, more precisely, from the partial coverage to the whole brain mean image. The same function was used in the third step to compute a coregistration between/from the mean whole brain image and/to the T1 weighted structural image, here the skullstripped image obtained through FreeSurfer within the structural preprocessing pipeline. Fourth, the same structural image as noted before was used to estimate a non-linear symmetric diffeomorphic mapping between each participant's native and the MNI template space. In more detail, the structural image and the ICBM 2009c Nonlinear Asymmetric version of the MNI template were used within ANTs' antsRegistration function. Fifth, the yielded transformation parameters from step 2 and 3 were concatenated used to transform the realigned functional images from all runs (partial coverage) to the functional whole brain and subsequently from the functional whole brain to the T1 weighted structural image space using FSL's FLIRT. Through SPM's Smooth function, these transformed images were then smoothed via a 2 mm FWHM 3D Gaussian kernel. The preprocessing of the data concluded with the creation of participant specific mask images that would allow to restrict the planned statistical analysis to voxels that are within regions of interest which are relevant to the study at hand, that is the auditory cortex. To this end, the mri_binarize function from FreeSurfer was used to create a structural mask of the temporal lobe via extracting voxels that were identified within the aparc+aseg.mgz volume (multi-tissue-class image yielded by FreeSurfer within the structural preprocessing pipeline) as belonging to one of the following labels: insula, transverse temporal, superior temporal, middle temporal or banks sts. Using FLS's ImageMaths this therefore created mask and the skullstripped functional whole brain mean image were subtracted from one another, leaving only gray matter voxels that were acquired during the functional sequences, restricted by anatomical boundaries of the temporal lobe. The thus created mask was dilated by one voxel and binarized through FreeSurfer's mri_binarize.

Statistical analysis

All statistical analyses were conducted in participants' native structural space using functionality from SPM implemented in a Nipype pipeline. Using Nipype's SpecifySPMModel and SPM's Level1Design a GLM was set up in order to estimate the voxel's spatiotemporal response induced via the presentation of the different music genres. Therefore, the 8 smoothed functional runs were included as separate sessions, along with the onsets and durations of each music genre (as generated from PsychoPy (please see 5.6)) as regressors of interest, as well as the run specific realignment parameters as regressors if no interest. This yielded a total of 26 regressors per run and thus 208 regressors per participant and GLM (48 regressors of no interest & 160 regressors of interest respectively). These participant specific GLMs were then estimated via SPM's EstimateModel function by convolving the defined regressors with the canonical HRF and applying a High-pass filter of 128 sec, modelling the serial autocorrelation with SPM's AR(1) option. The estimation was restricted to voxels that were within the temporal lobe mask, created during the preprocessing. This approach provides several advantages for the statistical analysis, as only gray matter voxels of regions of interest are estimated, therefore increasing sensitivity and specificity. This furthermore minimizes the influence of possible confounds introduced by voxels belonging for example to WM, CSF and ventricles, as well as limiting the computational cost and load. In case of traditional mass univariate analyses, it would furthermore drastically decrease the number of necessary multiple comparison corrections. As the aim of the here described study was to compare the representation of different music genres by means of a second moment matrix, or RDM, only the generated design matrix was used for further analyses as illustrated in the next chapter (5.10.1.5). Music genre specific baseline contrasts, accumulated across runs, were computed in order to inspect if the excerpts were able to evoke a neuronal response within the auditory cortex' regions of interest.

REPRESENTATIONAL SIMILARITY ANALYSIS

To assess the second moment matrix of the different response profiles associated with the different music genres and thus enabling a comparison between them, as well as between the respective RDMs of different ROIs and models, the framework of representational models, more precisely RSA, as described and motivated in 5.4, was used. The respective analysis was divided into 3 parts: additional data preprocessing, computation of RDMs and comparisons of RDMs across ROIs and models.

Transformation of ROIs

As functional runs were not spatially normalized, all analyses were conducted in participants' native space and auditory cortex ROI specific RDMs instead of the complete auditory cortex (or other regions) were of interest, the first necessary step of the RSA was the transformation of the respective ROIs from template to participants' native space. The auditory cortex constituting ROIs, as defined in 3.5, were identical to those used within the analyses of chapter 4: HG, PP, PT, STGa and STGp (LH and RH respectively) and are again depicted in Figure 5.11. They were extracted from the Harvard-Oxford atlas (Desikan et al., 2006; Frazier et al., 2005; Goldstein et al., 2007; Makris et al., 2006) as distributed with FSL, via Nipype's pickatlas and resampled to the resolution of the functional images (via nilearn's resample_to_img). Each ROI was transformed via ANTs' ApplyTransform function utilizing the inverse composite transform between the template and each participant's native space computed during the fourth coregistration step within the preprocessing pipeline. This resulted in ten participant specific ROIs which were used for the subsequent analysis steps. As in chapter 4 these steps were facilitated by using functionality from the ALPACA toolbox (please see 3.).

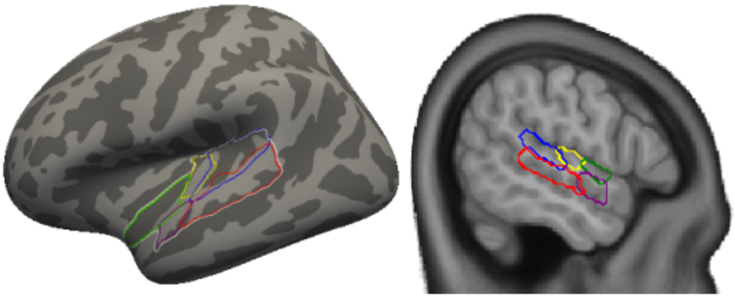


Figure 5.11: *Auditory cortex (white outlines) and its included ROIs: PAC (yellow outlines), PP (green outlines), PT (blue outlines), STGa (purple outlines) and STGp (red outlines) displayed in the LH, respectively on the inflated version of FreeSurfer's fsaverage surface (left) and on the ICBM152 template in MNI volume space (right). Note: Overlap of ROIs are not present in the data, but only in the graphical depiction due to sampling on the surface (left) and limited viewability of 3D data in 2D (right).*

Multivariate Noise Normalization

The in 5.5 addressed necessary multivariate noise normalization of the data was achieved through a three step procedure comprising time series extraction, GLM estimation and the noise normalization itself. The respective, hereinafter described steps were conducted per ROI and participant. Initially, the transformed ROIs were used to extract the time series from all voxels of the preprocessed images within a given ROI. Subsequently, time series were used to compute a new GLM specific for each ROI yielding beta estimates for each defined regressor as described in 5.10.1.4. The GLM outcomes, that is beta estimates, nuisance regressors and intercepts were then used to apply the multivariate noise normalization. In more detail, this was achieved through normalizing the data of all voxels through their standard deviation of its residuals, while also taking the noise covariance between voxels into account (Walther et al., 2016), implemented via the following equation where $\sum$ is the variance-covariance matrix obtained from the residuals of the GLM:

$$b_k^* = b_k \Sigma^{-\frac{1}{2}} \tag{5.1}$$

Voxel-wise multivariate noise normalization. Source: Walther et al. (2016, p. 190).

Using the information from the GLM, the aforementioned steps were applied in a run-wise manner instead of across all runs to consider run-specific effects and structures, as $\sum$ can differ tremendously between runs. This was furthermore important for the CV used to compute the RDMs or more precisely their constituting distances. $\sum$ was furthermore regularized using the optimal shrinkage factor as proposed by Ledoit & Wolf (2004) to address its rank-deficiency which is grounded in the imbalance of the $\sum$ underlying residuals timepoints x voxels matrix, having (in most cases) a substantially larger amount of voxels than timepoints (Walther et al., 2016). This resulted in pre-whitened beta estimates for all regressors of interest, i.e. the different music genres, per run, yielding a total of 160 noise normalized beta estimates (8 runs x 20 music genres). The multivariate noise normalization was achieved through the noiseNormalizeBeta function of the MATLAB based RSA toolbox (Nili et al., 2014).

Computation of representational dissimilarity matrices

Comparable to the multivariate noise normalization, the computation of the RDMs was conducted per ROI and as outlined in 5.5. As the beta estimates that were used to compute the squared Euclidean distance between music genres were noise normalized in a multivariate fashion, the squared Mahalanobis distance between the music genres was obtained:

$$\begin{aligned} d^2_{\text{Euclidean}}\left(\mathbf{b}_k^*, \mathbf{b}_j^*\right) &= \left(\mathbf{b}_j^* - \mathbf{b}_k^*\right)\left(\mathbf{b}_j^* - \mathbf{b}_k^*\right)^T \\ &= \left(\mathbf{b}_j \widehat{\Sigma}^{-\frac{1}{2}} - \mathbf{b}_k \widehat{\Sigma}^{-\frac{1}{2}}\right)\left(\mathbf{b}_j \widehat{\Sigma}^{-\frac{1}{2}} - \mathbf{b}_k \widehat{\Sigma}^{-\frac{1}{2}}\right)^T \\ &= \left(\mathbf{b}_j - \mathbf{b}_k\right) \widehat{\Sigma}^{-1} \left(\mathbf{b}_j - \mathbf{b}_k\right)^T \\ &= \mathbf{c}\mathbf{B}\widehat{\Sigma}^{-1}\mathbf{B}^T\mathbf{c}^T \\ &= d^2_{\text{Mahalanobis}}\left(\mathbf{b}_k, \mathbf{b}_j\right) \end{aligned} \tag{5.2}$$

Derivation of the squared Mahalanobis distances from the squared Euclidean distance computed on multivariate noise normalized beta estimates. Source: Walther et al. (2016, p. 190).

Following Walther et al. (2016), the distance computation was additionally cross- validated via using the runs of the experiment as partitions, i.e. estimating the beta coefficient for all music genres using all but one run and subsequently computing all item pair distances between the two resulting estimates. This was repeated until all runs were left out once. In order to yield a final RDM, the RDMs were averaged across CV folds. Thus, the utilized distance measure has to be characterized as the cross-validated squared Mahalanobis distance (Walther et al., 2016).

$$\begin{aligned} d^2_{\text{Mahalanobis, crossvalidated}}\left(\mathbf{b}_k, \mathbf{b}_j\right) &= \left(\mathbf{b}_j - \mathbf{b}_k\right)_A \Sigma_A^{-1} \left(\mathbf{b}_j - \mathbf{b}_k\right)_B^T \\ &= \mathbf{c}\mathbf{B}_A \Sigma_A^{-1} \mathbf{B}_B^T \mathbf{c}^T \\ &= \text{LDC}\left(\mathbf{b}_k, \mathbf{b}_j\right) \end{aligned} \tag{5.3}$$

Cross-validated Mahalanobis distance. Source: Walther et al. (2016, p. 191).

Only the final RDMs were used for further analysis which lead to 10 RDMs per participant (5 per hemisphere). The mentioned steps were carried out using the MATLAB based RSA toolbox (Nili et al., 2014).

Analyses of RDMs - models and ROIs

The resulting RDMs were further submitted to three different analysis approaches. The first entailed the comparison between the ROI specific target RDMs and created model RDMs (please see 5.8) in order to evaluate to what extent a given model RDM could provide a possibly explanation of the observed neuronal representation. To this end all model RDMs were tested per ROI target RDM using the method described in 5.5 and Kendall's τ_a. Notably, the model space was extended by a behavioral model which entailed the RDM yielded by the behavioral paradigm averaged across participants (via URIAL's rdm_avg). Next, neuronal RDMs were averaged across participants within each ROI and subsequently used in the same comparison approach as model RDMs with the behavioral RDMs acting as target RDM to enable insights into how well the observed behavior could be explained by the averaged representation of a given brain region. Lastly, the relationship between the averaged neuronal RMDs was investigated which allows the assessment of dissimilarities between representations of multiple brain regions and is therefore termed representational connectivity (Kriegeskorte et al., 2008). In more detail, this allows the inspection of how different brain regions represent information (Kriegeskorte et al., 2008) and thus can hint to different, as well as hierarchical processing strategies. The relationship among RDMs was assessed using Spearman's r as a distance measure, computing the distance between all pairs of RDMs using rdm_dist as implemented in URIAL. This yielded one RDM comprising the distance between brain region RDMs.

5.10.2. RESULTS

Hereinafter the results of the analysis described in the preceding chapter will be presented. Beginning with the outcome of the behavioral part of the experiment, the observed representations of music genres across ROIs will be depicted. This section then concludes with the presentation of insights obtained from the model comparisons.

Behavioral representations

Aiming to assess the cortical representations of music genres through a behavioral representation by means of the paradigm described in 5.7, all participants conducted the behavioral task, which yielded one RDM per participant, after the fMRI part. Participant specific RDMs were averaged to obtain one average behavioral RDM. This high dimensional RDM and its low dimensional visualization illustrated through MDS are depicted below in Figure 5.12. Focusing the RDM, a clear distinction between main genres could be observed for most of the genres which also extended into subgenres to a certain extent. In more detail, participants seemingly perceived the subgenres of a given main genre as highly similar to one another, even though not completely similar. This was most prominent for the classic genres, as they exhibit the highest similarity within a main genre. Besides rock 'n' roll, which was rated comparably dissimilar to all other subgenres, this was also true for the main genre rock. Interestingly, the main genre electro and african-american displayed a gradient of less prominent clustering, as their respective subgenres were apparently perceived more dissimilar to each other than those of rock and classic.

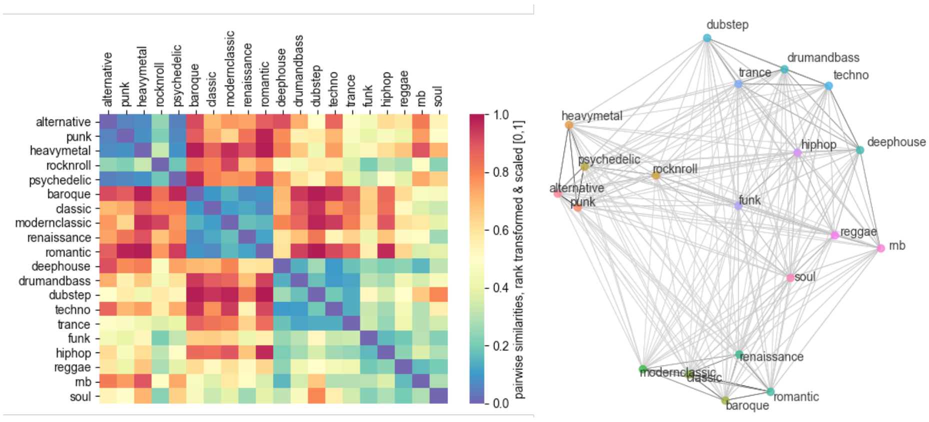


Figure 5.12: *Averaged behavioral RDM across 15 participants (left) and its low dimensional visualization via MDS (right). The RDM is indexed by the amount and category of music genre ordering subgenres within main genres and main genres along the following order: rock, classic, electronic music and afro-american. The respective subgenre are identified through ticks on both x and y axis. Colors in the RDM indicate dissimilarities where blueish colors reflect very small or no distance and reddish colors extremely large distances. Values range from 0 (no distance) to 1 (extreme distance) after rank transforming and scaling the distance values obtained from the behavioral paradigm. Notably, these transformations were only applied for visualization purposes, thus the underlying distances are not altered. The MDS arrangement includes one point in space for each music genre which are labeled accordingly. subgenres of one main genre are identified through a similar color.*

Especially african-american displayed relatively high dissimilarities between its subgenres, pointing to a potentially high internal diversity. These effects are also reflected in across main genres distances, as the coherently arranged main genres yielded the highest dissimilarities reciprocally. Being most prominent for rock and classic, this effect vanishes according to the inherent dissimilarity of subgenres within a given main genre, becoming more blurred for the comparison of rock and classic with electronic music and afro-american respectively. Notably, the across main genre distances of the last two displayed comparably low values. All of the aforementioned observations are also presented in the RDM corresponding MDS. The subgenres of rock and classic are placed very close together and their respective main genre clusters rather far apart. Furthermore, both are clearly less distant from electronic music and african-american. The MDS additionally stresses the uniqueness of classic as it yielded the highest dissimilarities to all other main genres. The addressed less similarities within and across the main genres electronic music and afro-american are also apparent as both are placed close together, while their respective subgenres are fairly further apart. Taken together, the behavioral arrangement yielded a clear distinction between main genres which was most pronounced for rock and classic and to a lesser extent for electronic music and african-american. Main genres that displayed a high internal similarity across subgenres exhibited high dissimilarities to the remaining main genres, that is the more similar subgenres of a main genre the more dissimilar this main genre to all others.

Brain representations across the ventral and dorsal pathway

RSA was used to assess the cortical representation of music genres within ROIs reflecting the auditory cortex in a hierarchical manner in both hemispheres. This yielded 10 RDMs per participants that were averaged across participants, but within ROI. The ten average RDMs and their respective MDS visualization are described hereinafter and are illustrated in Figure 5.13. The brain representations, i.e. categorization of music genres showed a high congruence between ROIs that however is characterized through rather subtle differences along the assumed hierarchy and between hemispheres. More precisely, RDMs varied as a function of these two factors, as their differences throughout the hierarchy were comparable, but yet distinguishable between hemispheres. In contrast to the behavioral RDM no clear distinction between sub or main genres was observable. However, upon a closer inspection, fine grained patterns became evident. This was most prominent for the main genres rock and african-american, with both, but especially the latter, displaying fairly small distances between their respective subgenres as compared to the other main genres. This clustering increased, by means of higher similarities between sub genes, from core, i.e. HG, towards the direction of parabelt regions (STGa/p). While a certain extent of categorization, particularly for african-american, is already visible in both HGs this further manifests in belt ROIs (PP and PT) and yields its peak in parabelt ROIs (STGa/p). Within subdivisions (either belt or parabelt ROIs) the anterior parts showed more pronounced similarities, while posterior regions showed an overall higher similarity. Interestingly, these effects were more prominent in the RH than in the LH. This is in line with previous research work pointing out the significance of the ventral stream and RH dominance during sound categorization. Within main genres, rock 'n' roll and R'n'B exhibited the largest distance among the subgenres of their respective main genres rock and african-american. Regarding classic, a comparable pattern was found however considerably less marked. Here also a cluster emerged along the ROIs with HG exhibiting the weakest and parabelt regions (STGa/p) the strongest expression.

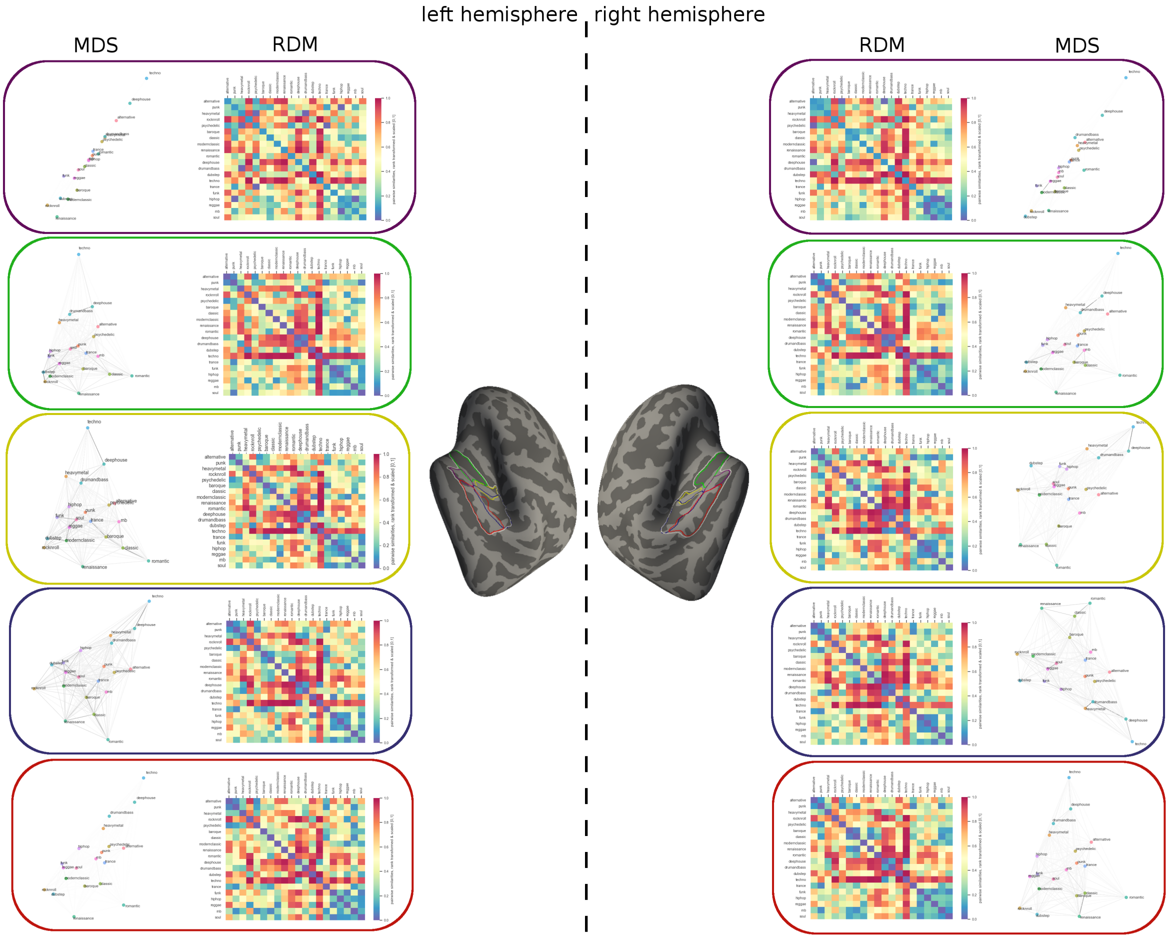


Figure 5.13: *Averaged neuronal RDMs for each ROI across 15 participants and their low dimensional visualization via MDS. Their arrangement reflects both the hierarchy and lateralization of ROIs. Colored frames reflect ROIs within a given hemisphere as indicated through the ROIs being displayed on the inflated surface in the middle of the figure: yellow - HG, green - PP, blue - PT, purple - STGa and red - STGp. The left column displays the RDMs of the LH ROIs and vice versa for the right column. Within each the RDM and their respective MDS visualization are displayed. The RDM is indexed by the amount and category of music genre ordering subgenres within main genres and main genres along the following order: rock, classic, electronic music and afro-american. The respective subgenre are identified through ticks on both x and y axis. Colors in the RDM indicate dissimilarities where blueish colors reflect very small or no distance and reddish colors extremely large distances. Values range from 0 (no distance) to 1 (extreme distance) after rank transforming and scaling the distance values. Notably, these transformations were only applied for visualization purposes, thus the underlying distances are not altered. The MDS arrangement includes one point in space for each music genre which are labeled accordingly. subgenres of one main genre are identified through a similar color.*

Comparing representation across hemispheres, differences with regard to extent and distribution of similarities could be observed. In more detail, the classic subgenres showed a more homogenous distances between one another in the LH, while distances were higher, but less equally distributed in the RH. Focusing the subgenres, renaissance, followed by modern classic stood out as the least similar genres in comparison with the remaining genres across all ROIs and independent of the hemisphere. Given the aforementioned observations, electronic music poses as an immense outlier, as it is very distinct from all other main genres, especially classic and the dissimilarities across its subgenre very tremendously. Remarkably, deephouse, drum and base and techno exhibit small distances among each other while large distances to dubstep and trance which are in turn more similar. This pattern is found across all ROIs and both hemisphere, being the least pronounced in both anterior parabelt ROIs (STGa in LH & RH). Furthermore, techno was the most distinct subgenre across all included genres, as it displayed very high distances when compared with the majority of other subgenres, independent of the hemisphere. Integrating the low dimensional MDS visualizations, the above addressed points become more clear. Even though main genres are not as clearly separated as in the behavioral equivalent, subgenres of a given main genre are placed close together, except the mentioned exceptions, that is rock ‘n’ roll for rock, modern classic and renaissance for classic, as well as dubstep and trance for electronic music. These subgenres are placed fairly outside the clusters of their respective main genres. African-american displays the most consistent and precise cluster, with only R ‘n’ B leaving it. As seen in the RDMs, techno finds itself further apart and isolated than any other subgenre. Taken the aforementioned points together, all ROIs exhibit comparable representations in which a clustering into main genres is present to a certain extent. However, as compared to the behavioral representation, these clusters are more blurred, as they, in most cases, include only a subset of subgenres and the distances between main genres are considerably small. To a limited amount the described representations are modulated as a function of hierarchy and hemisphere. In order to investigate the relationship between hierarchy and hemispheres, the representational connectivity between the representations of all ROIs was computed. This yielded a RDM containing the distances between all ROI specific RDMs. Aiming to address the above mentioned points, two versions of this RDM were visualized: one indexed by hierarchy, i.e. from to core (HG), over belt (PP/PT) to parabelt (STGa/p) and one indexed by hemisphere and hierarchy, i.e. the same order, but within hemispheres. Both are illustrated in Figure 5.14 together with the MDS visualization.

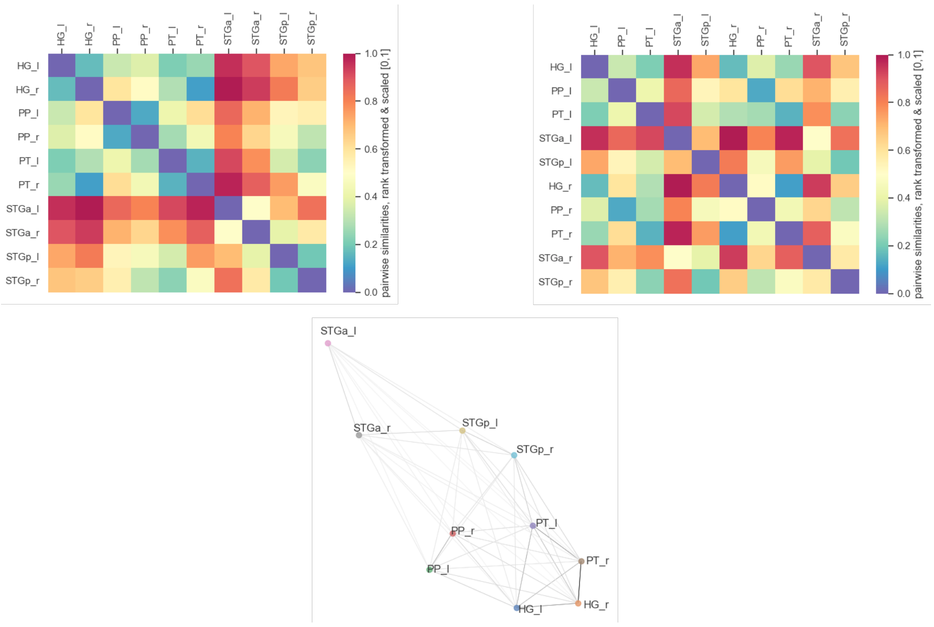


Figure 5.14: *Representational connectivity RDM indexed by the assumed processing hierarchy (left), by this hierarchy and hemisphere (right), as well as the corresponding MDS visualization. Note that as both RDMs contain the same data only differently arranged, the MDS solution would be identical. Hence, only one MDS arrangement is depicted. The respective ROIs are identified through ticks on both x and y axis. Colors in the RDM indicate dissimilarities where blueish colors reflect very small or no distance and reddish colors extremely large distances. Values range from 0 (no distance) to 1 (extreme distance) after rank transforming and scaling the distance values. Notably, these transformations were only applied for visualization purposes, thus the underlying distances are not altered. The MDS arrangement includes one point in space for each ROI, which are labeled accordingly.*

Here, based on the hierarchy indexed RDM, a clear distinction between core and belt ROIs on the one hand and parabelt ROIs on the other hand became obvious, with STGa and STGp, respectively in both hemispheres, displaying high dissimilarities between the remaining ROI representations. This was especially pronounced for STGa, which additionally exhibited fair dissimilarities to its posterior counterpart. Within the core/belt ROI cluster, PP, independent of the hemisphere, was the most dissimilar among the ROIs. With both being ROIs anterior to HG, this furthermore stresses the importance of the ventral stream within the categorization of sound. As already seen in Figure 5.13, besides STGa, all ROIs are highly similar across hemispheres. Even though also existent, the aforementioned patterns are not as clearly visible in the RDM that is indexed by hierarchy and hemisphere. Taking the MDS into account, the patterns further manifest while also showing a slight correspondence between ROIs within a given hemisphere which is however restricted to the core/belt ROI cluster. Summarizing both levels of RDM results, the ROI specific representations as well as their relationship, furthermore point to a special role of the anterior parts of the auditory cortex, i.e. the ventral stream.

Model comparisons

A broad range of model RDMs, covering simple computational models, i.e. models based on acoustic features and conceptual models, i.e. models based on hypotheses, as well as behavior, was tested with regard to their predictive ability towards the observed (neuronal) target RDMs. As outlined in 5.5 this was implemented by means of rank correlations between model and target RDMs, more precisely Kendall's τ_a. The significance of each model RDM relatedness was assessed through a one-sided signed-rank test, treating participants as a random effect and correcting for multiple comparisons across models using the FDR. The respective outcomes are depicted in Figure 5.15 and further presented in the following. Given the similarity between representations across ROIs as described in the preceding section, the model comparisons also exhibited a comparable pattern. Throughout all comparisons the behavioral, main and subgenre, timbre and tempogram models yielded the highest correlations with the target RDM. Even though being statistically significant, all of them fail to reach the noise ceiling by far. Conversely, the random, spectrum, pitch, chroma and dynamic tempo models always performed worse, also displaying negative correlations. Tables including all statistical characteristics of all comparisons (mean correlation, SD, p values) can be found in APPENDIX C1. Focusing the behavioral model, an interaction between hierarchy and hemisphere could be observed, that is correlations were higher for core and belt regions as those for parabelt regions, while additionally always being stronger in the RH as compared the LH. Hence, the core and belt regions of the RH showed the highest correlation with the behavior model. Despite the correlation values, the behavioral model was significantly correlated with all target RDMs. The conceptual models expressed comparable results, however displaying lower correlations values within all but the anterior and posterior parabelt ROIs. While no difference between the main and subgenre model was evident, both were always more related to the target RDM as compared to the random model. This pattern was pronounced in all but the anterior belt and parabelt ROIs in both hemispheres. However, with the exception of the LH STGa, they were all significantly correlated (please see APPENDIX C1). Throughout the complete hierarchy all conceptual models yielded higher correlations with RH representations. With respect to the acoustic feature models, more precisely the ones based on spectral properties, a clear distinction between the timbre and all other models was observed in that timbre provided the best explanation for the obtained representation, independent of hemisphere and processing stage that additionally was always significant. Moreover, the same holds true for the entire model space as timbre also outperformed the behavior and conceptual models in all but the bilateral HG and PT ROIs. Correlations of the timbre with the target RDM peaked in anterior belt and parabelt ROIs. The remaining spectral acoustic feature models, that is spectrum, pitch and chroma displayed a very limited relatedness, yielding even negative correlations within the majority of ROIs, with especially the chroma model exhibiting low predictive abilities. No striking difference between hemispheres or ROIs could be observed. This is in contrast with previous research work that suggested a prominent involvement of these ROIs in the processing of the respective features (Allen et al., 2017; Allen et al., 2018; Alluri et al., 2012; Alluri et al., 2013; Santoro et al., 2014; Santoro et al., 2017; Toiviainen et al., 2014). Considering the entirety of the cortical auditory system, spectral features provided better explanations for RH representations that in sum were more often significant. Dynamic tempo and tempogram as temporal acoustic feature models showed yet another divergence as dynamic tempo was always negatively correlated, while tempogram yielded significant high correlation values in all but the anterior belt and parabelt ROIs. Based on that it was the acoustic feature model with the second best predictive ability after timbre. Bilateral HG and PT posed as an exception to that, as within these ROIs both models reached comparable correlations. Regarding the noise ceiling, no prominent difference between ROIs could be observed, neither with respect to hemisphere nor processing stage. Furthermore, the noise ceiling of all representations were narrow and thus pointing to a high consistency of representations across participants, yet far below a correlation of one with none of them exceeding a correlation of 0.6. In sum, an interaction between the complexity of models, hierarchy and hemisphere could be observed, as complex models always outperformed low-level ones and all models showed an advantage for the RH, while their predictive ability differed between core, belt and parabelt ROIs within both the ventral and dorsal stream.

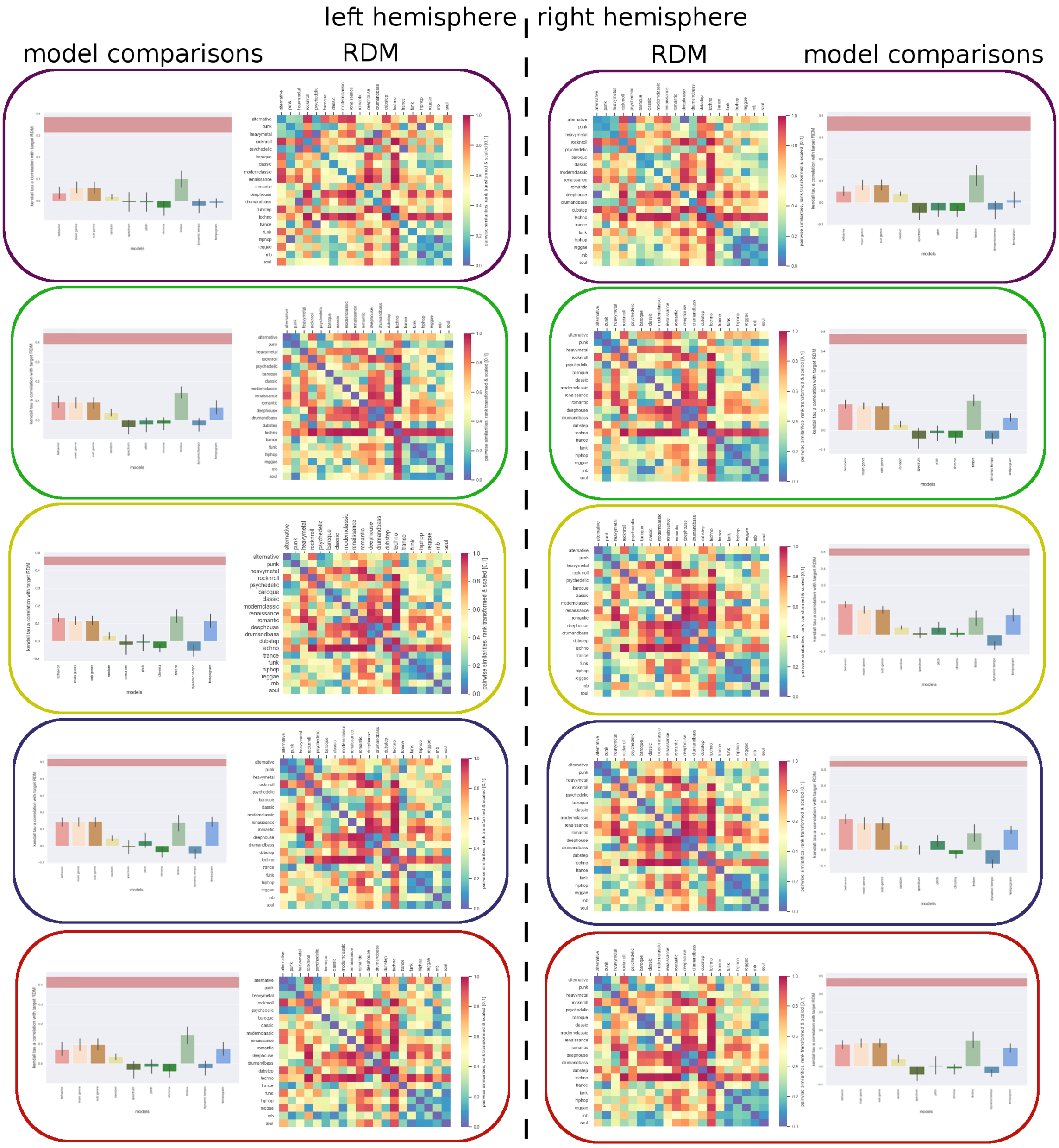


Figure 5.15: *Averaged neuronal RDMs for each ROI across 15 participants and their respective model comparison results. Their arrangement reflects both the hierarchy and lateralization of ROIs. Colored frames reflect ROIs within a given hemisphere: yellow - HG, green - PP, blue - PT, purple - STGa and red - STGp. The RDMs are identical to the ones displayed in Figure 42. The RDM is indexed by the amount and category of music genre ordering subgenres within main genres and main genres along the following order: rock, classic, electronic music and afro-american. The respective subgenre are identified through ticks on both x and y axis. Colors in the RDM indicate dissimilarities where blueish colors reflect very small or no distance and reddish colors extremely large distances. Values range from 0 (no distance) to 1 (extreme distance) after rank transforming and scaling the distance values. Notably, these transformations were only applied for visualization purposes, thus the underlying distances are not altered. Bar plots represent results of the model comparison, depicting all model RDMs' correlation with the corresponding target RDM. The order of the model RDMs is identical across all bar plots and as follows: behavioral model, main genre model, subgenre model, random model, four spectral acoustic feature models (spectrum, pitch, chroma, timbre) and two temporal acoustic feature models (dynamic tempo and tempogram). Black lines within bars represented error bars, while asterisks indicate that a respective model's correlation was significant at $p < 0.05$ (FDR). Red rectangles reflect the computed noise ceiling for each target RDM.*

Brain & behavioral representations

In order to evaluate the ability of brain representations to explain the observed behavior, the former were averaged within ROI across participants and used as model RDMs for the behavioral target RDM, using the same model comparison approach as described above. A graphical depiction of the respective outcome can be found in Figure 5.16. Even though all ROI specific representations yielded fairly low correlations with the behavioral representation, they were all significantly related (please see APPENDIX C1). A pattern comparable to that of the brain representation model comparison (preceding section) was observed, that can be described by the factors hemispheres, hierarchy and pathways. Across all ROI specific representations the RH versions exhibited a slight advantage as compared to the LH version, as they yielded higher correlations. This difference was similar throughout all brain representations. However, the predictive ability of brain representations varied as a function of hierarchy, as no prominent contrast between core and belt, but between parabelt and the remaining regions was observed. More precisely, correlation values dropped from belt to parabelt regions, while being comparable to those of the core region. Upon a closer look, the ROIs of the dorsal pathway systematically reached higher correlation than those of the ventral pathway. Thus model performance decreased towards parabelt regions and more drastically within the ventral pathway. Notably, all brain representations were far from reaching the noise ceiling. The behavioral representation were seemingly consistent across individuals, as indicated by the narrow noise ceiling, that nevertheless peaked at 0.6 and therefore displayed only a limited approximation of the true model.

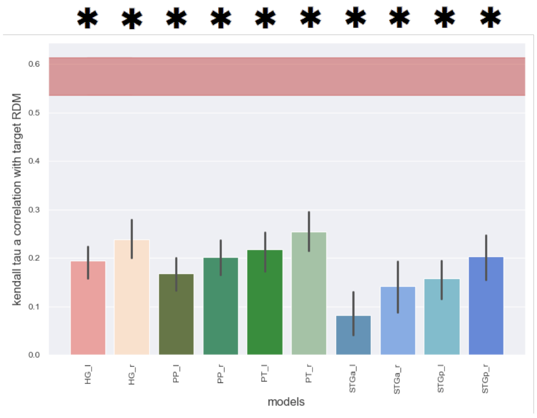


Figure 5.16: *Model comparison between behavior and brain representations. Bar plots represent the correlation between behavior and the respective brain representations, with their order and color indicating the assumed processing hierarchy: reddish - core, greenish - belt and blueish - parabelt. Black lines within bars represented error bars, while asterisks indicate that a respective model's correlation was significant at* $p < 0.05$ *(FDR). Red rectangles reflect the computed noise ceiling for the behavioral target RDM.*

5.11. NOTHING COMPARES TO YOU II - TIME COURSE OF MUSIC GENRE REPRESENTATIONS

5.11.1. MATERIALS & METHODS

This chapter presents details on the execution of the EEG part of the project, including participant characteristics, content and procedure, data acquisition and processing, as well as the application of the representational similarity framework.

PARTICIPANTS

In this part of the project, twelve participants (one left-handed, ten female) with a mean age of 22.08 years (SD=2.61) were included, with four of them currently making music (M=12.75 years, SD=1.79). All of them were culturalized within the western tonal system and indicated no hearing impairments, neurological or psychiatric disorders. All participants gave their informed consent for their participation in the experiment (APPENDIX B) and to publicly share their data, once anonymized (through the Open Brain Consent Form (APPENDIX A)). Participants received either no compensation or course credits. The ethics committee of the psychology department of the Philipps-University Marburg, Germany permitted all parts of the herein described project.

PROCEDURE

Initially, participants were informed about the experiment and possible risks of EEG investigations, while their ability to take part in the experiment was assessed beforehand (characteristics such as skin allergies, dreadlocks, freshly dyed hair, among others led to an exclusion). They were then asked to give their written informed consent to participate in the study and subsequently were prepared for the EEG measurement which took roughly 45-60 minutes. Afterwards, participants were seated in a quiet light dimmed room and equipped with headphones (Sennheiser HD 555). Before the experiment itself participants had the opportunity to adjust the volume of the to be presented auditory stimuli. More precisely, 4 examples (one from each main genre) were presented which volume could either be increased or decreased via buttons on a keyboard, with their preferred volume setting being indicated by another button press and logged for the experiment. Subsequently, the paradigm, as described in 5.5, was presented in 8 runs. The movie segments were presented on screen which was placed roughly 1 m in front of the participants. Pauses between each run were limited to 2 minutes, but participants could continue prematurely by pressing a button on the keyboard. These pauses were furthermore used to ask participants about their wellbeing. Depending on the breaks between runs, this part of the experiment lasted roughly 70-75 minutes. After the EEG measurement, participants were freed from the EEG cap and other sensors and given a short break of around 15 minutes to relax and wash their hair. For ca. one hour they then continued with the behavioral paradigm as depicted in 5.7 on a Lenovo laptop running Matlab (R2014a, MathWorks Inc., Natick, MA, USA) and using Sennheiser HD 380 pro headphones. The final part of the experiment consisted of the behavioral test battery (please see 5.9) leading to a duration of around 4 hours for the complete experiment.

EEG ACQUISITION

The entirety of EEG data was acquired at the department of Psychology of the Philipps-University Marburg on a BioSemi system (BioSemi B.V., Amsterdam, Netherlands, 1998), using Actiview (https://www.biosemi.com/download.htm, 2013, version 7.05) for recording. After cleaning those parts of the skin that would be covered by the EEG cap with Isopropanol (70%), a 64 channel cap following the 10-20 system (Electro Cap International Inc, Eaton, USA), together with 2 electrodes below and next to the right eye (vertical and horizontal EOG respectively), was used to record the electrophysiological data. Via measuring the participant's head circumference, the ideal cap size was assessed and the respective cap placed on participants' heads. In order ensure a central and fitting alignment, the distances between the nasion and inion, as well as left and right preauricular points were measured and the central vertex point defined by means of the intersection of the two lines (Klem, Lüders, Jasper, & Elger, 1999), which in turn served as the orientation for the EEG cap position. After the cap was fixated using an elastic chin strap, the electric impedance was diminished by reallocate hair and subsequently applying a conductive electrolyte gel to the parts covered by the electrodes. The impedance of all electrodes, scalp and reference, was set as sufficient at 25 µV. The EEG data was recorded at a sampling rate of 1024 Hz as one continuous measurement across all runs. Within Actiview a marker channel was used to additionally record onsets and durations of the different music genre excerpts, as well as pauses and run breaks.

