Investigating the Neural Correlates of Social and Emotional Processing Across Modalities

Jocelyne C. Whitehead

Integrated Program in Neuroscience
McGill University, Montreal
July 2021

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Table of Contents

Abstract

In an increasingly globalized world, we are witness to the growth of population diversity. It is custom to observe individuals of diverging cultures and experiences, expressing themselves uniquely. To best adapt, we must communicate, relate, and understand one another. Markedly, much of the social information gathered about one's intentions and affective state is non-verbal, relying on cues from the body, face, and voice, as well as abstract forms such as music. Importantly, individuals vary in how they perceive emotion, dependent upon biological sex, age, genotype, personality, or mood. Additionally, signs of extreme or irregular emotion perception may be indicative of major psychiatric disorders. Therefore, understanding perception variability in a healthy population is foundational to exploring adverse abnormalities.

The thesis aims to elucidate complexities of socio-emotional processing. Given that research is concentrated in visual domains, it explores auditory and across-modality perception using fMRI with fast multiband acquisition (TR=528ms). Evidence for widespread recruitment of heterogenous structures, influenced by individual differences and task-related attentional demands is reported. Unique methodological approaches to design, acquisition, and analyses are evaluated to improve reliability and validity of findings.

The first two studies use auditory perception to demonstrate how the brain functions in both a one-to-many (multifunctional region) and many-to-one (widespread, single-function) means. In *Study 1*, an intermediary singing condition is presented alongside music and speech, to investigate the neural relationship across domains with shared attributes. Findings demonstrate an anterior-to-posterior gradient in music-preferred regions, presenting sensitivity to diverging acoustic properties of vocal-to-instrumental music. *Study 2* assesses internal reliability of music-preferred regions, and influences of musical expertise. These regions likely respond to a unique domain-preferred weighted composition of acoustic parameters.

The proceeding two studies demonstrate multifunctionality and whole-brain processing of auditory and visual threat-related perception. In *Study 3*, multi-voxel pattern analysis is used to identify neural patterns in response to fearful versus neutral faces, bodies, vocalizations, and prosody. *Study 4* uses a perceptual decision-making task, observing neural recruitment in response to threat-related (fear-to-anger) morphs of faces and vocalizations. Findings show a distinction in processing ambiguous versus explicit threat, recruiting Salience, Frontoparietal Control and

Default Mode Networks in response to increasing task-difficulty. Finally, we report influences of state-anxiety on decision-making behaviour and underlying neural correlates.

Collectively, the thesis demonstrates intricacies of socio-emotional processing in the context of auditory and across-modality perception. The work outlines important methodological queries and solutions. It explores variability in healthy subjects, which may help direct the clinical community towards identifying and understanding outliers or provide insight to industry on development of facial and vocal recognition software.

Résumé

Dans un monde de plus en plus globalisé, nous assistons à une croissance de la diversité des populations. Il est de coutume d'observer des individus de cultures et d'horizons divergents, s'exprimant de manière unique. Pour mieux nous adapter, nous devons communiquer, établir des relations et se comprendre les uns les autres. Il est clair qu’une grande partie des informations sociales concernant les intentions et l'état affectif d'une personne se communiquent de manière non verbale et reposent sur des expressions issues du corps, du visage et de la voix, ainsi que sur des formes abstraites telles que la musique. Les individus varient dans la façon dont ils perçoivent les émotions, en fonction de leur sexe biologique, de leur l'âge, de leur génotype, de leur personnalité ou de leur humeur. De plus, des signes de perception émotionnelle extrême ou irrégulière peuvent indiquer des troubles psychiatriques majeurs. Par conséquent, il est fondamental de comprendre la variabilité de la perception dans une population en bonne santé pour explorer les anomalies indésirables.

La thèse vise à élucider les complexités du traitement socio-émotionnel. Alors que la majorité des travaux se fait dans la modalité visuelle, elle explore la perception auditive et intermodale en utilisant l'IRMf avec une acquisition multibande rapide (TR = 528 ms). Dans cet ouvrage, nous rapportons des preuves d'un recrutement généralisé de structures hétérogènes, influencées par des différences individuelles et des demandes d'attention liées aux tâches. Des approches méthodologiques uniques pour la conception, l'acquisition et les analyses que sont évaluées pour améliorer la fiabilité et la validité des résultats.

Les deux premières études utilisent la perception auditive pour démontrer comment le cerveau fonctionne à la fois dans un sens « one-to-many » (région multifonctionnelle) et « many-to-one » (largement répandu, à fonction unique). Dans *l'étude 1*, une condition de chant intermédiaire est présentée aux sujets accompagnés de musique et de parole afin d’étudier la relation neuronale entre des domaines pouvant avoir les attributs sont partagés. Les résultats démontrent un gradient antérieur-postérieur dans les régions musicales préférées, présentant une sensibilité aux propriétés acoustiques divergentes de la musique vocale. *L'étude 2* évalue la fiabilité interne des régions musicales préférées et les influences de l'expertise musicale. Ces régions répondent probablement à une combinaison unique de paramètres acoustiques, privilégiée par le domaine.