PREPROCESSING AND STATISTICAL ANALYSIS

Data preparation

Prior to any processing, the data from all participants was adapted to BIDS (Gorgolewski et al., 2017), following the respective extension proposal for EEG (Pernet et al., 2019).

Preprocessing

The preprocessing of the EEG data was conducted in python using MNE (version 0.15) (Gramfort et al., 2014) and comprised the following standard steps: filtering, re-referencing, artefact rejection and epoching. In order to clean the EEG data from certain types of noise and artefact related frequencies, finite impulse response filters of 0.1 Hz high-pass and 50 Hz low-pass were applied. Aiming at providing a more fitting offline reference, the cleaned EEG data of all channels was re-referenced using the average reference, that is, subtracting the mean across all channels from each channel (Dien, 1998). To further reduce potential artefacts, an Independent Component Analysis (ICA) was applied to the data aiming to unravel the EEG data into experiment (of interest) and artefact/noise (of no interest) components. The extend-infomax algorithm as implemented in MNE, along with a decimation rate of 3 was used to set up an ICA to obtain a preset amount of 35 components. The resulting components were inspected visually for each participant, with a subset being excluded from further analysis based on their characteristics such as frequency power across spectrum, among others. Finally, the EEG data was epoched defining a 500 ms pre- and 6000 ms post-stimulus onset, hence effectively capturing the complete stimulus duration, as well as a pre-stimulus time period. The latter was then used for a baseline correction of the obtained epochs to diminish edge artefacts.

Statistical analysis

As the epoched data was used for the RSA, no classic statistical analysis was necessary. However, in order to provide a brief sanity check, that is if the stimuli evoked a traceable response, music genre specific event related potentials were assessed and visually inspected using MNE.

REPRESENTATIONAL SIMILARITY ANALYSIS

As described and motivated in 5.5, RSA was used to assess the second moment matrix of the different music genre corresponding epoched data, as well as to provide a basis for a comparison between time points/windows and models. Overall this part of the analysis included 3 parts: computation of power spectral density and time window cropping, computation of RDMs and comparison of RDMs across time windows and models.

Power spectral density and time windows

Even though RSA has been applied within the context of highly time-resolved neurophysiological such as MEG and EEG its application is rapidly growing (Guggenmos, Sterzer, & Cichy, 2018), a prominent gap in terms of investigated modalities exists, as almost the entirety of studies focused on static visual stimuli, e.g., pictures of objects belonging to different categories (Cichy, Pantazis, & Oliva, 2014; Wardle, Kriegeskorte, Grootswagers, Khaligh-Razavi, & Carlson, 2016). In contrast, only a very limited amount of previous research work focused on auditory processing and if so, they utilized stimuli with a very short duration of around 500 ms (Su, Zulfiqar, Jamshed, Fonteneau, & Marslen-Wilson, 2014). Together with the fact that the aforementioned studies used either epoched or source reconstructed data, these circumstances pose as problems with regard to the most suited data processing approach for the stimuli, i.e. music genres, included in the here described study, as they had a duration of 6 sec. and therefore differed by a prominent magnitude from those of previous research work. This is especially motivated by the methodological aspects of the majority of studies that investigated music genre processing using electrophysiological data. In more detail, contrary to work in the visual domain where mainly epoched data was used, a broad range of time-frequency analysis has been applied (Banerjee et al., 2016; Bauer, Kreutz, & Herrmann, 2015; Daly et al., 2015; Hadjidimitriou & Hadjileontiadis, 2012; Hurless et al., 2013; Sammler, Grigutsch, Fritz, & Koelsch, 2007), which is based on the time-varying characteristics of music, that is a (most likely) non-linear combination of acoustic features which changes over time. The processing of music genres might therefore not be grounded in a static percept like it can be assumed within the visual domain (at least for pictures), but in an unstable percept that varies over the time course of listening to a musical piece and which is getting more demanding to grasp the more time passes. Hence, representations of genres are possibly not restricted to a set of electrodes or ERPs, but highly distributed across time and space, additionally forming a more stable representation during later time periods given an assumed genre related categorization that manifests as the processing progresses. In order to address these points, additionally considering the novel character of the project at hand, the following approach was conducted. Aiming to capture the variation of representations over time the epoched data was segmented into six equally spaced time windows of 1 sec. each as this partitioning provides a distinction between early, mid and late processing stages while at the same time windows of length that are sufficient to entail a stable percept. Subsequently, the global power spectral density (PSD) was computed for each music genre within each segmented time window to obtain the corresponding time-frequency information, providing a representation reflecting the temporal and spatial distribution. More precisely, a multitaper PSD with 7 discrete prolate spheroidal sequences (DPSS) windows was applied. Both the segmentation and PSD computation was achieved through functionality available in MNE. This approach yielded a n conditions PSD x n time time windows (20 music genres x 6 time windows) vector per participant that was used for further processing as described in the following section.

Multivariate noise normalization

As outlined in 5.5 multivariate noise normalization of the data prior to computing distances in order to assemble RDMs constitutes a crucial step, as it minimizes potential biases and increases the reliability of the estimates. Comparable to fMRI (please see 5.10.1.5), also the measurement channels of EEG, that is electrodes, exhibit a noise covariance between one another. Here, the variance-covariance matrix (electrodes x time points) was computed using the complete respective epoch, i.e. time window, as suggested by Guggenmos et al. (2018) and subsequently used to noise normalize the data in a multivariate manner without shrinkage. The above described computations were conducted via custom python code accompanying Guggenmos et al. (2018) available on GitHub (https://github.com/m-guggenmos/megmvpa/tree/master/tutorial_python), utilizing the python packages NumPy, SciPy and scikit-learn.

$$x^* = x\Sigma^{-\frac{1}{2}} \tag{5.4}$$

Multivariate noise normalization of electrophysiological data, where x indicates the observed patterns and $\sum$ *the obtained variance-covariance matrix.*

Computation of representational dissimilarity matrices RDMs were computed in a two step procedure, following Guggenmos et al. (2018). First, 5 pseudo-trials per condition were generated by randomly assigning multivariate noise normalized raw trials and averaging them. This was repeated 20 times in a random permutation. Subsequently, the cross-validated squared Euclidean distance was computed by the following equation:

$$d^2_{\text{Euclidean},c.v.}(x, y) = (x - y)_{[A]}(x - y)^T_{[B]} \tag{5.5}$$

Multivariate noise normalization of electrophysiological data, where x indicates the observed patterns and $\sum$ *the obtained variance-covariance matrix.*

Here A and B denote cross-validation folds with all but two left-out pseudo-trials and the corresponding left-out pseudo-trials respectively. The distance between each pair of music genres was computed for the above mentioned 20 permutations and then averaged across permutations to obtain a final distance. This procedure was repeated for each of the six time windows, thus yielding six time window specific RDMs containing all pair-wise distances for each participant.

Analyses of RDMs - models and time windows

In order to enable a precise investigation of the obtained RDMs, their representations were further analyzed by means of three properties: the predictive ability of model RDMs, their relationship with behavioral presentations and their relationship over time. Considering the first, all model RDMs (please see 5.8) extended by a behavioral model, that is the RDM based on the behavioral paradigm (5.6) averaged across participants (through rdm_avg as implemented in URIAL), were compared to all time window specific target RDMs as depicted in 5.5, using Kendall's τ_a. The goal of this comparison was to evaluate if certain model RDMS could provide an explanation of the neuronal representation observed over time. Aiming to unravel if the observed behavioral representation was grounded in rather early, mid or late phases of perception, time window RDMs were averaged across participants and served as model RDMs for the behavioral target RDM. In terms of parameters, this comparison was identical to the one outlined above. Furthermore, the temporal connectivity of the representations was assessed by comparing all time window specific RDMs with one another enabling the investigation of how dissimilar representations are during the time course of perception.

5.11.2. RESULTS

Hereinafter the results of the analysis described in the preceding chapter will be presented. Beginning with the outcome of the behavioral part of the experiment, the observed representations of music genres across time will be depicted. This section then concludes with the presentation of insights obtained from the model comparisons.

Behavioral representations

Using the paradigm described in 5.7, the behavioral representations, by means of a RDM, of all participants were obtained after the EEG part to yield a respective equivalent of their mental categorization. Figure 5.17 displays the average of this representation based on the individual participant specific RDM, as well as the corresponding MDS which aims to provide a low level representation. As the behavioral representation obtained in the preceding fMRI part, the RDM displayed a differentiation between main genres, as well as to a lesser extent between subgenres. Both of these observations varied as a function of genres, exhibiting a difference between rock and classic on the hand and electronic music and afro-american on the other, that is best described through their dissimilarity to other main genres and their within main genre dissimilarities between subgenres. Rock and classic displayed a prominent clustering as they both were highly dissimilar to all other genres, while their subgenres were very similar to one another. This effect was most prominent for classic music as it yielded the highest between main genre differences towards electronic music, followed by rock and least pronounced for afro-american, while its subgenres were perceived as tremendously similar. The same holds true for rock, but to a lesser extent, as main genre differences were diminished and within subgenre distances increased. Notably, rock'n'roll was very dissimilar to all other subgenres. In contrast to classic, rock exhibited the most significant difference towards the former and decreased further for electronic music and afro-american. These genres furthermore displayed an inversion of the aforementioned pattern, as their distinction towards other main genres and especially between one another was reduced. Together with comparably high distances between their respective subgenres, that were most pronounced in afro-american, this resulted in a rather fuzzy clustering. The corresponding MDS reflects the above noted pattern, as all main genres are placed apart from each other and their subclusters exhibiting varying distances.

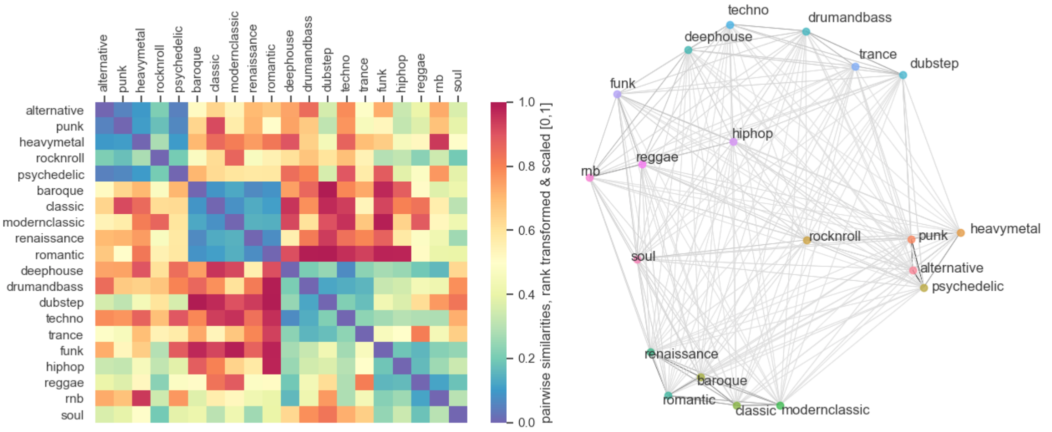


Figure 5.17: *Averaged behavioral RDM across 12 participants (left) and its low dimensional visualization via MDS (right). The RDM is indexed by the amount and category of music genre ordering subgenres within main genres and main genres along the following order: rock, classic, electronic music and afro-american. The respective subgenre are identified through ticks on both x and y axis. Colors in the RDM indicate dissimilarities where blueish colors reflect very small or no distance and reddish colors extremely large distances. Values range from 0 (no distance) to 1 (extreme distance) after rank transforming and scaling the distance values obtained from the behavioral paradigm. Notably, these transformations were only applied for visualization purposes, thus the underlying distances are not altered. The MDS arrangement includes one point in space for each music genre which are labeled accordingly. subgenres of one main genre are identified through a similar color.*

In more detail, rock and classic were far apart from each other, while their respective subgenres were placed in a close range. On the contrary, electronic music and afro-american were suited in near distance and their subgenres placed further apart. Furthermore, classic displayed the farthest distance to all other genres. In sum, the behavioral representation revealed a categorization into main genres that was defined through a high dissimilarity towards other main genres and a high similarity across main genre corresponding subgenres. It is worth mentioning, that the distance to other main genres increased with a decrease of distances between subgenres.

EEG time window representations & model comparisons

Through RSA, cortical representations of music genres across time were obtained that reflect early, mid and late stages of their perception respectively. For each participant, this resulted in six RDMs that captured the electrophysiological response to the complete stimuli in non-overlapping time windows of one second. Subsequently, a multitude of models, including such based on low level acoustic features and high level features like concepts and behavior, was evaluated with respect to how well they could explain the observed neuronal RDMs. This was conducted by means of the model comparison approach described in 5.5 and thus deployed the feature based models as model RMDs, while the neuronal RDMs posed as target RDMs. Both RDM types were compared through a Kendall's τ_a correlation approach, with their significance being assessed through one-sided sign-rank tests that defined participants as a random effect and applying the FDR to correct for multiple comparisons across models. Figure 5.18 illustrates the respective outcomes through the averaged time window corresponding RDMs, their low dimensional MDS visualization and the results of the model comparisons. RDMs varied tremendously across time points, with no clustering comparable to the behavioral representation, that is based on main or subgenres, becoming evident. However, certain factors can be utilized to characterize the observed representation that are overall defined through the absent of any precise categorization. Throughout all time windows, rock and especially alternative, displayed the highest distance to all other main genre and their respective subgenres, while its subgenres exhibited the highest degree of genre related clustering, yet showing very large distances. Both the large between main genre distance and the comparatively small within genre distance are most pronounced in the fourth time window from 3 - 4 seconds. Together with the high between genre difference displayed by afro-american this created a frame-like pattern of large distances on the "outside" of the RDM (rock and afro-american respectively) that framed smaller distances on the "inside" (classic and electronic music). With regard to the latter classic and electronic music displayed a highly diverse arrangement of distances, that however were mainly small. Additionally, they did not display a differentiation between main or subgenres, but an arrangement of small distances that cross genre borders. This pattern manifests throughout the early processing stages and peaks in the third time window of 2 - 3 seconds. Afterwards, in the second half of the stimulus perception, the high dissimilarity of rock and distributed cluster of small distances between classic and electronic music remains, while the frame creating high distance of afro-american to all other genres vanishes. This trend continues and concludes in the last time window of 5 - 6 seconds within which the representation appears to be close to random. The respective MDS visualizations underline this impression, as across all time windows, alternative is placed far apart from all other genres, but forms a distinctive cluster with the remaining subgenres of its main genre, except psychedelic, in the 3 - 4 seconds time window. As for the entirety of other genres a large cluster with small distances and widely overlapping genres could be observed, with only the last time window of 5 - 6 seconds displaying a broader, more distinct cluster with fewer overlapping, but yet again no genre related subdivisions. Notably, the diagonal of all RDMs, thus the dissimilarity of identical excerpts across trials and runs, do not exhibit the smallest distance, as a vast amount of pairwise distances were negativ. Focusing the relationship between time windows, thus the stability and evolution of the representations over time, their representational connectivity was computed by means of the Spearman correlation between time window corresponding RDMs.

This yielded a RDM that was indexed by the amount of time windows and included all pairwise distances. This RDM was additionally visualized through MDS, with both being depicted below in Figure 5.19. The representational connectivity, especially the MDS visualization, revealed a distinction between mid stages of the processing on the one hand and very early and late stages on the other. Hence, the representation is more comparable between early and late stages than between early and mid or mid and late stages. Considering the above described time window corresponding RDMs, mid stage representations seemingly differ systematically from those of early and particularly late stages. The predictive ability of the deployed models varied greatly across time windows and was restricted to very low correlation values that were frequently negative and on no account significantly related to the time window corresponding target RDM. Overall, the behavioral model performed best, followed by the random and tempogram model. On the contrary, the remaining conceptual and acoustic feature models performed worse. Statistical characteristics of the comparisons can be found in APPENDIX C2. The behavioral and the random model yielded comparable correlations with the target RDM across all time windows, with no clear advantage for either. They were therefore the most stable models, as all others showed a greater extent of variation. Considering the acoustic feature models that were based on spectral properties, pitch provided the best explanation followed by the closely related spectrum model. While the first was mostly positively correlated, except for the fifth time window from 5 - 6 seconds, the latter also displayed negative correlations in other processing stages. The more complex spectral acoustic features chroma and timbre always showed little to no correlation, in both positive and negative direction. The same accounts for the temporal acoustic feature models dynamic tempo and tempogram with the exception of the second (1 - 2 seconds) and sixth time window (5 - 6 seconds) in which tempogram performed nearly as good as the behavior and random model. These models furthermore reached the noise ceiling in the majority of the time windows, which is distorted by the extreme range of the noise ceiling across all processing stages and its extent to negative values in some cases, i.e. the third and fourth time window, that together points to a large variability of representations across participants. Taken together, music genre representations seem to vary greatly over time, not reaching any form of precise genre related clustering. Furthermore, with the exception of the behavior and random model, the tested model RDMs all fail to provide a sufficient explanation of the observed neuronal target RDM, with the abilities of the behavior and random model remaining questionable due to seemingly high interindividual variability.

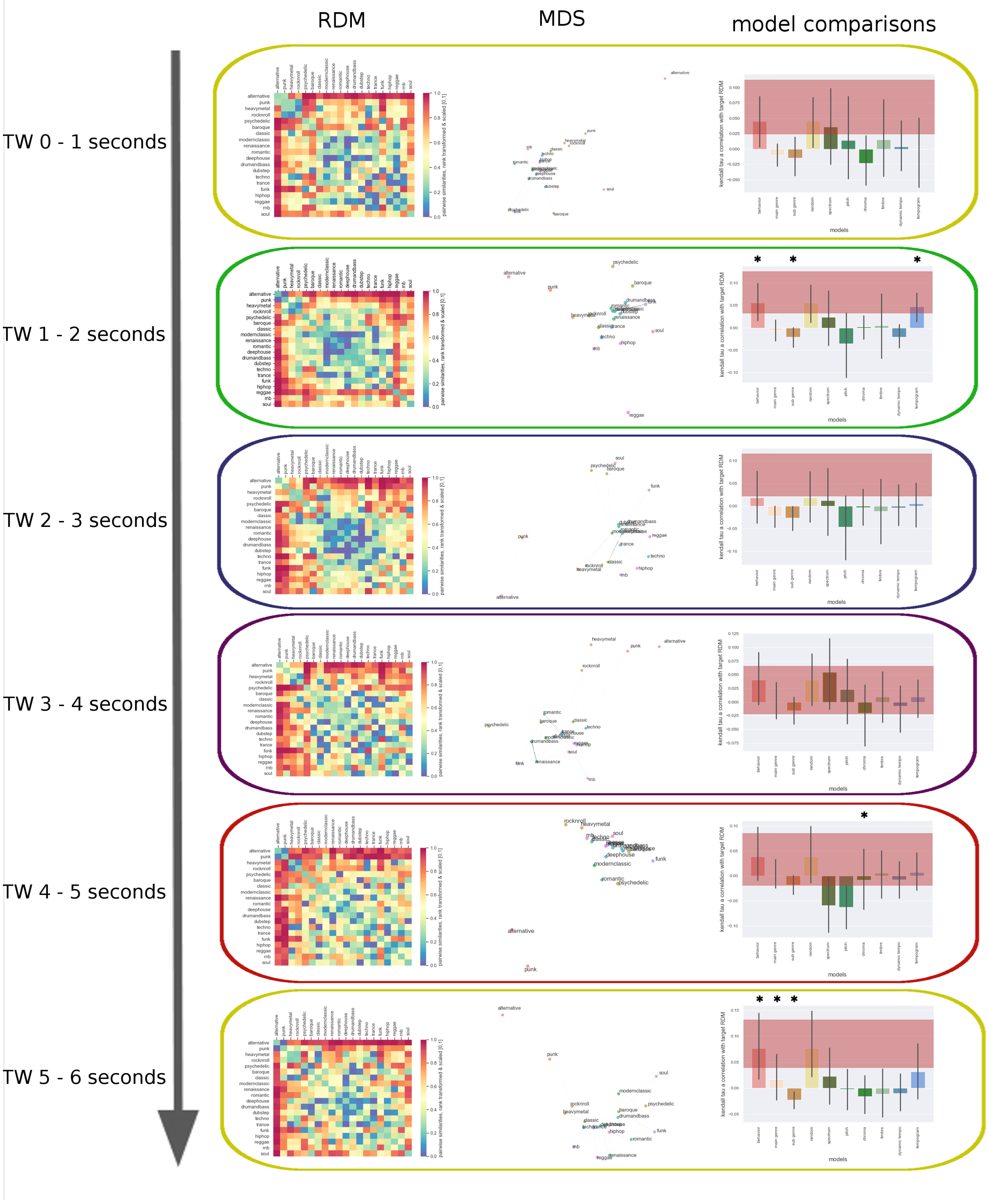


Figure 5.18: *Averaged neuronal RDMs for each time window across 12 participants and their low dimensional visualization via MDS, as well as corresponding model comparison results. Their arrangement reflects the time window based processing stages, that are furthermore highlighted by the colored frames. The left column displays the averaged RDMs, the middle column their MDS visualization and the right column the model comparison results. The RDM is indexed by the amount and category of music genre ordering subgenres within main genres and main genres along the following order: rock, classic, electronic music and afro-american. The respective subgenres are identified through ticks on both x and y axis. Colors in the RDM indicate dissimilarities where blueish colors reflect very small or no distance and reddish colors extremely large distances. Values range from 0 (no distance) to 1 (extreme distance) after rank transforming and scaling the distance values. Notably, these transformations were only applied for visualization purposes, thus the underlying distances are not altered. The MDS arrangement includes one point in space for each music genre which are labeled accordingly. subgenres of one main genre are identified through a similar color. The order of the model RDMs within the model comparison is identical across all bar plots and as follows behavioral model, main genre model, subgenre model, random model, four spectral acoustic feature models (spectrum, pitch, chroma, timbre) and two temporal acoustic feature models (dynamic tempo and tempogram). Black lines within bars represented error bars, while asterisks indicate that a respective model's correlation was significant at $p < 0.05$ (FDR). Red rectangles reflect the computed noise ceiling for each target RDM.*

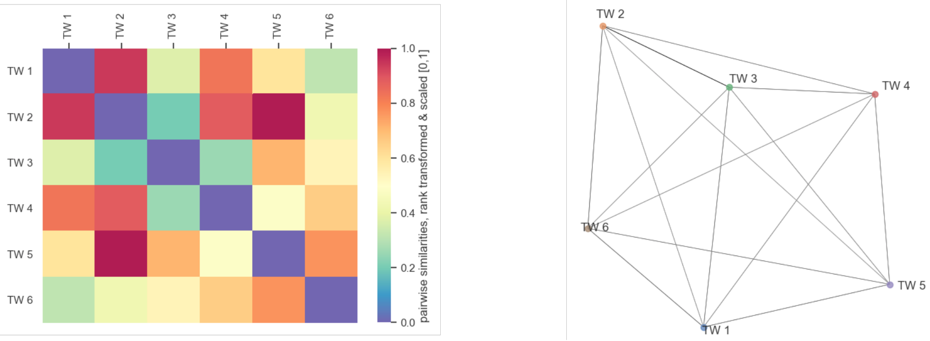


Figure 5.19: *Representational connectivity RDM indexed by the time windows (left) and its corresponding MDS visualization (right). The respective time windows are identified through ticks on both x and y axis. Colors in the RDM indicate dissimilarities where blueish colors reflect very small or no distance and reddish colors extremely large distances. Values range from 0 (no distance) to 1 (extreme distance) after rank transforming and scaling the distance values. Notably, these transformations were only applied for visualization purposes, thus the underlying distances are not altered. The MDS arrangement includes one point in space for each time window which are labeled accordingly.*

EEG time windows & behavior

Using the same model comparison approach as above, the time window specific RDMs were utilized as model RDMs to evaluate how well they could explain the behavioral target RDM, thus investigating if certain processing stages relate to the obtained behavioral categorization. The resulting correlation values were however fairly low across all time windows and did not display any marked differences between them. Furthermore, no time window specific representation reached the noise ceiling or was significantly related to the behavioral target RDM, which in turn appears to be stable across participants as indicated through a noise ceiling of intermediate range. Nevertheless, the noise ceiling peaked at 0.6. A graphical depiction of the above described outcome can be found in Figure 5.20.

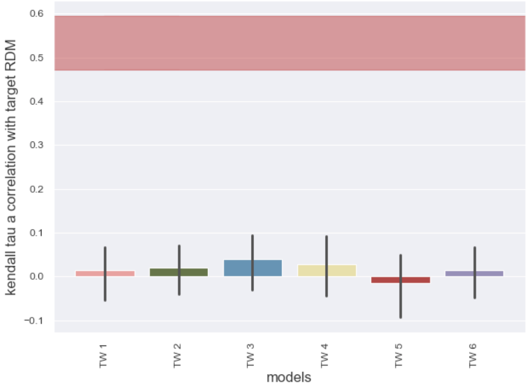


Figure 5.20: *Model comparison between behavior and time window specific representations. Bar plots represent the correlation between behavior and the respective time window representations, with their order indicating the assumed processing hierarchy. Black lines within bars represented error bars, while the red rectangle reflects the computed noise ceiling for the behavioral target RDM.*

5.12. NOTHING COMPARES TO YOU III - THE INFLUENCE OF MUSICAL TRAINING ON MUSIC GENRE PERCEPTION

This chapter will depict details on the purely behavioral part of this project that targeted a comparison between different populations, more precisely musicians and non-musicians. This encompasses information on data acquisition and analyses, as well as yielded results.

5.12.1. MATERIALS & METHODS

Starting with a description of the included participant sample and applied procedure, the following chapter will also entail details of the data acquisition and conclude with the different applied group comparison approaches.

PARTICIPANTS

As outlined in 5.1 & 5.2, this part aimed to test possible differences in the perception and categorization of music genres between musicians and non-musicians. To this end, two groups were included in this part of the study. Following previous research work (Intartaglia, White-Schwoch, Kraus, & Schön, 2017; Madsen, Whiteford, & Oxenham, 2017) that investigated potential differences between musicians and non-musicians, a precise set of characteristics were defined that participants had to fulfill in order to be assigned to one of both groups and therefore to be included in this part of the project. In more detail, musicians were defined as individuals that had at least 10 years of musical training with either an instrument or their voice and should currently actively engage in a musical project (e.g., playing in a band or singing in a choir) for at least one hour per week. On the other hand, non-musicians should not have received musical training that exceeded the duration of three years, as well as should not receive musical training currently. These criteria were communicated during the recruitment of the participants and the respective information assessed via self-disclosure. In order to verify their status, the scores of the MET subtests (please see 5.9), recorded within the behavioral test battery (as described in 5.9) were compared between the self-disclosure based groups by means of a Mann–Whitney U test as implemented in pingouin (Vallat, 2018). Participants in the musician group showed significantly higher scores for both tests of the MET. In more detail, musicians reached an average score of 42.9 (SD=4.20) in the melody subtest, while non-musicians yielded 31.8 (SD=4.85) (p = 0.000366). Focusing the rhythm subtest, the scores were 40.5 (SD=4.84) and 32.9 (SD=4.98) for musicians and non-musicans respectively (p = 0.008901). Per group 10 participants were included. The musician group consisted of 7 females and 3 males with a mean age of 27.5 years (SD=3.03) that had received musical training for 13.2 years (M, SD=3.43). In contrast, the participants in the non-musician group had 1.36 years of musical training (M, SD=1.63) and comprised 6 females and 4 males. Except for one participant, no neurological and/or psychiatric diseases were reported. One participant stated an auditory processing disorder within the behavioral test battery (please see 5.9), which lead to no difficulties related to auditory perception in the past or currently. Therefore, and as the participant's respective MET scores did not display outlier characteristics, the data was kept for further analyses. Two consent forms, one to provide their written informed consent (APPENDIX B) to participate in the study and one to enable the open sharing of their anonymized data in public repositories ("Open Brain Consent Form" (APPENDIX A)), were signed by all participants. Course credits were offered as compensation. This part of the project and its procedure, as well as contents, as described in 5.11.1.2, were approved by the ethics committee of the faculty of medicine of the Philipps-University Marburg, Germany.

PROCEDURE

Starting with explaining the content and instructions regarding the experiment, participants were then asked to provide their written informed consent to first, willingly take part and second, share their anonymized data. Afterwards their were seated in a quiet room and provided with headphones (Sennheiser HD 380 pro) to allow a high resolution auditory stimulation during the experiment. They then proceeded with the behavioral paradigm as described in 5.7, which lasted around 45-60 minutes. Subsequent to a short break, they completed the behavioral test battery (please see 5.9). Together, the complete experiment took about 2 hours.

REPRESENTATIONAL SIMILARITY ANALYSIS & GROUP COMPARISON APPROACHES

The behavioral categorization of the different music genres and its potential differences between musicians and non-musicians were assessed by RSA (for a description and motivation please see 5.5 & 5.7), as well as different uni- and multivariate approaches which can be divided into 3 parts: RDM computation and model comparisons, mean distances and correlations of RDMs across trials, RDM based machine learning. RDM, MDS and model fit plots were visualized via URIAL's plot_rdm, plot_mds and plot_model_fit respectively.

RDM computation and model comparisons

RDM computation

Participant specific RDMs were computed using the iMDS approach described in 5.7. For each participant, the final RDMs, as well as all trial corresponding RDMs were converted from mat to CSV files using the function rdm_mat2csv as included in the URIAL toolbox. This yielded one final RDM and a varying amount of trial corresponding RDMs (M=53.8, SD=11.2) per participant. The final RDMs were then concatenated within each group using URIAL's rdm_concat_df functionality. Additionally, this was conducted across groups to enable a 2nd level RSA comparison approach. In more detail, using Spearman correlation within the rdm_dist function from URIAL, the lower triangle, discarding the diagonal, from each participant's final RDM was correlated with the lower triangle from all other participants, thus yielding a RDM displaying the dissimilarity of RDMs between participants, as this allows a descriptive assessment of group corresponding RDM characteristics.

Model comparisons

In order to test if the models introduced and depicted in 5.8 could provide an explanation for the potentially different RDMs of musicians and non-musicians, the full range of models was compared to the iMDS RDMs using the model comparison approach outlined in 5.5 and implemented in URIAL as rdm_compare. More precisely, the model comparison was conducted in a group-specific manner, that is, once for all participants of the musician group and once for all participants included in the non-musician group. Within each model comparison, Kendall's τ_a was used as correlation coefficient.

Mean distance and RDM correlations across trials

As differences between the groups could also potentially occur not only in the final RDM, but further in the strategy applied during the task, two approaches to evaluate RDM characteristics over the time course of the experiment were applied: mean distance and correlations. These additional analyses extend the overall categorization as observed by the final iMDS RDM with insights about the variance and stability of RDMs and thus categorization over time. Especially, if the groups differed with regard to those parameters.

Mean distance

In order to calculate the mean distance of a given RDM, the lower triangle (discarding the diagonal) of each trial corresponding RDM was extracted for all trials using nilearn's sym_matrix_to_vec and the mean of the resulting vector computed through NumPy's mean function. Trial corresponding mean distances were then concatenated within and subsequently across participants. Group corresponding mean distances of all participants across trials were then plotted using a lineplot as implemented in seaborn, which provides details on how dissimilar the music genres were categorized in general across the duration of the experiment within each group.

RDM correlations

Using SciPy's implementation of Spearman's r, the correlation of succeeding RDMs was computed. As for the computation of the mean distance, only the lower triangle of the RDMs was used for correlations. After within and across participant concatenation, results per group and across trials were again visualized via a lineplot using seaborn functionality. The thus yielded information give rise to the stability of RDMs and therefore the arrangement and categorization of the different music genres, as well as how this might vary between the two groups.

Machine learning approaches

Supervised learning

To test if the obtained RDMs hold any predictive ability with regard to musicianship, they were utilized within a supervised learning approach. Therefore, the lower triangle of each participant's final iMDS RDM was extract as described above and labeled, according to the respective participant's group, as either belonging to a musician or non-musician. The scikit-learn implementation of a linear SVM was then used to try to predict musicianship based on RDM values. Training and testing of the classifier was conducted within a 5-fold stratified Kfold CV across participants, leaving 80 % (16 participants, 8 per group) as training and 20 % (4 participants, 2 per group) as test set. A permutation test with 10000 permutations, randomly shuffling group labels, was run in order to test the classifier's significance by comparing the initial accuracy score against the distribution of accuracy scores based on shuffled labels. The outcome of this analysis was visualized using Matplotlib.

Hierarchical clustering

Along the line of thought of the supervised learning approach, a supplementary analysis applying an unsupervised learning method to the data was used to investigate if the RDM values, without any corresponding group labels, would allow for a segmentation, respectively allocation into different groups. To this end, a hierarchical clustering of all participant specific RDM values was computed using the Ward variance minimization algorithm within SciPy's linkage function. The obtained clustering was then visualized as a dendrogram through SciPy's dendrogram.

5.12.2. RESULTS

In the following, the results of the third part of this project, which aimed to investigate potential differences in the categorization between musicians and non-musicians, will be presented. After the description of the obtained behavioral RDMs and their relationship between the included participants and groups, as well as with the tested model RDMs, details of the outcomes of the different group comparison approaches will be given.

Behavior

As described in 5.7, an iMDS based paradigm was used to obtain behavioral representations of cortical processes related to the categorization of music genre, in order to investigate if these respective processes differ between groups of musicians and non-musicians, thus if musical training has an influence on the perception of music genres. All participants conducted the identical paradigm which resulted in one RDM for each participant. The RDMs were subsequently averaged across participants but within the respective groups. The thereby yielded group corresponding high dimensional RDMs and their low dimensional MDS visualization are illustrated in Figure 5.21. The obtained behavioral representations are highly similar between the groups and exhibit the same pattern that was also found in their counterparts of the other parts of this project, that is fMRI and EEG. More precisely, a prominent clustering of main genres was observed that was characterized through large distances between main genres and small distances between subgenres of a given main genre. This was most prominent for rock and classic, with the latter displaying the biggest difference to all other genres. Furthermore, this pattern deteriorated towards electronic music and afro-american, as they revealed smaller distances to each other and greater distinctions between their respective subgenres. However, upon a closer look, it were also these factors that differed between the two groups.

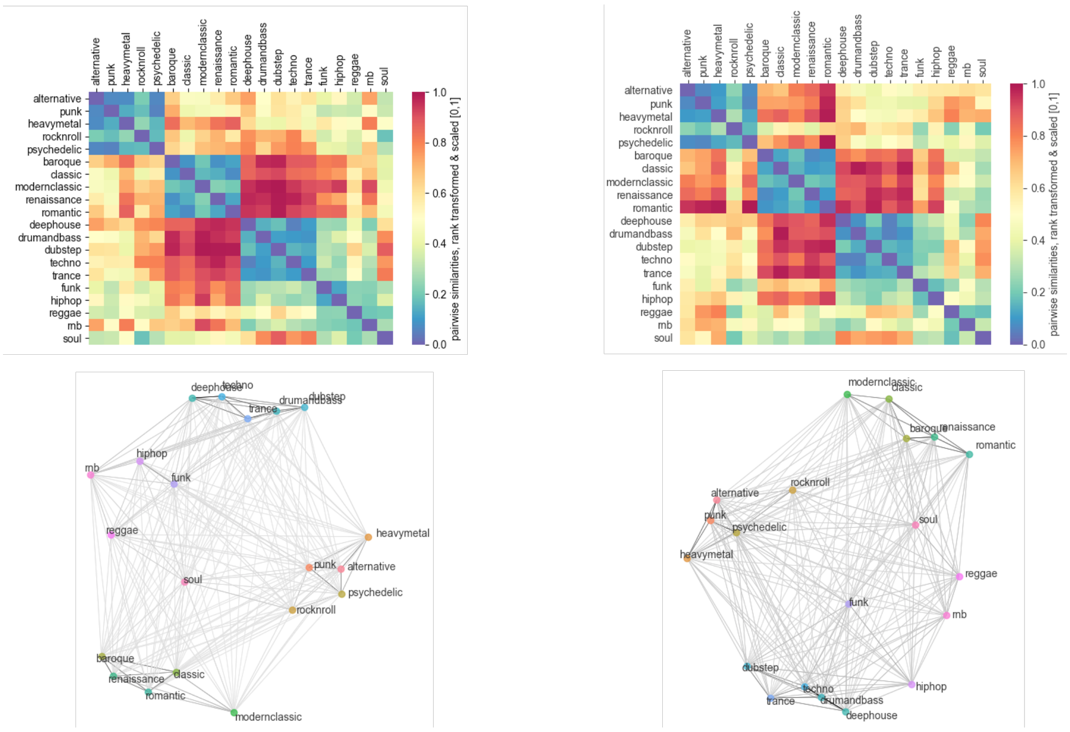


Figure 5.21: *Averaged behavioral RDMs (top row) for musicians (left) and non-musicians (right), as well as their low dimensional visualization via MDS (bottom row). The RDM is indexed by the amount and category of music genre ordering subgenres within main genres and main genres along the following order: rock, classic, electronic music and afro-american. The respective subgenre are identified through ticks on both x and y axis. Colors in the RDM indicate dissimilarities where blueish colors reflect very small or no distance and reddish colors extremely large distances. Values range from 0 (no distance) to 1 (extreme distance) after rank transforming and scaling the distance values obtained from the behavioral paradigm. Notably, these transformations were only applied for visualization purposes, thus the underlying distances are not altered. The MDS arrangement includes one point in space for each music genre which are labeled accordingly. subgenres of one main genre are identified through a similar color.*

While both the musicians' group RDM (mgRDM) and non-musicians' group RDM (nmgRDM) displayed the largest distance of main and corresponding subgenres between classic and electronic music, the distinction by means of larger distances between classic and rock was more pronounced in the nmgRDM. Thus non-musicians seemingly perceived rock and classic as more miscellaneous than musicians. Focusing the vanishing clustering towards electronic music and afro-american, the former was equally precise defined, whereas the latter was delimited to a greater extent in the mgRDM. In contrast, the nmgRDM showed larger distances between the subgenres included in afro-american and smaller distances to those of other genres, especially electronic music. It therefore appears that non-musicians perceived afro-american as more divers in itself and less distinctive towards other main genres as musicians. Another difference between groups became evident in the comparison of rock with electronic music and afro-american, as well as classic and afro-american. Regarding the first, the mgRDM yielded a greater distinction between rock and electronic music than between rock and afro-american which was reversed for the nmgRDM. Concerning the second, the difference between classic and afro-american was more pronounced in musicians than in non-musicians. As for the other parts of this project, the corresponding MDS present an intuitive visualization as they stress the above described patterns even further. This is especially true for the broader perception of afro-american genres, as they are placed further apart in the non-musician MDS. Taken together, the obtained representation is highly comparable between groups, yet displays fine-grained differences. In order to investigate if the relationship between participant corresponding RDMs already enables a differentiation between groups, the RDM of each participant was correlated with the RDMs of all other tested individuals, independent of their group status. The resulting values based on Spearman's r were used to indicate the dissimilarity between all participants' representation within a respective RDM, which is depicted in Figure 5.22.

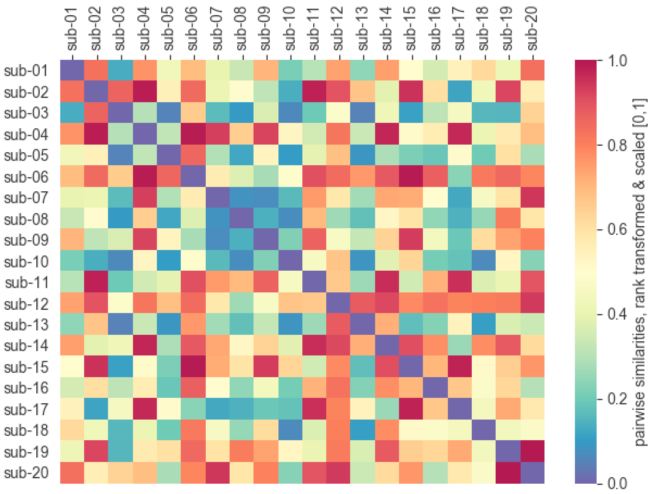


Figure 5.22: *RDM indicating the dissimilarity between the representations of all tested individuals. The RDM is indexed by participant labels, with sub-01 till sub-10 reflecting musicians and sub-11 till sub-20 reflecting non-musicians. Colors in the RDM indicate dissimilarities where bluish colors reflect very small or no distance and reddish colors extremely large distances. Values range from 0 (no distance) to 1 (extreme distance) after rank transforming and scaling the distance values obtained from the behavioral paradigm. Notably, these transformations were only applied for visualization purposes, thus the underlying distances are not altered.*

A clear distinction between groups by means of large distances between musicians and non-musicians and small distances within groups could not be observed, thus reflecting the highly similar representations as described in the preceding section. Nevertheless, the RDM provides group respective information which is related to the consistency of representations between individuals of a given group. It becomes evident that the distances between participants of the musician group are smaller than those of the non-musician group. They furthermore cluster slightly more together, especially sub-07 to sub-10, whereas non-musicians in contrast exhibit a more diverse pattern.

Behavioral representations & models

Aiming to examine if the behavioral representations of the different groups are best explained by different models, the model comparison approach (please see 5.5), including the same set of models, was conducted for each group respectively. The outcomes of this Kendall's τ_a correlation based comparison of model and target RDMs is illustrated in Figure 5.23. As expected due to high similarity between the behavioral representations, the model comparison yielded almost identical outcomes in both groups.

In more detail, a clear distinction with respect to the predictive ability of models could be observed between the higher level conceptual models and the low level acoustic feature models. The main genre and subgenre model explained the obtained behavioral representations best, with both of them reaching the noise ceiling, being significantly related to the behavioral target RDM and no performance difference between one another or groups. The acoustic feature models performed tremendously worse, as they all yielded drastically lower correlation values and failed to reach the noise ceiling, with pitch being the best of the spectral models in both groups, followed by timbre. Regarding the temporal models, tempogram yielded higher correlations in both groups than dynamic tempo. Together with chroma, it was among the only acoustic feature models to be significantly related to the behavioral target RDM. On a broader

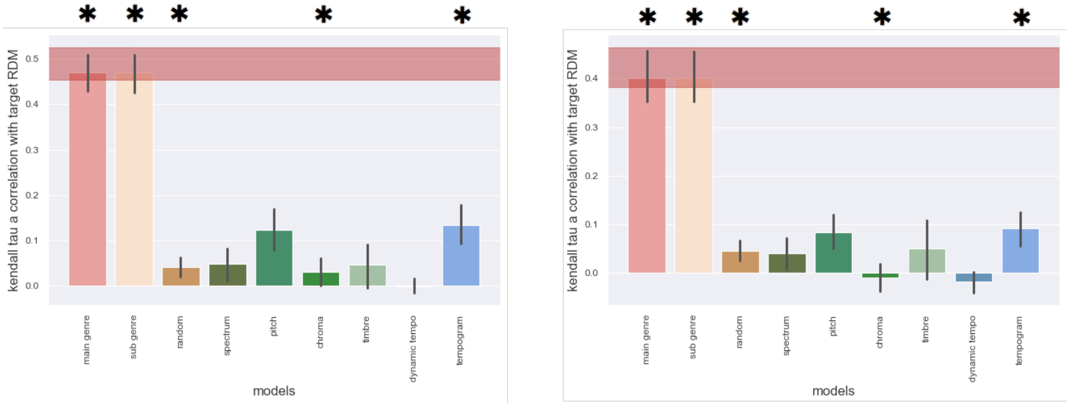


Figure 5.23: *Comparison of model and behavioral target RDMs for musicians (left) and non-musicians (right). The order of the model RDMs within the model comparison is identical across all bar plots and as follows: main genre model, subgenre model, random model, four spectral acoustic feature models (spectrum, pitch, chroma, timbre) and two temporal acoustic feature models (dynamic tempo and tempogram). Black lines within bars represented error bars, while asterisks indicate that a respective model's correlation was significant at $p < 0.05$ (FDR). Red rectangles reflect the compute noise ceiling for each target RDM.*

scale, models seem to display a slight advantage in the musicians group, as correlation values were marginally higher. Reflecting the greater consistency of representations between individuals in the musicians' group, their noise ceiling was more narrow and peaked at a higher value as compared to non-musicians. Please see APPENDIX C3 for tables containing the respective statistical values of the model comparisons.

Group comparison approaches

To enable the investigation if not only the final outcome, that is the behavioral representation but also the strategy that lead to it differs between groups, furthermore the characteristics of the RDMs across trials in the behavioral paradigm were compared. To this end, their stability, i.e. correlation between RDMs of adjacent trials and the mean distance of RDMs across trials were computed. Figure 5.24 displays the results of both aforementioned analyses. Focusing stability first, the RDMs of all trials were highly correlated to one another, not only between neighboring trials and independent of the group. Representations of musicians were slightly more stable, especially in earlier trials. which was however only marginal given the range of the correlation values from 1 to 0.98 (Spearman's r). Comparably, the mean distance of RDMs increased in a similar extent in both groups. Albeit the mean distance was always greater and less variable in the non-musician than the musician group, an interpretation is limited based on the small scale of mean distance change and difference across trials. Further testing if the obtained behavioral representations hold a group predictive value themselves, two different machine learning approaches were conducted using the pairwise similarities, one supervised learning and one unsupervised learning method. The first utilized an SVM and included a permutation test to assess the significance of the outcome. As visible in Figure 5.25, the groups could not be predicted based on the behavioral presentations as the obtained classification score was identical to the chance level of 0.5 % and yielded a p-value of 0.579. The second applied a hierarchical clustering that was subsequently visualized as a dendrogram in Figure 5.26. Here also no difference between the two groups were evident, as the resulting clusters were always mixed and thus included individuals of both groups. Taken the aforementioned points together, neither the strategy nor the final behavioral presentation seems to be diverging between musicians and non-musicians.

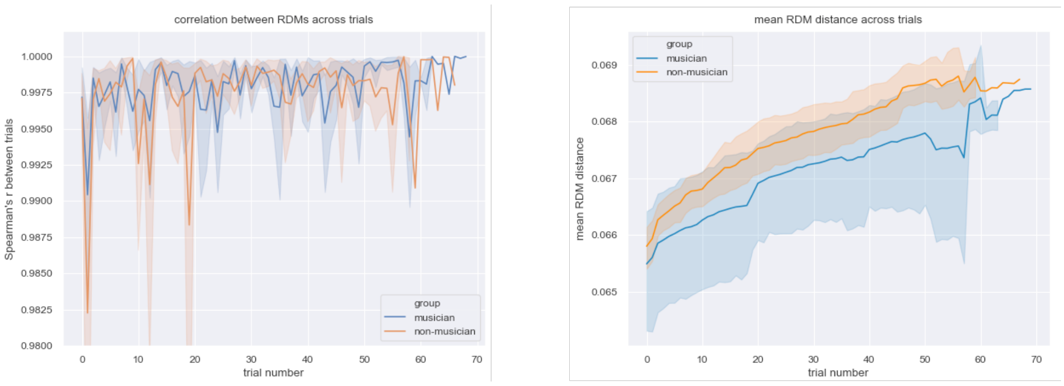


Figure 5.24: *RDM correlations (left) and mean distance (right) across trials for musicians (blue) and non-musicians (orange). Solid lines indicated mean and shades standard deviation across participants. Note that due to a varying number of total trials per participant shades diminish and disappear after around 60 trials.*

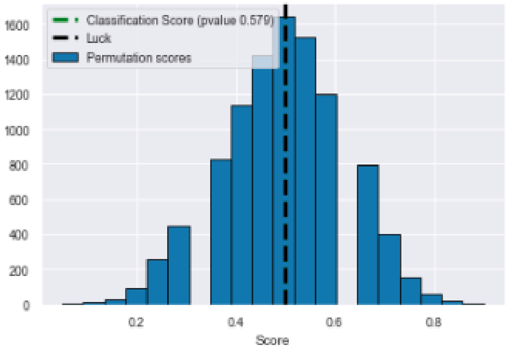


Figure 5.25: *Results of the SVM permutation test, indicating the obtained classification score (green dotted line), chance level (black dotted line) and distribution of the permutation scores (blue rectangles).*

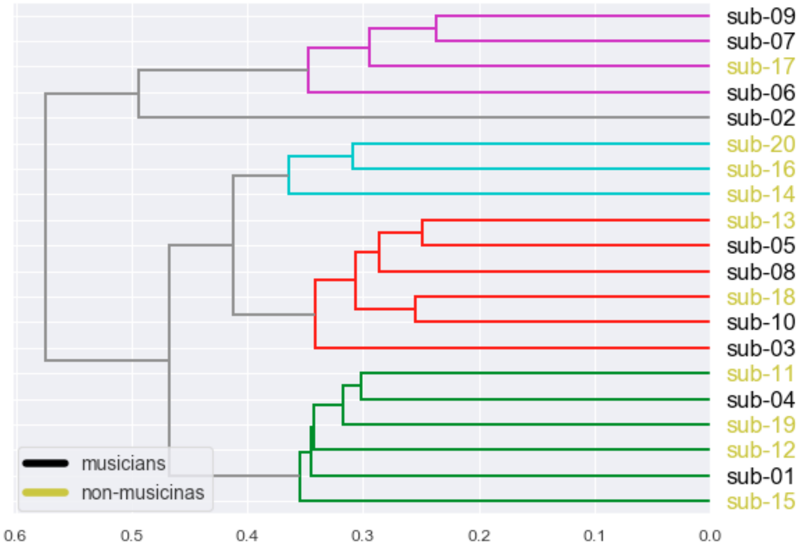


Figure 5.26: *Dendrogram illustrating the results of the applied unsupervised learning approach, i.e. hierarchical clustering. The y axis displays participants (that is their RDM), with the color reflecting the two groups: musicians (black) and non-musicians (yellow). The x axis indicates the distance between participants and their clusters. The later are reflected by the different colors of the dendrogram lines.*

5.13. DISCUSSION

In the following, the results of the above described project, including all subparts will be discussed. Starting with the relationship between neuronal and behavioral representations, this chapter will proceed with the predictive ability of high and low level models. Finally, and after an assessment of representations across space and time, the potential influence of musical training on the categorization of musical genres will be evaluated and limitations, as well as possible extensions will be presented.

Through the framework of representational models, this part of the here presented thesis targeted the processing of the high dimensional and complex representation of music by means of its categorization, as well as how these representations could be explained by a broad range of models. It builds upon and extends prior research work through four aspects. Beginning with the creation of a theory driven stimulus set that was additionally validated by an independent group, the resulting music genres were utilized within a highly multimodal approach, that covered not only spatial, but also temporal neuronal correlates by means of fMRI and EEG. Furthermore, the model space of previous studies was extended to include low level, i.e. acoustic features and high level models, i.e. conceptual models. Notably, more spectral and temporal feature based low level models were added, along with a novel complex behavioral model. Finally, the framework was applied to different populations to provide first insights on the potential effect of musical training on music genre categorization. While the first part of the project focused on the categorization process within the light of the auditory cortex' dual pathway model to examine if the representation of music genre through the computation of a respective percept is achieved predominantly in the ventral pathway as suggested by prior research work, the second part aimed to provide novel insights from the temporal domain, that is categorization over time. The third project on the contrary was purely based on behavior and tested individuals with different levels of music training. In order to capture the complexity of music genre perception as sufficient as possible, the identical set of models that represented both bottom-up and top-down influences was used within the same analyses approach throughout all project parts.

Behavioral music representations as an estimate of mental categorization

Throughout all analyses of all project parts it became evident that the utilized iMDS based paradigm was able to obtain and provide a meaningful behavioral presentation of the mental categorization of music genres. Notably, the behavioral representations were highly similar across all project parts and apparently also across trials as revealed by the third project. Considering its performance within the model comparisons of part I and II, it provided the highest predictive ability of all tested models in the majority of brain regions (part I) and time windows (II), while additionally exhibiting the most reliable and stable performance. This is in contrast to the other models, as their correlation with the respective target RDMs varied rather drastically. Together, behavioral representations therefore seem to capture a combination and weighting of features that closely resembles those of the neuronal representations, thus pointing to the importance of high level models to describe sound, here music, categorization. More precisely it offers an attractive opportunity to evaluate the top-down influence of abstract mental representations on cortical music processing. This is in line with Casey (2017), who reported a high correlation between behavioral representations and neuronal representations in HG, STGp and STGa. However, several distinctions have to be noted. First, the respective study used a smaller and more broad stimulus set in an unspecified music genre categorization task, that appears to might have been a forced choice task based on the output being a confusion matrix, as used commonly for the assessment of behavioral representations (Kriegeskorte

& Mur, 2012). Conversely, the here presented study allowed a more fine-grained categorization due to the characteristics of the stimulus set, while also avoiding potential paradigm confounds, like strong context effects, through obtaining stimulus arrangements in a higher and more complex space that was additionally designed to equally weight each pairwise relationship. Second, hemispheres were included as a factor, which was absent in previous research work. For example, the anatomical ROIs from Casey (2017, 2012) were always bilateral, thus neglecting assumed hemispheric differences in the processing of music. Indeed, the here presented results are supporting a computation that diverges between hemispheres as a slight advantage for the behavioral model in the RH was observed and RH neuronal representations were better able to explain the observed behavioral categorization. In contrast to Casey (2017), correlations were always limited and rather fairy low, while also decreasing from core towards parabelt regions. This might be based on the differentiation between sub and main genre that was made in the here presented study, but absent in the former, as there behavioral and neuronal representations were compared on a broader level and a smaller amount of music genres. Thus, the tremendously increased complexity might have resulted in a diminished correspondence between behavioral and neuronal representations. However, this effect is seemingly more grounded in the neuronal data, as behavioral representations revealed a precise cluster. Additionally, Casey (2017) used Spearman's r to compare the modality specific representation, while the study at hand applied Kendall's A due to the inclusion of conceptual models that contained identical distances for multiple stimuli pairs, which based on its inherent assumptions could also have a certain influence on model performance (Nili et al., 2014). Another prominent difference is the task and paradigm structure. Still assuming Casey (2017) conducted a version of a forced choice task, genre information was already present in the task, which might have driven the participants' strategy and categorization itself, whereas no such task and information was given in the behavioral paradigm used here. While supporting the correspondence between brain and behavioral representations that might also influence one another, it has to be noted that this only holds true for neuronal representations obtained via fMRI, but not EEG. Here, its consistently good and high performance is motivated exclusively by the even worse correlations of the remaining models, which as furthermore supported by the basically nonexistent relation between all time window specific representations and the behavioral ones. As this project was the first to relate behavioral and electrophysiological categorization over time, no data basis exists, hence limiting assumptions and render them highly speculative in nature. However, one major confound could be reflected in the high temporal resolution of the EEG. Initially thought to be an advantage, the variation of music genre representations over time might pose a prominent contrast to the stability of the behavioral representation that is based on a longer time scale and thus more holistic percepts of music genres and their categorization. This is based on the categorization of music genres in EEG being based on an almost real time computation in response to the stimuli, thus evolving and always changing. Considering both, fMRI and EEG further advantages and disadvantages of the behavioral paradigm at hand are evident. As the behavioral representation was utilized as both high level feature model RDM and target RDM these parts of the project furthermore allow a sanity check. In more detail, the observed pattern of model comparisons where the neuronal representations acted as target and the behavioral representations as model RDMs were highly similar to model comparison results where their roles were switched. Given the averaging of model RDMs, that are subsequently compared to individual target RDMs, the thereby resulting variance and mean correlation thus provide a complementary point of view. One factor that negatively affects the comparison of brain and behavioral representations is the existence and absence of a task in the behavioral and imaging paradigm respectively. Whereas in the imaging paradigm participants were not given a task, the behavioral paradigm explicitly required an arrangement according to the perceived similarity of the stimuli, which might have triggered a certain strategy. Even though intended, this categorization of music genre might therefore be not present in the neuronal data which renders the influence of the task unresolvable and thus limits the capacity to compare thereby obtained representations. However, as demonstrated by other studies investigating the relationship between brain and behavioral representations (Casey, 2017; Casey et al., 2012; Kriegeskorte et al., 2008), a sufficient assimilation is nearly impossible, as a categorization task in imaging settings could only be implemented in a very restricted way and the absence of a task in behavioral settings is simply not possible. Furthermore, the utilized behavioral paradigm itself could evoke a certain extent of variance and confounds, as the only instruction that was given was to arrange the stimuli based on their perceived similarity. The rather unrestricted character of such an instruction might prevented a certain basis in strategy, which in turn can also be problematic as all participants might define similarity differently and thus or in general apply diverging strategies to arrange the stimuli. Beside assessing a given participant's strategy through respective questions after the paradigm, repeated measurements with different instructions could be conducted to allow further insights. For example, after an initial version with the same instruction as applied here, further versions that e.g., put an emphasis on certain features ("Please arrange the stimuli according to their timbre.") could be conducted and subsequently compared to the initial version. Going back to the model performance, besides in part II (EEG), the behavioral representations exhibited predictive abilities that were comparable to those of the main and subgenre models. As the behavioral representations were empirically assessed, but the conceptual models motivated through music genre descriptions that in turn were based on semantic features obtained through musicology and computer science, this points to high correspondence between both representations and the influence such concepts have on behavior or in other words how they guide it. The results of project III provide even more evidence for this effect, as here both informative conceptual models yielded by far the highest correlation with the behavioral target RDM, thus providing the best explanation for the observed behavioral categorization as an estimate of mental representations through concepts that are based on music genre definitions. The absent performance difference between the main and subgenre model might furthermore point to a level of categorization that is rather broad and most prominent between main genres, while also allowing a categorization into subgenres. The genre dependency of such categorization levels is for example displayed through the varying distances between subgenres of a given main genre, as observed for e.g., classic compared to afro-american. Taken the aforementioned points together, the obtained results furthermore stress the predictive ability and importance of behavioral representations in order to investigate mental categorization and its interplay with neuronal representations, especially to which extent they influence and drive each other along different processing stages.

Music representations and conceptual models

Overall, the conceptual models had two objectives. First, they aimed to provide high level models (main and subgenre) that entail a certain pattern of music genre categorization, that followed abstract semantic properties like genre definitions which in turn were based on and motivated by a combination of various features as suggested by musicological and computer science/signal processing research. Second, the included random model acted as a control model that assumed no categorization and therefore an absence of meaningful representations and information. Hence, their utilization as model RDMs and resulting diverging performance allowed an assessment if neuronal and behavioral representations can be considered meaningful or not and in case of the first, if they follow a certain principle. As noted above, the conceptual models main and subgenre were highly similar to the behavioral representation, in both their categorization pattern and thus also performance in the model comparisons, thus pointing to their feasibility and superiority in providing principles that guide music genre categorization. This especially holds true for project I and III, as both yielded the highest correlations in the majority of the comparisons. In project I their performance was comparable to that of the behavior model throughout all ROIs with no prominent difference between them. Providing a first motivation for mental categorization processes that seemingly follow certain genre definitions based on complex features spaces, the conceptual models therefore support the assumption that meaningful representations and information related to the perception of music genres are evident throughout the auditory cortex and the different processing stages and streams it includes. This is further underlined by the results of part III within which an even stronger distinction between the conceptual and the acoustic feature models became evident. In more detail, they not only exhibited by far the highest correlations, but also reached the noise ceiling. Thus they provide a sufficient explanation of the observed behavioral representation. As for all other project parts, no difference between the main and subgenre model could be observed. Given their properties, it appears that the behavioral categorization of music genres and the mental representation this might reflect are at least partially guided by and grounded in a pronounced differentiation between main genres, that is genres that differ on a larger and broader scale based on the expression of their underlying acoustic features. Here, certain further factors need to be considered. The subgenre model could be too similar to the main genre model in terms of their overall structure in order to yield a distinct performance, especially as a rank correlation approach was used for model comparisons. Respective subgenre models with increasingly larger distances between subgenres of a given main genre could help to reveal a gradient of categorization by showing at which abstract distance structure, thus differentiation between subgenres, the underlying principles fail to capture the cortical representation. This could provide insights in how fine-grained music is represented by means of how pronounced its genre are distinguished. However, in light of the obtained results, a more detailed representation of subgenres seems to be dependent on the music genre space and the participants' experience with regard to the included excerpts. This is especially motivated by the precise clustering of rock and classic and loose, more divers clustering of electronic music and afro-american in the behavior representations. While pointing to a potentially higher within main genre variability, this could also be based on the exposure participants had to these genres. Furthermore, the amount and extent of clustering is based on a summary statistic and could therefore change by including more or less genres. In order to tackle music genre specific representations, conceptual models that assume a different clustering for certain genres could be utilized, which would additionally allow to incorporate factors such as music experience and preference. Through the high representational similarity between conceptual models and the behavioral categorization, as well as their comparable predictive ability in part I, the aforementioned points can, at least to a certain extent, also be applied to neuronal representations. Given the slight performance differences between core, belt and parabelt on the one hand and anterior and posterior ROIs on the other, the conceptual models point to a cortical representation that differs as a function of these two factors and therefore potentially diverging processing strategies. Their fairly low correlations with the time window corresponding correlations in project II further support the assumption that music genre representations are rather stable, holistic and achieved in later processing stages. The computation of a stable percept over a longer time scale appears counterintuitive as humans can identify music genres in only a couple of seconds (Gjerdingen & Perrott, 2008), however with regard to this, a certain influence of the applied task that potentially alters the processing strategy might render a more complex representation and model as used here impractical. Focusing the random model, it is worth mentioning that even though it was significantly related to the target RDM in a multitude of comparisons, its performance as a control model was highly insightful. This based on the persistent difference found in its predictive ability as compared to the other conceptual models. Within all comparisons of part I and II, it performed drastically worse, thus providing evidence that the observed representations more likely follow a precise structure and categorization and therefore contain meaningful information about music genres, as a random pattern with no evident information. However, its in some cases high or significant correlation could be based on its randomness that captures certain aspects of the representation, as well as the applied model comparison approach and the included rank correlation. In summary, the utilized conceptual models provide evidence that the cortical representation of music genres by means of their categorization entails meaningful information that follows principles which are based on a combination of complex features and reflect a prominent division in main and only little differentiation in subgenres. While these principles seem to guide behavior, they do not appear to be time sensitive, but variable throughout the cortical auditory processing hierarchy.

Music representations and acoustic feature models

At the core of theories that propose a rather purely bottom-up approach that already starts in subcortical regions of the auditory system, acoustic features were long assumed to solely drive cortical processing and thus categorization. Albeit recent evidence for the contrary and the influence of top-down computations have been shown, the multidimensional character and complexity of acoustic features render an adequate investigation that is able to disentangle this interaction almost impossible. The here presented project aimed to tackle this problem in a first attempt via including models that capture both approaches. As noted in the preceding section, the higher level feature models that reflected a potential top-down influence performed reasonably well in all but the second part of the project. With the exception of timbre and tempogram, the contrary was observed for the low level acoustic feature models that represent bottom-up approaches. Their performance additionally differed drastically between the project parts. With regard to part I and in line with previous research that stressed the importance of timbre as a factor to describe and differentiate between sounds and therefore also categories, the timbre model showed a clear advantage in most comparisons. It was the only acoustic feature model to perform nearly as well as the behavioral and conceptual model and the only model to increase its predictive ability towards anterior regions. As timbre is defined through multiple other acoustic features, this points to two assumptions. First, also within the space of acoustic feature models, more complex models provide a better explanation for the observed representations than simpler ones. This is supported by Casey (2012) who also obtained the highest correlation values for their timbre model as compared to other spectral models. Also, the vast majority of research work that somehow included or directly tested timbre with regard to its neuronal correlates, reported more consistent and broader distributed patterns for the encoding of timbre than for other acoustic features (Allen et al., 2018; Moerel et al., 2013; Santoro et al., 2014; Santoro et al., 2017). Second, a potentially high dimensional and generalized representation of music, thus percept, is found in the anterior ROIs and computed along the ventral pathway. Building on top of the importance for music perception, prior studies proposed that especially the bilateral PP has a crucial role in the processing and representation of timbre that additionally seems to be music specific (Leaver & Rauschecker, 2010; Norman-Haignere et al., 2015; Whitehead & Armony, 2018). The tempogram model also yielded higher predictive abilities as the remaining models, but in no case outperformed timbre. This might be based on timbre including feature components that are partially also entailed in the tempogram feature. Thus timbre potentially also included aspects of tempogram and therefore reached overall higher correlations. Furthermore, even though timbre comprises spectral and temporal features, it is more based on spectral components. Considering the low temporal resolution of fMRI, the respective measurements might be more sensitive to spectral than temporal components. Even though, a difference with regard to feature sensitivity cannot be ruled out, it might not account for the observed differentiation, as for example the tempogram model yielded higher correlations towards posterior ROIs, which supports the preferential encoding and representation of temporal features in these regions, especially the PT, that was found and argued for in prior studies (Timothy D. Griffiths & Warren, 2002). In combination with the model performance of the timbre representation, this furthermore points to a diverging specialization of the ventral and dorsal stream in that both are involved in the categorization of music genres based on timbral features, with the dorsal stream additionally incorporating temporal information. In contrast to other studies that outlined the importance of the remaining acoustic features, especially the spectral ones (Casey et al., 2012), their observed predictive ability in the here presented study was rather limited or non-existent. While this further supports the advantage of more complex models it also reveals differences that can be attributed to the analysis approach or stimulus material. A vast amount of studies that used complex natural stimuli applied methods that model stimuli response on the voxel level, e.g. voxel-wise encoding models (Allen et al., 2018; Santoro et al., 2014; Santoro et al., 2017) or decomposition methods (Norman-Haignere et al., 2015). A comparable fine-grained tuning was not assessed in the here presented study, as a given stimulus response was assessed via the multivoxel pattern it evoked in different ROIs and evaluated using RDMs as a summary statistic. Another substantial part of studies included stimuli that were specifically created to display systematic differences with respect to certain stimuli (Timothy D. Griffiths & Warren, 2002; Leaver & Rauschecker, 2010) and thus investigated the respective feature more isolated. Here, the emphasis was put on the combination of acoustic features necessary for the processing and possibly categorization of music genres. Thus they were all present and exhibited their natural spectrotemporal modulation which resulted in their perception as part of a complex excerpt and thus percept, that in turn might prohibited a subsequent disentanglement through only a certain component of its feature space. Or in other words, single low level features like pitch are not able to explain representations that are based on a combination of multiple features. In order to provide more insights, computed acoustic feature models could be combined in a linear and non-linear fashion in order to investigate if a certain weighted combination of acoustic features would be able to explain a respective representation or more generally, if the predictive ability of acoustic feature models increases with the amount of included features. This could also account for the observed model comparison pattern in part III, as there the entirety of acoustic feature models, including the complex models timbre and tempogram, performed tremendously worse than the conceptual models. As already addressed in the preceding chapter, this might reflect the combination of features that is present in an abstract level in the conceptual models, as they are largely based on semantic information that in turn relies on the expression of acoustic features. Another factor that could have drastically influenced the obtained results is the way the acoustic features were extracted. Albeit the acoustic features, as well as the applied music information retrieval methods that were used to extract them, are widely accepted and heavily used, the underlying parameters, that is the type of fourier transformation or the window length of segments, define the resulting feature vector. Continuing with the usage of summary statistics that rely on the similarity between these feature vectors, a single small change of the parameters during the extraction of a given feature could result in a complete alteration of the representation. The same holds true for the similarity measure used to compute acoustic feature model RDMs, as each method will result in a different representation. This could also be present in the here presented thesis, as the acoustic feature model RDMs were based on the euclidean distance between the respective feature vectors of the music excerpts, whereas e.g., Casey et al. (2012), Casey (2017) and Güçlü et al. (2016) utilized Spearman's r. The feature extraction method could furthermore be responsible for the low correlations of the acoustic feature models with the time window specific representations. While the latter capture the categorization within a certain time frame and thus possibly processing stage, the former lose their temporal precision, as features are compared within segments and then averaged over time to obtain one distance value for the entire stimulus duration. The thus holistic representation might not be sufficient to explain the evolution of the representation over the time course of the stimuli. In sum, low level acoustic features provided very restricted predictive abilities in all project parts. They nevertheless enabled insights, especially in the neuronal representations obtained through fMRI, as here complex features that entailed a combination of more simple features performed best throughout all ROIs and additionally revealed a distinction between the ventral and dorsal pathway. More precisely, complex spectral features (timbre) yielded higher correlations in the anterior belt and parabelt ROIs while the opposite was observed for complex temporal features. The results therefore render the bottom-up role of independent single acoustic features in the computation of categories and therefore music representations questionable.

Music representation across space and time

Aiming to provide detailed insights into the perception and representation of music genres by means of their categorization, the study at hand applied a highly multimodal approach to capture the respective cognitive processes across multiple stages and as precise as possible. This involved fMRI for spatial representations along the two auditory pathways (part I) and EEG for temporal categorization over the time course of the stimuli (part II). Additionally, high level models that represent top-down and low level models that reflect bottom-up influences and thus potential principles that guide the computation of the respective representations were evaluated. Overall, both representations and model comparison results differed drastically between the project parts. While fMRI yielded stable music genre representations that varied as a function of hemisphere, processing hierarchy and pathway, the EEG results provided only limited insights based on instable representations. However, neither of both revealed a categorization that somewhat follows genre borders. As outlined in the preceding sections, throughout space and time, the behavioral model, which reflected a behavioral correlate of the mental representation (Kriegeskorte & Mur, 2012), explained the observed neuronal representations best, followed by the conceptual models that were based on semantic features (i.e., genre descriptions). Together with a general low model performance of all acoustic feature models, this points to a clear advantage of complex models. Focusing fMRI, the obtained representations were highly similar across all investigated regions, which is in line with the results of Casey (2017) who also showed comparable categorization patterns but in slightly different regions that were not separated by hemisphere and utilizing a different stimulus set and task, as well as analysis approach. While the comparison between the two studies is limited due to the aforementioned difference, the observed correspondence between ROIs in both might point to a general and abstract principle of music representation that is best captured through multidimensional spaces (Casey, 2017). The here applied framework in combination with a more fine-grained spatial investigation could however reveal further characteristics. This is especially grounded in the latter, as the investigated regions reflected different hierarchical processing stages, pathways and hemispheres, whereas previous research work focused on rather broad regions that rendered subtleties unresolvable. Considering the different processing stages that were assumed to run from core towards belt and parabelt regions, a gradient of representations that runs along the same axis was observed. More precisely, representations changed from core over belt to parabelt regions, with the most prominent difference being present in the comparison between core and belt regions at the one hand and parabelt regions on the other. This effect was furthermore modulated by the respective pathway, as the representation diverged more towards belt and parabelt regions within the ventral pathway, that is PP and STGa and additionally stressed by the difference between the PP and the remaining representations of the core/belt regions cluster. Additionally, the predictive ability of nearly all tested models decreased along the aforementioned gradient but was more prominent within the ventral pathway. Here the acoustic feature model timbre was an exception to that, as it was the only low level model to yield higher correlations than the high level models, which was yet again most pronounced in the ventral pathway. Together these observations provide support for a variety of proposed auditory cortex functionality and organization with regard to the perception and categorization of music. First, it appears that the entire auditory cortex, although to a variable extent, is involved in the perception and categorization, thus representation of music through complex computations (Formisano, Martino, et al., 2008; Patterson et al., 2002; Staeren et al., 2009). This also entails the dorsal pathway and its regions (Timothy D. Griffiths & Warren, 2002; Güçlü et al., 2016), as their representations and model comparisons not differed tremendously with respect to those of the core (Kilian-Hütten et al., 2011) and ventral pathway regions. Second, the ventral pathway apparently exhibits a special role in the representation of music as suggested by numerous other studies (Angulo-Perkins et al., 2014; Armony, Aubé, Angulo-Perkins, Peretz, & Concha, 2015; Aube et al., 2015; Fedorenko et al., 2012; Leaver & Rauschecker, 2010; Norman-Haignere et al., 2015; Rogalsky et al., 2011; Whitehead & Armony, 2018), as here representations differed to a greater extent and the predictive ability of all models, but timbre diminished. In line with prior research work pointing to the importance of timbre in sound and music perception (Allen et al., 2017; Leaver & Rauschecker, 2010), as well as its seemingly prominent encoding an anterior regions of the auditory cortex (Leaver & Rauschecker, 2010), this suggests that the ventral pathway might lead to an abstract music representation, that cannot be explained by low or high level features alone. Third, by incorporating hemispheres as a factor, the representation of music, even though highly bilateral, seems to have a slight advantage in the RH, as throughout the entire processing hierarchy and both pathways the aforementioned effects were more pronounced in the RH than the LH, thus supporting previous research work (Angulo-Perkins et al., 2014; Armony et al., 2015; Fedorenko et al., 2012; Leaver & Rauschecker, 2010; Norman-Haignere et al., 2015; Rogalsky et al., 2011; Whitehead & Armony, 2018). It is furthermore worth mentioning that the above outlined characteristics and the resulting music representations are seemingly evident and shared between individuals, as the noise ceilings of the behavioral representations were rather narrow, thus exhibiting no large variance between participants. While the same holds true for the neuronal representations obtained via fMRI, the stability of representations marks the difference between EEG and all other tested modalities. In more detail, stability refers to two factors: across participants and over time. All investigated time windows yielded large noise ceilings that indicate a tremendous variation between the representations of individuals. Additionally, representations varied more across time windows, as for example those of ROIs in fMRI. By further integrating the performance of the models, a possible common cause for both becomes evident. Even though the perception of music as a dynamic signal is a time sensitive process, the corresponding representation might be a more holistic concept that is more abstract and grounded in longer time scales. The perception of individuals might hence be as dynamic as music itself and diverging throughout time, but yet converging over longer time scales. Further support is provided by the fact that the percept or the modelled response in both behavior and fMRI is based on the complete stimulus duration, thus over a longer time scale that reflects a more holistic representation and conceptual models, that entail a complex and abstract representation, performing insufficiently only in the EEG part of the project. However, another very likely source of the observed EEG patterns could be inherent to the method. So far, RSA was mainly used in the visual domain, especially pictures (Cichy et al., 2014; Wardle et al., 2016) and thus static images or if in the auditory domain applied to very short sounds of less than a second (Su et al., 2014). In both cases, the percept is rather fixed and stable and if changing, only within a small time scale. As this was the first study to utilize RSA on electrophysiological data which entailed responses to longer auditory stimuli, no empirical point of comparison exists. Therefore, the applied methods, that is time frequency and time windows could also be not suited for the intended investigation. In sum, the cortical representation of music and its genres reflect rather stable and holistic concepts and is achieved through complex computations within the entire auditory cortex that are most specialized along the ventral pathway.

The influence of musicianship on music representations

Given the immense plasticity musical training can evoke in the human brain, by means of e.g., feature processing and its resulting behavioral correlates, the third part of the project aimed to unravel if a respective effect is also present in the categorization of music into genres and thus its representation. To this end, musicians and non-musicians conducted the same behavioral paradigm with the corresponding results subsequently being compared between groups through various approaches. This entailed the examination if differences in models could account for potential contrasts, as well as a set of analyses that focused on the stability of representations over time and machine learning approaches. Even though musicians and non-musicians exhibited visually different representations, none of the applied group comparison approaches yielded significant differences. This points to a computation and existence of representations that are comparable between musicians and non-musicians and hence the absent influence of musical training on its categorization and representation. However, considering the processes that were shown to be influenced by musical training and the here utilized task, potential explanations for the observed absence of group differences become evident. The task that is evoked through the behavioral paradigm, that is an arrangement and thus categorization of stimuli might be too abstract to reveal any differences between the two groups. This is based on difference between musicians and non-musicians being mostly described by means of certain features, but not complex naturalistic musical excerpts. Most prominently, this includes the perception and processing of melodies (Koelsch, Schröger, & Tervaniemi, 1999; Schellenberg & Moreno, 2010), pitch (C. Marie et al., 2012; Micheyl et al., 2006; Parbery-Clark et al., 2009; Schellenberg & Moreno, 2010; Strait et al., 2010; Tervaniemi et al., 2005), tempo (Jeon & Fricke, 1997; Rammsayer & Altenmüller, 2006) and timbre (Chartrand & Belin, 2006). The task of the here presented study was however not designed to evoke a focus on a certain feature (albeit this could have happened, depending on a given participant's strategy), but to perceive and arrangement a given music genre as a whole. Thus it might have included a combination of all of the aforementioned feature specific tasks. Furthermore, there exists, so far, no evidence that a potentially altered processing of certain features influences the categorization and therefore representation. In turn, musicians might have processed the presented music excerpts differently, but categorized them in a comparable manner. This is further supported by the outcomes of the model comparison approach, as here models yielded similar correlation values in both groups and displayed a clear advantage for the conceptual models. Their representation is therefore apparently more driven by higher level and complex features, that might include a multitude of information of the low level acoustic features, as all acoustic features, also the rather complex ones like timbre and tempogram, provided a low predictive ability. Notably, the acoustic feature models that reached the highest correlations, that is pitch, timbre and tempogram, include nearly all of the aforementioned features which processing is proposed to differ between musicians and non-musicians. Hence, that might have been equally important for both groups. A possible approach to disentangle altered feature processing and its potential influence on categorization could be to apply a repeated-measurements design within which participants arrange the music excerpts based on certain stimuli (pitch, timbre, tempo, etc.) and evaluate if the resulting representations differ between groups. Other descriptions of distinctions between musicians and non-musicians famously indicate an increased memory capacity for musicians due to their training (R. M. Brown, Zatorre, & Penhune, 2015; Chaffin & Logan, 2006; Logan, 2000; Peretz & Zatorre, 2005; Strait & Kraus, 2011; Strait et al., 2010). Aiming to investigate if different memory capacities could have an influence on the task and the respective outcomes, the stability of categorization across trials was computed. As no meaningful differentiation was observed, it appears unlikely that e.g., musicians remembered the music excerpts and their arrangement better and therefore produced more stable representations. However, as for the altered feature processing, this difference could simply not manifest itself within the utilized task. In contrast to the above discussed points, the results of the model comparison and comparison of representations across individuals hinted to differences between the groups. In more detail, the noise ceiling of the musician group was more narrow and reached a higher value, thus the representation was more coherent between individuals. Hence, musicians might more often use a comparable strategy, which nevertheless has to be evaluated in further studies with larger sample sizes. Finally, as the most prominent differences were visible for certain genres only, the genre and thus comparison space could be adapted, via including more, less and other genre or restrict the comparison to a subset of genre representations. Taken together, there seems to be no influence of musical training on the perception, computation and representation of music genres, as both musicians and non-musicians apparently apply the same set of high level models, that is top-down principles.

Music representations - more than the sum of its features?

As noted in the introduction: music is at least as diverse as its many definitions. By applying a multimodal approach in combination with a multitude of models, the here presented study targeted to tackle this complexity in order to provide new insights into the representation of music and its genres by means of their categorization. At the core of this investigating were the computations and transformations the human brain might apply to obtain a specialized percept and how these processes can be explained by models that cover bottom-up and top-down factors, as well as if they are altered through musical training. Taken the outcomes of all project parts together, it appears that music categorization is a cognitive process that varies highly over the time course of a given music excerpt and is distributed across the entire auditory cortex, thus covering assumed early and late processing stages along the cortical hierarchy of the auditory system. Notably, these representations are guided by high level feature models that entail mental representations as reflected in behavior and conceptual representations based on semantic and descriptive features, whereas low level models that are based on acoustic features do not seem to be heavily involved. Here, timbre and tempogram as higher dimensional acoustic features pose as exceptions, which further underlines both, their importance in the perception and categorization of music genres, as well as the necessity of multidimensional concepts to capture at least a small portion of music's complexity. Being seemingly guided by a holistic and stable mental representation that follows main and to a lesser extent, subgenre distinctions which in turn might be based on music experience and enculturation, the categorization and representation of music is achieved through a slightly right lateralized network that covers the ventral and dorsal auditory pathway. Within that, the regions of the ventral pathway apparently lead to a general representation of music that is different from all others and cannot be explained sufficiently by low or high level features and concepts. The representations of music might therefore follow a prototype theory in which genres are defined based on concepts and the combination of underlying features that constitute them. Yet to be discovered computations and transformations, in terms of an interaction of bottom-up and top-down processes then lead to the categorization and subsequently stable representations of music and its genres. Based on that and in the light of the absence of alteration through musical training, it points to the potentially prominent role of music enculturation.

Limitations and outlook

The complexity of music is both one of its characteristics and reason for the comparably slow advancement of its understanding. Within the light of the here presented study, two main limitations have to be addressed and included. First, stimuli and paradigm and second, analysis approaches. First and foremost, as the representations are completely based on the included music genres and excerpts, due to the use of summary statistics in RSA, the entire study and its outcomes is dependent on these factors. This affects all levels of the stimulus set, as changing a main genre, a subgenre within a main genre or yet the representative excerpt of a subgenre could potentially lead to drastically different results. This is stressed by the results of the behavioral paradigm which revealed systematic differences between genres. Thus further studies should include diverging sets of music genres in order to test to which degree the here obtained results are reliable and generalizable. Next, in contrast to previous research work, the here utilized stimulus set exclusively comprised instrumental music. While this was intended to evoke a rather feature based processing with spectrotemporal modulations typically present in music, it could also have distorted the results, as most of contemporary music genres contain vocals as a defining factor. Thus, some excerpts were inherently more representative than others. For example, classic and electronic music are vastly non-vocal whereas the opposite is true for afro-american. As the importance of vocals and potentially lyrics cannot be ruled out as factors that are included in the generation of the mentioned concepts, further research work that tackles systematic differences between vocal and non-vocal music is necessary. Considering the paradigm, the absence of a task in the imaging parts has to be evaluated with respect to two aspects. The first refers to the influence certain tasks can have on auditory processing and thus representations. Here more insights on the independence of different tasks on music genre representations are needed, which also have to include an assessment if a representation can be willingly altered. The second refers to the absence and presence of a task in the imaging and behavioral paradigm respectively, as this difference could introduce prominent confounds when comparing both representations, also with regard to other models. Focusing on analysis related problems, the number and complexity of ROIs in the fMRI part need to be considered. While utilizing a set of ROIs that was more fine-grained and allowed the assessment of hierarchical, pathway and hemisphere specific processing, it could have been both too broad to unravel subtle effects and too narrow to reveal larger differences, as the representation is grounded in a multidimensional space that is spanned by the number of voxels. Comparably, the applied EEG approach appears to be problematic. Here, analysis types other than time frequency based ones should be explored within the generation of RDMs. Also EEG, in contrast to fMRI, did not include any kind of feature selection as the signal of all electrodes was used. Furthermore, other time segments, larger and smaller need to be investigated. Even though the model space was extended as compared to previous studies, it was still rather limited. A respective extension should include more low and high level features, as well as combinations of them. This especially refers to deep or convolutional neural networks which have been shown to capture cognitive processes incredibly well (Güçlü et al., 2016). Thus they could be used and adapted to resemble music categorizations and representations, in turn helping to gain further insights into the involved computational principles. Additionally, the multimodal character of the here presented study is far from being sufficiently used, as the representational model framework contains a tremendous number of possibilities for data integration. One example would be cross-modal searchlights that for example could provide information on where and when behavioral representations can explain their neuronal counterparts best. This also includes fMRI informed EEG and vice versa, that could help to shed light on sequential processing. Within the light of part III, the investigation of music genre representations needs to be extended to different populations. Not only through musicians with different training, but especially through individuals with varying music enculturation, music preferences and listening habits, as all these factors might contribute to and thus alter the mental concepts, that is prototypes, of music and its genres.

6

GENERAL DISCUSSION

The last chapter of this thesis contitutes a general discussion within which, after a summary of the findings, the results of the respective projects will be integrated and evaluated with regard to the general aim of the here presented work: the investigation of music representations along the processing hierarchy of the cortical auditory pathway. To this end, the first part of this chapter will focus on methodological considerations necessary to address in music processing research as a section of auditory neuroscience. Subsequently, representations of music as asssessed throughout the different projects of this thesis will be evaluated. The chapter's final part will outline important limitations of the conducted research work and propose possible ways to address them in further studies. All of the above remarks will be situated within the literature and previous research work on the neuroscience of music.

The aim of the here presented thesis was to provide further and novel insights of how perceived sound waves are transformed to specialized and stable representations of music along the hierarchy of cortical auditory processing and how bottom-up and top-down influences affect or alter these computations. To this end, three projects were conducted.

The first two projects were concerned with common problems found in auditory neuroscience, that is aggravated measurements based on MRI settings and the high variability of the human auditory cortex, more precisely its localization and parcellation. As these two factors, together with the absence of data and analysis sharing have to be held responsible for a large portion of the drastic inconsistency that is found across studies and thus the slow advancement, two open-source toolboxes were developed in order to tackle these problems. The first implemented a possibility to assess hearing thresholds and abilities in MRI settings, that furthermore allows the evaluation of possible confounds. The second generated a framework for the robust and reliable localization of the human auditory cortex through divers functional and structural approaches. The outcome of both studies revealed the strong influence different MRI settings can have on hearing thresholds and abilities which in turn pose as an immense confound in studies interested in auditory processing, as well as the tremendous inconsistency of the human auditory cortex' structure and organization due to a myriad of analysis approaches, that leads to diverging definitions of its regions which might render the results of different studies only marginally comparable. Together, they furthermore stress the need and necessity of standardization and automation, that in combination with resource sharing pave the way to find a common ground and further enhance methodological aspects in auditory neuroscience. Additionally, the second toolbox was used the define a working model of the auditory cortex and its regions that was used throughout the subsequent parts of this thesis, thus allowing an in depth and versatile characterization of the auditory cortex, that is its functionality with regard to the generation of cortical music representations.