Les deux études précédentes démontrent la multifonctionnalité et le traitement cérébral global de la perception auditive et visuelle liée aux menaces. Dans *l'étude 3,* l'analyse des schémas multi-voxels fut utilisée pour identifier les réseaux neuronaux impliqués dans la perception d'expressions facials, corporelles, vocales et prosodiques effrayantes et neutres. *L'étude 4* utilise une tâche de prise de décision perceptive pour mettre en évidence l'activité cérébrale en réponse à des morphes de visages et de vocalisations liées à la menace (de la peur à la colère). Les résultats montrent une différence dans le traitement des menaces ambiguës par rapport aux menaces explicites, en recrutant des réseaux de saillance, de contrôle fronto-pariétal et de mode par défaut en réponse à la difficulté croissante des tâches. Enfin, nous rapportons les influences de l'état d'anxiété sur le comportement de prise de décision et les corrélats neuronaux sous-jacents.

Dans son ensemble, la thèse démontre les subtilités du traitement socio-émotionnel dans le contexte de la perception auditive et intermodale. L'ouvrage présente d'importantes questions et solutions méthodologiques. Celui-ci explore la variabilité chez les sujets sains, ce qui peut aider la communauté clinique à identifier et à comprendre des valeurs aberrantes ou à fournir des informations à l'industrie sur le développement de logiciels de reconnaissance faciale et vocale.

Acknowledgements

The completion of this dissertation has been a feat, requiring a great deal of patience and perseverance of which I extend my undying gratitude to a select number of individuals who have been imperative to its success. First, and foremost, my supervisor and all-around mentor, Dr. Jorge Armony. Through countless lessons, corrections, discussions, and debates, you have challenged me to think harder, be resourceful and ultimately, bring passion to every aspect of our work. I am forever grateful for your guidance; you have made a profound impact on my career, and on my life. I am grateful to my co-supervisor, Dr. Pierre Jolicoeur, and committee members, Drs. Alain Dagher, Boris Bernhardt, Alexandre Lehmann, and Marc Schoenwiesner, whose invaluable insight and advice evoked critical thought, and held me accountable to high standards of scientific rigor. I would like to extend my sincere thanks to the MRI technologists at the MNI, Ron Lopez, David Costa, and Louise Marcotte, as well as to Mike Ferreira and Ilana Leppert. You welcomed our "push-it-to-the-limit" philosophy with patience, support, and enthusiasm. You brought kindness and good humour in times of stress and uncertainty, and I cannot thank you enough for your generosity.

To my collaborators, lab mates and above-all, great friends, Dr. Ignacio Spiousas, Hanjian Xu, Fumika Kondo, Kyle Logie-Hagan, Dr. Simon Rigoulot, Lucas Peek, and Dr. Peer Herholz. It has been a privilege to work alongside each of you. Thank you for the stimulating discussions both on and off topic. On that note, a special thank you to friends and colleagues at BRAMS, who together, fostered an environment rich in knowledge, creativity, and encouragement. I would especially like to thank Drs. Ramona Kaiser, Sébastien Paquette, Carola Türk, and Duha Ahmed who guided me through the Dos and Don'ts of graduate school and offered great friendship along the way. To my peers and friends of the IPN program, with whom I shared many hours learning, studying, celebrating, and commiserating, your presence made all the difference. Additionally, I would like to give a special thank you to the administrators of IPN, GPS, as well as Drs. Josephine Nalbantoglu and Joseph Rochford. I am greatly indebted to you for providing opportunities for translational growth and advancement in my career. In that way, I must thank those with whom I have been fortunate to work for, and learn from during my PhD; Drs. Jennifer Wild, Catherine Paquet, Laurette Dubé and David Syncox. You pushed me out of my comfort zone and encouraged cultivation of new skills. To Drs. Richard Brown, Susan Howlett, Jennifer Ryan and associate lab members – I thank you for giving me ownership and autonomy on my projects over the years and having great faith in me. I would also like to thank a string of fierce women in neuroscience, who have been great mentors to me, with or without their knowledge, Drs. Megha Sharda, Yvonne Yau, Jessica Shrouff, Dominique Vuvan and Claire Gizowski.

I am extremely grateful to my close friends, near and far, who have spent countless hours coaching and supporting me through the twists and turns of this PhD. I would like to give particular mention to Laurence Desbiens, Emily Maclean, Mel Desroshers, Eleanor McGroarty, Vicky Woo, Allie Smith, Piero Sovrani, Austin Johnson, and Lena Slanisky, who were there with me through the thick of it. I would also like to thank my greatest cheerleaders; my extended family, Hoolis. You are unapologetically determined and offer a constant reminder of why I push myself every day. A special thank you to my aunts, uncles, cousins, and grandparents who make it their business to be with me every step of the way. I would like to thank my sisters, Jillian and Emily, who always have an opinion and I am ever so grateful for it. Your words of wisdom and love have kept me grounded. To my partner and my best friend, Martin Nguyen, thank you for keeping my brain busy, my heart content and my stomach full. You have helped me find joy, in even the toughest of times. I could not have done it without you. Last, but certainly not least, I would like to thank my parents Gail and Stephen Whitehead. I have thought about this moment many times over the years, and always seem to fall short in finding the right words to express my gratitude. You are my greatest teachers, and I am grateful for every day that I can learn from you.