The organization and function of the cortical auditory system with regard to the processing of different auditory categories was investigated in the third project. More precisely, it targeted to unravel if the cortical auditory system is, to a certain extent, specialized for music, or in other words if music evokes a representation that is distinct and unique. To this end, an fMRI study was conducted in which participants were present with music, speech and singing in order to create a gradient of musicality that first, extended the previous rather dichotomous comparison of music and other auditory categories and second, allowed to disentangle music specific components, independent of instrumental or sung. Additionally, the utilized stimuli were controlled for low level acoustic features to probe theories that attribute those features the most important role in driving differences in cortical processing. Aiming to provide a detailed characterization, two complementary analysis approaches were applied: spatially highly sensitive multivariate pattern analyses and computational connectivity analysis. While the first tackled the segregation of auditory cortex ROIs through examining where in the auditory cortex different patterns are evoked by the categories, the second aimed to investigate how these ROIs are integrated with one another and how their respective network structure might change in response to the categories, therefore leading to the aforementioned patterns. The obtained results point to the tremendous specificity of the auditory cortex with respect to the processing of music. Even though resulting a bilateral network, music predominantly modulated the connections between auditory cortex ROIs of the RH, with the intrinsic connectivity of the RH HG seemingly posing as a hub that is heavily involved in distributing information in the network and generating the former through the specific recruitment of ROIs necessary to process a given auditory category. Additionally, connections to anterior regions were more important than those to posterior regions, despite both being modulated. The special role of anterior regions was further supported by the searchlight results that revealed that regions within the PP and STGa distinguish between music and speech, regardless of instrumental or sung, but not between instrumental music and singing. It thus appears that the auditory system generates a highly specialized processing of music through modulated connections of its regions in the RH, that becomes more precise towards anterior regions. Furthermore, as prominent differences in the acoustic features of the category constituting stimuli were not present, it provides further support for the assumption that these features alone are not responsible for the observed differentiated processing, but that the auditory system combines them in a yet to be discovered weighted manner.

In an initial effort to investigate how this seemingly abstract cortical representation of music is achieved throughout auditory processing and how it might be explained through different models that comprise low and high level features, the fourth and final project applied a multimodal approach that entailed fMRI, EEG, behavior and computational models to provide insights on how music in itself is categorized by the human brain. Therefore, participants were presented with a variety of music genres and their respective neuronal representation obtained either via fMRI or EEG. Aiming to probe behavioral representations as a meaningful correlate of mental representations, all participants furthermore conducted a behavioral paradigm with the identical excerpts. Furthermore, a vast amount of acoustic features was extracted from the stimuli and conceptual representations were formed based on genre definitions and theories. The behavioral, acoustic feature and conceptual representations served as models aiming to test if they could provide a sufficient explanation of the observed neuronal representations along the auditory system (fMRI) or multiple time windows (EEG). In order to enable this comparison on the assumed rather abstract level of music representations, RSA was applied in all parts of the project. Additionally, musical training was evaluated as one possible factor to modulate music representations in a purely behavioral setting by acquiring data from musicians and non-musicians. The respective outcomes revealed that the categorization of music, thus its representation, is rather unstable during the processing of a given stimulus and is most likely achieved over long time scales, as EEG representations varied to a larger extent between time windows and participants than fMRI representations between auditory cortex ROIs and participants. Concerning the latter, an interaction between hemisphere and processing hierarchy was observed, that is models exhibited a greater predictive ability in RH representations and manifested towards non-primary, i.e. parabelt regions in both anterior and posterior direction. However, the predictive ability of the utilized models diminished drastically in anterior regions, that additionally were very dissimilar from the other ROIs' representations. Thus, the anterior regions of the auditory cortex seemingly encode an abstract representation of music that cannot be explained by contemporary models. Focusing the models, complex models such as behavior, conceptual ones and certain acoustic features always outperformed simpler acoustic feature models and therefore hinting to the significance of top-down processes and their possible interaction with complex acoustic features as bottom-up aspects in the generation of music representations. Together with the absence of differences between musicians and non-musicians that renders the influence of musical training on music representations questionable, cortical music representations might be the result of a prototype theory and processing within which incoming sounds are transformed and grouped according to templates that are based on experience and prior knowledge, hence including music enculturation.

In sum, further advancement, standardization and automation is needed in the research field of auditory neuroscience in order to enhance measurement settings and building the basis for a common ground that facilitates further endeavors. One of them certainly is the further investigation of cortical music representations, which so far seems to be achieved through a specialized processing within the RH and a gradient towards the anterior regions of the auditory cortex that is guided by the interaction between complex low and high level models.

6.1. METHODOLOGICAL CONSIDERATIONS IN THE REALM OF NEUROSCIENTIFIC MUSIC RESEARCH

The different parts of the here presented thesis made several problematic circumstances of music research within the realm of auditory neuroscience abundantly clear: it is prone to errors and can achieve only a restricted validity and generalizability. As outlined in project 1, two of the main factors that contribute to the aggravated investigation of cortical auditory processes, thus also music representations are the measurement settings and the localization and parcellation of the auditory cortex. Here subproject 1 could show on an example data set that hearing thresholds and abilities vary drastically as a function of the applied DAP and the resulting ASN. In more detail, the higher the amount of the ASN introducing factors gradient switch and RF pulse, the more prominent the deterioration. It appears that this effect furthermore interacts with the frequency spectrum of sounds, as it was intense for frequencies that are most prominent in DAPs, that is 4 kHZ and below. Therefore, not only the three principle confounds of ASN, auditory cortex activation, stimulus degradation and participant discomfort are present, but also the more pronounced masking and alteration of only certain stimulus aspects. Given the proposed importance of acoustic features in the processing of music (Allen et al., 2017; Behroozmand, Ibrahim, Korzyukov, Robin, & Larson, 2014; Casey, 2017; Casey et al., 2012; Schön et al., 2010; Stern, 2017; Tierney et al., 2013; Whitehead & Armony, 2018), this poses as an additionally possible confound that needs to be considered. As unfortunately this project was not concluded before project 2 and 3 were conducted, a respective influence due to the continuous DAP utilized in both cannot be ruled out. Even though Whitehead & Armony (2018) provided evidence that continuous and non-continuous DAPs introduced no difference in the outcomes of the dataset of project 2, this comparison was limited to the type of analyses they applied and cannot be adapted to the spatially and time course sensitive analyses of the here presented study. Furthermore, the altered perception of the music excerpts in subproject 1 of project 3 (fMRI) could also have altered the cortical representations and thus rendered the comparison between them and the acoustic feature models inaccurate. However, given the comparable model performance in subproject 2 and 3 (EEG and behavior, respectively) this seems unlikely, yet an influence on the proposed combination of features might be present. The second subproject of the first project pointed to the tremendous variety of auditory cortex localizations and parcellations, that for example can range from 1 to 12 ROIs within a comparable spatial extent. Based on that and the underlying properties (structural and functional), as well as the respective cognitive processes they might entail, the results of project 2 and 3 are potentially highly ROI specific. Even though the applied working model of the auditory cortex comprised ROIs that follow main principles like the differentiation in primary and non-primary regions (Glasser et al., 2016; Moerel et al., 2014; Norman-Haignere et al., 2015; Norman-Haignere & McDermott, 2018) and were additionally motivated by previous research work (Allen et al., 2018; Regev et al., 2013), they could have been too broad or too narrow in both spatial extent and number. Here, the outcomes could be used to further inform subsequent characterizations with novel insights and given its tradeoff between accuracy and complexity, additionally serve as a basis to explore more broad and/or fine-grained differences. Concerning other factors that have a prominent influence the utilized stimuli and paradigm need to be addressed. Even though project 2 and 3 included a versatile and controlled stimulus set, their generalizability and ecological validity is rather limited in the light of the sheer endlessly complex myriad of sounds humans are surrounded by, also within a given category like music. While this is less pronounced in project 2, both do not include music excerpts with vocals, as project 2 entailed instrumental music and singing and project 3 exclusively instrumental music, thus in turn are missing the probably most frequent type of contemporary music. Especially project 3 is affected by these aspects, as here only one music excerpt is utilized to represent a complete genre and additionally covers only a subset of the immense versatile music space (Ajoodha et al., 2015; Gjerdingen & Perrott, 2008; Tzanetakis & Cook, 2002). A change of excerpt, sub- or main genre could therefore result in completely diverging results as the representations are based on the multidimensional space they comprise. The stimuli that were used throughout this thesis in all projects had furthermore a fairly short duration that does not reflect the dynamic evolution of spectrotemporal features typically found in music (Alluri et al., 2012; Alluri et al., 2013; Toiviainen et al., 2014). As reflected in the results of subproject 2 of project 3, the processing of music and its representations seems to vary to a large extent over time. Going a step further, the paradigm the stimuli are presented in has an eminently influence on the thereby evoked cortical processing, with especially the task and the corresponding attention introducing potential confounds, like for example the focus on certain features (Alho et al., 2014; J. B. Fritz et al., 2007; J. B. Fritz et al., 2005; Häkkinen et al., 2015; Häkkinen & Rinne, 2018; Hjortkjær et al., 2018; McPherson & McDermott, 2018; Otazu et al., 2009; Petkov et al., 2004; Rinne et al., 2008; Scheich et al., 2007; Woods & Alain, 2009), with this alteration already being present in subcortical portions of the auditory system (Rinne et al., 2008). In order to avoid an altered processing and thus representation, project 2 and 3 applied a passive listening paradigm that did not require any task, but to stay awake. Nevertheless, it cannot be ruled out that participants conducted a certain task or focused on certain features of the stimuli only. Comparably, the applied paradigms do not allow to disentangle the role of attention. Taken the aforementioned points together, numerous circumstances that should be considered for music and other auditory neuroscience research become evident. First, non-continuous DAPs like ISSS should be preferred to avoid ASN related confounds while at the same time acquiring a sufficient amount of data that allows to capture the time course of perception and processing. Second, a working model of the auditory cortex should be motivated, precisely described and justify its use case. Third, the stimuli selection and preparation should be done carefully, be based on previous research work and include natural and versatile excerpts, comparable to experiments that investigated natural sounds (Allen et al., 2018; Giordano et al., 2013; Moerel et al., 2013; Santoro et al., 2014; Santoro et al., 2017) but within music. Furthermore, musical excerpts should have a sufficient length to capture their inherent variation over time (Alluri et al., 2012; Alluri et al., 2013; Hoefle et al., 2018; Toiviainen et al., 2014). Fourth, if possible, the investigation of cortical music representations should always be multimodal and incorporate low and high level features.

6.2. REPRESENTATIONS OF MUSIC WITHIN THE CORTICAL AUDITORY SYSTEM

A wave. This is how it all starts. A wave that is transformed to one of humanity's most defining characteristics: music. But how is this sound wave transformed to a seemingly specialized and stable percept that thereinafter unfolds its immense impact? The results of the here presented thesis suggest that this process is achieved along the hierarchy of the cortical auditory system that leads from rather simple to highly complex and abstract representations (Zatorre & Salimpoor, 2013) and has to be described through four factors: hemisphere, network, hierarchy and features. After reaching the primary auditory cortex within HG as the first point of processing in the cortical portion of the auditory system, the incoming signal is initially analyzed and subsequently distributed in a bilateral network that comprises regions within the STP and STG. As revealed by project 3, this distribution is apparently conducted by the RH HG and results in modulated connections within the RH for music, within which the HG recruits both the PP and the PT, with the former also inhibiting the HG. Being the only auditory category that evokes purely right lateralized modulated connections, as compared to singing and speech, this provides further evidence for the specialization of the auditory system for music processing, that is a bilateral network with a right hemispheric weighting, as suggested by numerous previous studies (Cha et al., 2016; X. Chen et al., 2018; M. E. Klein &

Zatorre, 2011; Koelsch, 2011a; Martínez-Molina et al., 2016; Mišic et al., 2018; Peretz & Zatorre, 2005; Zatorre & Baum, 2012; Zatorre et al., 2002; Zatorre & Salimpoor, 2013). The results of project 4 support this assumption as here representations were more precise in the RH ROIs and could be better explained by all models, independent of low level acoustic or high level features. Additionally, RH representations throughout the entire auditory cortex exhibited a higher predictive ability with respect to the obtained behavioral representations that are assumed to reflect mental representations and concepts to a high degree (Kriegeskorte & Mur, 2012). The aforementioned distribution thereby follows the hierarchy of the auditory system from core over belt to parabelt, i.e. from primary to no-primary regions (Chevillet et al., 2011; Humphries, Sabri, Lewis, & Liebenthal, 2014; Kell et al., 2018; Kumar, Stephan, Warren, Friston, & Griffiths, 2007; Norman-Haignere et al., 2015; Norman-Haignere & McDermott, 2018; Okada et al., 2010; Upadhyay et al., 2008; Zatorre & Salimpoor, 2013). As revealed by the outcomes of project 3, the HG recruits other regions and thus disseminates information seemingly in a purely feedforward manner towards belt and parabelt regions in anterior and posterior direction. This is further underlined by the searchlight maps showing that these regions contain the highest amount of information that allow to distinguish different sound categories, while the HG regions and their surrounding portions exhibited the opposite effect. The representations obtained in project 4 support this assumption, as there representations became more precise along the same gradient and differed between core, belt and parabelt regions. Notably, even though these effects were present in both the ventral pathway towards anterior regions and the dorsal pathway towards posterior regions, thus stressing that also the dorsal pathway and the entire auditory cortex is involved in complex processing steps by means of auditory object analyses and categorization (Formisano, Martino, et al., 2008; Timothy D. Griffiths & Warren, 2002; Kilian-Hütten et al., 2011; Staeren et al., 2009), especially the ROIs of the ventral pathway appear to be responsible for the generation of abstract music representations. Project 3 and 4 both contribute results favorable for this proposition. In project 3 music was the only auditory category to evoke an inhibition towards the HG, more precisely from the PP, thus limiting further incoming information. Additionally, the searchlight maps pointed to the importance of anterior regions with respect to the representations of music, because they apparently encode information about music that is independent of its type. This is based on music and singing compared to speech yielding clusters that contain high informational content within the PP and STGa, whereas this effect vanished when comparing music and singing. The comparison of representations among one another and with the applied models not only showed that representations changed along the above described gradient, but that the anterior representation were more distinct, also to their respective posterior counterparts and that the predictive ability of the models decreased towards non-primary regions, but only in anterior direction. It is worth mentioning, that the results furthermore support the hypotheses that even though all ROIs of the auditory cortex are involved in complex processing steps, a preferred feature encoding exists for certain regions. In more detail, fine-grained spectral and broad temporal modulations are supposedly encoded in anterior regions and broad spectral and fine-grained temporal modulations in posterior regions (Häkkinen & Rinne, 2018; Whitehead & Armony, 2018). Project 3 showed that singing and speech, both containing the latter expression of modulations, lead to a recruitment of both plana temporale and that meaningful information about voice or speech was found in posterior regions either when distinguishing music and speech or music and singing. In project 4, the predictive ability of the tempogram model was most prominent in the bilateral PT ROIs. In contrast and as already addressed above, based on the outcomes of project 3, music, containing the first type of modulations, is apparently more preferentially encoded in anterior regions. Project 4 supports this claim, as the timbre model, being a complex acoustic feature model that entails spectral and temporal aspects, however with a more pronounced weight on spectral components, is the only model that not exhibits a drastically diminished correlation with anterior ROIs. Together this points to a highly abstract and specialized representation of music in anterior regions of the auditory cortex and thus supports previous research work (Angulo-Perkins et al., 2014; Armony et al., 2015; Norman-Haignere et al., 2015; Whitehead & Armony, 2018). This is further supported by incorporating the role of acoustic features. Albeit the majority of proposed auditory processing models emphasize the important role of acoustic features within rather purely bottom-up approaches, recent insights point to the necessity of models that entail a most likely non-linear combination of acoustic features, as well as high level features in order to sufficiently explain the observed auditory cortex functionality and thus representations (Häkkinen & Rinne, 2018; Whitehead & Armony, 2018). The results of project 3 and 4 provide supplementary evidence for this assumption. As the stimuli of the different auditory categories in project 3 were controlled for their acoustic features and largely displayed no significant difference between the categories, it appears unlikely that the observed distinctions in the connectivity pattern or searchlight maps originate from differences in single acoustic features or a simple combination of them. An effect that has also been observed and suggested by Whitehead & Armony (2018), who utilized the same data set, but different analysis approaches. In line with this presumption, project 4 revealed that models based on rather simple acoustic features like spectrum, pitch, etc. are always outperformed by more complex acoustic feature models like timbre and tempogram. As noted above, these models comprise several aspects of other acoustic features and therefore seemingly capture more of the abstract representation their combination evokes. Nevertheless, all acoustic feature models are outperformed by the high level models, especially behavior. Reflecting a mental representation and concepts of music that are based on experience and a multitude of semantic information (e.g., genre definitions) that in turn are grounded in the combination of acoustic features, these models hint to the tremendous influence of high level features through top-down processes. This is further stressed by the observation that the conceptual models can fully explain the obtained behavioral representations. Together with the high variability of representations over time and between participants visible in part II of project 4, this hints to the existence of holistic, stable and abstract mental representations. As already addressed, these representations furthermore appear to be driven by experience and music enculturation, as representations were similar across participants within and across the parts of project 4 and additionally are not varied by an altered processing of music, as no difference between musicians and non-musicians was observed. This again stresses the insufficient explanation single acoustic features provide. Taken all of the aforementioned points together, cortical representations of music are achieved through a transformation of incoming signals from simple to complex and abstract representation along the gradient of hierarchical auditory processing within a bilateral, slightly right lateralized network (Bizley & Cohen, 2013; Chevillet et al., 2011; Garcia, Hall, & Plack, 2010; Häkkinen & Rinne, 2018; Humphries et al., 2014; Rauschecker & Scott, 2009; Recanzone & Cohen, 2010; Zatorre & Salimpoor, 2013). These transformations are conducted in the entire auditory cortex, but reach their full extent in anterior regions of the STP and STG and are implemented through an interaction between bottom-up and top-down processes within which acquired templates or prototypes are utilized to guide the analysis of acoustic features and thus lead to or are already conducted in multidimensional spaces (Casey, 2017).

6.3. LIMITATIONS & OUTLOOK

As already addressed, the nearly infinite complexity and dimensionality of music is both its greatest feature and the main reason it might never be fully understood. Even the most extensive and exhaustive research on music will be bound by the restricted possibility to tackle its versatility with the necessary ecological validity. As the here presented thesis is by no means an exception to that, prominent limitations are present and comprise three main aspects: music itself, the auditory system and cognitive processing. Focusing the first, the thesis' introduction motivated the restriction to music typically found in western cultures and the inclusion of mainly individuals that had a respective music enculturation, based on the available resources. However, this resulted in yet another WEIRD fallacy (Henrich et al., 2010), as large portions of the available music and thus culture space was left out (Mehr et al., 2018). Given the results of previous research work that underline the influence of music enculturation on music cognition (Jacoby & McDermott, 2017; Jacoby et al., 2019; J. H. McDermott et al., 2016) the obtained results and formulated assumptions will, if at all, only hold true for the music culture investigated. Thus, further investigations will need to include other music cultures to probe the generalizability of its proposed processing mechanisms. Given the projects at hand, this especially concerns music spaces and templates / prototypes. Concerning the second point, all investigations included in this thesis focused on the cortical portions of the auditory system only. Hence, no further and novel insights with regard to the remaining parts, particularly non-neocortex structures like MGB and IC are provided. In the light of the proposed interaction between bottom-up and top-down processes and the prominent reciprocal connections between the auditory nerve and the auditory cortex (Bajo et al., 2010; Rinne et al., 2008; Slee & David, 2015; Suga & Ma, 2003; Winer & Lee, 2007), only a small subset of the auditory processing was covered in the thesis at hand and therefore important aspects of the generation of music representations missed by an order of magnitude. As recent studies could show that these regions apply complex processing steps (Bartlett, 2013; De Martino et al., 2013), future endeavors should consider their crucial function to investigate if they already entail specialized music representations. The third point refers to the exclusive focus on the perception of music, thus ignoring the complete other half of processing, that is production. With respect to that, the here presented thesis offers the basis for a framework, in terms of experiments and analysis methods that can be utilized to address the same question asked here with a focus on production: how is the seemingly specialized and stable representation of music transformed to produce it? In other words, subsequent studies need to investigate how the representations of music are translated to precise motor actions during singing and playing an instrument. Here, especially the interactions of the auditory cortex with other brain regions are of interest. Following the idea of a framework, the work of this thesis could and should be extended to other populations and clinical contexts. While both would provide further insights on how the brain is tuned to music, the latter would offer the possibility for the extension of existing use cases of music as medicine. For example, longitudinal studies that comprise infants to adolescents could shed light on the development of the brain's specialized representation of music and how it might be influenced or utilized within cases of developmental disorders and diseases. Here, again the essential factor of music enculturation could furthermore be included. Conversely, through the better understanding of how music representations become special and are formed along the auditory processing hierarchy, the therapy of neurodegenerative or acquired disorders via music could be enhanced, including Parkinson's Disease (Raglio, 2015), Dementia (Y. Zhang et al., 2017) and stroke (Karmonik et al., 2016). However, one of the most immediate applications could be the improvement of cochlear implants for the perception of music, as here the observed neuronal and behavioral representations and how they might be driven by bottom-up and top-influences provide the basis to further fine-tune their functionality which currently is more prominently focused on speech (H. J. McDermott, 2004; Paquette et al., 2018).

REFERENCES

BIBLIOGRAPHY

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., … Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*. doi:10.3389/fninf.2014.00014

Abrams, D. A., Bhatara, A., Ryali, S., Balaban, E., Levitin, D. J., & Menon, V. (2011). Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cerebral Cortex (New York, NY)*, *21*(7), 1507–1518. doi:10.1093/cercor/bhq198

Agus, T. R., Paquette, S., Suied, C., Pressnitzer, D., & Belin, P. (2017). Voice selectivity in the temporal voice area despite matched low-level acoustic cues. *Scientific Reports*, *7*(1), 11526. doi:10.1038/s41598-017-11684-1

Ahmed, S., & Shellock, F. G. (2001). Magnetic resonance imaging safety: Implications for cardiovascular patients. *Journal of Cardiovascular Magnetic Resonance: Official Journal of the Society for Cardiovascular Magnetic Resonance*, *3*(3), 171–182.

Ahveninen, J., Hämäläinen, M., Jääskeläinen, I. P., Ahlfors, S. P., Huang, S., Lin, F.-H., … Belliveau, J. W. (2011). Attention-driven auditory cortex short-term plasticity helps segregate relevant sounds from noise. *Proceedings of the National Academy of Sciences*, *108*(10), 4182–4187. doi:10.1073/pnas.1016134108

Ahveninen, J., Huang, S., Nummenmaa, A., Belliveau, J. W., Hung, A.-Y., Jääskeläinen, I. P., … Raij, T. (2013). Evidence for distinct human auditory cortex regions for sound location versus identity processing. *Nature Communications*, *4*, 2585. doi:10.1038/ncomms3585

Ahveninen, J., Jääskeläinen, I. P., Raij, T., Bonmassar, G., Devore, S., Hämäläinen, M., … Belliveau, J. W. (2006). Task-modulated "what" and "where" pathways in human auditory cortex. *Proceedings of the National Academy of Sciences*, *103*(39), 14608–14613. doi:10.1073/pnas.0510480103

Ajoodha, R., Klein, R., & Rosman, B. (2015). Single-labelled music genre classification using content-based features. In *2015 pattern recognition association of south africa and robotics and mechatronics international conference (PRASA-RobMech)* (pp. 66–71). 2015 pattern recognition association of south africa and robotics and mechatronics international conference (PRASA-RobMech). doi:10.1109/RoboMech.2015.7359500

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *Molecular biology of the cell* (4th). Garland Science.

Alho, K., Rinne, T., Herron, T. J., & Woods, D. L. (2014). Stimulus-dependent activations and attention-related modulations in the auditory cortex: A meta-analysis of fMRI studies. *Hearing Research*, *307*, 29–41. doi:10.1016/j.heares.2013.08.001

Allen, E. J., Burton, P. C., Olman, C. A., & Oxenham, A. J. (2017). Representations of pitch and timbre variation in human auditory cortex. *Journal of Neuroscience*, *37*(5), 1284–1293. doi:10.1523/JNEUROSCI.2336-16.2016

Allen, E. J., Moerel, M., Lage-Castellanos, A., De Martino, F., Formisano, E., & Oxenham, A. J. (2018). Encoding of natural timbre dimensions in human auditory cortex. *NeuroImage*, *166*(2017), 60–70. Publisher: Elsevier Ltd. doi:10.1016/j.neuroimage.2017.10.050

Alley, T. R., & Greene, M. E. (2008). The relative and perceived impact of irrelevant speech, vocal music and non-vocal music on working memory. *Current Psychology: A Journal for*

Diverse Perspectives on Diverse Psychological Issues, *27*(4), 277–289. doi:10.1007/s12144-008-9040-z

Alluri, V., Toiviainen, P., Burunat, I., Kliuchko, M., Vuust, P., & Brattico, E. (2017). Connectivity patterns during music listening: Evidence for action-based processing in musicians. *Human Brain Mapping*, *38*(6). ISBN: 1097-0193 (Electronic) 1065-9471 (Linking). doi:10.1002/hbm.23565

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. Publisher: Elsevier Inc. ISBN: 1095-9572 (Electronic)\r1053-8119 (Linking). doi:10.1016/j.neuroimage.2011.11.019

Alluri, V., Toiviainen, P., Lund, T. E., Wallentin, M., Vuust, P., Nandi, A. K., . . . Brattico, E. (2013). From vivaldi to beatles and back: Predicting lateralized brain responses to music. *NeuroImage*, *83*, 627–636. Publisher: Elsevier Inc. ISBN: 1053-8119. doi:10.1016/j.neuroimage.2013.06.064

Altmann, C. F., Doehrmann, O., & Kaiser, J. (2007). Selectivity for animal vocalizations in the human auditory cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, *17*(11), 2601–2608. doi:10.1093/cercor/bhl167

Andoh, J., Matsushita, R., & Zatorre, R. J. (2015). Asymmetric interhemispheric transfer in the auditory network: Evidence from TMS, resting-state fMRI, and diffusion imaging. *Journal of Neuroscience*, *35*(43), 14602–14611. doi:10.1523/JNEUROSCI.2333-15.2015

Andoh, J., Matsushita, R., & Zatorre, R. J. (2018). Insights into auditory cortex dynamics from non-invasive brain stimulation. *Frontiers in Neuroscience*, *12*. doi:10.3389/fnins.2018.00469

Angenstein, N., Stadler, J., & Brechmann, A. (2016). Auditory intensity processing: Effect of MRI background noise. *Hearing Research*, *333*, 87–92. doi:10.1016/j.heares.2016.01.007

Angulo-Perkins, A., Aubé, W., Peretz, I., Barrios, F. A., Armony, J. L., & Concha, L. (2014). Music listening engages specific cortical regions within the temporal lobes: Differences between musicians and non-musicians. *Cortex; a journal devoted to the study of the nervous system and behavior*, *59*, 126–37. ISBN: 1973-8102 (Electronic)\r0010-9452 (Linking). doi:10.1016/j.cortex.2014.07.013

ANSI/ASA S1.1-2013. (2018), In *Wikipedia*. Page Version ID: 861522227. Retrieved January 1, 2019, from https://en.wikipedia.org/w/index.php?title=ANSI/ASA_S1.1-2013&oldid=861522227

Armony, J. L., Aubé, W., Angulo-Perkins, A., Peretz, I., & Concha, L. (2015). The specificity of neural responses to music and their relation to voice processing: An fMRI-adaptation study. *Neuroscience Letters*, *593*, 1–5. Publisher: Elsevier Ireland Ltd. doi:10.1016/j.neulet.2015.03.011

Asano, R., & Boeckx, C. (2015). Syntax in language and music: What is the right level of comparison? *Frontiers in Psychology*, *6*, 1–16. doi:10.3389/fpsyg.2015.00942

Association (ASHA), A. S.-L.-H. (2005). Guidelines for manual pure-tone threshold audiometry [American speech-language-hearing association]. doi:10.1044/policy.GL2005-00014

Aube, W., Angulo-Perkins, A., Peretz, I., Concha, L., & Armony, J. L. (2015). Fear across the senses: Brain responses to music, vocalizations and facial expressions. *Social Cognitive and Affective Neuroscience*, *10*(3), 399–407. doi:10.1093/scan/nsu067

Avants, B., Tustison, N., & Song, G. (2009). Advanced normalization tools, ANTS 1.0.

Bajo, V. M., Nodal, F. R., Moore, D. R., & King, A. J. (2010). The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nature Neuroscience*, *13*(2), 253–260. doi:10.1038/nn.2466

Balkwill, L. L., Thompson, W. F., & Matsunaga, R. (2004). Recognition of emotion in japanese, western, and hindustani music by japanese listeners. *Japanese Psychological Research*, *46*(4), 337–349. doi:10.1111/j.1468-5584.2004.00265.x

Bandettini, P. A., Jesmanowicz, A., Van Kylen, J., Birn, R. M., & Hyde, J. S. (1998). Functional MRI of brain activation induced by scanner acoustic noise. *Magnetic Resonance in Medicine*, *39*(3), 410–416.

Banerjee, A., Sanyal, S., Patranabis, A., Banerjee, K., Guhathakurta, T., Sengupta, R., … Ghose, P. (2016). Study on brain dynamics by non linear analysis of music induced EEG signals. *Physica A: Statistical Mechanics and its Applications*, *444*, 110–120. doi:10.1016/j.physa.2015.10.030

Bartlett, E. L. (2013). The organization and physiology of the auditory thalamus and its role in processing acoustic features important for speech perception. *Brain and language*, *126*(1), 29–48. doi:10.1016/j.bandl.2013.03.003

Barton, B., Venezia, J. H., Saberi, K., Hickok, G., & Brewer, A. A. (2012). Orthogonal acoustic dimensions define auditory field maps in human cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(50), 20738–20743. doi:10.1073/pnas.1213381109

Bauer, A.-K. R., Kreutz, G., & Herrmann, C. S. (2015). Individual musical tempo preference correlates with EEG beta rhythm. *Psychophysiology*, *52*(4), 600–604. doi:10.1111/psyp.12375

Baumann, S., Petkov, C. I., & Griffiths, T. D. (2013). A unified framework for the organization of the primate auditory cortex. *Frontiers in Systems Neuroscience*, *7*, 1–8. ISBN: 1662-5137 (Electronic)\r1662-5137 (Linking). doi:10.3389/fnsys.2013.00011

Beck, E. (1928). Die myeloarchitektonische felderung des in der sylvischen furche gelegenen teils des menschlichen schlafenlappen. *36*, 1–21.

Behroozmand, R., Ibrahim, N., Korzyukov, O., Robin, D. A., & Larson, C. R. (2014). Left-hemisphere activation is associated with enhanced vocal pitch error detection in musicians with absolute pitch. *Brain and Cognition*, *84*(1), 97–108. doi:10.1016/j.bandc.2013.11.007

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101. doi:10.1016/j.neuroimage.2007.04.042

Bekinschtein, T. A., Davis, M. H., Rodd, J. M., & Owen, A. M. (2011). Why clowns taste funny: The relationship between humor and semantic ambiguity. *Journal of Neuroscience*, *31*(26), 9665–9671. doi:10.1523/JNEUROSCI.5058-10.2011

Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *NeuroReport*, *14*(16), 2105. Retrieved from https://journals.lww.com/neuroreport/Abstract/2003/11140/Adaptation_to_speaker_s_voice_in_right_anterior.19.aspx

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, *13*(1), 17–26. doi:10.1016/S0926-6410(01)00084-2

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309–312. doi:10.1038/35002078

Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, *436*(7054), 1161–1165. doi:10.1038/nature03867

Besson, M., & Schön, D. (2001). Comparison between language and music. *Annals of the New York Academy of Sciences*, *930*, 232–258.

Bidelman, G. M., & Grall, J. (2014). Functional organization for musical consonance and tonal pitch hierarchy in human auditory cortex. *NeuroImage*, *101*, 204–214. ISBN: 1095-9572 (Electronic)\r1053-8119 (Linking). doi:10.1016/j.neuroimage.2014.07.005

Bilecen, D., Scheffler, K., Schmid, N., Tschopp, K., & Seelig, J. (1998). Tonotopic organization of the human auditory cortex as detected by BOLD-FMRI. *Hearing research*, *126*(1), 19–27. ISBN: 0378-5955 (Print)\r0378-5955 (Linking). doi:10.1016/S0378-5955(98)00139-7

Birn, R. M., Saad, Z. S., & Bandettini, P. A. (2001). Spatial heterogeneity of the nonlinear dynamics in the FMRI BOLD response. *NeuroImage*, *14*(4), 817–826. doi:10.1006/nimg.2001.0873

Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nature reviews. Neuroscience*, *14*(10), 693–707. doi:10.1038/nrn3565

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot International*, *5*. Retrieved from http://dare.uva.nl/search?metis.record.id=200596

Bohland, J. W., Bokil, H., Allen, C. B., & Mitra, P. P. (2009). The brain atlas concordance problem: Quantitative comparison of anatomical parcellations. *PLOS ONE*, *4*(9), e7200. doi:10.1371/journal.pone.0007200

Bonte, M., Hausfeld, L., Scharke, W., Valente, G., & Formisano, E. (2014). Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *Journal of Neuroscience*, *34*(13), 4548–4557. doi:10.1523/JNEUROSCI.4339-13.2014

Boubela, R. N., Kalcher, K., Nasel, C., & Moser, E. (2014). Scanning fast and slow: Current limitations of 3 tesla functional MRI and future potential. *Frontiers in Physics*, *2*. doi:10.3389/fphy.2014.00001

Braak, H. (1978). The pigment architecture of the human temporal lobe. *Anatomy and Embryology*, *154*(2), 213–240.

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*. doi:10.3389/fpsyg.2011.00308

Brattico, E., Bogert, B., Alluri, V., Tervaniemi, M., Eerola, T., & Jacobsen, T. (2016). It's sad but i like it: The neural dissociation between musical emotions and liking in experts and laypersons. *Frontiers in Human Neuroscience*, *9*, 1–21. doi:10.3389/fnhum.2015.00676

Brauer, J., & Friederici, A. D. (2007). Functional neural networks of semantic and syntactic processes in the developing brain. *Journal of Cognitive Neuroscience*, *19*(10), 1609–1623. doi:10.1162/jocn.2007.19.10.1609

Brechmann, A., & Scheich, H. (2005). Hemispheric shifts of sound representation in auditory cortex with conceptual listening. *Cerebral Cortex (New York, N.Y.: 1991)*, *15*(5), 578–587. doi:10.1093/cercor/bhh159

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. Taylor & Francis.

Brett, M., Johnsrude, I. S., & Owen, A. M. (2002). The problem of functional localization in the human brain. *Nature Reviews Neuroscience*, *3*(3), 243–249. doi:10.1038/nrn756

Bridge, H. (2011). Mapping the visual brain: How and why. *Eye*, *25*(3), 291–296. doi:10.1038/eye.2010.166

Brodmann, K. (1909). *Vergleichende lokalisationslehre der grosshirnrinde in ihren prinzipien dargestellt auf grund des zellenbaues* (). Leipzig : Barth. Retrieved from http://archive.org/details/b28062449

Bronzino, J. D., & Peterson, D. R. (2006). *Biomedical engineering fundamentals*. Google-Books-ID: C2ladV9aI8MC. CRC Press.

Brown, J. C. (1991). Calculation of a constant q spectral transform. *The Journal of the Acoustical Society of America*, *89*(1), 425–434. doi:10.1121/1.400476

Brown, R. M., Zatorre, R. J., & Penhune, V. B. (2015). Expert music performance: Cognitive, neural, and developmental bases. In E. Altenmüller, S. Finger, & F. Boller (Eds.), *Progress in* *brain research* (Vol. 217, pp. 57–86). Music, Neurology, and Neuroscience: Evolution, the Musical Brain, Medical Conditions, and Therapies. doi:10.1016/bs.pbr.2014.11.021

Brown, S. (2000). Evolutionary models of music: From sexual selection to group selection. In F. Tonneau & N. S. Thompson (Eds.), *Perspectives in ethology: Evolution, culture, and behavior* (pp. 231–281). Perspectives in Ethology. doi:10.1007/978-1-4615-1221-9_9

Brummett, R. E., Talbot, J. M., & Charuhas, P. (1988). Potential hearing loss resulting from MR imaging. *Radiology*, *169*(2), 539–540. doi:10.1148/radiology.169.2.3175004

Bruns, A., Mueggler, T., Künnecke, B., Risterucci, C., Prinssen, E. P., Wettstein, J. G., & von Kienlin, M. (2015). "Domain gauges
: A reference system for multivariate profiling of brain fMRI activation patterns induced by psychoactive drugs in rats. *NeuroImage*, *112*, 70–85. Publisher: Elsevier Inc. ISBN: 1053-8119. doi:10.1016/j.neuroimage.2015.02.032

Cal, R., & Bahmad Jr, F. (2009). Vestibular evoked myogenic potentials: An overview. *Brazilian Journal of Otorhinolaryngology*, *75*(3), 456–462. Retrieved from http://www.scielo.br/scielo.php?script=sci_abstract&pid=S1808-86942009000300023&lng=en&nrm=iso&tlng=en

Callan, D. E., Tsytsarev, V., Hanakawa, T., Callan, A. M., Katsuhara, M., Fukuyama, H., & Turner, R. (2006). Song and speech: Brain regions involved with perception and covert production. *NeuroImage*, *31*(3), 1327–42. doi:10.1016/j.neuroimage.2006.01.036

Cammoun, L., Thiran, J. P., Griffa, A., Meuli, R., Hagmann, P., & Clarke, S. (2015). Intrahemispheric cortico-cortical connections of the human auditory cortex. *Brain Structure and Function*, *220*(6), 3537–3553. doi:10.1007/s00429-014-0872-z

Campain, R., & Minckler, J. (1976). A note on the gross configurations of the human auditory cortex. *Brain and Language*, *3*(2), 318–323. doi:10.1016/0093-934X(76)90026-2

Campbell, P. S. (2010). *Musical enculturation: Sociocultural influences and meanings of children's experiences in and through music*. Oxford University Press. Retrieved from http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199214389.001.0001/acprof-9780199214389-chapter-4

Carcea, I., Insanally, M. N., & Froemke, R. C. (2017). Dynamics of auditory cortical activity during behavioural engagement and auditory perception. *Nature Communications*, *8*, 14412. doi:10.1038/ncomms14412

Casey, M. A. (2017). Music of the 7ts: Predicting and decoding multivoxel fMRI responses with acoustic, schematic, and categorical music features. *Frontiers in Psychology*, *8*. doi:10.3389/fpsyg.2017.01179

Casey, M. A., Thompson, J., Kang, O., Raizada, R., & Wheatley, T. (2012). Population codes representing musical timbre for high-level fMRI categorization of music genres. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7263 LNAI*, 34–41. doi:10.1007/978-3-642-34713-9_5

Catani, M., & Thiebaut de Schotten, M. (2008). A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex*. Special Issue on "Brain Hodology - Revisiting disconnection approaches to disorders of cognitive function", *44*(8), 1105–1132. doi:10.1016/j.cortex.2008.05.004

Celesia, G. G. (1976). Organization of auditory cortical areas in man. *Brain: A Journal of Neurology*, *99*(3), 403–414.

Cha, K., Zatorre, R. J., & Schönwiesner, M. (2016). Frequency selectivity of voxel-by-voxel functional connectivity in human auditory cortex. *Cerebral Cortex*, *26*(1), 211–224. doi:10.1093/cercor/bhu193

Chaffin, R., & Logan, T. (2006). Practicing perfection: How concert soloists prepare for performance. *Advances in Cognitive Psychology*, *2*(2), 113–130. doi:10.2478/v10053-008-0050-z

Chambers, J., Akeroyd, M. A., Summerfield, A. Q., & Palmer, A. R. (2001). Active control of the volume acquisition noise in functional magnetic resonance imaging: Method and psychoacoustical evaluation. *The Journal of the Acoustical Society of America, 110*(6), 3041–3054.

Chambers, J., Bullock, D., Kahana, Y., Kots, A., & Palmer, A. (2007). Developments in active noise control sound systems for magnetic resonance imaging. *Applied Acoustics*. Application of Acoustics to Medicine, *68*(3), 281–295. doi:10.1016/j.apacoust.2005.10.008

Chanda, M. L., & Levitin, D. J. (2013). The neurochemistry of music. *Trends in Cognitive Sciences, 17*(4), 179–191. ISBN: 1879-307X (Electronic)\r1364-6613 (Linking). doi:10.1016/j.tics.2013.02.007

Chartrand, J.-P., & Belin, P. (2006). Superior voice timbre processing in musicians. *Neuroscience Letters, 405*(3), 164–167. doi:10.1016/j.neulet.2006.06.053

Chen, C., Halpern, A. R., Bly, B., Edelman, R., & Schlaug, G. (2000). Planum temporale asymmetry and absolute pitch. *NeuroImage, 5*(5), S114.

Chen, G., Saad, Z. S., Adleman, N. E., Leibenluft, E., & Cox, R. W. (2015). Detecting the subtle shape differences in hemodynamic responses at the group level. *Frontiers in Neuroscience, 9*, 375. doi:10.3389/fnins.2015.00375

Chen, X., Zhao, Y., Zhong, S., Cui, Z., Li, J., Gong, G., … Nan, Y. (2018). The lateralized arcuate fasciculus in developmental pitch disorders among mandarin amusics: Left for speech and right for music. *Brain Structure and Function, 0*(0), 0. Publisher: Springer Berlin Heidelberg ISBN: 0123456789. doi:10.1007/s00429-018-1608-2

Chevillet, M., Riesenhuber, M., & Rauschecker, J. P. (2011). Functional correlates of the anterolateral processing hierarchy in human auditory cortex. *Journal of Neuroscience, 31*(25), 9345–9352. doi:10.1523/JNEUROSCI.1448-11.2011

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience, 17*(3), 455–462. doi:10.1038/nn.3635

Clarke, S., & Morosan, P. (2012). Architecture, connectivity, and transmitter receptors of human auditory cortex. In D. Poeppel, T. Overath, A. N. Popper, & R. R. Fay (Eds.), *The human auditory cortex* (pp. 11–38). Springer Handbook of Auditory Research. doi:10.1007/978-1-4614-2314-0_2

Clarke, S., & Rivier, F. (1998). Compartments within human primary auditory cortex: Evidence from cytochrome oxidase and acetylcholinesterase staining. *The European Journal of Neuroscience, 10*(2), 741–745.

Cohen-Adad, J., Polimeni, J. R., Helmer, K. G., Benner, T., McNab, J. A., Wald, L. L., … Mainero, C. (2012). T* mapping and b orientation-dependence at 7 t reveal cyto- and myeloarchitecture organization of the human cortex. *NeuroImage, 60*(2), 1006–1014. doi:10.1016/j.neuroimage.2012.01.053

Cohen, J. R., & D'Esposito, M. (2016). The segregation and integration of distinct brain networks and their relationship to cognition. *The Journal of Neuroscience, 36*(48), 12083–12094. doi:10.1523/JNEUROSCI.2965-15.2016

Cohen, Y. E., Bennur, S., Christison-Lagay, K., Gifford, A. M., & Tsunada, J. (2016). Functional organization of the ventral auditory pathway. *Advances in Experimental Medicine and Biology, 894*, 381–388. doi:10.1007/978-3-319-25474-6_40

Coombs, C. H. (1954). A method for the study of interstimulus similarity. *Psychometrika, 19*(3), 183–194. doi:10.1007/BF02289183

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1990). Attentional modulation of neural processing of shape, color, and velocity in humans. *Science (New York, N.Y.) 248*(4962), 1556–1559.

Correia, J. M., Jansma, B. M. B., & Bonte, M. (2015). Decoding articulatory features from fMRI responses in dorsal speech regions. *Journal of Neuroscience*, *35*(45), 15015–15025. doi:10.1523/JNEUROSCI.0977-15.2015

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, an International Journal*, *29*(3), 162–173.

Cross, I. (2001). Music, cognition, culture, and evolution. *Annals of the New York Academy of Sciences*, *930*, 28–42. ISBN: 1749-6632. doi:10.1111/j.1749-6632.2001.tb05723.x

Da Costa, S., van der Zwaag, W., Marques, J. P., Frackowiak, R. S. J., Clarke, S., & Saenz, M. (2011). Human primary auditory cortex follows the shape of heschl's gyrus. *Journal of Neuroscience*, *31*(40), 14067–14075. ISBN: 1529-2401 (Electronic)\n0270-6474 (Linking). doi:10.1523/JNEUROSCI.2000-11.2011

Da Costa, S., van der Zwaag, W., Miller, L. M., Clarke, S., & Saenz, M. (2013). Tuning in to sound: Frequency-selective attentional filter in human primary auditory cortex. *Journal of Neuroscience*, *33*(5), 1858–1863. doi:10.1523/JNEUROSCI.4405-12.2013

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. i. segmentation and surface reconstruction. *NeuroImage*, *9*(2), 179–194. doi:10.1006/nimg.1998.0395

Dale, A. M., & Sereno, M. I. (1993). Improved localizadon of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. *Journal of Cognitive Neuroscience*, *5*(2), 162–176. doi:10.1162/jocn.1993.5.2.162

Daly, I., Williams, D., Hallowell, J., Hwang, F., Kirke, A., Malik, A., … Nasuto, S. J. (2015). Music-induced emotions can be predicted from a combination of brain activity and acoustic features. *Brain and Cognition*, *101*, 1–11. doi:10.1016/j.bandc.2015.08.003

Darrow, A.-A., Haack, P., & Kuribayashi, F. (1987). Descriptors and preferences for eastern and western musics by japanese and american nonmusic majors. *Journal of Research in Music Education*, *35*(4), 237–248. doi:10.2307/3345076

Darwin, C. (1871). *The descent of man*. D. Appleton and Company. Retrieved from https://books.google.de/books?id=ZvsHAAAAIAAJ

Davis, K. A. (2005). Contralateral effects and binaural interactions in dorsal cochlear nucleus. *Journal of the Association for Research in Otolaryngology: JARO*, *6*(3), 280–296. doi:10.1007/s10162-005-0008-5

Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through "top–down" processes? evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience*, *23*(12), 3914–3932. doi:10.1162/jocn_a_00084

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *23*(8), 3423–3431.

Dawson, M. R. W. (2018). *Connectionist representations of tonal music: Discovering musical patterns by interpreting artificial neural networks*. Google-Books-ID: PgFRDwAAQBAJ. Athabasca University Press.

De Angelis, V., De Martino, F., Moerel, M., Santoro, R., Hausfeld, L., & Formisano, E. (2018). Cortical processing of pitch: Model-based encoding and decoding of auditory fMRI responses to real-life sounds. *NeuroImage*, *180*, 291–300. doi:10.1016/j.neuroimage.2017.11.020

De Lucia, M., Camen, C., Clarke, S., & Murray, M. M. (2009). The role of actions in auditory object discrimination. *NeuroImage*, *48*(2), 475–485. Publisher: Elsevier Inc. ISBN: 1053-8119. doi:10.1016/j.neuroimage.2009.06.041

De Martino, F., Moerel, M., Ugurbil, K., Formisano, E., & Yacoub, E. (2015). Less noise, more activation: Multiband acquisition schemes for auditory functional MRI. *Magnetic resonance in medicine*, *74*(2), 462–7. doi:10.1002/mrm.25408

De Martino, F., Moerel, M., van de Moortele, P.-F., Ugurbil, K., Goebel, R., Yacoub, E., & Formisano, E. (2013). Spatial organization of frequency preference and selectivity in the human inferior colliculus. *Nature communications*, *4*, 1386. Publisher: Nature Publishing Group ISBN: 2041-1723 (Electronic)\r2041-1723 (Linking). doi:10.1038/ncomms2379. arXiv: NIHMS150003

De Martino, F., Moerel, M., Xu, J., Van De Moortele, P. F., Ugurbil, K., Goebel, R., … Formisano, E. (2015). High-resolution mapping of myeloarchitecture in vivo: Localization of auditory areas in the human brain. *Cerebral Cortex*, *25*(10), 3394–3405. ISBN: 1460-2199 (Electronic)\r1047-3211 (Linking). doi:10.1093/cercor/bhu150

De Martino, F., Schmitter, S., Moerel, M., Tian, J., Ugurbil, K., Formisano, E., … de Moortele, P. F. v. (2012). Spin echo functional MRI in bilateral auditory cortices at 7t: An application of b1 shimming. *NeuroImage*, *63*(3), 1313–1320. Publisher: Elsevier Inc. ISBN: 1095-9572 (Electronic)\r1053-8119 (Linking). doi:10.1016/j.neuroimage.2012.08.029

de Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L., & Theunissen, F. E. (2017). The hierarchical cortical organization of human speech processing. *The Journal of Neuroscience*, *37*(27), 6539–6557. doi:10.1523/JNEUROSCI.3267-16.2017

Deco, G., Tononi, G., Boly, M., & Kringelbach, M. L. (2015). Rethinking segregation and integration: Contributions of whole-brain modelling. *Nature reviews. Neuroscience*, *16*(7), 430–439. ISBN: 1471-0048 (Electronic)\r1471-003X (Linking). doi:10.1038/nrn3963

Decock, L., & Douven, I. (2011). Similarity after goodman. *Review of Philosophy and Psychology*, *2*(1), 61–75. doi:10.1007/s13164-010-0035-y

Definition of music. (n.d.), In *Wikipedia*. Page Version ID: 913110457. Retrieved November 30, 2018, from https://en.wikipedia.org/w/index.php?title=Definition_of_music&oldid=913110457

Degerman, A., Rinne, T., Salmi, J., Salonen, O., & Alho, K. (2006). Selective attention to sound location or pitch studied with fMRI. *Brain Research*, *1077*(1), 123–134. doi:10.1016/j.brainres.2006.01.025

Dehaene-Lambertz, G., Montavont, a., Jobert, a., Allirol, L., Dubois, J., Hertz-Pannier, L., & Dehaene, S. (2010). Language or music, mother or mozart? structural and environmental influences on infants' language networks. *Brain and Language*, *114*(2), 53–65. Publisher: Elsevier Inc. ISBN: 1090-2155 (Electronic)\n0093-934X (Linking). doi:10.1016/j.bandl.2009.09.003

Demorest, S. M., Morrison, S. J., Stambaugh, L. A., Beken, M., Richards, T. L., & Johnson, C. (2009). An fMRI investigation of the cultural specificity of music memory. *Social Cognitive and Affective Neuroscience*, *5*(2), 282–291. ISBN: 1749-5016. doi:10.1093/scan/nsp048

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., … Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980. doi:10.1016/j.neuroimage.2006.01.021

DeSimone, K., Viviano, J. D., & Schneider, K. A. (2015). Population receptive field estimation reveals new retinotopic maps in human subcortex. *Journal of Neuroscience*, *35*(27), 9836–9847. doi:10.1523/JNEUROSCI.3840-14.2015

Devonshire, I. M., Papadakis, N. G., Port, M., Berwick, J., Kennerley, A. J., Mayhew, J. E. W., & Overton, P. G. (2012). Neurovascular coupling is brain region-dependent. *NeuroImage*, *59*(3), 1997–2006. doi:10.1016/j.neuroimage.2011.09.050

Dewilde, J., Grainger, D., Price, D., & Renaud, C. (2007). Magnetic resonance imaging safety issues including an analysis of recorded incidents within the UK. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *51*(1), 37–48. doi:10.1016/j.pnmrs.2007.01.003

DeYoe, E. A., Carman, G. J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., … Neitz, J. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(6), 2382–2386.

Di Salle, F., Esposito, F., Scarabino, T., Formisano, E., Marciano, E., Saulino, C., … Seifritz, E. (2003). fMRI of the auditory system: Understanding the neural basis of auditory gestalt. *Magnetic Resonance Imaging*, *21*(10), 1213–1224. ISBN: 0730-725X (Print)\r0730-725X (Linking). doi:10.1016/j.mri.2003.08.023

Dick, F., Tierney, A. T., Lutti, A., Josephs, O., Sereno, M. I., & Weiskopf, N. (2012). In vivo functional and myeloarchitectonic mapping of human primary auditory areas. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(46), 16095–16105. doi:10.1523/JNEUROSCI.1712-12.2012

Diedrichsen, J., & Kriegeskorte, N. (2017). Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLOS Computational Biology*, *13*(4), e1005508. doi:10.1371/journal.pcbi.1005508

Dien, J. (1998). Issues in the application of the average reference: Review, critiques, and recommendations. *Behavior Research Methods, Instruments, & Computers*, *30*(1), 34–43. doi:10.3758/BF03209414

Doehrmann, O., Naumer, M. J., Volz, S., Kaiser, J., & Altmann, C. F. (2008). Probing category selectivity for environmental sounds in the human auditory brain. *Neuropsychologia*, *46*(11), 2776–2786. doi:10.1016/j.neuropsychologia.2008.05.011

Doehrmann, O., Weigelt, S., Altmann, C. F., Kaiser, J., & Naumer, M. J. (2010). Audiovisual functional magnetic resonance imaging adaptation reveals multisensory integration effects in object-related sensory cortices. *Journal of Neuroscience*, *30*(9), 3370–3379. doi:10.1523/JNEUROSCI.5074-09.2010

Dorsaint-Pierre, R., Penhune, V. B., Watkins, K. E., Neelin, P., Lerch, J. P., Bouffard, M., & Zatorre, R. J. (2006). Asymmetries of the planum temporale and heschl's gyrus: Relationship to language lateralization. *Brain: A Journal of Neurology*, *129*, 1164–1176. doi:10.1093/brain/awl055

Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences*, *12*(3), 99–105. doi:10.1016/j.tics.2008.01.001

Douglas, P., Lau, E., Anderson, A., Kerr, W., Head, A., Wollner, M. A., … Cohen, M. S. (2013). Single trial decoding of belief decision making from EEG and fMRI data using independent components features. *Frontiers in Human Neuroscience*, *7*. doi:10.3389/fnhum.2013.00392

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *NeuroImage*, *39*(2), 647–660. doi:10.1016/j.neuroimage.2007.09.034

Edelman, S. (1998). Representation is representation of similarities. *The Behavioral and Brain Sciences*, *21*(4), 449–467, discussion 467–498.

Eden, G. F., Joseph, J. E., Brown, H. E., Brown, C. P., & Zeffiro, T. A. (1999). Utilizing hemodynamic delay and dispersion to detect fMRI signal change without auditory interference: The behavior interleaved gradients technique. *Magnetic Resonance in Medicine*, *41*(1), 13–20. doi:10.1002/(SICI)1522-2594(199901)41:1<13::AID-MRM4>3.0.CO;2-T

Eerola, T., & Vuoskoski, J. (2013). A review of music and emotion studies: Approaches, emotion models, and stimuli. *Music perception*, *30*(3). doi:10.1525/MP.2012.30.3.307

Elias, L. J., Bryden, M. P., & Bulman-Fleming, M. B. (1998). Footedness is a better predictor than is handedness of emotional lateralization. *Neuropsychologia*, *36*(1), 37–43. doi:10.1016/S0028-3932(97)00107-3

Elmer, S., Hänggi, J., & Jäncke, L. (2016). Interhemispheric transcallosal connectivity between the left and right planum temporale predicts musicianship, performance in temporal speech processing, and functional specialization. *Brain Structure and Function*, *221*(1), 331–344. doi:10.1007/s00429-014-0910-x

Emmorey, K., Allen, J. S., Bruss, J., Schenker, N., & Damasio, H. (2003). A morphometric analysis of auditory brain regions in congenitally deaf adults. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(17), 10049–10054. doi:10.1073/pnas.1730169100

Engel, A., & Keller, P. E. (2011). The perception of musical spontaneity in improvised and imitated jazz performances. *Frontiers in Psychology*, *2*. doi:10.3389/fpsyg.2011.00083

Engel, L. R., Frum, C., Puce, A., Walker, N. A., & Lewis, J. W. (2009). Different categories of living and non-living sound-sources activate distinct cortical networks. *NeuroImage*, *47*(4), 1778–1791. Publisher: Elsevier B.V. ISBN: 2122633255. doi:10.1016/j.neuroimage.2009.05.041. arXiv: NIHMS150003

Engel, S. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex*, *7*(2), 181–192. doi:10.1093/cercor/7.2.181

Erickson, R. (1975). *Sound structure in music*. Google-Books-ID: t3j6_ShXeWYC. University of California Press.

Esparza, T. M., Bello, J. P., & Humphrey, E. J. (2015). From genre classification to rhythm similarity: Computational and musicological insights. *Journal of New Music Research*, *44*(1), 39–57. doi:10.1080/09298215.2014.929706

Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLOS ONE*, *12*(9), e0184661. doi:10.1371/journal.pone.0184661

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., … Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111. doi:10.1038/s41592-018-0235-4

Etzel, J. a., Zacks, J. M., & Braver, T. S. (2013). Searchlight analysis: Promise, pitfalls, and potential. *NeuroImage*, *78*, 261–269. doi:10.1016/j.neuroimage.2013.03.041

Ewender, T., Hoffmann, S., & Pfister, B. (2009). Nearly perfect detection of continuous f0 contour and frame classification for TTS synthesis. In *Proceedings of interspeech 2009 : Speech and intelligence ; 6 - 10 september, 2009, brighton, UK* (pp. 100–103). 10th annual conference of the international speech communication association (INTERSPEECH 2009). ISCA. Retrieved from https://www.research-collection.ethz.ch/handle/20.500.11850/156555

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., … Jiang, T. (2016). The human brainnetome atlas: A new brain atlas based on connectional architecture. *Cerebral Cortex (New York, N.Y.: 1991)*, *26*(8), 3508–3526. doi:10.1093/cercor/bhw157

Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2004). Priming of non-speech vocalizations in male adults: The influence of the speaker's gender. *Brain and Cognition*, *55*(2), 300–302. doi:10.1016/j.bandc.2004.02.024

Fedorenko, E., McDermott, J. H., Norman-Haignere, S., & Kanwisher, N. (2012). Sensitivity to musical structure in the human brain. *Journal of Neurophysiology*, *108*(12), 3289–300. ISBN: 1522-1598 (Electronic) 0022-3077 (Linking). doi:10.1152/jn.00209.2012

Fedorenko, E., Patel, A. D., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. *Memory & cognition*, *37*(1), 1–9. doi:10.3758/MC.37.1.1

Feldmann, H. (1992). [history of instrumental measuring of hearing acuity: The first acumeter]. *Laryngo- Rhino- Otologie*, *71*(9), 477–482. doi:10.1055/s-2007-997336

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, *1*(1), 1–47.

Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(20), 11050–11055. doi:10.1073/pnas.200033797

Fischl, B., Liu, A., & Dale, A. M. (2001). Automated manifold surgery: Constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE transactions on medical imaging*, *20*(1), 70–80. doi:10.1109/42.906426

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., … Dale, A. M. (2002). Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, *33*(3), 341–355.

Fischl, B., Salat, D. H., van der Kouwe, A. J. W., Makris, N., Ségonne, F., Quinn, B. T., & Dale, A. M. (2004). Sequence-independent segmentation of magnetic resonance images. *NeuroImage*, *23 Suppl 1*, S69–84. doi:10.1016/j.neuroimage.2004.07.016

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage*, *9*(2), 195–207. doi:10.1006/nimg.1998.0396

Fischl, B., Sereno, M. I., Tootell, R. B., & Dale, A. M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, *8*(4), 272–284.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., … Dale, A. M. (2004). Automatically parcellating the human cerebral cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, *14*(1), 11–22.

Formisano, E., De Martino, F., & Valente, G. (2008). Multivariate analysis of fMRI time series: Classification and regression of brain responses using machine learning. *Magnetic Resonance Imaging*, *26*(7), 921–934. ISBN: 0730-725X. doi:10.1016/j.mri.2008.01.052

Formisano, E., Kim, D.-s., Salle, F. D., Ugurbil, K., Goebel, R., Maastricht, U., & Policlinico, N. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, *40*, 859–869. ISBN: 0896-6273 (Print)\n0896-6273 (Linking). doi:10.1016/S0896-6273(03)00669-X

Formisano, E., Martino, F. D., Bonte, M., & Goebel, R. (2008). "Who"is saying "what"? brain-based decoding of human voice and speech. *Science*, *322*(5903), 970–973. doi:10.1126/science.1164318

Foster, J. R., Hall, D. A., Summerfield, A. Q., Palmer, A. R., & Bowtell, R. W. (2000). Sound-level measurements and calculations of safe noise dosage during EPI at 3 t. *Journal of Magnetic Resonance Imaging*, *12*(1), 157–163. doi:10.1002/1522-2586(200007)12:1<157::AID-JMRI17>3.0.CO;2-M

Frässle, S., Paulus, F. M., Krach, S., Schweinberger, S. R., Stephan, K. E., & Jansen, A. (2016). Mechanisms of hemispheric lateralization: Asymmetric interhemispheric recruitment in the face perception network. *NeuroImage*, *124*, 977–988. doi:10.1016/j.neuroimage.2015.09.055

Frässle, S., Stephan, K. E., Friston, K. J., Steup, M., Krach, S., Paulus, F. M., & Jansen, A. (2015). Test-retest reliability of dynamic causal modeling for fMRI. *NeuroImage*, *117*, 56–66. Publisher: Elsevier Inc. doi:10.1016/j.neuroimage.2015.05.040

Frazier, J. A., Chiu, S., Breeze, J. L., Makris, N., Lange, N., Kennedy, D. N., … Biederman, J. (2005). Structural brain magnetic resonance imaging of limbic and thalamic volumes in pediatric bipolar disorder. *The American Journal of Psychiatry*, *162*(7), 1256–1265. doi:10.1176/appi.ajp.162.7.1256

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn Sci*, *6*(2), 78–84. ISBN: 1364-6613 (Print). doi:S1364661300018398[pii]

Friederici, A. D., Kotz, S. A., Scott, S. K., & Obleser, J. (2009). Disentangling syntax and intelligibility in auditory language comprehension. *Human Brain Mapping*, *457*, NA–NA. doi:10.1002/hbm.20878

Friederici, A. D., Rüschemeyer, S.-A., Hahne, A., & Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. *Cerebral cortex (New York, N.Y. : 1991)*, *13*(2), 170–177. ISBN: 1047-3211 (Print). doi:10.1093/cercor/13.2.170

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*(4), 1273–1302. ISBN: 9780122648410. doi:10.1016/S1053-8119(03)00202-7

Friston, K. J., Jezzard, P., & Turner, R. (1994). Analysis of functional MRI time-series. *Human Brain Mapping*, *1*(2), 153–171. doi:10.1002/hbm.460010207

Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S., & Turner, R. (1996). Movement-related effects in fMRI time-series. *Magnetic Resonance in Medicine*, *35*(3), 346–355.

Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention—focusing the searchlight on sound. *Current Opinion in Neurobiology*. Sensory systems, *17*(4), 437–455. doi:10.1016/j.conb.2007.07.011

Fritz, J. B., Elhilali, M., & Shamma, S. A. (2005). Active listening: Task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hearing Research*. 3rd Symp. on Molec. Mechanisms in Central Auditory Function, Plasticity and Disorder, *206*(1), 159–176. doi:10.1016/j.heares.2005.01.015

Fullerton, B. C., & Pandya, D. N. (2007). Architectonic analysis of the auditory-related areas of the superior temporal region in human brain. *The Journal of Comparative Neurology*, *504*(5), 470–498. doi:10.1002/cne.21432

Gaab, N., Gabrieli, J. D. E., & Glover, G. H. (2007). Assessing the influence of scanner background noise on auditory processing. II. an fMRI study comparing auditory processing in the absence and presence of recorded scanner noise using a sparse design. *Human Brain Mapping*, *28*(8), 721–732. ISBN: 1065-9471 (Print)\n1065-9471 (Linking). doi:10.1002/hbm.20299

Gage, N. M., & Baars, B. (2018). *Fundamentals of cognitive neuroscience: A beginner's guide*. Google-Books-ID: q6RBDwAAQBAJ. Academic Press.

Gage, N. M., Juranek, J., Filipek, P. A., Osann, K., Flodman, P., Isenberg, A. L., & Spence, M. A. (2009). Rightward hemispheric asymmetries in auditory language cortex in children with autistic disorder: An MRI investigation. *Journal of Neurodevelopmental Disorders*, *1*(3), 205–214. doi:10.1007/s11689-009-9010-2

Galaburda, A., & Sanides, F. (1980). Cytoarchitectonic organization of the human auditory cortex. *The Journal of Comparative Neurology*, *190*(3), 597–610. doi:10.1002/cne.901900312

Galati, G., Committeri, G., Spitoni, G., Aprile, T., Di Russo, F., Pitzalis, S., & Pizzamiglio, L. (2008). A selective representation of the meaning of actions in the auditory mirror system. *NeuroImage*, *40*(3), 1274–1286. doi:10.1016/j.neuroimage.2007.12.044

Garcia, D., Hall, D. A., & Plack, C. J. (2010). The effect of stimulus context on pitch representations in the human auditory cortex. *NeuroImage*, *51*(2), 808–816. Publisher: Elsevier Inc. ISBN: 1053-8119. doi:10.1016/j.neuroimage.2010.02.079

Gebauer, L., Kringelbach, M. L., & Vuust, P. (2012). Ever-changing cycles of musical pleasure: The role of dopamine and anticipation. *Psychomusicology: Music, Mind, and Brain*, *22*(2), 152–167. ISBN: 2162-1535\n0275-3987. doi:10.1037/a0031126\n10.1037/a0031126.supp(Supplemental)

Geretsegger, M., Mössler, K. A., Bieleninik, Ł., Chen, X.-J., Heldal, T. O., & Gold, C. (2017). Music therapy for people with schizophrenia and schizophrenia-like disorders. *Cochrane Database of Systematic Reviews*, (5). doi:10.1002/14651858.CD004025.pub4

Ghosal, A., Chakraborty, R., Dhara, B. C., & Saha, S. K. (2015). Perceptual feature-based song genre classification using RANSAC. *Int. J. Comput. Intell. Stud. 4*(1), 31–49. doi:10.1504/IJCISTUDIES.2015.069831

Gil-Loyzaga, P., Pujol, R., & Irving, S. (2016). Journey into the world of hearing. Retrieved January 5, 2019, from http://www.cochlea.eu/en

Gilmore, R. O., Diaz, M. T., Wyble, B. A., & Yarkoni, T. (2017). Progress toward openness, transparency, and reproducibility in cognitive neuroscience. *Annals of the New York Academy of Sciences*, *1396*(1), 5–18. doi:10.1111/nyas.13325

Giordano, B. L., McAdams, S., Zatorre, R. J., Kriegeskorte, N., & Belin, P. (2013). Abstract encoding of auditory objects in cortical activity patterns. *Cerebral Cortex*, *23*(9), 2025–2037. ISBN: 1047-3211. doi:10.1093/cercor/bhs162

Gjerdingen, R. O., & Perrott, D. (2008). Scanning the dial: The rapid recognition of music genres. *Journal of New Music Research*, *37*(2), 93–100. ISBN: 0929-8215. doi:10.1080/09298210802479268

Glasberg, B. R., & Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*(1), 103–138.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., … Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–8. ISBN: 0008-5472 (Print)\r0008-5472 (Linking). doi:10.1038/nature18933. arXiv: NIHMS150003

Glasser, M. F., & Van Essen, D. C. (2011). Mapping human cortical areas in vivo based on myelin content as revealed by t1- and t2-weighted MRI. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(32), 11597–11616. doi:10.1523/JNEUROSCI.2180-11.2011

Glorig, A. (1965). *Audiometry: Principles and practices*. Google-Books-ID: XNVsAAAAMAAJ. Williams & Wilkins Co.

Goldstein, J. M., Seidman, L. J., Makris, N., Ahern, T., O'Brien, L. M., Caviness, V. S., … Tsuang, M. T. (2007). Hypothalamic abnormalities in schizophrenia: Sex effects and genetic vulnerability. *Biological Psychiatry*, *61*(8), 935–945. doi:10.1016/j.biopsych.2006.06.027

Goldstone, R. L., Medin, D. L., & Halberstadt, J. (1997). Similarity in context. *Memory & Cognition*, *25*(2), 237–255.

Goll, J. C., Crutch, S. J., & Warren, J. D. (2010). Central auditory disorders: Toward a neuropsychology of auditory objects. *Current opinion in neurology*, *23*(6), 617–627. doi:10.1097/WCO.0b013e32834027f6

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25.

Goodman, N. (1972). Seven strictures on similarity. In *Problems and projects*. Bobs-Merril.

Gorgolewski, K. J., Alfaro-Almagro, F., Auer, T., Bellec, P., Capotă, M., Chakravarty, M. M., … Poldrack, R. A. (2017). BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLOS Computational Biology*, *13*(3), e1005209. doi:10.1371/journal.pcbi.1005209

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., … Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*, 160044. doi:10.1038/sdata.2016.44

Gorgolewski, K. J., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics*, *5*, 13. doi:10.3389/fninf.2011.00013

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the p600. *Language and Cognitive Processes*, *25*(2), 149–188. doi:10.1080/01690960902965951

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., … Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, *86*, 446–460. Publisher: Elsevier Inc. ISBN: 2122633255. doi:10.1016/j.neuroimage.2013.10.027. arXiv: NIHMS150003

Greenberg, D. M., Kosinski, M., Stillwell, D. J., Monteiro, B. L., Levitin, D. J., & Rentfrow, P. J. (2016). The song is you: Preferences for musical attribute dimensions reflect personality. *Social Psychological and Personality Science*, *7*(6), 597–605. doi:10.1177/1948550616641473

Griffiths, T. D. [T. D.], Johnsrude, I., Dean, J. L., & Green, G. G. (1999). A common neural substrate for the analysis of pitch and duration pattern in segmented sound? *Neuroreport*, *10*(18), 3825–3830.

Griffiths, T. D. [Timothy D.], & Hall, D. A. (2012). Mapping pitch representation in neural ensembles with fMRI. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(39), 13343–13347. doi:10.1523/JNEUROSCI.3813-12.2012

Griffiths, T. D. [Timothy D.], & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences*, *25*(7), 348–353.

Groen, M. A., Whitehouse, A. J. O., Badcock, N. A., & Bishop, D. V. M. (2013). Associations between handedness and cerebral lateralisation for language: A comparison of three measures in children. *PLoS ONE*, *8*(5). doi:10.1371/journal.pone.0064876

Grosche, P., Muller, M., & Kurth, F. (2010). Cyclic tempogram - a mid-level tempo representation for musicsignals. *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. Retrieved from http://www.academia.edu/30814095/CYCLIC_TEMPOGRAM_A_MID-LEVEL_TEMPO_REPRESENTATION_FOR_MUSICSIGNALS

Güçlü, U., Thielen, J., Hanke, M., & van Gerven, M. (2016). Brains on beats. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in neural information processing systems 29* (pp. 2101–2109). Curran Associates, Inc. Retrieved from http://papers.nips.cc/paper/6222-brains-on-beats.pdf

Guggenmos, M., Sterzer, P., & Cichy, R. M. (2018). Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. *NeuroImage*, *173*, 434–447. doi:10.1016/j.neuroimage.2018.02.044

Habib, M., & Besson, M. (2009). What do music training and musical experience teach us about brain plasticity? *Music Perception: An Interdisciplinary Journal*, *26*(3), 279–285. ISBN: 0520057295. doi:10.1525/mp.2009.26.3.279

Hackett, T. A., Preuss, T. M., & Kaas, J. H. (2001). Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *The Journal of Comparative Neurology*, *441*(3), 197–222.

Hackett, T. A., Stepniewska, I., & Kaas, J. H. (1998). Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *The Journal of Comparative Neurology*, *394*(4), 475–495.

Hadjidimitriou, S. K., & Hadjileontiadis, L. J. (2012). Toward an EEG-based recognition of music liking using time-frequency analysis. *IEEE transactions on bio-medical engineering*, *59*(12), 3498–3510. doi:10.1109/TBME.2012.2217495

Häkkinen, S., Ovaska, N., & Rinne, T. (2015). Processing of pitch and location in human auditory cortex during visual and auditory tasks. *Frontiers in Psychology*, *6*. doi:10.3389/fpsyg.2015.01678

Häkkinen, S., & Rinne, T. (2018). Intrinsic, stimulus-driven and task-dependent connectivity in human auditory cortex. *Brain Structure & Function*, *223*(5), 2113–2127. doi:10.1007/s00429-018-1612-6

Hall, D. A., Chambers, J., Akeroyd, M. A., Foster, J. R., Coxon, R., & Palmer, A. R. (2009). Acoustic, psychophysical, and neuroimaging measurements of the effectiveness of active cancellation during auditory functional magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *125*(1), 347–359. doi:10.1121/1.3021437

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Summerfield, A. Q., Palmer, A. R., Elliott, M. R., & Bowtell, R. W. (2000). Modulation and task effects in auditory processing measured using fMRI. *Human Brain Mapping*, *10*(3), 107–119.

Haller, S., & Bartsch, A. (2009). *Pitfalls in fMRI*. doi:10.1007/s00330-009-1456-9

Halpern, A. R. (2001). Cerebral substrates of musical imagery. *Annals of the New York Academy of Sciences*, *930*, 179–192.

Halpern, A. R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia*, *42*(9), 1281–1292. ISBN: 1570577129. doi:10.1016/j.neuropsychologia.2003.12.017

Hamilton, L. S., & Huth, A. G. (2018). The revolution will not be controlled: Natural stimuli in speech neuroscience. *Language, Cognition and Neuroscience*, *0*(0), 1–10. doi:10.1080/23273798.2018.1499946

Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., … Fischl, B. (2006). Reliability of MRI-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer. *NeuroImage*, *32*(1), 180–194. doi:10.1016/j.neuroimage.2006.02.051

Hanke, M., Dinga, R., Häusler, C., Guntupalli, J. S., Casey, M., Kaule, F. R., & Stadler, J. (2015). High-resolution 7-tesla fMRI data on the perception of musical genres – an extension to the studyforrest dataset. *F1000Research*, (0). doi:10.12688/f1000research.6679.1

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V., & Pollmann, S. (2009). PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics*, *7*(1), 37–53. ISBN: 1559-0089 (Electronic)\n1539-2791 (Linking). doi:10.1007/s12021-008-9041-y

Harms, M. P., & Melcher, J. R. (2002). Sound repetition rate in the human auditory pathway: Representations in the waveshape and amplitude of fMRI activation. *Journal of Neurophysiology*, *88*(3), 1433–1450. doi:10.1152/jn.2002.88.3.1433

Harnsberger, L. C. (1996). *Essential dictionary of music: Definitions, composers, theory, instrument & vocal ranges*. Google-Books-ID: lrjbxAuZYeYC. Alfred Music Publishing.

Harwood, D. L. (1976). Universals in music: A perspective from cognitive psychology. *Ethnomusicology*, *20*(3), 521–533. doi:10.2307/851047

Hasson, U., Nusbaum, H. C., & Small, S. L. (2007). Brain networks subserving the extraction of sentence information and its encoding to memory. *Cerebral Cortex*, *17*(12), 2899–2913. doi:10.1093/cercor/bhm016

Hattori, Y., Fukatsu, H., & Ishigaki, T. (2007). Measurement and evaluation of the acoustic noise of a 3 tesla MR scanner. *Nagoya Journal of Medical Science*, *69*(1), 23–28.

Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: The early beginnings. *NeuroImage*, *62*(2), 852–855. Publisher: Elsevier Inc. ISBN: 1095-9572 (Electronic)\r1053-8119 (Linking). doi:10.1016/j.neuroimage.2012.03.016

Haxby, J. V., Connolly, A. C., & Guntupalli, S. J. (2014). Decoding neural representational spaces using multivariate pattern analysis. *Annual Review of Neuroscience*, *37*(1), 435–456. doi:10.1146/annurev-neuro-062012-170325

Haynes, J. D. (2015). A primer on pattern-based approaches to fMRI: Principles, pitfalls, and perspectives. *Neuron*, *87*(2), 257–270. ISBN: 1097-4199 (Electronic)\r0896-6273 (Linking). doi:10.1016/j.neuron.2015.05.025

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*(2), 61–83, discussion 83–135. doi:10.1017/S0140525X0999152X

Herdener, M., Esposito, F., Scheffler, K., Schneider, P., Logothetis, N. K., Uludag, K., & Kayser, C. (2013). Spatial representations of temporal and spectral sound cues in human auditory cortex. *Cortex*, *49*(10), 2822–2833. ISBN: 1973-8102 (Electronic)\r0010-9452 (Linking). doi:10.1016/j.cortex.2013.04.003

Herholz, S. C., Lappe, C., Knief, A., & Pantev, C. (2008). Neural basis of music imagery and the effect of musical expertise. *European Journal of Neuroscience*, *28*(11), 2352–2360. ISBN: 1460-9568. doi:10.1111/j.1460-9568.2008.06515.x

Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: Behavior, function, and structure. *Neuron*, *76*(3), 486–502. doi:10.1016/j.neuron.2012.10.011

Hertz, U., & Amedi, A. (2010). Disentangling unisensory and multisensory components in audiovisual integration using a novel multifrequency fMRI spectral analysis. *NeuroImage*, *52*(2), 617–632. doi:10.1016/j.neuroimage.2010.04.186

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. doi:10.1038/nrn2113

Hill, R. W., Wyse, G. A., & Anderson, M. (2016). *Animal physiology*. Oxford University Press. Retrieved from https://global.oup.com/ushe/product/animal-physiology-9781605354712?cc=de&lang=en&

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science (New York, N.Y.) 182*(4108), 177–180.

Hjortkjær, J., Kassuba, T., Madsen, K. H., Skov, M., & Siebner, H. R. (2018). Task-modulated cortical representations of natural sound source categories. *Cerebral Cortex*, *28*(1), 295–306. doi:10.1093/cercor/bhx263

Hoefle, S., Engel, A., Basilio, R., Alluri, V., Toiviainen, P., Cagy, M., & Moll, J. (2018). Identifying musical pieces from fMRI data using encoding and decoding models. *Scientific Reports*, *8*(1), 2266. Publisher: Nature Publishing Group. doi:10.1038/s41598-018-20732-3

Holt, F. (2007). *Genre in popular music*. Google-Books-ID: Ek8ssV0pv_IC. University of Chicago Press.

Hopf, A. (1954). Die myeloarchitektonik des isocortex temporalis beim menschen. *1*, 443–496.

Hubl, D., Dougoud-Chauvin, V., Zeller, M., Federspiel, A., Boesch, C., Strik, W., ... Koenig, T. (2010). Structural analysis of heschl's gyrus in schizophrenia patients with auditory hallucinations. *Neuropsychobiology*, *61*(1), 1–9. doi:10.1159/000258637

Humphries, C., Liebenthal, E., & Binder, J. R. (2010). Tonotopic organization of human auditory cortex. *NeuroImage*, *50*(3), 1202–1211. Publisher: Elsevier Inc. ISBN: 1095-9572 (Electronic) 1053-8119 (Linking). doi:10.1016/j.neuroimage.2010.01.046. arXiv: NIHMS150003

Humphries, C., Sabri, M., Lewis, K., & Liebenthal, E. (2014). Hierarchical organization of speech perception in human auditory cortex. *Frontiers in Neuroscience*, *8*. doi:10.3389/fnins.2014.00406

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science Engineering*, *9*(3), 90–95. doi:10.1109/MCSE.2007.55

Hurless, N., Mekic, A., Peña, S., Humphries, E., Gentry, H., & Nichols, D. F. (2013). Music genre preference and tempo alter alpha and beta waves in human nonmusicians .

Intartaglia, B., White-Schwoch, T., Kraus, N., & Schön, D. (2017). Music training enhances the automatic neural processing of foreign speech sounds. *Scientific Reports*, *7*(1), 12631. doi:10.1038/s41598-017-12575-1

Istók, E., Brattico, E., Jacobsen, T., Ritter, A., & Tervaniemi, M. (2013). 'i love rock 'n' roll'-music genre preference modulates brain responses to music. *Biological Psychology*, *92*(2), 142–151. ISBN: 0301-0511. doi:10.1016/j.biopsycho.2012.11.005

Jacoby, N., & McDermott, J. H. (2017). Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Current biology: CB*, *27*(3), 359–370. doi:10.1016/j.cub.2016.12.031

Jacoby, N., Undurraga, E. A., McPherson, M. J., Valdés, J., Ossandón, T., & McDermott, J. H. (2019). Universal and non-universal features of musical pitch perception revealed by singing. *Current Biology*. doi:10.1016/j.cub.2019.08.020

Jamini, D. (2005). *Harmony and composition: Basics to intermediate*. Google-Books-ID: uluN-ngEACAAJ. Trafford Publishing.

Jäncke, L. (2009). Music drives brain plasticity. *F1000 Biology Reports*, *1*. doi:10.3410/B1-78

Jäncke, L., Buchanan, T. W., Lutz, K., & Shah, N. J. (2001). Focused and nonfocused attention in verbal and emotional dichotic listening: An FMRI study. *Brain and Language*, *78*(3), 349–363. doi:10.1006/brln.2000.2476

Javitt, D. C., & Sweet, R. A. (2015). Auditory dysfunction in schizophrenia: Integrating clinical and basic features. *Nature reviews. Neuroscience*, *16*(9), 535–550. doi:10.1038/nrn4002

Jeon, J. Y., & Fricke, F. R. (1997). Duration of perceived and performed sounds. *Psychology of Music*, *25*(1), 70–83. doi:10.1177/0305735697251006

Jewett, D. L., & Williston, J. S. (1971). Auditory-evoked far fields averaged from the scalp of humans. *Brain: A Journal of Neurology*, *94*(4), 681–696.

Joanisse, M. F., & DeSouza, D. D. (2014). Sensitivity of human auditory cortex to rapid frequency modulation revealed by multivariate representational similarity analysis. *Frontiers in Neuroscience*, *8*, 1–10. doi:10.3389/fnins.2014.00306

Jones, E., Oliphant, E., & Peterson, P. (2001). Scipy: Open source scientific tools for python.

Jovicich, J., Czanner, S., Greve, D., Haley, E., van der Kouwe, A., Gollub, R., . . . Dale, A. (2006). Reliability in multi-site structural MRI studies: Effects of gradient non-linearity correction on phantom and human data. *NeuroImage*, *30*(2), 436–443. doi:10.1016/j.neuroimage.2005.09.046

Juslin, P. N., & Sloboda, J. (Eds.). (2011). *Handbook of music and emotion: Theory, research, applications*. Oxford, New York: Oxford University Press.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The p600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*(2), 159–201. doi:10.1080/016909600386084

Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience*, *15*(1), 98–110. doi:10.1162/089892903321107855

Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(22), 11793–11799. doi:10.1073/pnas.97.22.11793

Kaas, J. H., Hackett, T. A., & Tramo, M. J. (1999). Auditory processing in primate cerebral cortex. *Current Opinion in Neurobiology*, *9*(2), 164–170.

Kajikawa, Y., Frey, S., Ross, D., Falchier, A., Hackett, T. A., & Schroeder, C. E. (2015). Auditory properties in the parabelt regions of the superior temporal gyrus in the awake macaque monkey: An initial survey. *The Journal of Neuroscience*, *35*(10), 4140–4150. doi:10.1523/JNEUROSCI.3556-14.2015

Kaplan, J. T., & Meyer, K. (2012). Multivariate pattern analysis reveals common neural patterns across individuals during touch observation. *NeuroImage*, *60*(1), 204–212. Publisher: Elsevier Inc. ISBN: 1053-8119. doi:10.1016/j.neuroimage.2011.12.059. arXiv: NIHMS150003

Karmonik, C., Brandt, A., Anderson, J., Brooks, F., Lytle, J., Silverman, E., & Frazier, J. T. (2016). Music listening modulates functional connectivity and information flow in the human brain. *Brain Connectivity*. doi:10.1089/brain.2016.0428

Kauramäki, J., Jääskeläinen, I. P., & Sams, M. (2007). Selective attention increases both gain and feature selectivity of the human auditory cortex. *PLOS ONE*, *2*(9), e909. doi:10.1371/journal.pone.0000909

Kell, A. J., Yamins, D. L., Shook, E. N., Norman-Haignere, S. V., & McDermott, J. H. (2018). A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. *Neuron*, *98*(3), 630–644.e16. doi:10.1016/j.neuron.2018.03.044

Kemp, D. T. (2002). Otoacoustic emissions, their origin in cochlear function, and use. *British Medical Bulletin*, *63*(1), 223–241. doi:10.1093/bmb/63.1.223

Keshavan, A., Datta, E., M McDonough, I., Madan, C. R., Jordan, K., & Henry, R. G. (2018). Mindcontrol: A web application for brain segmentation quality control. *NeuroImage*, *170*, 365–372. doi:10.1016/j.neuroimage.2017.03.055

Kilian-Hütten, N., Valente, G., Vroomen, J., & Formisano, E. (2011). Auditory cortex encodes the perceptual interpretation of ambiguous sound. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *31*(5), 1715–1720. ISBN: 0270-6474\r1529-2401. doi:10.1523/JNEUROSCI.4572-10.2011

Kim, J.-J., Crespo-Facorro, B., Andreasen, N. C., O'Leary, D. S., Zhang, B., Harris, G., & Magnotta, V. A. (2000). An MRI-based parcellation method for the temporal lobe. *NeuroImage*, *11*(4), 271–288. Publisher: Academic Press. doi:10.1006/NIMG.2000.0543

Klapuri, A., & Davy, M. (2007). *Signal processing methods for music transcription*. Google-Books-ID: AF30yR41GIAC. Springer Science & Business Media.

Klauschen, F., Goldman, A., Barra, V., Meyer-Lindenberg, A., & Lundervold, A. (2009). Evaluation of automated brain MR image segmentation and volumetry methods. *Human Brain Mapping*, *30*(4), 1310–1327. doi:10.1002/hbm.20599

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., … Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLOS Computational Biology*, *13*(2), e1005350. doi:10.1371/journal.pcbi.1005350

Klein, M. E., & Zatorre, R. J. (2011). A role for the right superior temporal sulcus in categorical perception of musical chords. *Neuropsychologia*, *49*(5), 878–887. Publisher: Elsevier Ltd ISBN: 1873-3514 (Electronic)\r0028-3932 (Linking). doi:10.1016/j.neuropsychologia.2011.01.008

Klem, G. H., Lüders, H. O., Jasper, H. H., & Elger, C. (1999). The ten-twenty electrode system of the international federation. the international federation of clinical neurophysiology. *Electroencephalography and Clinical Neurophysiology. Supplement*, *52*, 3–6.

Koelsch, S. (2009a). Music-syntactic processing and auditory memory: Similarities and differences between ERAN and MMN. *Psychophysiology*, *46*(1), 179–190. doi:10.1111/j.1469-8986.2008.00752.x

Koelsch, S. (2009b). Neural substrates of processing syntax and semantics in music. *Music that Works: Contributions of Biology, Neurophysiology, Psychology, Sociology, Medicine and Musicology*, 143–153. ISBN: 9783211751206. doi:10.1007/978-3-211-75121-3_9

Koelsch, S. (2011a). Toward a neural basis of music perception - a review and updated model. *Frontiers in Psychology, 2,* 1–20. ISBN: 1664-1078 (Electronic). doi:10.3389/fpsyg.2011.00110

Koelsch, S. (2011b). Towards a neural basis of processing musical semantics. *Physics of Life Reviews*, *8*(2), 89–105. Publisher: Elsevier B.V. ISBN: 1571-0645. doi:10.1016/j.plrev.2011.04.004

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nature reviews. Neuroscience*, *15*(3), 170–180. ISBN: 1471-003X. doi:10.1038/nrn3666

Koelsch, S. (2015). Music-evoked emotions: Principles, brain correlates, and implications for therapy. *Annals of the New York Academy of Sciences*, *1337*(1), 193–201. doi:10.1111/nyas.12684

Koelsch, S., & Friederici, A. D. (2003). Toward the neural basis of processing structure in music. comparative results of different neurophysiological investigation methods. *Annals of the New York Academy of Sciences*, *999*, 15–28. ISBN: 0077-8923 (Print)\r0077-8923 (Linking). doi:10.1196/annals.1284.002

Koelsch, S., Fritz, T., Schulze, K., Alsop, D., & Schlaug, G. (2005). Adults and children processing music: An fMRI study. *NeuroImage*, *25*(4), 1068–1076. ISBN: 1053-8119 (Print)\r1053-8119 (Linking). doi:10.1016/j.neuroimage.2004.12.050

Koelsch, S., Gunter, T. C., von Cramon, D. Y., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: A cortical "language-network"serves the processing of music. *NeuroImage*, *17*(2), 956–966. ISBN: 1053-8119 (Print) 1053-8119. doi:10.1016/S1053-8119(02)91154-7

Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and in music: An ERP study. *Journal of cognitive neuroscience*, *17*(10), 1565–1577. ISBN: 0898-929X. doi:10.1162/089892905774597290

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience*, *7*(3), 302–307. ISBN: 1097-6256 (Print) 1097-6256. doi:10.1038/nn1197

Koelsch, S., Rohrmeier, M., Torrecuso, R., & Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(38), 15443–8. ISBN: 1091-6490 (Electronic)\r0027-8424 (Linking). doi:10.1073/pnas.1300272110

Koelsch, S., Schröger, E., & Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *Neuroreport*, *10*(6), 1309–1313.

Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences*, *9*(12), 578–584. ISBN: 1364-6613 (Print)\r1364-6613 (Linking). doi:10.1016/j.tics.2005.10.001

Kolossa, A., & Kopp, B. (2018). Data quality over data quantity in computational cognitive neuroscience. *NeuroImage*, *172*, 775–785. doi:10.1016/j.neuroimage.2018.01.005

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience needs behavior: Correcting a reductionist bias. *Neuron*, *93*(3), 480–490. doi:10.1016/j.neuron.2016.12.041

Krause, J. (2012). Human origins and the search for “missing links”. *PLoS Biology*, *10*(5). doi:10.1371/journal.pbio.1001333

Krauss, T., & Galloway, H. (1982). Melodic intonation therapy with language delayed apraxic children. *Journal of Music Therapy*, *19*(2), 102–113. doi:10.1093/jmt/19.2.102

Kreutz, G., Ott, U., Teichmann, D., Osawa, P., & Vaitl, D. (2008). Using music to induce emotions: Influences of musical preference and absorption. *Psychology of Music*, *36*(1), 101–126. doi:10.1177/0305735607082623

Kriegeskorte, N. (2011). Pattern-information analysis: From stimulus decoding to computational-model testing. *NeuroImage*, *56*(2), 411–21. doi:10.1016/j.neuroimage.2011.01.061

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868. ISBN: 0027-8424 (Print)\r0027-8424 (Linking). doi:10.1073/pnas.0600244103

Kriegeskorte, N., & Mur, M. (2012). Inverse MDS: Inferring dissimilarity structure from multiple item arrangements. *Frontiers in Psychology*, *3*, 1–13. doi:10.3389/fpsyg.2012.00245

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, *2*, 4. doi:10.3389/neuro.06.004.2008

Kriegstein, K. V., & Giraud, A.-L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *NeuroImage*, *22*(2), 948–955. doi:10.1016/j.neuroimage.2004.02.020

Kulesza, R. J. (2007). Cytoarchitecture of the human superior olivary complex: Medial and lateral superior olive. *Hearing Research*, *225*(1), 80–90. doi:10.1016/j.heares.2006.12.006

Kumar, S., Kriegstein, K. v., Friston, K. J., & Griffiths, T. D. (2012). Features versus feelings: Dissociable representations of the acoustic features and valence of aversive sounds. *Journal of Neuroscience*, *32*(41), 14184–14192. doi:10.1523/JNEUROSCI.1759-12.2012

Kumar, S., Stephan, K. E., Warren, J. D., Friston, K. J., & Griffiths, T. D. (2007). Hierarchical processing of auditory objects in humans. *PLoS Computational Biology*, *3*(6), e100. Publisher: Public Library of Science. doi:10.1371/journal.pcbi.0030100

Kwoun, S.-J. (2009). An examination of cue redundancy theory in cross-cultural decoding of emotions in music. *Journal of Music Therapy*, *46*(3), 217–237.

Langers, D. R. M. (2014). Assessment of tonotopically organised subdivisions in human auditory cortex using volumetric and surface-based cortical alignments. *Human Brain Mapping*, *35*(4), 1544–1561. doi:10.1002/hbm.22272

Langers, D. R. M., Backes, W. H., & van Dijk, P. (2007). Representation of lateralization and tonotopy in primary versus secondary human auditory cortex. *NeuroImage*, *34*(1), 264–273. ISBN: 1053-8119 (Print)\r1053-8119 (Linking). doi:10.1016/j.neuroimage.2006.09.002

Langers, D. R. M., de Kleine, E., & van Dijk, P. (2012). Tinnitus does not require macroscopic tonotopic map reorganization. *Frontiers in Systems Neuroscience*, *6*. doi:10.3389/fnsys.2012.00002

Langers, D. R. M., & Van Dijk, P. (2011). Robustness of intrinsic connectivity networks in the human brain to the presence of acoustic scanner noise. *NeuroImage*, *55*(4), 1617–1632. Publisher: Elsevier Inc. ISBN: 1053-8119. doi:10.1016/j.neuroimage.2011.01.019

Langers, D. R. M., Van Dijk, P., & Backes, W. H. (2005). Interactions between hemodynamic responses to scanner acoustic noise and auditory stimuli in functional magnetic resonance imaging. *Magnetic Resonance in Medicine*, *53*(1), 49–60. ISBN: 0740-3194 (Print)\r0740-3194. doi:10.1002/mrm.20315

Lartillot, O., Toiviainen, P., & Eerola, T. (2008). A matlab toolbox for music information retrieval. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, & R. Decker (Eds.), *Data analysis, machine learning and applications* (pp. 261–268). Studies in Classification, Data Analysis, and Knowledge Organization. Springer Berlin Heidelberg.

Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: Effects of acoustic features and auditory object category. *The Journal of neuroscience : the* *official journal of the Society for Neuroscience*, *30*(22), 7604–7612. doi:10.1523/JNEUROSCI.0296-10.2010

Leaver, A. M., & Rauschecker, J. P. (2016). Functional topography of human auditory cortex. *Journal of Neuroscience*, *36*(4), 1416–1428. doi:10.1523/JNEUROSCI.0226-15.2016

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*(2), 365–411. doi:10.1016/S0047-259X(03)00096-4

Lena, J. C., & Peterson, R. A. (2008). Classification as culture: Types and trajectories of music genres. *American Sociological Review*, *73*(5), 697–718. Retrieved from https://www.jstor.org/stable/25472554

Leon, F. A. d., & Martinez, K. (2014). Music genre classification using polyphonic timbre models. In *2014 19th international conference on digital signal processing* (pp. 415–420). 2014 19th international conference on digital signal processing. doi:10.1109/ICDSP.2014.6900697

Leonard, C. M., Eckert, M. A., Lombardino, L. J., Oakland, T., Kranzler, J., Mohr, C. M., … Freeman, A. (2001). Anatomical risk factors for phonological dyslexia. *Cerebral Cortex (New York, N.Y.: 1991)*, *11*(2), 148–157.

Levitin, D. J., & Grafton, S. T. (2016). Measuring the representational space of music with fMRI: A case study with sting. *Neurocase*, *00*(0), 1–10. Publisher: Routledge. doi:10.1080/13554794.2016.1216572

Levitin, D. J., & 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), 2142–2152. ISBN: 1053-8119 (Print). doi:10.1016/j.neuroimage.2003.08.016

Li, H.-C., Wang, H.-H., Chou, F.-H., & Chen, K.-M. (2015). The effect of music therapy on cognitive functioning among older adults: A systematic review and meta-analysis. *Journal of the American Medical Directors Association*, *16*(1), 71–77. doi:10.1016/j.jamda.2014.10.004

Li, M., Rudd, B., Lim, T. C., & Lee, J.-H. (2011). In situ active control of noise in a 4 t MRI scanner. *Journal of magnetic resonance imaging: JMRI*, *34*(3), 662–669. doi:10.1002/jmri.22694

Li, X., Morgan, P. S., Ashburner, J., Smith, J., & Rorden, C. (2016). The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *Journal of Neuroscience Methods*, *264*, 47–56. doi:10.1016/j.jneumeth.2016.03.001

Liao, C., Worsley, K., Poline, J.-B., Aston, J., Duncan, G., & Evans, A. (2002). Estimating the delay of the fMRI response. *NeuroImage*, *16*(3), 593–606. doi:10.1006/nimg.2002.1096

Lindeberg, T., & Friberg, A. (2015). Idealized computational models for auditory receptive fields. *PLOS ONE*, *10*(3), e0119032. doi:10.1371/journal.pone.0119032

Linnemann, A., Ditzen, B., Strahler, J., Doerr, J. M., & Nater, U. M. (2015). Music listening as a means of stress reduction in daily life. *Psychoneuroendocrinology*, *60*, 82–90. Publisher: Elsevier Ltd ISBN: 1873-3360 (Electronic)\r0306-4530 (Linking). doi:10.1016/j.psyneuen.2015.06.008

Linnemann, A., Strahler, J., & Nater, U. M. (2016). The stress-reducing effect of music listening varies depending on the social context. *Psychoneuroendocrinology*, *72*, 97–105. Publisher: Elsevier Ltd. doi:10.1016/j.psyneuen.2016.06.003

Livingstone, S. R., Peck, K., & Russo, F. A. (2013). Acoustic differences in the speaking and singing voice. *Proceedings of Meetings on Acoustics*, *19*(1), 035080. doi:10.1121/1.4799460

Logan, B. (2000). Mel frequency cepstral coefficients for music modeling. In *In international symposium on music information retrieval.*

Luo, C., Guo, Z. w., Lai, Y. x., Liao, W., Liu, Q., Kendrick, K. M., … Li, H. (2012). Musical training induces functional plasticity in perceptual and motor networks: Insights from resting-state fMRI. *PLoS ONE*, *7*(5), 1–10. ISBN: 1932-6203 (Electronic) 1932-6203 (Linking). doi:10.1371/journal.pone.0036568

Lutti, A., Dick, F., Sereno, M. I., & Weiskopf, N. (2014). Using high-resolution quantitative mapping of r1 as an index of cortical myelination. *NeuroImage*, *93 Pt 2*, 176–188. doi:10.1016/j.neuroimage.2013.06.005

Madsen, S. M. K., Whiteford, K. L., & Oxenham, A. J. (2017). Musicians do not benefit from differences in fundamental frequency when listening to speech in competing speech backgrounds. *Scientific reports*, *7*(1), 12624. doi:10.1038/s41598-017-12937-9

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in broca's area: An MEG study. *Nature Neuroscience*, *4*(5), 540–545. ISBN: 1097-6256 (Print)\r1097-6256 (Linking). doi:10.1038/87502

Makris, N., Goldstein, J. M., Kennedy, D., Hodge, S. M., Caviness, V. S., Faraone, S. V., … Seidman, L. J. (2006). Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophrenia Research*, *83*(2), 155–171. doi:10.1016/j.schres.2005.11.020

Maratos, A., Gold, C., Wang, X., & Crawford, M. (2008). Music therapy for depression. *Cochrane Database of Systematic Reviews*, (1). doi:10.1002/14651858.CD004517.pub2

Marie, C., Kujala, T., & Besson, M. (2012). Musical and linguistic expertise influence pre-attentive and attentive processing of non-speech sounds. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *48*(4), 447–457. doi:10.1016/j.cortex.2010.11.006

Marie, D., Jobard, G., Crivello, F., Perchey, G., Petit, L., Mellet, E., … Tzourio-Mazoyer, N. (2015). Descriptive anatomy of heschl's gyri in 430 healthy volunteers, including 198 left-handers. *Brain Structure & Function*, *220*(2), 729–743. doi:10.1007/s00429-013-0680-x

Marie, D., Maingault, S., Crivello, F., Mazoyer, B., & Tzourio-Mazoyer, N. (2016). Surface-based morphometry of cortical thickness and surface area associated with heschl's gyri duplications in 430 healthy volunteers. *Frontiers in Human Neuroscience*, *10*. doi:10.3389/fnhum.2016.00069

Martínez-Molina, N., Mas-Herrero, E., Rodríguez-Fornells, A., Zatorre, R. J., & Marco-Pallarés, J. (2016). Neural correlates of specific musical anhedonia. *Proceedings of the National Academy of Sciences*, 201611211. doi:10.1073/PNAS.1611211113

Mas-Herrero, E., Dagher, A., & Zatorre, R. J. (2018). Modulating musical reward sensitivity up and down with transcranial magnetic stimulation. *Nature Human Behaviour*, *2*(1), 27. doi:10.1038/s41562-017-0241-z

Mas-Herrero, E., Zatorre, R. J., Rodriguez-Fornells, A., & Marco-Pallarés, J. (2014). Dissociation between musical and monetary reward responses in specific musical anhedonia. *Current Biology*, *24*(6), 699–704. ISBN: 1879-0445 (Electronic)\n0960-9822 (Linking). doi:10.1016/j.cub.2014.01.068

McAdams, S., & Bregman, A. (1979). Hearing musical streams. *Computer Music Journal*, *3*(4), 26–43+60. Retrieved from http://www.jstor.org/stable/4617866

McClellan, R. (2000). *The healing forces of music: History, theory, and practice*. Google-Books-ID: Dq1LAQAAIAAJ. ToExcel.

McDermott, H. J. (2004). Music perception with cochlear implants: A review. *Trends in Amplification*, *8*(2), 49–82. doi:10.1177/108471380400800203

McDermott, J. H., & Hauser, M. D. (2004). Are consonant intervals music to their ears? spontaneous acoustic preferences in a nonhuman primate. *Cognition*, *94*(2), B11–B21. doi:10.1016/j.cognition.2004.04.004

McDermott, J. H., & Hauser, M. D. (2007). Nonhuman primates prefer slow tempos but dislike music overall. *Cognition*, *104*(3), 654–668. ISBN: 0010-0277. doi:10.1016/j.cognition.2006.07.011

McDermott, J. H., Schultz, A. F., Undurraga, E. A., & Godoy, R. A. (2016). Indifference to dissonance in native amazonians reveals cultural variation in music perception. *Nature*, *535*(7613), 547–550. Publisher: Nature Publishing Group ISBN: 0028-0836. doi:10.1038/nature18635

McFee, B., Raffel, C., Liang, D., Ellis, D., McVicar, M., Battenberg, E., & Nieto, O. (2015). Librosa: Audio and music signal analysis in python. (pp. 18–24). Python in science conference. doi:10.25080/Majora-7b98e3ed-003

McKinney, W. (2010). Data structures for statistical computing in python. (pp. 51–56). Proceedings of the 9th python in science conference. Retrieved January 10, 2019, from http://conference.scipy.org/proceedings/scipy2010/mckinney.html

McPherson, M. J., & McDermott, J. H. (2018). Diversity in pitch perception revealed by task dependence. *Nature Human Behaviour*, *2*(1), 52–66. doi:10.1038/s41562-017-0261-8

Medial geniculate nucleus. (2019), In *Wikipedia*. Page Version ID: 920069502. Retrieved January 5, 2019, from https://en.wikipedia.org/w/index.php?title=Medial_geniculate_nucleus&oldid=920069502

Mehr, S. A., Singh, M., York, H., Glowacki, L., & Krasnow, M. M. (2018). Form and function in human song. *Current biology: CB*, *28*(3), 356–368.e5. doi:10.1016/j.cub.2017.12.042

Mendoza, J. E. (2011). Trapezoid body. In J. S. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of clinical neuropsychology* (pp. 2549–2549). doi:10.1007/978-0-387-79948-3_807

Merkel, D. (2014). Docker: Lightweight linux containers for consistent development and deployment. *Linux J. 2014*(239). Retrieved from http://dl.acm.org/citation.cfm?id=2600239.2600241

Merrill, J., Sammler, D., Bangert, M., Goldhahn, D., Lohmann, G., Turner, R., & Friederici, A. D. (2012). Perception of words and pitch patterns in song and speech. *Frontiers in Psychology*, *3*. doi:10.3389/fpsyg.2012.00076

Merzenich, M. M., & Brugge, J. F. (1973). Representation of the cochlear partition of the superior temporal plane of the macaque monkey. *Brain Research*, *50*(2), 275–296.

Meyer, L. B. (1973). *Explaining music: Essays and explorations*. Google-Books-ID: 3_9zjvU2pyYC. University of California Press.

Meyer, L. B. (1989). *Style and music: Theory, history, and ideology*. Google-Books-ID: hYLuAAAAMAAJ. University of Pennsylvania Press.

Meyer, L. B., Narmour, E., & Solie, R. A. (1988). *Explorations in music, the arts, and ideas: Essays in honor of leonard b. meyer*. Google-Books-ID: Ym2LglVhl1QC. Pendragon Press.

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, *219*(1), 36–47. doi:10.1016/j.heares.2006.05.004

Middlebrooks, J. C., & Bremen, P. (2013). Spatial stream segregation by auditory cortical neurons. *The Journal of Neuroscience*, *33*(27), 10986–11001. doi:10.1523/JNEUROSCI.1065-13.2013

Millman, K. J., Brett, M., Barnowski, R., & Poline, J.-B. (2018). Teaching computational reproducibility for neuroimaging. *Frontiers in Neuroscience*, *12*. doi:10.3389/fnins.2018.00727

Minati, L., Rosazza, C., D'Incerti, L., Pietrocini, E., Valentini, L., Scaioli, V., … Bruzzone, M. G. (2008). FMRI/ERP of musical syntax: Comparison of melodies and unstructured note sequences. *Neuroreport*, *19*(14), 1381–5. doi:10.1097/WNR.0b013e32830c694b

Mišic, B., Betzel, R. F., Griffa, A., de Reus, M. A., He, Y., Zuo, X.-N., … Zatorre, R. J. (2018). Network-based asymmetry of the human auditory system. *Cerebral Cortex (New York, N.Y.: 1991)*, *28*(7), 2655–2664. doi:10.1093/cercor/bhy101

Mithen, S., Morley, I., Wray, A., Tallerman, M., & Gamble, C. (2006). The singing neanderthals: The origins of music, language, mind and body. *Cambridge Archaeological Journal*, *16*(1), 97–112. ISBN: 0297643177. doi:10.1017/S0959774306000060

Moelker, A., Maas, R. A. J. J., Lethimonnier, F., & Pattynama, P. M. T. (2002). Interventional MR imaging at 1.5 t: Quantification of sound exposure. *Radiology*, *224*(3), 889–895. doi:10.1148/radiol.2243010978

Moelker, A., & Pattynama, P. M. T. (2003). Acoustic noise concerns in functional magnetic resonance imaging. *Human Brain Mapping*, *20*(3), 123–141. doi:10.1002/hbm.10134

Moerel, M., De Martino, F., & Formisano, E. (2014). An anatomical and functional topography of human auditory cortical areas. *Frontiers in Neuroscience*, *8*(8), 1–14. ISBN: 1662-453X. doi:10.3389/fnins.2014.00225

Moerel, M., De Martino, F., Santoro, R., Ugurbil, K., Goebel, R., Yacoub, E., & Formisano, E. (2013). Processing of natural sounds : Characterization of multipeak spectral tuning in human auditory cortex. *33*(29), 11888–11898. doi:10.1523/JNEUROSCI.5306-12.2013

Moerel, M., De Martino, F., Uğurbil, K., Yacoub, E., & Formisano, E. (2015). Processing of frequency and location in human subcortical auditory structures. *Scientific reports*, *5*, 17048. Publisher: Nature Publishing Group. doi:10.1038/srep17048

Moerel, M., Martino, F. D., & Formisano, E. (2012). Processing of natural sounds in human auditory cortex: Tonotopy, spectral tuning, and relation to voice sensitivity. *Journal of Neuroscience*, *32*(41), 14205–14216. doi:10.1523/JNEUROSCI.1388-12.2012

Moerel, M., Martino, F. D., Uğurbil, K., Formisano, E., & Yacoub, E. (2018). Evaluating the columnar stability of acoustic processing in the human auditory cortex. *Journal of Neuroscience*, 3576–17. doi:10.1523/JNEUROSCI.3576-17.2018

Mollasadeghi, A., Mehrparvar, A. H., Atighechi, S., Davari, M. H., Shokouh, P., Mostaghaci, M., & Bahaloo, M. (2013). Sensorineural hearing loss after magnetic resonance imaging. *Case Reports in Radiology*, *2013*. doi:10.1155/2013/510258

Moore, J. K. (2000). Organization of the human superior olivary complex. *Microscopy Research and Technique*, *51*(4), 403–412. doi:10.1002/1097-0029(20001115)51:4<403::AID-JEMT8>3.0.CO;2-Q

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., & Zilles, K. (2001). Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system. *NeuroImage*, *13*(4), 684–701. ISBN: 1053-8119 (Print)\r1053-8119 (Linking). doi:10.1006/nimg.2000.0715

Morrison, S. J., Demorest, S. M., & Stambaugh, L. A. (2008). Enculturation effects in music cognition: The role of age and music complexity. *Journal of Research in Music Education*, *56*(2), 118–129. doi:10.1177/0022429408322854

Mueller, K., Mildner, T., Fritz, T., Lepsien, J., Schwarzbauer, C., Schroeter, M. L., & Möller, H. E. (2011). Investigating brain response to music: A comparison of different fMRI acquisition schemes. *NeuroImage*, *54*(1), 337–343. Publisher: Elsevier Inc. ISBN: 1095-9572 (Electronic)\r1053-8119 (Linking). doi:10.1016/j.neuroimage.2010.08.029

Muller, M., Ewert, S., & Kreuzer, S. (2009). Making chroma features more robust to timbre changes. In *2009 IEEE international conference on acoustics, speech and signal processing* (pp. 1877–1880). 2009 IEEE international conference on acoustics, speech and signal processing. doi:10.1109/ICASSP.2009.4959974

Murray, M. M., Camen, C., Andino, S. L. G., Bovet, P., & Clarke, S. (2006). Rapid brain discrimination of sounds of objects. *Journal of Neuroscience*, *26*(4), 1293–1302. doi:10.1523/JNEUROSCI.4511-05.2006

Nater, U., Krebs, M., & Ehlert, U. (2005). Sensation seeking, music preference, and psychophysiological reactivity to music. *Musicae Scientiae*, *IX*, 239–254. Retrieved from https://www.zora.uzh.ch/id/eprint/65496/

Nicholls, J. G., Martin, A. R., Fuchs, P. A., Brown, D. A., Diamond, M. E., & Weisblat, D. A. (2012). *From neuron to brain* (5th ed. 2012). Sunderland, Mass: Sinauer.

Nichols, T. E., Das, S., Eickhoff, S. B., Evans, A. C., Glatard, T., Hanke, M., … Yeo, B. T. (2017). Best practices in data analysis and sharing in neuroimaging using MRI. *Nature Neuroscience*, *20*(3). ISBN: 9788578110796. doi:10.1038/nn.4500. arXiv: 1011.1669v3

Nieuwenhuys, R. (2013). The myeloarchitectonic studies on the human cerebral cortex of the vogt-vogt school, and their significance for the interpretation of functional neuroimaging data. *Brain Structure & Function*, *218*(2), 303–352. doi:10.1007/s00429-012-0460-z

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. *PLoS computational biology*, *10*(4), e1003553. ISBN: 1553-7358 (Electronic)\n1553-734X (Linking). doi:10.1371/journal.pcbi.1003553

Norman-Haignere, S., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron*, *88*(6), 1281–1296. Publisher: Elsevier Inc. doi:10.1016/j.neuron.2015.11.035

Norman-Haignere, S., Kanwisher, N., & McDermott, J. H. (2013). Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. *Journal of Neuroscience*, *33*(50), 19451–19469. ISBN: 1529-2401 (Electronic)\r0270-6474 (Linking). doi:10.1523/JNEUROSCI.2880-13.2013

Norman-Haignere, S., & McDermott, J. H. (2016). Distortion products in auditory fMRI research: Measurements and solutions. *NeuroImage*, *129*, 401–413. doi:10.1016/j.neuroimage.2016.01.050

Norman-Haignere, S., & McDermott, J. H. (2018). Neural responses to natural and model-matched stimuli reveal distinct computations in primary and nonprimary auditory cortex. *PLOS Biology*, *16*(12), e2005127. doi:10.1371/journal.pbio.2005127

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430. doi:10.1016/j.tics.2006.07.005

Norton, A., Zipse, L., Marchina, S., & Schlaug, G. (2009). Melodic intonation therapy. *Annals of the New York Academy of Sciences*, *1169*(1), 431–436. doi:10.1111/j.1749-6632.2009.04859.x

Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *NeuroImage*, *55*(2), 713–723. doi:10.1016/j.neuroimage.2010.12.020

Obleser, J., Wise, R. J. S., Dresner, M. A., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *Journal of Neuroscience*, *27*(9), 2283–2289. doi:10.1523/JNEUROSCI.4663-06.2007

Ogg, M., Slevc, L. R., & Idsardi, W. J. (2017). The time course of sound category identification: Insights from acoustic features. *The Journal of the Acoustical Society of America*, *142*(6), 3459–3473. doi:10.1121/1.5014057

Ojala, M., & Garriga, G. C. (2010). Permutation tests for studying classifier performance. *J. Mach. Learn. Res. 11*, 1833–1863. Retrieved from http://dl.acm.org/citation.cfm?id=1756006.1859913

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., … Hickok, G. (2010). Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex (New York, NY)*, *20*(10), 2486–2495. doi:10.1093/cercor/bhp318

Oldfield, R. C. (1971). The assessment and analysis of handedness: The edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. doi:10.1016/0028-3932(71)90067-4

Oline, S. N., Ashida, G., & Burger, R. M. (2016). Tonotopic optimization for temporal processing in the cochlear nucleus. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *36*(32), 8500–8515. doi:10.1523/JNEUROSCI.4449-15.2016

Oliphant, T. E. (2006). A guide to NumPy.

Oliver, D. L., Beckius, G. E., & Shneiderman, A. (1995). Axonal projections from the lateral and medial superior olive to the inferior colliculus of the cat: A study using electron microscopic autoradiography. *The Journal of Comparative Neurology*, *360*(1), 17–32. doi:10.1002/cne.903600103

Oliver, D. L., Cant, N. B., Fay, R. R., & Popper, A. N. (2018). *The mammalian auditory pathways: Synaptic organization and microcircuits*. Google-Books-ID: XJxQDwAAQBAJ. Springer.

Otazu, G. H., Tai, L.-H., Yang, Y., & Zador, A. M. (2009). Engaging in an auditory task suppresses responses in auditory cortex. *Nature neuroscience*, *12*(5), 646–654. doi:10.1038/nn.2306

Overath, T., McDermott, J. H., Zarate, J. M., & Poeppel, D. (2015). The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. *Nature Neuroscience*, *18*(6), 903–911. ISBN: 1097-6256. doi:10.1038/nn.4021. arXiv: 15334406

Oxenham, A. J. (2012). Pitch perception. *Journal of Neuroscience*, *32*(39), 13335–13338. doi:10.1523/JNEUROSCI.3815-12.2012

Pachet, F., & Cazaly, D. (2000). A taxonomy of musical genres. In *Content-based multimedia information access - volume 2* (pp. 1238–1245). RIAO '00. Paris, France, France: LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE. Retrieved January 10, 2019, from http://dl.acm.org/citation.cfm?id=2856151.2856177

Pagallo, U., Corrales, M., Fenwick, M., & Forgo, N. (2018). The rise of robotics & AI: Technological advances & normative dilemmas. (pp. 1–13). doi:10.1007/978-981-13-2874-9_1

Pálmason, H., Bjoern, Jónsson, ó., Schedl, M., & Knees, P. (2017). Music genre classification revisited: An in-depth examination guided by music experts. In *CMMR*. doi:10.1007/978-3-030-01692-0_4

Paltoglou, A. E., Sumner, C. J., & Hall, D. A. (2009). Examining the role of frequency specificity in the enhancement and suppression of human cortical activity by auditory selective attention. *Hearing Research*, *257*(1), 106–118. doi:10.1016/j.heares.2009.08.007

Paltoglou, A. E., Sumner, C. J., & Hall, D. A. (2011). Mapping feature-sensitivity and attentional modulation in human auditory cortex with functional magnetic resonance imaging. *The European Journal of Neuroscience*, *33*(9), 1733–1741. doi:10.1111/j.1460-9568.2011.07656.x

Pantev, C., & Herholz, S. C. (2011). Plasticity of the human auditory cortex related to musical training. *Neuroscience and Biobehavioral Reviews*, *35*(10), 2140–2154. Publisher: Elsevier Ltd ISBN: 1873-7528 (Electronic) 0149-7634 (Linking). doi:10.1016/j.neubiorev.2011.06.010

Paquette, S., Ahmed, G. D., Goffi-Gomez, M. V., Hoshino, A. C. H., Peretz, I., & Lehmann, A. (2018). Musical and vocal emotion perception for cochlear implants users. *Hearing Research*, *370*, 272–282. doi:10.1016/j.heares.2018.08.009

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing*, *30*(6), 653–661. doi:10.1097/AUD.0b013e3181b412e9

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature neuroscience*, *6*(7), 674–681. ISBN: 1097-6256 (Print)\n1097-6256 (Linking). doi:10.1038/nn1082

Patel, A. D. (2008). *Music, language, and the brain*. Oxford University Press, USA. Retrieved from https://books.google.de/books?id=EkItxyZqNecC

Patel, A. D. (2014). The evolutionary biology of musical rhythm: Was darwin wrong? *PLoS Biology*, *12*(3), 1–6. ISBN: 10.1371/journal.pbio.1001821. doi:10.1371/journal.pbio.1001821

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. *Journal of Cognitive Neuroscience*, *10*(6), 717–733. ISBN: 0898-929X (Print)\r0898-929X (Linking). doi:10.1162/089892998563121

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, *36*(4), 767–776. ISBN: 0896-6273 (Print). doi:10.1016/S0896-6273(02)01060-7

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825-2830. Retrieved from http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

Peelen, M. V., & Downing, P. E. (2007). Using multi-voxel pattern analysis of fMRI data to interpret overlapping functional activations. *Trends in cognitive sciences*, *11*(1), 4–5. doi:10.1016/j.tics.2006.10.009

Peelle, J. E. (2014). Methodological challenges and solutions in auditory functional magnetic resonance imaging. *Frontiers in Neuroscience*, *8*(8), 1–13. ISBN: 1662-4548 (Print)\r1662-453x. doi:10.3389/fnins.2014.00253

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing*, *39*(2), 204–214. doi:10.1097/AUD.0000000000000494

Peelle, J. E., Eason, R. J., Schmitter, S., Schwarzbauer, C., & Davis, M. H. (2010). Evaluating an acoustically quiet EPI sequence for use in fMRI studies of speech and auditory processing. *NeuroImage*, *52*(4), 1410–1419. Publisher: Elsevier B.V. ISBN: 1095-9572 (Electronic)\r1053-8119 (Linking). doi:10.1016/j.neuroimage.2010.05.015

Peirce, J. W. (2007). PsychoPy—psychophysics software in python. *Journal of Neuroscience Methods*, *162*(1), 8–13. doi:10.1016/j.jneumeth.2006.11.017

Pelletier, C. L. (2004). The effect of music on decreasing arousal due to stress: A meta-analysis. *Journal of Music Therapy*, *41*(3), 192–214.

Penhune, V. B., Zatorre, R. J., MacDonald, J. D., & Evans, A. C. (1996). Interhemispheric anatomical differences in human primary auditory cortex: Probabilistic mapping and volume measurement from magnetic resonance scans. *Cerebral Cortex (New York, N.Y.: 1991)*, *6*(5), 661–672.

Penny, W. D., Stephan, K. E., Daunizeau, J., Rosa, M. J., Friston, K. J., Schofield, T. M., & Leff, A. P. (2010). Comparing families of dynamic causal models. *PLOS Computational Biology*, *6*(3), e1000709. doi:10.1371/journal.pcbi.1000709

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*(7), 688–691. doi:10.1038/nn1083

Peretz, I., Vuvan, D., Lagrois, M.-É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philos. Trans. R. Soc. B*, *370*, 20140090. ISBN: 10.1098/rstb.2014.0090. doi:10.1098/rstb.2014.0090

Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. *Annual review of psychology*, *56*, 89–114. ISBN: 0066-4308 (Print)\n0066-4308 (Linking). doi:10.1146/annurev.psych.56.091103.070225

Pernet, C. R., Appelhoff, S., Gorgolewski, K. J., Flandin, G., Phillips, C., Delorme, A., & Oostenveld, R. (2019). EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Scientific Data*, *6*(1), 1–5. doi:10.1038/s41597-019-0104-8

Pernet, C. R., McAleer, P., Latinus, M., Gorgolewski, K. J., Charest, I., Bestelmeyer, P. E., … Belin, P. (2015). The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. *NeuroImage*, *119*, 164–174. Publisher: Elsevier B.V. ISBN: 1095-9572 (Electronic)\r1053-8119 (Linking). doi:10.1016/j.neuroimage.2015.06.050

Petkov, C. I., Kang, X., Alho, K., Bertrand, O., Yund, E. W., & Woods, D. L. (2004). Attentional modulation of human auditory cortex. *Nature Neuroscience*, *7*(6), 658–663. doi:10.1038/nn1256

Pfeifer, R. (1920). *Myelogenetisch-anatomische untersuchungen über das kortikale ende der hörleitung*. BG Teubner.

Picciotto, M. (2018). Analytical transparency and reproducibility in human neuroimaging studies. *Journal of Neuroscience*, *38*(14), 3375–3376. doi:10.1523/JNEUROSCI.0424-18.2018

Pickles, J. O. (2015). Chapter 1 - auditory pathways: Anatomy and physiology. In M. J. Aminoff, F. Boller, & D. F. Swaab (Eds.), *Handbook of clinical neurology* (Vol. 129, pp. 3–25). The Human Auditory System. doi:10.1016/B978-0-444-62630-1.00001-9

Pizzamiglio, L., Aprile, T., Spitoni, G., Pitzalis, S., Bates, E., D'Amico, S., & Di Russo, F. (2005). Separate neural systems for processing action- or non-action-related sounds. *NeuroImage*, *24*(3), 852–861. doi:10.1016/j.neuroimage.2004.09.025

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication*. The Nature of Speech Perception, *41*(1), 245–255. doi:10.1016/S0167-6393(02)00107-3

Poliva, O. (2017). From where to what: A neuroanatomically based evolutionary model of the emergence of speech in humans. *F1000Research*, *4*. doi:10.12688/f1000research.6175.3

Price, D. L., De Wilde, J., Papadaki, A., Curran, J. S., & Kitney, R. (2001). Investigation of acoustic noise on 15 MRI scanners from 0.2 t to 3 t. *Journal of magnetic resonance imaging : JMRI*, *13*, 288–93. doi:10.1002/1522-2586(200102)13:23.3.CO;2-G

Prickett, C. A., & Moore, R. S. (1991). The use of music to aid memory of alzheimer's patients. *Journal of Music Therapy*, *28*(2), 101–110. doi:10.1093/jmt/28.2.101

Pulman, M. (2006). Popular music genres: An introduction. by stuart borthwick and ron moy. edinburgh: Edinburgh university press ltd, 2004. 246z pp. ISBN 0-74861745-0 (paperback). *Popular Music*, *25*(2), 340–342. doi:10.1017/S026114300625092X

Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A.-S., McNamara, J. O., & Williams, S. M. (2001). The external ear. *Neuroscience. 2nd edition.* Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK10908/

Raamana, P. R., Weiner, M. W., Wang, L., & Beg, M. F. (2015). Thickness network features for prognostic applications in dementia. *Neurobiology of Aging*. Novel Imaging Biomarkers for Alzheimer's Disease and Related Disorders (NIBAD), *36*, S91–S102. doi:10.1016/j.neurobiolaging.2014.05.040

Rademacher, J., Galaburda, A. M., Kennedy, D. N., Filipek, P. A., & Caviness, V. S. (1992). Human cerebral cortex: Localization, parcellation, and morphometry with magnetic resonance imaging. *Journal of Cognitive Neuroscience*, *4*(4), 352–374. doi:10.1162/jocn.1992.4.4.352

Raglio, A. (2015). Music therapy interventions in parkinson's disease: The state-of-the-art. *Frontiers in Neurology*, *6*. doi:10.3389/fneur.2015.00185

Rammsayer, T., & Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians. *Music Perception: An Interdisciplinary Journal*, *24*(1), 37–48. doi:10.1525/mp.2006.24.1.37

Ranaweera, R. D., Kwon, M., Hu, S., Tamer, G. G., Luh, W.-M., & Talavage, T. M. (2016). Temporal pattern of acoustic imaging noise asymmetrically modulates activation in the auditory cortex. *Hearing research*, *331*, 57–68. doi:10.1016/j.heares.2015.09.017

Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Current Opinion in Neurobiology*, *8*(4), 516–521.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–724. doi:10.1038/nn.2331

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proceedings of the National Academy of Sciences*, *97*(22), 11800–11806. doi:10.1073/pnas.97.22.11800

Ravicz, M. E., Melcher, J. R., & Kiang, N. Y.-S. (2000). Acoustic noise during functional magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *108*(4), 1683–1696. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2270941/

Reale, R. A., & Imig, T. J. (1980). Tonotopic organization in auditory cortex of the cat. *The Journal of Comparative Neurology*, *192*(2), 265–291. doi:10.1002/cne.901920207

Recanzone, G. H., & Cohen, Y. E. (2010). Serial and parallel processing in the primate auditory cortex revisited. *Behavioural Brain Research*, *206*(1), 1–7. doi:10.1016/j.bbr.2009.08.015

Regev, M., Honey, C. J., Simony, E., & Hasson, U. (2013). Selective and invariant neural responses to spoken and written narratives. *Journal of Neuroscience*, *33*(40), 15978–15988. doi:10.1523/JNEUROSCI.1580-13.2013

Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi’s of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, *84*(6), 1236–1256. ISBN: 0022-3514. doi:10.1037/0022-3514.84.6.1236

Reuter, M., Rosas, H. D., & Fischl, B. (2010). Highly accurate inverse consistent registration: A robust approach. *NeuroImage*, *53*(4), 1181–1196. doi:10.1016/j.neuroimage.2010.07.020

Reuter, M., Schmansky, N. J., Rosas, H. D., & Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage*, *61*(4), 1402–1418. doi:10.1016/j.neuroimage.2012.02.084

Reybrouck, M., Eerola, T., & Podlipniak, P. (2018). *Music and the functions of the brain: Arousal, emotions, and pleasure*. Google-Books-ID: QGlVDwAAQBAJ. Frontiers Media SA.

Riecke, L., Opstal, A. J. v., Goebel, R., & Formisano, E. (2007). Hearing illusory sounds in noise: Sensory-perceptual transformations in primary auditory cortex. *Journal of Neuroscience*, *27*(46), 12684–12689. doi:10.1523/JNEUROSCI.2713-07.2007

Riecke, L., Peters, J. C., Valente, G., Kemper, V. G., Formisano, E., & Sorger, B. (2017). Frequency-selective attention in auditory scenes recruits frequency representations throughout human superior temporal cortex. *Cerebral Cortex*, *27*(5), 3002–3014. doi:10.1093/cercor/bhw160

Rinne, T., Balk, M. H., Koistinen, S., Autti, T., Alho, K., & Sams, M. (2008). Auditory selective attention modulates activation of human inferior colliculus. *Journal of Neurophysiology*, *100*(6), 3323–3327. doi:10.1152/jn.90607.2008

Rivier, F., & Clarke, S. (1997). Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex: Evidence for multiple auditory areas. *NeuroImage*, *6*(4), 288–304. doi:10.1006/nimg.1997.0304

Rodd, J. M., Johnsrude, I. S., & Davis, M. H. (2012). Dissociating frontotemporal contributions to semantic ambiguity resolution in spoken sentences. *Cerebral Cortex*, *22*(8), 1761–1773. doi:10.1093/cercor/bhr252

Rogalsky, C., Rong, F., Saberi, K., & Hickok, G. (2011). Functional anatomy of language and music perception: Temporal and structural factors investigated using functional magnetic resonance imaging. *Journal of Neuroscience*, *31*(10), 3843–3852. ISBN: 0270-6474. doi:10.1523/jneurosci.4515-10.2011. arXiv: NIHMS150003

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neuroscience*, *2*(12), 1131–1136. doi:10.1038/16056

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., ... Rudner, M. (2013). The ease of language understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*. doi:10.3389/fnsys.2013.00031

Rudolf Carnap. (1928). *Der logische aufbau der welt*. Retrieved January 10, 2019, from http://archive.org/details/RudolfCarnapDerLogischeAufbauDerWeltPrint

Sachs, M. E., Damasio, A., & Habibi, A. (2015). The pleasures of sad music: A systematic review. *Frontiers in Human Neuroscience*, *9*, 1–12. doi:10.3389/fnhum.2015.00404

Saenz, M., & Langers, D. R. M. (2014). Tonotopic mapping of human auditory cortex. *Hearing Research*, *307*, 42–52. Publisher: Elsevier B.V. doi:10.1016/j.heares.2013.07.016

Saitou, T., Goto, M., Unoki, M., & Akagi, M. (2007). Speech-to-singing synthesis: Converting speaking voices to singing voices by controlling acoustic features unique to singing voices. In *2007 IEEE workshop on applications of signal processing to audio and acoustics* (pp. 215–218). 2007 IEEE workshop on applications of signal processing to audio and acoustics. doi:10.1109/ASPAA.2007.4393001

Salimpoor, V. N., & Zatorre, R. J. (2013). Neural interactions that give rise to musical pleasure. *Psychology of Aesthetics, Creativity, and the Arts*, *7.1*(1), 62–75. ISBN: 1931-3896. doi:10.1037/a0031819

6

Sammler, D., Grigutsch, M., Fritz, T., & Koelsch, S. (2007). Music and emotion: Electrophysiological correlates of the processing of pleasant and unpleasant music. *Psychophysiology*, *44*(2), 293–304. doi:10.1111/j.1469-8986.2007.00497.x

Santoro, R., Moerel, M., De Martino, F., Goebel, R., Ugurbil, K., Yacoub, E., & Formisano, E. (2014). Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. *PLoS Comput Biol*, *10*(1), e1003412. ISBN: 10.1371/journal.pcbi.1003412. doi:10.1371/journal.pcbi.1003412

Santoro, R., Moerel, M., De Martino, F., Valente, G., Ugurbil, K., Yacoub, E., & Formisano, E. (2017). Reconstructing the spectrotemporal modulations of real-life sounds from fMRI response patterns. *Proceedings of the National Academy of Sciences*, *114*(18), 4799–4804. ISBN: 1091-6490 (Electronic) 0027-8424 (Linking). doi:10.1073/pnas.1617622114. arXiv: 1408.1149

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The p600-as-p3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, *137*, 29–39. doi:10.1016/j.bandl.2014.07.010

Scarff, C. J., Dort, J. C., Eggermont, J. J., & Goodyear, B. G. (2004). The effect of MR scanner noise on auditory cortex activity using fMRI. *Human Brain Mapping*, *22*(4), 341–349. ISBN: 1065-9471 (Print). doi:10.1002/hbm.20043

Scheich, H., Brechmann, A., Brosch, M., Budinger, E., & Ohl, F. W. (2007). The cognitive auditory cortex: Task-specificity of stimulus representations. *Hearing Research*. Auditory Cortex 2006 - The Listening Brain, *229*(1), 213–224. doi:10.1016/j.heares.2007.01.025

Schellenberg, E. G., & Moreno, S. (2010). Music lessons, pitch processing, and g. *Psychology of Music*, *38*(2), 209–221. doi:10.1177/0305735609339473

Schirmer, A., Fox, P. M., & Grandjean, D. (2012). On the spatial organization of sound processing in the human temporal lobe: A meta-analysis. *NeuroImage*, *63*(1), 137–147. Publisher: Elsevier Inc. doi:10.1016/j.neuroimage.2012.06.025

Schlaug, G. (2015). Musicians and music making as a model for the study of brain plasticity. *Progress in brain research*, *217*, 37–55. doi:10.1016/bs.pbr.2014.11.020

Schlaug, G., Marchina, S., & Norton, A. (2008). From singing to speaking: Why singing may lead to recovery of expressive language function in patients with broca's aphasia. *Music Perception: An Interdisciplinary Journal*, *25*(4), 315–323. doi:10.1525/mp.2008.25.4.315

Schmitter, S., Diesch, E., Amann, M., Kroll, A., Moayer, M., & Schad, L. R. (2008). Silent echo-planar imaging for auditory FMRI. *Magma (New York, N.Y.) 21*(5), 317–325. doi:10.1007/s10334-008-0132-4

Schneider, P., Andermann, M., Wengenroth, M., Goebel, R., Flor, H., Rupp, A., & Diesch, E. (2009). Reduced volume of heschl's gyrus in tinnitus. *NeuroImage*, *45*(3), 927–939. Publisher: Elsevier Inc. ISBN: 1095-9572 (Electronic) 1053-8119 (Linking). doi:10.1016/j.neuroimage.2008.12.045

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience*, *5*(7), 688–694. doi:10.1038/nn871

Schön, D., Gordon, R., Campagne, A., Magne, C., Astésano, C., Anton, J.-L., & Besson, M. (2010). Similar cerebral networks in language, music and song perception. *NeuroImage*, *51*(1), 450–61. doi:10.1016/j.neuroimage.2010.02.023

Schönwiesner, M., Dechent, P., Voit, D., Petkov, C. I., & Krumbholz, K. (2015). Parcellation of human and monkey core auditory cortex with fMRI pattern classification and objective detection of tonotopic gradient reversals. *Cerebral Cortex (New York, NY)*, *25*(10), 3278–3289. doi:10.1093/cercor/bhu124

Schreiner, C. E., & Winer, J. A. (2007). Auditory cortex mapmaking: Principles, projections, and plasticity. *Neuron*, *56*(2), 356–365. doi:10.1016/j.neuron.2007.10.013

Schwarzbauer, C., Davis, M. H., Rodd, J. M., & Johnsrude, I. (2006). Interleaved silent steady state (ISSS) imaging: A new sparse imaging method applied to auditory fMRI. *NeuroImage*, *29*(3), 774–782. ISBN: 1053-8119 (Print)\r1053-8119 (Linking). doi:10.1016/j.neuroimage.2005.08.025

scikit learn - How are feature importances in RandomForestClassifier determined? (2013, April 4). Retrieved October 15, 2018, from https://stackoverflow.com/questions/15810339/how-are-feature-importances-in-randomforestclassifier-determined

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain: A Journal of Neurology*, *123 Pt 12*, 2400–2406.

Scott, S. K., & McGettigan, C. (2013). The neural processing of masked speech. *Hearing Research*. Annual Reviews 2013, *303*, 58–66. doi:10.1016/j.heares.2013.05.001

Ségonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D., Hahn, H. K., & Fischl, B. (2004). A hybrid approach to the skull stripping problem in MRI. *NeuroImage*, *22*(3), 1060–1075. doi:10.1016/j.neuroimage.2004.03.032

Seifritz, E., Di Salle, F., Esposito, F., Herdener, M., Neuhoff, J. G., & Scheffler, K. (2006). Enhancing BOLD response in the auditory system by neurophysiologically tuned fMRI sequence. *NeuroImage*, *29*(3), 1013–1022. ISBN: 1053-8119 (Print)\r1053-8119 (Linking). doi:10.1016/j.neuroimage.2005.08.029

Sereno, M. I., Lutti, A., Weiskopf, N., & Dick, F. (2013). Mapping the human cortical surface by combining quantitative t(1) with retinotopy. *Cerebral Cortex (New York, N.Y.: 1991)*, *23*(9), 2261–2268. doi:10.1093/cercor/bhs213

Setsompop, K., Cohen-Adad, J., Gagoski, B. A., Raij, T., Yendiki, A., Keil, B., … Wald, L. L. (2012). Improving diffusion MRI using simultaneous multi-slice echo planar imaging. *NeuroImage*, *63*(1), 569–580. doi:10.1016/j.neuroimage.2012.06.033

Shah, N. J., Jäncke, L., Grosse-Ruyken, M.-L., & Müller-Gärtner, H. W. (1999). Influence of acoustic masking noise in fMRI of the auditory cortex during phonetic discrimination. *Journal of Magnetic Resonance Imaging*, *9*(1), 19–25. doi:10.1002/(SICI)1522-2586(199901)9:1<19::AID-JMRI3>3.0.CO;2-K

Sharda, M., Tuerk, C., Chowdhury, R., Jamey, K., Foster, N., Custo-Blanch, M., . . . Hyde, K. (2018). Music improves social communication and auditory–motor connectivity in children with autism. *Translational Psychiatry*, *8*(1), 231. doi:10.1038/s41398-018-0287-3

Shepard, R. N. (1958). Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology*, *55*(6), 509–523.

Sheppard, A., Chen, Y.-C., & Salvi, R. (2018). MRI noise and hearing loss. *The Hearing Journal*, *71*(4), 30. doi:10.1097/01.HJ.0000532395.75558.2d

Siedenburg, K., & Müllensiefen, D. (2017). Modeling timbre similarity of short music clips. *Frontiers in Psychology*, *8*. doi:10.3389/fpsyg.2017.00639

Sigalovsky, I. S., Fischl, B., & Melcher, J. R. (2006). Mapping an intrinsic MR property of gray matter in auditory cortex of living humans: A possible marker for primary cortex and hemispheric differences. *NeuroImage*, *32*(4), 1524–1537. doi:10.1016/j.neuroimage.2006.05.023

Sigalovsky, I. S., Hawley, M. L., Harms, M. P., & Melchert, J. R. (2001). Sound level representations in the human auditory pathway investigated using fMRI. *NeuroImage*, *13*(6), 3255.

Signal Spectra. (n.d.). Retrieved October 20, 2018, from https://www.music.mcgill.ca/~gary/307/week1/spectra.html

Sinex, D. G., Guzik, H., Li, H., & Henderson Sabes, J. (2003). Responses of auditory nerve fibers to harmonic and mistuned complex tones. *Hearing Research*, *182*(1), 130–139.

Slee, S. J., & David, S. V. (2015). Rapid task-related plasticity of spectrotemporal receptive fields in the auditory midbrain. *Journal of Neuroscience*, *35*(38), 13090–13102. doi:10.1523/JNEUROSCI.1671-15.2015

Smith, D. F., & Gedayt, J. (2001). Clomipramine challenge studied by [150] h20 PET : Focus on the thalamus . (6), 2001.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., . . . Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*. Mathematics in Brain Imaging, *23*, S208–S219. doi:10.1016/j.neuroimage.2004.07.051

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience*, *32*(25), 8443–8453. doi:10.1523/JNEUROSCI.5069-11.2012

Soley, G., & Hannon, E. E. (2010). Infants prefer the musical meter of their own culture: A cross-cultural comparison. *Developmental Psychology*, *46*(1), 286–292. doi:10.1037/a0017555

Soltysik, D. A., Peck, K. K., White, K. D., Crosson, B., & Briggs, R. W. (2004). Comparison of hemodynamic response nonlinearity across primary cortical areas. *NeuroImage*, *22*(3), 1117–27. ISBN: 1053-8119 (Print)\r1053-8119 (Linking). doi:10.1016/j.neuroimage.2004.03.024

Staeren, N., Renvall, H., De Martino, F., Goebel, R., & Formisano, E. (2009). Sound categories are represented as distributed patterns in the human auditory cortex. *Current Biology*, *19*(6), 498–502. ISBN: 1879-0445 (Electronic)\n0960-9822 (Linking). doi:10.1016/j.cub.2009.01.066

Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): Random permutations and cluster size control. *NeuroImage*, *65*, 69–82. Publisher: Elsevier Inc. ISBN: 1095-9572 (Electronic)\r1053-8119 (Linking). doi:10.1016/j.neuroimage.2012.09.063

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, *46*(4), 1004–1017. doi:10.1016/j.neuroimage.2009.03.025

Stern, P. (2017). Brain mechanisms of pitch perception. *Science*, *357*(6353), 768–769. doi:10.1126/science.357.6353.768-i

Strait, D. L., & Kraus, N. (2011). Playing music for a smarter ear: Cognitive, perceptual and neurobiological evidence. *Music perception*, *29*(2), 133–146. doi:10.1525/MP.2011.29.2.133

Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research*, *261*(1), 22–29. doi:10.1016/j.heares.2009.12.021

Striem-Amit, E., Hertz, U., & Amedi, A. (2011). Extensive cochleotopic mapping of human auditory cortical fields obtained with phase-encoding fMRI. *PLOS ONE*, *6*(3), e17832. doi:10.1371/journal.pone.0017832

Sturm, B. L. (2013). Classification accuracy is not enough: On the evaluation of music genre recognition systems. *Journal of Intelligent Information Systems*, *41*(3), 371–406. doi:10.1007/s10844-013-0250-y

Su, L., Zulfiqar, I., Jamshed, F., Fonteneau, E., & Marslen-Wilson, W. (2014). Mapping tonotopic organization in human temporal cortex: Representational similarity analysis in EMEG source space. *Frontiers in Neuroscience*, *8*. doi:10.3389/fnins.2014.00368

Suga, N., & Ma, X. (2003). Multiparametric corticofugal modulation and plasticity in the auditory system. *Nature Reviews Neuroscience*, *4*(10), 783–794. doi:10.1038/nrn1222

Swaroop Guntupalli, J. (2013). *Whole brain hyperalignment : Inter-subject hyperalignment of local representational spaces* (Doctoral dissertation).

Sweet, R. A., Dorph-Petersen, K.-A., & Lewis, D. A. (2005). Mapping auditory core, lateral belt, and parabelt cortices in the human superior temporal gyrus. *Journal of Comparative Neurology*, *491*(3), 270–289. doi:10.1002/cne.20702

Talavage, T. M., & Edmister, W. B. (2004). Nonlinearity of FMRI responses in human auditory cortex. *Human Brain Mapping*, *22*(3), 216–228. doi:10.1002/hbm.20029

Talavage, T. M., Edmister, W. B., Ledden, P. J., & Weisskoff, R. M. (1999). Quantitative assessment of auditory cortex responses induced by imager acoustic noise. *Human Brain Mapping*, *7*(2), 79–88. doi:10.1002/(SICI)1097-0193(1999)7:2<79::AID-HBM1>3.0.CO;2-R

Talavage, T. M., & Hall, D. A. (2012). How challenges in auditory fMRI led to general advancements for the field. *Neuroimage*, *62*(2), 641–647. doi:10.1016/j.neuroimage.2012.01.006

Talavage, T. M., Sereno, M. I., Melcher, J. R., Ledden, P. J., Rosen, B. R., & Dale, A. M. (2004). Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *Journal of Neurophysiology*, *91*(3), 1282–1296. doi:10.1152/jn.01125.2002

Teo, T., Hargreaves, D. J., & Lee, J. (2008). Musical preference, identification, and familiarity: A multicultural comparison of secondary students from singapore and the united kingdom. *Journal of Research in Music Education*, *56*(1), 18–32. doi:10.1177/0022429408322953

Tervaniemi, M., & Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain Research Reviews*, *43*(3), 231–246. doi:10.1016/j.brainresrev.2003.08.004

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: An event-related potential and behavioral study. *Experimental Brain Research*, *161*(1), 1–10. doi:10.1007/s00221-004-2044-5

Thaut, M., & Hoemberg, V. (2014). *Handbook of neurologic music therapy*. Google-Books-ID: 5Gb0AwAAQBAJ. Oxford University Press.

Thoma, M. V., La Marca, R., Brönnimann, R., Finkel, L., Ehlert, U., & Nater, U. M. (2013). The effect of music on the human stress response. *PLoS ONE*, *8*(8), 1–12. ISBN: 10.1371/journal.pone.0070156. doi:10.1371/journal.pone.0070156

Thomas, J. M., Huber, E., Stecker, G. C., Boynton, G. M., Saenz, M., & Fine, I. (2015). Population receptive field estimates of human auditory cortex. *NeuroImage*, *105*, 428–39. Publisher: Elsevier Inc. doi:10.1016/j.neuroimage.2014.10.060

Tian, L., Ma, L., & Wang, L. (2016). Alterations of functional connectivities from early to middle adulthood: Clues from multivariate pattern analysis of resting-state fMRI data. *NeuroImage, 129*, 389–400. Publisher: Elsevier Inc. doi:10.1016/j.neuroimage.2016.01.039

Tierney, A., Dick, F., Deutsch, D., & Sereno, M. (2013). Speech versus song: Multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. *Cerebral Cortex, 23*(2), 249–254. doi:10.1093/cercor/bhs003

Tillmann, B. (2012). Music and language perception: Expectations, structural integration, and cognitive sequencing. *Topics in Cognitive Science, 4*(4), 568–584. doi:10.1111/j.1756-8765.2012.01209.x

Toiviainen, P., Alluri, V., Brattico, E., Wallentin, M., & Vuust, P. (2014). Capturing the musical brain with lasso: Dynamic decoding of musical features from fMRI data. *NeuroImage, 88*, 170–180. Publisher: Elsevier Inc. doi:10.1016/j.neuroimage.2013.11.017

Torgerson, W. S. (1958). *Theory and methods of scaling*. Wiley.

Torgerson, W. S. (1965). Multidimensional scaling of similarity. *Psychometrika, 30*(4), 379–393.

Torresen, J. (2018). A review of future and ethical perspectives of robotics and AI. *Frontiers in Robotics and AI, 4*. doi:10.3389/frobt.2017.00075

Tsunada, J., & Cohen, Y. E. (2014). Neural mechanisms of auditory categorization: From across brain areas to within local microcircuits. *Frontiers in Neuroscience, 8*. doi:10.3389/fnins.2014.00161

Turner, B. O., Paul, E. J., Miller, M. B., & Barbey, A. K. (2018). Small sample sizes reduce the replicability of task-based fMRI studies. *Communications Biology, 1*(1), 62. doi:10.1038/s42003-018-0073-z

Tużnik, P., Augustynowicz, P., & Francuz, P. (2018). Electrophysiological correlates of timbre imagery and perception. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology, 129*, 9–17. doi:10.1016/j.ijpsycho.2018.05.004

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*(4), 327–352. doi:10.1037/0033-295X.84.4.327

Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing, 10*(5), 293–302. ISBN: 1063-6676. doi:10.1109/TSA.2002.800560

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage, 15*(1), 273–289. doi:10.1006/nimg.2001.0978

Upadhyay, J., Silver, A., Knaus, T. A., Lindgren, K. A., Ducros, M., Kim, D.-S., & Tager-Flusberg, H. (2008). Effective and structural connectivity in the human auditory cortex. *Journal of Neuroscience, 28*(13), 3341–3349. doi:10.1523/JNEUROSCI.4434-07.2008

Vallat, R. (2018). Pingouin: Statistics in python. *The Journal of Open Source Software*. doi:10.21105/joss.01026

Varoquaux, G. (2017). Cross-validation failure: Small sample sizes lead to large error bars. *NeuroImage*, 1–10. Issue: April. doi:10.1016/j.neuroimage.2017.06.061. arXiv: 1706.07581

Varoquaux, G., Gramfort, A., Pedregosa, F., Michel, V., & Thirion, B. (2011). Multi-subject dictionary learning to segment an atlas of brain spontaneous activity. In G. Székely & H. K. Hahn (Eds.), *Information processing in medical imaging* (pp. 562–573). Lecture Notes in Computer Science. Springer Berlin Heidelberg.

Von Economo, C., & Koskinas, G. (1925). *The cytoarchitectonics of the adult human cortex*. Berlin: Springer Verlag;

von Economo, C., & Horn, L. (1930). Über Windungsrelief, Maße und Rindenarchitektonik der Supratemporalfläche, ihre individuellen und ihre Seitenunterschiede. *Zeitschrift für die gesamte Neurologie und Psychiatrie*, *130*(1), 678–757. doi:10.1007/BF02865945

von Helmholtz, H., & Ellis, A. (1954). *On the sensations of tone as a physiological basis for the theory of music*. Dover Books on History of Science and Classics of Science. Dover Publications. Retrieved from https://books.google.de/books?id=o6YSpMn-h3UC

Walker, K. M. M., Bizley, J. K., King, A. J., & Schnupp, J. W. H. (2011). Multiplexed and robust representations of sound features in auditory cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(41), 14565–14576. doi:10.1523/JNEUROSCI.2074-11.2011

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The musical ear test, a new reliable test for measuring musical competence. *Learning and Individual Differences*, *20*(3), 188–196. ISBN: 1041-6080. doi:10.1016/j.lindif.2010.02.004

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage*, *137*, 188–200. doi:10.1016/j.neuroimage.2015.12.012

Wardle, S. G., Kriegeskorte, N., Grootswagers, T., Khaligh-Razavi, S. M., & Carlson, T. A. (2016). Perceptual similarity of visual patterns predicts dynamic neural activation patterns measured with MEG. *NeuroImage*, *132*, 59–70. Publisher: Elsevier Inc. ISBN: 1053-8119. doi:10.1016/j.neuroimage.2016.02.019. arXiv: 1506.02208

Warrier, C., Wong, P., Penhune, V., Zatorre, R. J., Parrish, T., Abrams, D., & Kraus, N. (2009). Relating structure to function: Heschl's gyrus and acoustic processing. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(1), 61–69. doi:10.1523/JNEUROSCI.3489-08.2009

Waskom, M., Botvinnik, O., O'Kane, D., Hobson, P., Ostblom, J., Lukauskas, S., … Qalieh, A. (2018). Mwaskom/seaborn: V0.9.0 (july 2018). doi:10.5281/zenodo.1313201

Wasserthal, C., Brechmann, A., Stadler, J., Fischl, B., & Engel, K. (2014). Localizing the human primary auditory cortex in vivo using structural MRI. *NeuroImage*, *93 Pt 2*, 237–51. Publisher: Elsevier Inc. ISBN: 1095-9572 (Electronic)\n1053-8119 (Linking). doi:10.1016/j.neuroimage.2013.07.046

Weiss, M. W., Trehub, S. E., Schellenberg, E. G., & Habashi, P. (2016). Pupils dilate for vocal or familiar music. *Journal of Experimental Psychology. Human Perception and Performance*, *42*(8), 1061–1065. doi:10.1037/xhp0000226

Weiss, M. W., Vanzella, P., Schellenberg, E. G., & Trehub, S. E. (2015). Pianists exhibit enhanced memory for vocal melodies but not piano melodies. *Quarterly Journal of Experimental Psychology (2006)*, *68*(5), 866–877. doi:10.1080/17470218.2015.1020818

Whitehead, J. C., & Armony, J. L. (2018). Singing in the brain: Neural representation of music and voice as revealed by fMRI. *Human Brain Mapping*, *39*(12), 4913–4924. doi:10.1002/hbm.24333

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*(6), 80–83. doi:10.2307/3001968

Wilding, T., McKay, C., Baker, R., & Kluk, K. (2012). Auditory steady state responses in normal-hearing and hearing-impaired adults: An analysis of between-session amplitude and latency repeatability, test time, and f ratio detection paradigms. *Ear and hearing*, *33*(2), 267–278. doi:10.1097/AUD.0b013e318230bba0

Wilson, A. C., & Sarich, V. M. (1969). A molecular time scale for human evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *63*(4), 1088–1093. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC223432/

Winer, J. A., & Lee, C. C. (2007). The distributed auditory cortex. *Hearing Research*. Auditory Cortex 2006 - The Listening Brain, *229*(1), 3–13. doi:10.1016/j.heares.2007.01.017

Winer, J. A., & Schreiner, C. E. (Eds.). (2011). *The auditory cortex*. Springer US. Retrieved from https://www.springer.com/us/book/9781441900739

Wingfield, A., Tun, P. A., & McCoy, S. L. (2005). Hearing loss in older adulthood: What it is and how it interacts with cognitive performance. Retrieved January 1, 2019, from https://journals.sagepub.com/doi/abs/10.1111/j.0963-7214.2005.00356.x?casa_token=dfI9PWumgCkAAAAA%3AuPCqEbrpGHFdKRWEPlEHWBlnhEqogitSTpwbptfpJ61cWgl0M8NegcX7FM-rDi2OKEgybmzpueXMS7U

Wong, P. C. M., Warrier, C. M., Penhune, V. B., Roy, A. K., Sadehh, A., Parrish, T. B., & Zatorre, R. J. (2008). Volume of left heschl's gyrus and linguistic pitch learning. *Cerebral Cortex (New York, N.Y.: 1991)*, *18*(4), 828–836. doi:10.1093/cercor/bhm115

Woods, D. L., & Alain, C. (2009). Functional imaging of human auditory cortex. *Current opinion in otolaryngology & head and neck surgery*, *17*(5), 407–11. ISBN: 10689508 (ISSN). doi:10.1097/MOO.0b013e3283303330

Woods, D. L., Herron, T. J., Cate, A. D., Yund, E. W., Stecker, G. C., Rinne, T., & Kang, X. (2010). Functional properties of human auditory cortical fields. *Frontiers in Systems Neuroscience*, *4*. ISBN: 1662-5137. doi:10.3389/fnsys.2010.00155

Woods, D. L., Stecker, G. C., Rinne, T., Herron, T. J., Cate, A. D., Yund, E. W., … Kang, X. (2009). Functional maps of human auditory cortex: Effects of acoustic features and attention. *PLoS ONE*, *4*(4). ISBN: 1932-6203 (Electronic)\n1932-6203 (Linking). doi:10.1371/journal.pone.0005183

Yoo, S., Chung, J.-Y., Jeon, H.-A., Lee, K.-M., Kim, Y.-B., & Cho, Z.-H. (2012). Dual routes for verbal repetition: Articulation-based and acoustic–phonetic codes for pseudoword and word repetition, respectively. *Brain and Language*, *122*(1), 1–10. doi:10.1016/j.bandl.2012.04.011

Yosinski, J. (2018). Training and understanding deep neural networks for robotics, design, and visual perception. doi:https://doi.org/10.7298/X4513WGB

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

Zatorre, R. J., & Baum, S. R. (2012). Musical melody and speech intonation: Singing a different tune. *PLoS Biology*, *10*(7), 5. ISBN: 1545-7885 (Electronic)\r1544-9173 (Linking). doi:10.1371/journal.pbio.1001372

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, *11*(10), 946–953. doi:10.1093/cercor/11.10.946

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex:music and speech. *TRENDS in Cogniti ve S ciences*, *6*(1), 37–46.

Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the mind's ear: A PET investigation of musical imagery and perception. *Journal of Cognitive Neuroscience*, *8*(1), 29–46. doi:10.1162/jocn.1996.8.1.29

Zatorre, R. J., & Salimpoor, V. N. (2013). From perception to pleasure: Music and its neural substrates. *Proceedings of the National Academy of Sciences of the United States of America*, *110 Suppl*, 10430–7. ISBN: 0027-8424. doi:10.1073/pnas.1301228110

Zatorre, R. J., & Zarate, J. M. (2012). *The human auditory cortex*. doi:10.1007/978-1-4614-2314-0

Zhang, J., Yang, T., Bao, Y., Li, H., Pöppel, E., & Silveira, S. (2018). Sadness and happiness are amplified in solitary listening to music. *Cognitive Processing*, *19*(1), 133–139. doi:10.1007/s10339-017-0832-7

Zhang, Y., Cai, J., An, L., Hui, F., Ren, T., Ma, H., & Zhao, Q. (2017). Does music therapy enhance behavioral and cognitive function in elderly dementia patients? a systematic review and meta-analysis. *Ageing Research Reviews*, *35*, 1–11. doi:10.1016/j.arr.2016.12.003

APPENDIX

A. OPEN BRAIN CONSENT FORM

Philipps Universität Marburg

ProbandInnen-Information *Open Brain Consent Documentation*

Wir möchten die in dieser Studie erhobenen Daten, inklusive der fMRT-Bilddaten, der Allgemeinheit zur Verfügung stellen (über eine öffentliche Internetdatenbank). Es ist daher möglich, dass die Daten der Untersuchung, an der Sie aktuell teilnehmen, zukünftig auch für andere Forschungsprojekte genutzt werden. Diese Projekte könnten andere Zielsetzungen verfolgen als die aktuelle Studie.

Alle Daten, die wir teilen, werden vollständig pseudonymisiert. Ihr Name wird nicht gespeichert, sondern nur eine Codenummer, so dass niemand weiß, welche Daten zu Ihnen gehören. Zur Pseudonymisierung Ihrer Daten bitten wir Sie sechs ganzzahlige Zahlen Ihrer Wahl mit möglichen Ziffern von 0 bis 9 mit dem Anfangs- und Endbuchstaben ihrer Geburtsstadt in Großbuchstaben zu kombinieren (z.B. 340692 MN). Die Verbindung zwischen den Messdaten und Ihrer Identität kann dann nur mit einer Schlüsselliste hergestellt werden. Diese wird getrennt von den Messdaten in einem verschlossenen Schrank in der Klinik für Psychiatrie und Psychotherapie aufbewahrt und zwei Jahre nach Abschluss der Untersuchung gelöscht. Zusätzlich dazu werden wir Daten entfernen, die unter Umständen dazu führen könnten, dass Rückschlüsse auf Ihre Person gezogen werden könnten (wie zum Beispiel Gesichtsmerkmale oder das Datum der Teilnahme). Dazu wird ein sogenanntes Defacing vorgenommen, bei welchem Ihr Gesicht und damit Ihre Erkennungsmerkmale aus den fMRT-Daten entfernt werden. Damit ist eine Zuordnung Ihrer Daten durch Gesicht oder Zähne nicht mehr möglich.

Wenn Sie Ihre Meinung ändern und die Zustimmung zur Teilnahme zurückziehen möchten, können Sie dies tun (dazu können Sie Peer Herholz unter 06421/58-63913 erreichen). In diesem Fall werden wir keine zusätzlichen Daten erheben und die bereits erhobenen Daten werden gelöscht, bevor diese in die Datenbank hochgeladen werden. Allerdings können Daten und Forschungsergebnisse, die bereits auf die öffentliche Datenbank geladen und mit anderen Forschern geteilt wurden, nicht wieder gelöscht oder zurückgezogen werden.

Es ist möglich, dass Studien, die Ihre Informationen nutzen, zu neuen Entwicklungen zur Erforschung und zum Verständnis des Gehirns, zu neuen Methoden, neuen Diagnosemethoden oder neuen (pharmazeutischen) Produkten führen. Durch die Zustimmung zur freiwilligen Weitergabe der Daten gibt es keine Möglichkeit, an den Profiten durch diese Produkte teilzuhaben. Auch die Rechte an diesen Methoden oder Produkten werden nicht weitergegeben. Die Weitergabe Ihrer Daten ist freiwillig. Sie können trotzdem an der Studie teilnehmen, auch wenn Sie nicht wollen, dass die Daten weitergegeben werden. Mit Ihrer Unterschrift bestätigen Sie, dass Sie die Daten für zukünftige Forschungen bereitstellen. Sie stimmen zu, dass diese global mit anderen Forschern und Instituten geteilt werden. Die Einzelheiten, Resultate und Implikationen dieser Studien sind unbekannt.

Marburg, der ______________________________
Datum, Unterschrift der/des ProbandIn

Quelle: Halchenko, Y. (2017). Open Brain Consent Documentation. Release 0.0.2.dev.
https://media.readthedocs.org/pdf/open-brain-consent/latest/open-brain-consent.pdf (zuletzt abgerufen am 06.02.2018, 10:06 Uhr)

B. CONSENT FORM PROJECT 4

Projektleiter:
Prof. Dr. Andreas Jansen
Klinik für Psychiatrie und Psychotherapie
Philipps-Universität Marburg
Rudolf-Bultmann-Straße 8, 35039 Marburg
Tel: 06421-58-65273
Fax: 06421-58-68939

Probanden-Information für ein Forschungsvorhaben mit behavioraler Testung zur Vorbereitung der mündlichen Aufklärung durch den Studienleiter für die Studie:

Nothing compares to you – Eine Pilotstudie zur Verarbeitung unterschiedlicher Musikeigenschaften und -genres

Sehr geehrte Frau, sehr geehrter Herr,

wir möchten Sie bitten, als Proband* (*Im Rahmen dieses Textes schließt die männliche Bezeichnung stets die weibliche Bezeichnung mit ein) an der Pilotstudie „Nothing compares to you - Eine Pilotstudie zur Verarbeitung unterschiedlicher Musikeigenschaften und -genres“ teilzunehmen. Dabei werden wir Sie mittels einer verhaltenspsychologischen Testung untersuchen, die Aussagen über die subjektive Wahrnehmung von musikalischen Stimuli erlaubt und dabei helfen soll, die Grundlagen und Zusammenhänge der Verarbeitung verschiedener Musikgenres besser zu verstehen.
In diesem Informationsblatt beschreiben wir die geplanten Untersuchungen, konkretisieren den Umgang mit den erhobenen Daten und nennen abschließend Ausschlusskriterien für die Teilnahme an der Studie.

(1) Informationen zu den Untersuchungen:

Ziel der Pilotstudie ist es, die Verarbeitung verschiedener Musikgenres in unterschiedlichen Probandengruppen zu untersuchen. Studienteilnehmer führen dabei einen verhaltenspsychologischen Test durch. Im ersten Untersuchungsteil werden am Computer eine Reihe von Fragebögen zu Musikpräferenz und musikalischem Training bearbeitet. Danach folgt ebenfalls computergestützt eine Verhaltensaufgabe, in der unterschiedliche Beispielstücke in Relation zueinander gesetzt werden sollen.

(2) Hiermit bitten wir Sie um Zustimmung zu folgender Untersuchung:

Erster Teil: Vorbereitung

Die Vorbereitung vor der eigentlichen Studie umfasst vier Tests (ca. 65 Minuten)

1. Musikpräferenz-Fragebogen (MPQ-R): Gibt Aufschluss über persönliche Musikvorlieben und musikalische Übung.
2. Short Test of Music Preference (STOMP): Gibt Aufschluss über die persönliche Musikpräferenz.

3. Montreal Battery of Evaluation of Amusia (MBEA): Fragebogen zum Ausschluss von Amusie (Unfähigkeit Tonfolgen und/oder Rhythmen zu erkennen und zu reproduzieren).
4. Musical Ear Test (MET): Feststellung des musikalischen Hörvermögens.

Zweiter Teil: Verhaltensstudie

In einem ruhigen, ablenkungsfreien Raum wird mithilfe eines Laptops und hochwertiger Kopfhörer eine sogenannte multidimensionale Skalierung durchgeführt. Dazu werden 20 6-sekündige Musikstimuli in unterschiedlicher Kombination präsentiert, die jeweils angehört und basierend auf ihrer Ähnlichkeit in einen räumlichen Zusammenhang gesetzt werden sollen. Durchschnittlich werden ca. 40 Minuten für das Experiment veranschlagt.

Die Teilnahme an diesen Untersuchungen ist freiwillig, Sie können jederzeit und ohne Angabe von Gründen Ihre Einwilligung zurückziehen, ohne dass Ihnen daraus Nachteile entstehen. Für die Teilnahme an den Testungen erhalten Sie eine finanzielle Aufwandsentschädigung (10€/h) oder eine Vergütung in Versuchspersonenstunden.

(3) Datenschutz

Ihre personenbezogenen Daten werden maschinell gespeichert und weiterverarbeitet. Die Weitergabe, Speicherung und Auswertung dieser studienbezogenen Daten erfolgt nach gesetzlichen Bestimmungen ohne Namensnennung und setzt vor Teilnahme an der Studie Ihre freiwillige Einwilligung voraus. Zur Pseudonymisierung Ihrer Daten bitten wir Sie fünf Zahlen Ihrer Wahl von 0 bis 100 mit zwei frei gewählten Großbuchstaben zu kombinieren (***** XX). Diese Kombination, die Sie auf jedem auszufüllenden Formular angeben, kann nur noch mithilfe einer Schlüsselliste zu Ihrer Person zugeordnet werden. Diese Schlüsselliste wird getrennt von den Messdaten in einem verschlossenen Schrank in der Klinik für Psychiatrie und Psychotherapie aufbewahrt und zwei Jahre nach Abschluss der Untersuchung gelöscht.

(4) Ausschlusskriterien

Da ein visuell gestützter Test zur Verarbeitung auditorischer Musikstimuli durchgeführt werden soll, sind Personen mit Hör- und/oder starken Sehschädigungen nicht für die Teilnahme geeignet.

Wenn weitere Fragen bestehen, werden diese gerne vom Untersuchungsleiter beantwortet.

Ihre Ansprechpartner:

- Peer Herholz (herholz@staff.uni-marburg.de, Telefon: 06421 / 5863913)
- Dr. Jens Sommer (jens.sommer@staff.uni-marburg.de, Telefon: 06421 / 5865581)

Studienleiter:

- Prof Dr. Andreas Jansen (jansena2@staff.uni-marburg.de, Telefon: 06421/5865273)

Einwilligungserklärung zur Teilnahme an dem Forschungsvorhaben

Nothing compares to you – Eine Pilotstudie zur Verarbeitung unterschiedlicher Musikeigenschaften und -genres

(behaviorale Testung)

Bei Ihrer Bereitschaft zur Teilnahme bitten wir Sie, die Einwilligungserklärung vor der Untersuchung vollständig auszufüllen und zu unterschreiben.

Ich bestätige hiermit, dass ich durch den Untersucher, Herrn/Frau .... ……………... über Wesen, Bedeutung, Risiken und Tragweite der beabsichtigten Untersuchung aufgeklärt wurde und für meine Entscheidung genügend Bedenkzeit hatte. Ich habe die Probandeninformation gelesen. Ich fühle mich ausreichend informiert und habe verstanden, worum es geht. Der Untersucher hat mir ausreichend Gelegenheit gegeben, Fragen zu stellen, die alle für mich ausreichend beantwortet wurden. Ich hatte genügend Zeit, mich zu entscheiden.

Ich habe verstanden, dass bei wissenschaftlichen Studien persönliche und verhaltenspsychologische Daten erhoben werden. Die Weitergabe, Speicherung und Auswertung dieser studienbezogenen Daten erfolgt nach gesetzlichen Bestimmungen und setzt vor Teilnahme an der Studie meine freiwillige Einwilligung voraus. Ich erkläre mich damit einverstanden, dass im Rahmen dieser Studie erhobene Daten auf Fragebögen und elektronischen Datenträgern aufgezeichnet und ohne Namensnennung zum Zwecke wissenschaftlicher Auswertung analysiert werden.

Ich habe eine Kopie der Probandeninformation und dieser unterschriebenen Einwilligungserklärung erhalten. Meine Einwilligung, an diesem Forschungsvorhaben als Proband teilzunehmen, erfolgt freiwillig. Ich wurde darauf hingewiesen, dass ich meine Einwilligung jederzeit ohne Angabe von Gründen und ohne Nachteile widerrufen kann.

Ich willige hiermit ein, als Proband an dem Forschungsvorhaben “Nothing compares to you – Eine Pilotstudie zur Verarbeitung unterschiedlicher Musikeigenschaften und -genres” teilzunehmen. Die Teilnahme umfasst die Bearbeitung von vier Fragebögen und eine verhaltenspsychologische Testung.

VOM PROBANDEN AUSZUFÜLLEN:

Name: ……………………………………….

Geburtsdatum: ……………………………..

Datum: ……………………….. Uhrzeit: ………………………

Ort: ……………………………. Unterschrift: ………………………………

VOM UNTERSUCHER AUSZUFÜLLEN:

Ich habe den Probanden mündlich über Wesen, Bedeutung, Reichweite und Risiken des Forschungsvorhabens aufgeklärt.

Datum: ……………………….. Uhrzeit: ………………………….

Ort: ……………………………..

Untersucher: …………………………………….…

Projektleiter:
Prof. Dr. Andreas Jansen
Klinik für Psychiatrie und Psychotherapie
Philipps-Universität Marburg
Rudolf-Bultmann-Straße 8, 35039 Marburg
Tel: 06421-58-65273
Fax: 06421-58-68939

Probanden-Information für Forschungsvorhaben mit MRT-Untersuchungen

zur Vorbereitung der mündlichen Aufklärung durch den Studienleiter für die Studie:

Nothing compares to you – Eine Pilotstudie zur neuronalen Verarbeitung unterschiedlicher Musikeigenschaften und -genres

Sehr geehrte Frau, sehr geehrter Herr,

wir möchten Sie bitten, als Proband* (*Im Rahmen dieses Textes schließt die männliche Bezeichnung stets die weibliche Bezeichnung mit ein) an der Pilotstudie „Nothing compares to you - Eine Studie zur neuronalen Verarbeitung unterschiedlicher Musikeigenschaften und -genres“ teilzunehmen. Dabei werden wir Sie mit bildgebenden Techniken (MRT) untersuchen, die Aussagen über Funktion und Struktur des Gehirns erlauben, und die dabei helfen sollen, die neuronalen Grundlagen und Zusammenhänge der Verarbeitung verschiedener Musikgenres besser zu verstehen.
In diesem Informationsblatt beschreiben wir die geplanten bildgebenden Untersuchungen, klären über mögliche Risiken auf, beschreiben den Umgang mit sog. „Zufallsbefunden“, konkretisieren den Umgang mit den erhobenen Daten und nennen abschließend Ausschlusskriterien für die Teilnahme an der Studie.

(1) Informationen zu den bildgebenden Untersuchungen:

Ziel der Pilotstudie ist es, die neuronale Verarbeitung verschiedener Musikgenres in unterschiedlichen Probandengruppen zu untersuchen. Studienteilnehmer werden dabei sowohl behavioral als auch mit funktioneller und struktureller Magnetresonanztomographie (MRT) untersucht. Dabei sollen im MRT während des Musikhörens neuronale Korrelate erfasst werden, die veranschaulichen, wie unterschiedliche Musikgenres verarbeitet werden.

(2) Hiermit bitten wir Sie um Zustimmung zu folgender Untersuchung:

Erster Teil: Vorbereitung

Die Vorbereitung vor der eigentlichen Studie umfasst vier Tests (ca. 65 Minuten)

1. Musikpräferenz-Fragebogen (MPQ-R): Gibt Aufschluss über persönliche Musikvorlieben und musikalische Übung.

2. Short Test of Music Preference (STOMP): Gibt Aufschluss über die persönliche Musikpräferenz.
3. Montreal Battery of Evaluation of Amusia (MBEA): Fragebogen zum Ausschluss von Amusie (Unfähigkeit Tonfolgen und/oder Rhythmen zu erkennen und zu reproduzieren).
4. Musical Ear Test (MET): Feststellung des musikalischen Hörvermögens.

Zweiter Teil: MRT-Studie

Bevor die Messung im Scanner beginnt, erhalten Sie außerhalb des Scanners eine Einführung und Aufklärung, die Ihnen das Vorgehen und Ihre Aufgabe im Detail erklärt. Danach startet die Untersuchung mit einer anatomischen Aufnahme ihres Gehirns. Wir bitten Sie dafür ca. 5 Minuten still liegen zu bleiben. Sie müssen währenddessen keine Aufgabe bearbeiten.

Anschließend werden die funktionellen Bilder Ihres Gehirns aufgenommen. Dazu werden Ihnen über Kopfhörer 6-sekündige Musikstimuli in Blöcken präsentiert, die jeweils von kurzen Pausen unterbrochen werden. Begleitend schauen Sie einen Spielfilm ohne Ton, aber mit Untertiteln an. Auch hier bitten wir Sie, möglichst still liegen zu bleiben und Sie müssen keine weitere Aufgabe bearbeiten.

Mit Vor- und Nachbereitung wird die MRT-Studie ca. 60 Minuten in Anspruch nehmen.

Dritter Teil: Verhaltensstudie

In einem ruhigen, ablenkungsfreien Raum wird mithilfe eines Laptops und hochwertiger Kopfhörer eine sogenannte multidimensionale Skalierung durchgeführt. Dazu werden 20 6-sekündige Musikstimuli in unterschiedlicher Kombination präsentiert, die jeweils angehört und basierend auf ihrer Ähnlichkeit in einen räumlichen Zusammenhang gesetzt werden sollen. Durchschnittlich werden ca. 40 Minuten für das Experiment veranschlagt.

Die Teilnahme an diesen Untersuchungen ist freiwillig. Sie können jederzeit und ohne Angabe von Gründen Ihre Einwilligung zurückziehen, ohne dass Ihnen daraus Nachteile entstehen. Für die Teilnahme an den Testungen erhalten Sie eine finanzielle Aufwandsentschädigung (10€/h) oder eine Vergütung in Versuchspersonenstunden sowie auf Wunsch eine CD mit Bildern Ihres Gehirns.

(3) Ablauf der Kernspintomographieuntersuchung und mögliche Gefahren

Das menschliche Gehirn besteht aus einer Vielzahl von Molekülen, die jeweils spezifische magnetische Eigenschaften aufweisen (sog. Kernspinresonanz). Bei Anwendung von starken Magnetfeldern können diese gemessen und im Anschluss daran die Konzentration von Molekülen im Gehirn festgestellt werden. Auf diese Weise lässt sich die Durchblutung des Gehirns ermitteln, genauer die kontinuierlichen Veränderungen des Sauerstoffgehalts im Blut innerhalb weniger Sekunden. Diese Veränderungen ergeben spezifische Muster, je nachdem, welche Sinnesreize oder Vorstellungen die Versuchsperson während der Messung innerlich verarbeitet. Auf diese Weise können unterschiedliche Funktionen des Gehirns sichtbar gemacht werden. Man spricht hier von funktioneller Magnetresonanztomographie (fMRT).

Ablauf der Untersuchung:

Sie liegen während der Untersuchung auf einem Tisch, der Sie mit dem Oberkörper in die Öffnung des MR-Tomographen hineinfährt. Zusätzlich wird ein Magnetspulenrahmen, ähnlich einem Helm, um Ihren Kopf gelegt.
Während der Messung entstehen unterschiedlich laute Geräusche des MR-Tomographen, die von elektrischen Umschaltvorgängen der Magnetfelder herrühren. Um deren Einfluss gering zu halten, tragen Sie einen Gehörschutz über den Kopfhörern. Sie haben während der Untersuchung jederzeit die Möglichkeit, mit den Untersuchern über eine Wechselsprechanlage in Kontakt zu treten. Zusätzlich bekommen Sie einen gesonderten Alarmknopf für Notfälle, sodass Sie auf Ihren Wunsch hin jederzeit aus dem MR-Tomographen herausgefahren werden können.

Mögliche Gefahren:

Die MRT-Technologie ist für den Körper nach heutigem Erkenntnisstand unschädlich. Sie basiert auf mehr als 20 Jahren Erfahrung und wird täglich in allen größeren Kliniken eingesetzt. Bekannte Risiken ergeben sich ausschließlich durch metallische Gegenstände oder Stoffe mit magnetischen Eigenschaften, die sich am oder im Körper befinden. Diese können sich erhitzen und zu Verbrennungen führen. Lose Metallteile können durch das Magnetfeld beschleunigt werden und dann zu Verletzungen führen. Daher sind Personen von der Teilnahme an der Studie ausgeschlossen, die elektrische Geräte (z.B. Herzschrittmacher, Medikamentenpumpen) oder Metallteile (z.B. Schrauben nach Knochenbruch oder eine Spirale zur Verhütung) im oder am Körper haben.
Risiken unabhängig von den Genannten sind bislang nicht bekannt. Abgesehen von möglichen Unbequemlichkeiten, die vom stillen Liegen in der engen fMRT-Röhre resultieren, und der Lautstärke des Tomographen, sollten keine Beschwerden während der Untersuchung auftreten.
Wir möchten Sie allerdings darauf hinweisen, dass über mögliche langfristige Risiken bei wiederholten MRT-Messungen bisher keine wissenschaftlich abgesicherten Ergebnisse vorliegen.

(4) Umgang mit Zufallsbefunden

Bei der Studie handelt es sich um eine Forschungsstudie. Eine neuroradiologische Befundung der MR-Bilder im Sinne einer klinisch orientierten Diagnostik findet daher nicht statt. Dennoch kann es vorkommen, dass in den MR-Bildern Signalauffälligkeiten entdeckt werden, die eine mögliche klinische Relevanz haben („Zufallsbefund“). Falls sich bei der Untersuchung Anhaltspunkte für einen Zufallsbefund ergeben, die eine fachärztliche neuroradiologische Diagnostik empfehlenswert erscheinen lassen, werden Sie persönlich darüber informiert und Ihnen eine fachärztlich neuroradiologische Diagnostik empfohlen.

Falls Sie über einen Zufallsbefund nicht informiert werden wollen, stellt dies ein Ausschlusskriterium für die Teilnahme an der Studie dar.

(5) Datenschutz

Ihre personenbezogenen Daten werden maschinell gespeichert und weiterverarbeitet. Die Weitergabe, Speicherung und Auswertung dieser studienbezogenen Daten erfolgt nach gesetzlichen Bestimmungen ohne Namensnennung und setzt vor Teilnahme an der Studie Ihre freiwillige Einwilligung voraus. Zur Pseudonymisierung Ihrer Daten bitten wir Sie fünf Zahlen Ihrer Wahl von 0 bis 100 mit zwei frei gewählten Großbuchstaben zu kombinieren (***** XX). Diese Kombination, die Sie auf jedem auszufüllenden Formular angeben, kann nur noch mithilfe einer Schlüsselliste zu Ihrer Person zugeordnet werden. Diese Schlüsselliste wird getrennt von den Messdaten in einem verschlossenen Schrank in der Klinik für Psychiatrie und Psychotherapie aufbewahrt und zwei Jahre nach Abschluss der Untersuchung gelöscht.

(6) Ausschlusskriterien

Die Anwendung von Magnetfeldern bei der MRT-Untersuchung schließt die Teilnahme von Personen aus, die elektrische Geräte (z.B. Herzschrittmacher, Medikamentenpumpen usw.) oder Metallteile (z.B. Schrauben nach Knochenbruch) im oder am Körper haben. Frauen, die eine Spirale haben oder schwanger sind, werden nicht als Probandinnen zugelassen. Probanden, die nicht über mögliche Zufallsbefunde informiert werden wollen, können ebenfalls nicht an der Studie teilnehmen.

Wenn weitere Fragen bestehen, werden diese gerne vom Untersuchungsleiter beantwortet.

Ihre Ansprechpartner:

- Peer Herholz (herholz@staff.uni-marburg.de, Telefon: 06421 / 5863913)
- Dr. Jens Sommer (jens.sommer@staff.uni-marburg.de, Telefon: 06421 / 5865581)

Studienleiter:

- Prof Dr. Andreas Jansen (jansena2@staff.uni-marburg.de, Telefon: 06421/5865273)

Philipps Universität Marburg

Einwilligungserklärung zur Teilnahme an dem Forschungsvorhaben

Nothing compares to you – Eine Pilotstudie zur neuronalen Verarbeitung unterschiedlicher Musikeigenschaften und -genres

(behaviorale Testung & fMRT-Untersuchung)

Bei Ihrer Bereitschaft zur Teilnahme bitten wir Sie, die Einwilligungserklärung vor der Untersuchung vollständig auszufüllen und zu unterschreiben.

Ich bestätige hiermit, dass ich durch den Untersucher, Herrn/Frau .... ………………... über Wesen, Bedeutung, Risiken und Tragweite der beabsichtigten Untersuchung aufgeklärt wurde und für meine Entscheidung genügend Bedenkzeit hatte. Ich habe die Probandeninformation gelesen. Ich fühle mich ausreichend informiert und habe verstanden, worum es geht. Der Untersucher hat mir ausreichend Gelegenheit gegeben, Fragen zu stellen, die alle für mich ausreichend beantwortet wurden. Ich hatte genügend Zeit, mich zu entscheiden.

Ich wurde darauf hingewiesen, dass es sich bei der Studie um eine Forschungsstudie handelt. Eine neuroradiologische Befundung der MR-Bilder im Sinne einer klinisch orientierten Diagnostik findet daher nicht statt. Dennoch kann es vorkommen, dass in den MR-Bildern Signalauffälligkeiten entdeckt werden, die eine mögliche klinische Relevanz haben („Zufallsbefund“). Mir ist bekannt, dass der Versuchsleiter mich informieren würde, falls sich bei der Untersuchung Anhaltspunkte für einen Zufallsbefund ergeben, die eine fachärztliche neuro-radiologische Diagnostik empfehlenswert erscheinen lassen.

Ich habe verstanden, dass bei wissenschaftlichen Studien persönliche Daten und medizinische Befunde erhoben werden. Die Weitergabe, Speicherung und Auswertung dieser studienbezogenen Daten erfolgt nach gesetzlichen Bestimmungen und setzt vor Teilnahme an der Studie meine freiwillige Einwilligung voraus. Ich erkläre mich damit einverstanden, dass im Rahmen dieser Studie erhobene Daten auf Fragebögen und elektronischen Datenträgern aufgezeichnet und ohne Namensnennung zum Zwecke wissenschaftlicher Auswertung analysiert werden.

Ich habe eine Kopie der Probandeninformation und dieser unterschriebenen Einwilligungserklärung erhalten. Meine Einwilligung, an diesem Forschungsvorhaben als Proband teilzunehmen, erfolgt freiwillig. Ich wurde darauf hingewiesen, dass ich meine Einwilligung jederzeit ohne Angabe von Gründen und ohne Nachteile widerrufen kann.

Ich willige hiermit ein, als Proband an dem Forschungsvorhaben “Nothing compares to you – Eine Pilotstudie zur neuronalen Verarbeitung unterschiedlicher Musikeigenschaften und -genres” teilzunehmen. Die Teilnahme umfasst die Bearbeitung von vier Fragebögen sowie eine Verhaltenstestung und eine Untersuchung mit dem MR-Scanner.

VOM PROBANDEN AUSZUFÜLLEN:

Name: ……………………………………….

Geburtsdatum: ……………………………..

Datum: ……………………….. Uhrzeit: ………………………

Ort: ……………………………. Unterschrift: ………………………………

VOM UNTERSUCHER AUSZUFÜLLEN:

Ich habe den Probanden mündlich über Wesen, Bedeutung, Reichweite und Risiken des Forschungsvorhabens aufgeklärt.

Datum: ……………………….. Uhrzeit: ………………………….

Ort: ……………………………..

Untersucher: ……………………………………….…

Philipps Universität Marburg

ProbandInnen-Information *Open Brain Consent Documentation*

Wir möchten die in dieser Studie erhobenen Daten, inklusive der fMRT-Bilddaten, der Allgemeinheit zur Verfügung stellen (über eine öffentliche Internetdatenbank). Es ist daher möglich, dass die Daten der Untersuchung, an der Sie aktuell teilnehmen, zukünftig auch für andere Forschungsprojekte genutzt werden. Diese Projekte könnten andere Zielsetzungen verfolgen, als die aktuelle Studie.
Alle Daten, die wir teilen, werden vollständig pseudonymisiert. Ihr Name wird nicht gespeichert, sondern nur eine Codenummer, so dass niemand weiß, welche Daten zu Ihnen gehören. Zur Pseudonymisierung Ihrer Daten bitten wir Sie fünf Zahlen Ihrer Wahl von 0 bis 100 mit zwei frei gewählten Großbuchstaben zu kombinieren (***** XX). Die Verbindung zwischen den Messdaten und Ihrer Identität kann dann nur mit einer Schlüsselliste hergestellt werden. Diese wird getrennt von den Messdaten in einem verschlossenen Schrank in der Klinik für Psychiatrie und Psychotherapieaufbewahrt und zwei Jahre nach Abschluss der Untersuchung gelöscht. Zusätzlich dazu werden wir Daten entfernen, die unter Umständen dazu führen könnten, dass Rückschlüsse auf Ihre Person gezogen werden könnten (wie zum Beispiel Gesichtsmerkmale oder das Datum der Teilnahme). Dazu wird ein sogenanntes Defacing vorgenommen, bei welchem Ihr Gesicht und damit Ihre Erkennungsmerkmale aus den fMRT-Daten entfernt werden. Damit ist eine Zuordnung Ihrer Daten durch Gesicht oder Zähne nicht mehr möglich.

Wenn Sie Ihre Meinung ändern und die Zustimmung zur Teilnahme zurückziehen möchten, können Sie dies tun (dazu können Sie Peer Herholz unter 06421/58-63913 erreichen). In diesem Fall werden wir keine zusätzlichen Daten erheben und die bereits erhobenen Daten werden gelöscht, bevor diese in die Datenbank hochgeladen werden. Allerdings können Daten und Forschungsergebnisse, die bereits auf die öffentliche Datenbank geladen und mit anderen Forschern geteilt wurden, nicht wieder gelöscht oder zurückgezogen werden. Es besteht weiterhin ein Beschwerderecht bei einer Datenschutz- Aufsichtsbehörde, dem hessischen Datenschutz-Beauftragten (Der Hessische Datenschutzbeauftragte, Postfach 3163, 65021 Wiesbaden).

Es ist möglich, dass Studien, die Ihre Informationen nutzen, zu neuen Entwicklungen zur Erforschung und zum Verständnis des Gehirns, zu neuen Methoden, neuen Diagnosemethoden oder neuen (pharmazeutischen) Produkten führen. Durch die Zustimmung zur freiwilligen Weitergabe der Daten gibt es keine Möglichkeit, an den Profiten durch diese Produkte teilzuhaben. Auch die Rechte an diesen Methoden oder Produkten werden nicht weitergegeben. Die Weitergabe Ihrer Daten ist freiwillig. Sie können trotzdem an der Studie teilnehmen, auch wenn Sie nicht wollen, dass die Daten weitergegeben werden. Mit Ihrer Unterschrift bestätigen Sie, dass Sie die Daten für zukünftige Forschungen bereitstellen. Sie stimmen zu, dass diese global mit anderen Forschern und Instituten geteilt werden. Die Einzelheiten, Resultate und Implikationen dieser Studien sind unbekannt. Die für die Datenverarbeitung dieses Projektes verantwortliche Person ist Peer Herholz.

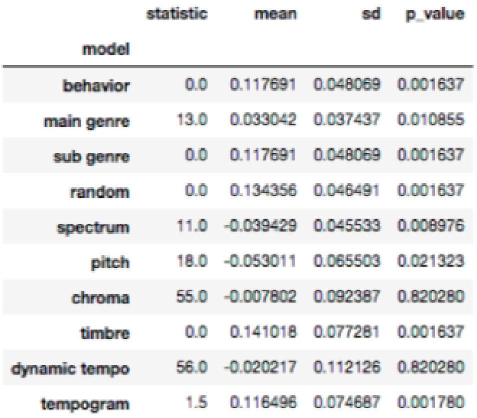
Marburg, der ______________________________
Datum, Unterschrift der/des ProbandIn

Quelle: Halchenko, Y. (2017). Open Brain Consent Documentation. Release 0.0.2.dev.
https://media.readthedocs.org/pdf/open-brain-consent/latest/open-brain-consent.pdf (zuletzt abgerufen am 01.11.2017, 10:15 Uhr)

C. MODEL COMPARISONS PROJECT 4

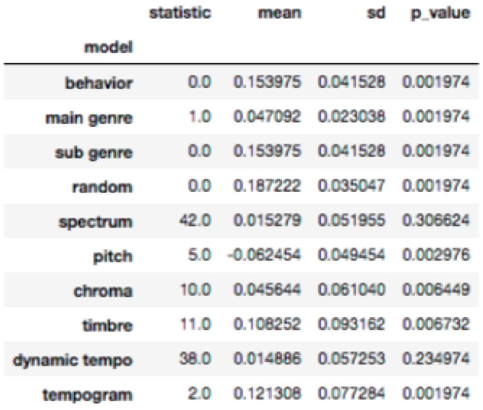
C.1. MODEL COMPARISONS PROJECT 4 "NOTHING COMPARES TO YOU I"

HG LH



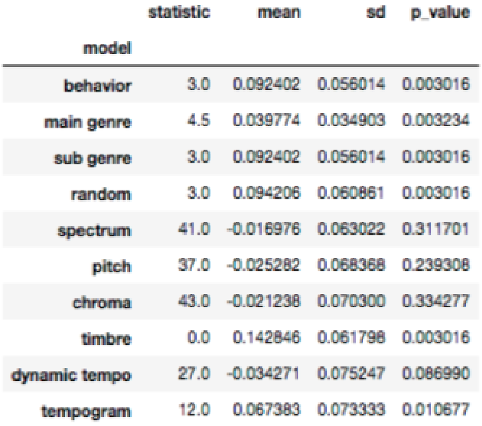
	statistic	mean	sd	p_value
model				
behavior	0.0	0.117691	0.048069	0.001637
main genre	13.0	0.033042	0.037437	0.010855
sub genre	0.0	0.117691	0.048069	0.001637
random	0.0	0.134356	0.046491	0.001637
spectrum	11.0	-0.039429	0.045533	0.008976
pitch	18.0	-0.053011	0.065503	0.021323
chroma	55.0	-0.007802	0.092387	0.820280
timbre	0.0	0.141018	0.077281	0.001637
dynamic tempo	56.0	-0.020217	0.112126	0.820280
tempogram	1.5	0.116496	0.074687	0.001780

HG RH



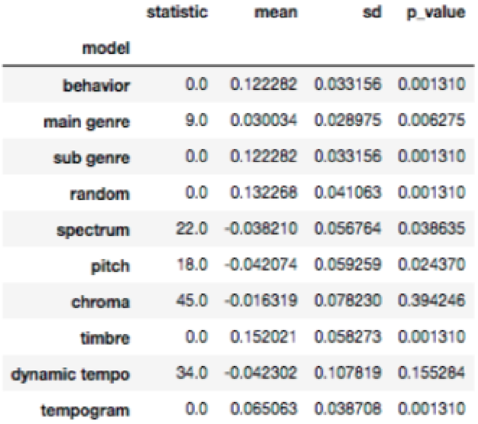
	statistic	mean	sd	p_value
model				
behavior	0.0	0.153975	0.041528	0.001974
main genre	1.0	0.047092	0.023038	0.001974
sub genre	0.0	0.153975	0.041528	0.001974
random	0.0	0.187222	0.035047	0.001974
spectrum	42.0	0.015279	0.051955	0.306624
pitch	5.0	-0.062454	0.049454	0.002976
chroma	10.0	0.045644	0.061040	0.006449
timbre	11.0	0.108252	0.093162	0.006732
dynamic tempo	38.0	0.014886	0.057253	0.234974
tempogram	2.0	0.121308	0.077284	0.001974

PP LH



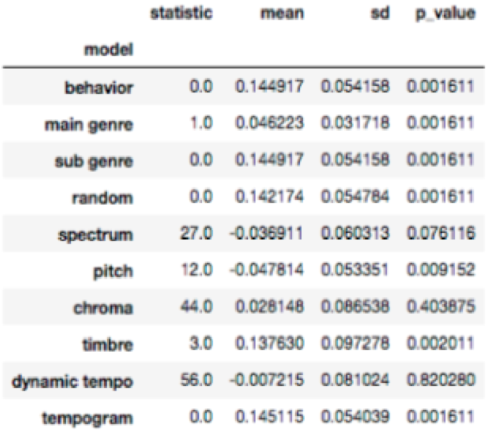
	statistic	mean	sd	p_value
model				
behavior	3.0	0.092402	0.056014	0.003016
main genre	4.5	0.039774	0.034903	0.003234
sub genre	3.0	0.092402	0.056014	0.003016
random	3.0	0.094206	0.060861	0.003016
spectrum	41.0	-0.016976	0.063022	0.311701
pitch	37.0	-0.025282	0.068368	0.239308
chroma	43.0	-0.021238	0.070300	0.334277
timbre	0.0	0.142846	0.061798	0.003016
dynamic tempo	27.0	-0.034271	0.075247	0.086990
tempogram	12.0	0.067383	0.073333	0.010677

PP RH



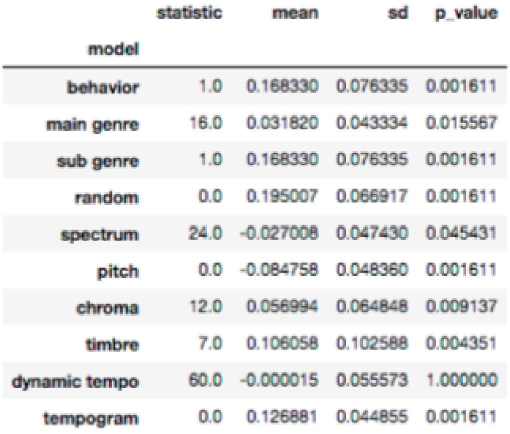
	statistic	mean	sd	p_value
model				
behavior	0.0	0.122282	0.033156	0.001310
main genre	9.0	0.030034	0.028975	0.006275
sub genre	0.0	0.122282	0.033156	0.001310
random	0.0	0.132268	0.041063	0.001310
spectrum	22.0	-0.038210	0.056764	0.038635
pitch	18.0	-0.042074	0.059259	0.024370
chroma	45.0	-0.016319	0.078230	0.394246
timbre	0.0	0.152021	0.058273	0.001310
dynamic tempo	34.0	-0.042302	0.107819	0.155284
tempogram	0.0	0.065063	0.038708	0.001310

PT LH



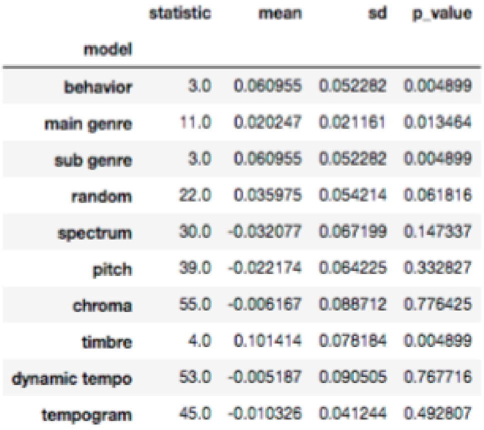
model	statistic	mean	sd	p_value
behavior	0.0	0.144917	0.054158	0.001611
main genre	1.0	0.046223	0.031718	0.001611
sub genre	0.0	0.144917	0.054158	0.001611
random	0.0	0.142174	0.054784	0.001611
spectrum	27.0	-0.036911	0.060313	0.076116
pitch	12.0	-0.047814	0.053351	0.009152
chroma	44.0	0.028148	0.086538	0.403875
timbre	3.0	0.137630	0.097278	0.002011
dynamic tempo	56.0	-0.007215	0.081024	0.820280
tempogram	0.0	0.145115	0.054039	0.001611

PT RH



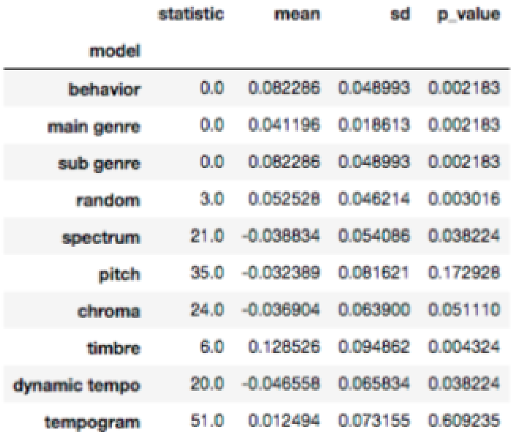
model	statistic	mean	sd	p_value
behavior	1.0	0.168330	0.076335	0.001611
main genre	16.0	0.031820	0.043334	0.015567
sub genre	1.0	0.168330	0.076335	0.001611
random	0.0	0.195007	0.066917	0.001611
spectrum	24.0	-0.027008	0.047430	0.045431
pitch	0.0	-0.084758	0.048360	0.001611
chroma	12.0	0.056994	0.064848	0.009137
timbre	7.0	0.106058	0.102588	0.004351
dynamic tempo	60.0	-0.000015	0.055573	1.000000
tempogram	0.0	0.126881	0.044855	0.001611

STGa LH



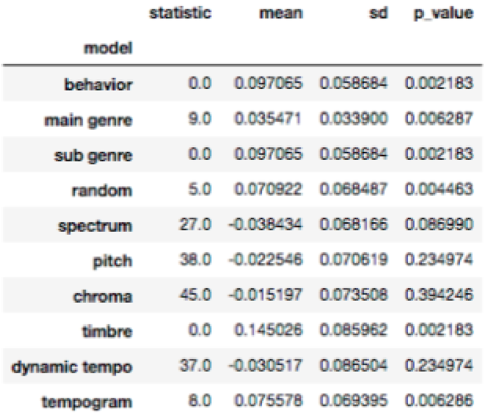
model	statistic	mean	sd	p_value
behavior	3.0	0.060955	0.052282	0.004899
main genre	11.0	0.020247	0.021161	0.013464
sub genre	3.0	0.060955	0.052282	0.004899
random	22.0	0.035975	0.054214	0.061816
spectrum	30.0	-0.032077	0.067199	0.147337
pitch	39.0	-0.022174	0.064225	0.332827
chroma	55.0	-0.006167	0.088712	0.776425
timbre	4.0	0.101414	0.078184	0.004899
dynamic tempo	53.0	-0.005187	0.090505	0.767716
tempogram	45.0	-0.010326	0.041244	0.492807

STGa RH



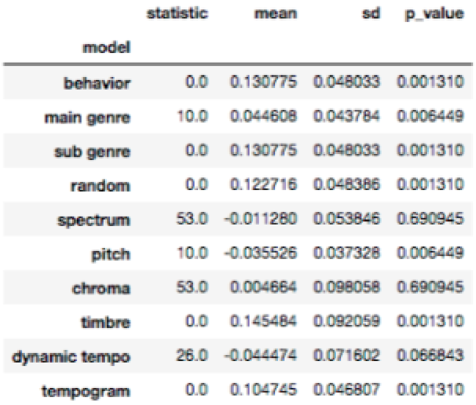
model	statistic	mean	sd	p_value
behavior	0.0	0.082286	0.048993	0.002183
main genre	0.0	0.041196	0.018613	0.002183
sub genre	0.0	0.082286	0.048993	0.002183
random	3.0	0.052528	0.046214	0.003016
spectrum	21.0	-0.038834	0.054086	0.038224
pitch	35.0	-0.032389	0.081621	0.172928
chroma	24.0	-0.036904	0.063900	0.051110
timbre	6.0	0.128526	0.094862	0.004324
dynamic tempo	20.0	-0.046558	0.065834	0.038224
tempogram	51.0	0.012494	0.073155	0.609235

STGp LH



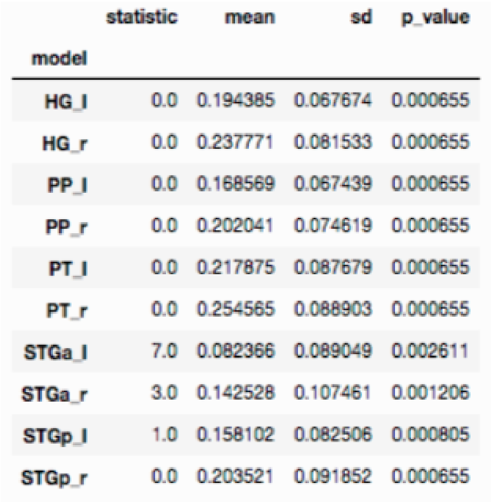
model	statistic	mean	sd	p_value
behavior	0.0	0.097065	0.058684	0.002183
main genre	9.0	0.035471	0.033900	0.006287
sub genre	0.0	0.097065	0.058684	0.002183
random	5.0	0.070922	0.068487	0.004463
spectrum	27.0	-0.038434	0.068166	0.086990
pitch	38.0	-0.022546	0.070619	0.234974
chroma	45.0	-0.015197	0.073508	0.394246
timbre	0.0	0.145026	0.085962	0.002183
dynamic tempo	37.0	-0.030517	0.086504	0.234974
tempogram	8.0	0.075578	0.069395	0.006286

STGp RH



model	statistic	mean	sd	p_value
behavior	0.0	0.130775	0.048033	0.001310
main genre	10.0	0.044608	0.043784	0.006449
sub genre	0.0	0.130775	0.048033	0.001310
random	0.0	0.122716	0.048386	0.001310
spectrum	53.0	-0.011280	0.053846	0.690945
pitch	10.0	-0.035526	0.037328	0.006449
chroma	53.0	0.004664	0.098058	0.690945
timbre	0.0	0.145484	0.092059	0.001310
dynamic tempo	26.0	-0.044474	0.071602	0.066843
tempogram	0.0	0.104745	0.046807	0.001310

Behavior

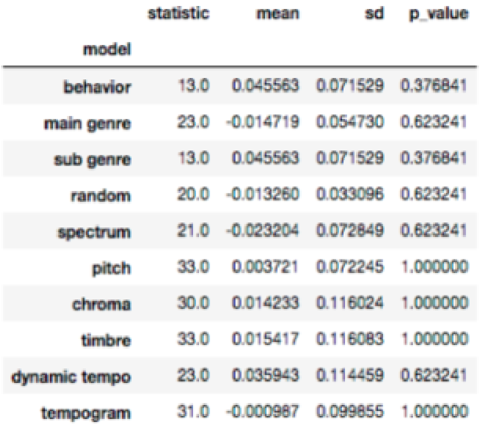


model	statistic	mean	sd	p_value
HG_l	0.0	0.194385	0.067674	0.000655
HG_r	0.0	0.237771	0.081533	0.000655
PP_l	0.0	0.168569	0.067439	0.000655
PP_r	0.0	0.202041	0.074619	0.000655
PT_l	0.0	0.217875	0.087679	0.000655
PT_r	0.0	0.254565	0.088903	0.000655
STGa_l	7.0	0.082366	0.089049	0.002611
STGa_r	3.0	0.142528	0.107461	0.001206
STGp_l	1.0	0.158102	0.082506	0.000805
STGp_r	0.0	0.203521	0.091852	0.000655

Figure 1: *Model comparison results of project 4 “Nothing compares to you I”. Displayed are the statistical values for the model comparison within each ROI, as well as behavior. ROIs and behavior are indicated through captions. Each row indicates one model. Tables contain the statistic of the correlation, the mean correlation value, the standard deviation and the p value (corrected for multiple comparisons at $p < 0.05$).*

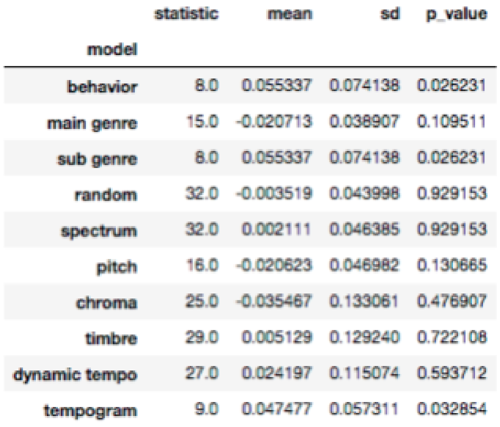
C.2. MODEL COMPARISONS PROJECT 4 "NOTHING COMPARES TO YOU II"

time window 1



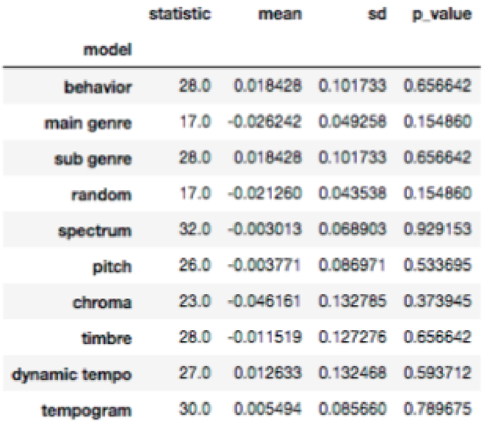
	statistic	mean	sd	p_value
model				
behavior	13.0	0.045563	0.071529	0.376841
main genre	23.0	-0.014719	0.054730	0.623241
sub genre	13.0	0.045563	0.071529	0.376841
random	20.0	-0.013260	0.033096	0.623241
spectrum	21.0	-0.023204	0.072849	0.623241
pitch	33.0	0.003721	0.072245	1.000000
chroma	30.0	0.014233	0.116024	1.000000
timbre	33.0	0.015417	0.116083	1.000000
dynamic tempo	23.0	0.035943	0.114459	0.623241
tempogram	31.0	-0.000987	0.099855	1.000000

time window 2



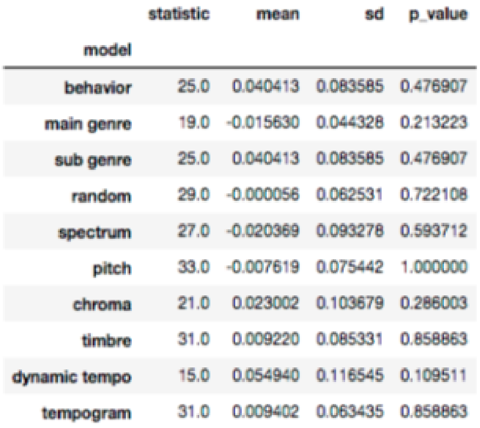
	statistic	mean	sd	p_value
model				
behavior	8.0	0.055337	0.074138	0.026231
main genre	15.0	-0.020713	0.038907	0.109511
sub genre	8.0	0.055337	0.074138	0.026231
random	32.0	-0.003519	0.043998	0.929153
spectrum	32.0	0.002111	0.046385	0.929153
pitch	16.0	-0.020623	0.046982	0.130665
chroma	25.0	-0.035467	0.133061	0.476907
timbre	29.0	0.005129	0.129240	0.722108
dynamic tempo	27.0	0.024197	0.115074	0.593712
tempogram	9.0	0.047477	0.057311	0.032854

time window 3

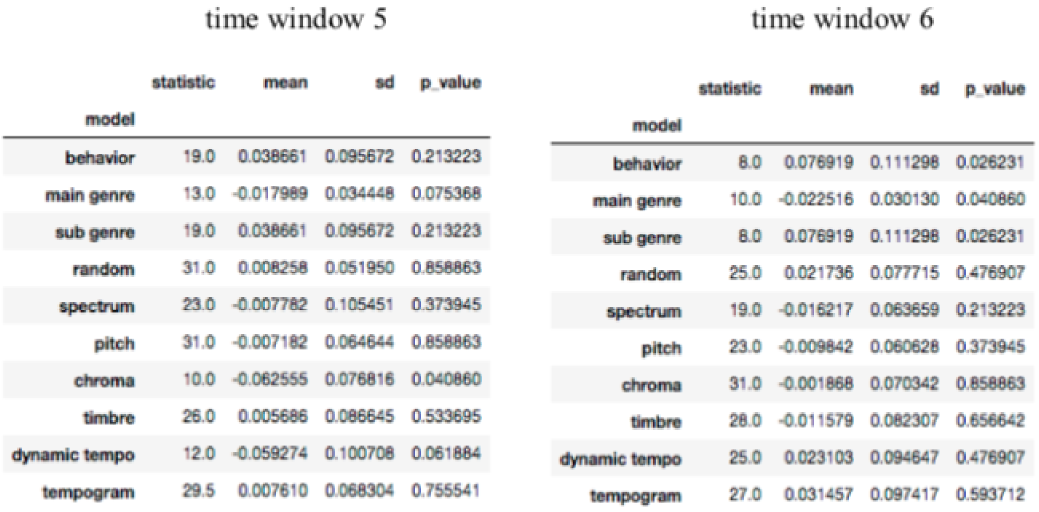


	statistic	mean	sd	p_value
model				
behavior	28.0	0.018428	0.101733	0.656642
main genre	17.0	-0.026242	0.049258	0.154860
sub genre	28.0	0.018428	0.101733	0.656642
random	17.0	-0.021260	0.043538	0.154860
spectrum	32.0	-0.003013	0.068903	0.929153
pitch	26.0	-0.003771	0.086971	0.533695
chroma	23.0	-0.046161	0.132785	0.373945
timbre	28.0	-0.011519	0.127276	0.656642
dynamic tempo	27.0	0.012633	0.132468	0.593712
tempogram	30.0	0.005494	0.085660	0.789675

time window 4



	statistic	mean	sd	p_value
model				
behavior	25.0	0.040413	0.083585	0.476907
main genre	19.0	-0.015630	0.044328	0.213223
sub genre	25.0	0.040413	0.083585	0.476907
random	29.0	-0.000056	0.062531	0.722108
spectrum	27.0	-0.020369	0.093278	0.593712
pitch	33.0	-0.007619	0.075442	1.000000
chroma	21.0	0.023002	0.103679	0.286003
timbre	31.0	0.009220	0.085331	0.858863
dynamic tempo	15.0	0.054940	0.116545	0.109511
tempogram	31.0	0.009402	0.063435	0.858863



time window 5

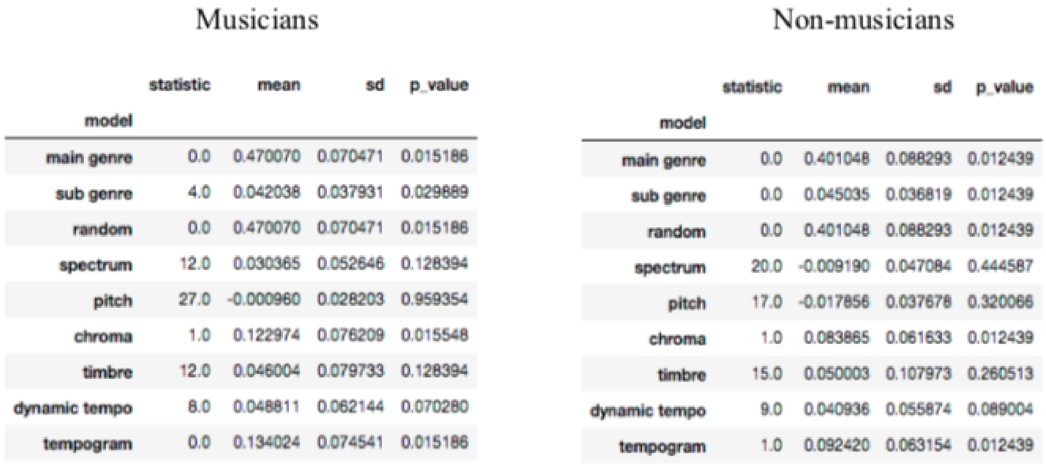
model	statistic	mean	sd	p_value
behavior	19.0	0.038661	0.095672	0.213223
main genre	13.0	-0.017989	0.034448	0.075368
sub genre	19.0	0.038661	0.095672	0.213223
random	31.0	0.008258	0.051950	0.858863
spectrum	23.0	-0.007782	0.105451	0.373945
pitch	31.0	-0.007182	0.064644	0.858863
chroma	10.0	-0.062555	0.076816	0.040860
timbre	26.0	0.005686	0.086645	0.533695
dynamic tempo	12.0	-0.059274	0.100708	0.061884
tempogram	29.5	0.007610	0.068304	0.755541

time window 6

model	statistic	mean	sd	p_value
behavior	8.0	0.076919	0.111298	0.026231
main genre	10.0	-0.022516	0.030130	0.040860
sub genre	8.0	0.076919	0.111298	0.026231
random	25.0	0.021736	0.077715	0.476907
spectrum	19.0	-0.016217	0.063659	0.213223
pitch	23.0	-0.009842	0.060628	0.373945
chroma	31.0	-0.001868	0.070342	0.858863
timbre	28.0	-0.011579	0.082307	0.656642
dynamic tempo	25.0	0.023103	0.094647	0.476907
tempogram	27.0	0.031457	0.097417	0.593712

Figure 2: *Model comparison results of project 4 “Nothing compares to you II”. Displayed are the statistical values for the model comparison within each time window, as well as behavior. Time windows and behavior are indicated through captions. Each row indicates one model. Tables contain the statistic of the correlation, the mean correlation value, the standard deviation and the p value (corrected for multiple comparisons at p < 0.05).*

C.3. MODEL COMPARISONS PROJECT 4 "NOTHING COMPARES TO YOU III"



Musicians

model	statistic	mean	sd	p_value
main genre	0.0	0.470070	0.070471	0.015186
sub genre	4.0	0.042038	0.037931	0.029889
random	0.0	0.470070	0.070471	0.015186
spectrum	12.0	0.030365	0.052646	0.128394
pitch	27.0	-0.000960	0.028203	0.959354
chroma	1.0	0.122974	0.076209	0.015548
timbre	12.0	0.046004	0.079733	0.128394
dynamic tempo	8.0	0.048811	0.062144	0.070280
tempogram	0.0	0.134024	0.074541	0.015186

Non-musicians

model	statistic	mean	sd	p_value
main genre	0.0	0.401048	0.088293	0.012439
sub genre	0.0	0.045035	0.036819	0.012439
random	0.0	0.401048	0.088293	0.012439
spectrum	20.0	-0.009190	0.047084	0.444587
pitch	17.0	-0.017856	0.037678	0.320066
chroma	1.0	0.083865	0.061633	0.012439
timbre	15.0	0.050003	0.107973	0.260513
dynamic tempo	9.0	0.040936	0.055874	0.089004
tempogram	1.0	0.092420	0.063154	0.012439

Figure 3: *Model comparison results of project 4 “Nothing compares to you III”. Displayed are the statistical values for the model comparison within each group. Groups are indicated through captions. Each row indicates one model. Tables contain the statistic of the correlation, the mean correlation value, the standard deviation and the p value (corrected for multiple comparisons at p < 0.05).*

CURRICULUM VITÆ

The CV was removed from the final print version.

The CV was removed from the final print version.

The CV was removed from the final print version.

The CV was removed from the final print version.

The CV was removed from the final print version.

The CV was removed from the final print version.

REGISTER OF LECTURERS

Meine akademischen Lehrer waren die Damen/Herren in Marburg:
Berger
Borchers
Bremmer
Decher
Domahs
Endres
Franz
Frobel
Homberg
Jansen
de Jong-Lendle
Kauschke
Künzel
Nagels
Oberwinkler
Oertel
Oliver
Peper
Plant
Rust
Schachtner
Schubö
Schwarting
Sommer
Thieme
Thorwart
Weihe
Westermann
Wiese

In Leipzig:
Awagana
Dölling
Fairhurst
Glück
Heck
Hoppe-Graff
Keller
Müller
Opitz
Pappert
Pechmann

Stiebels
Trommer
Zimmermann

In Montréal, Kanada:
Armony
Jerbi
Peretz

ACKNOWLEDGEMENTS

The acknowledgments were removed from the final print version.

The acknowledgments were removed from the final print version.

The acknowledgments were removed from the final print version.

The acknowledgments were removed from the final print version.

EHRENWOERTLICHE ERKLAERUNG

EHRENWOERTLICHE EKLAERUNG
Ich erklaere ehrenwoertlich, dass ich die dem Fachbereich Medizin Marburg zur Promotionspruefung eingereichte Arbeit mit dem Titel “The neurobiology of cortical music representations” in der Klinik fur Psychiatrie und Psychotherapue unter Leitung von Prof. Dr. Andreas Jansen ohne sonstige Hilfe selbst durchgefuehrt und bei der Abfassung der Arbeit keine anderen als die in der Dissertation aufgefuehrten Hilfsmittel benutzt habe. Ich habe bisher an keinem in- oder auslaendischen Medizinischen Fachbereich ein Gesuch um Zulassung zur Promotion eingereicht, noch die vorliegende oder eine andere Arbeit als Dissertation vorgelegt. Ich versichere, dass ich saemtliche woertlichen oder sinngemaeßen Uebernahmen und Zitate kenntlich gemacht habe. Mit dem Einsatz von Software zur Erkennung von Plagiaten bin ich einverstanden.

Ort, Datum, Unterschrift

Die Hinweise zur Erkennung von Plagiaten habe ich zur Kenntnis genommen.

Ort, Datum, Unterschrift Betreuer