Contributions of original knowledge

1. We have demonstrated strong support for the existence of regions in the right planum polare and planum temporale that respond preferentially to musical stimuli, regardless of whether produced by instruments or the human voice.
2. Through assessments of intrasubject reliability, we have demonstrated regions in the right posterior superior temporal gyrus and supramarginal gyrus sensitive to musical expertise (years of musical training), with regions in the left superior temporal gyrus/sulcus, middle temporal gyrus, and planum temporale to language expertise (number of languages spoken).
3. We have shown that the neural response to expressions of fear and neutral emotion, across-modality (faces, bodies, vocalizations and prosody) can be distinguished by differences of widespread cortical and subcortical supramodal recruitment, including, but not limited to the amygdala, hippocampus, thalamus, putamen, insula, ventromedial orbitofrontal gyrus, precuneus, and posterior cingulate gyrus.
4. We have shown recruitment for Task-positive (Salience and Frontoparietal) and Default Mode Networks in response to increasing task difficulty and the resolution of ambiguity, during a task of emotion-based perceptual decision-making.
5. We have demonstrated that the amygdala not only responds to explicit threat-related stimuli (fear and anger), across-modality (facial expressions and vocalizations), but also to the categorical identification of ambiguous threat-related (morphed fear-anger) expressions.
6. We have shown that increasing subject-specific state anxiety scores correlate with shorter reaction times, greater recruitment of the amygdala, and regions of the Task-positive Network in response to ambiguous threat-related expressions.
7. We have shown valuable methodological insight into the use of full-spectrum stimulus sets, internal consistency assessments, complementary univariate and multivariate analyses, and the use of continuous versus sparse acquisition sequences, particularly in the case of complex, auditory socially-relevant stimuli.

Contribution of authors

This dissertation includes the proceeding 4 manuscripts (Chapters 2-5), of which I (JCW) am the first author:

- **Chapter 2:** 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, 4913-4924.
- **Chapter 3:** Whitehead, J.C.*, & Armony. J.L. (under review). Intra-individual reliability of voice- and music-elicited responses and their modulation by expertise. *Neuroscience.*
- **Chapter 4:** Whitehead, J.C.*, & Armony, J.L. (2019). Multivariate fMRI pattern analysis of fear perception across modalities. *European Journal of Neuroscience,* 49, 1552-1563.
- **Chapter 5:** Whitehead, J.C.*, Spiousas, I., & Armony, J.L. (pre-submission). Neural correlates of socio-emotional ambiguity and threat-related processing across modality.

In each of the 4 proceeding manuscripts, the experimental design was developed collaboratively between JCW and JLA, while IS was also involved in Chapter 5. JCW recruited subjects and collected questionnaire and neuroimaging (fMRI) data in Chapters 2 through 5, with assistance from IS in Chapter 5. Scripts for behavioural data collection were written by IS in Chapter 5, with assistance from JCW and JLA, while data was acquired by JCW and IS. Neuroimaging data was analyzed by JCW in Chapters 2 to 5, while behavioural data was analyzed by JCW and IS in Chapter 5, all with guidance from JLA. The first draft of each manuscript in Chapters 2-5 was written by JCW, while subsequent drafts were reviewed and edited by JLA and JCW. The final draft of each manuscript was approved by JLA. All authors approved inclusion of Chapters 2 to 5 in the current thesis.

List of abbreviations

2AFC	Two-alternative forced-choice
AAL	Automated anatomical labeling
AC	Auditory cortex
ACC	Anterior cingulate cortex
AIns/AI	Anterior insula
ANOVA	Analysis of variance
AON	Action observation network
AUC	Area under the curve
BLA	Basolateral nucleus
BOLD	Blood-oxygen-level-dependent
CM	Centromedial nucleus
CO	Central operculum
DAN	Dorsal attention network
DMN	Default mode network
dLPFC	Dorsolateral prefrontal cortex
EBA	Extrastriate body area
ECOC	Error-correcting output codes
ER	Event-related
ERP	Event-related potential
F0	Fundamental frequency
FBA	Fusiform body area
FFA	Fusiform face area
FG	Fusiform gyrus
FIR	Finite impulse response
fMRI	Functional magnetic resonance imaging
FOV	Field-of-view
FPCN	Frontoparietal control network
FWE	Family-wise error
FWHM	Full-width-half-maximum
HRF	Hemodynamic response function
ICA	Independent component analysis
ICC	Intraclass correlation coefficient
IFG	Inferior frontal gyrus
IPL	Inferior parietal lobe
IPS	Intraparietal Sulcus
ISI	Inter-stimulus interval
ISSS	Interleaved silent steady state
ITG	Inferior temporal gyrus
LOSO	Leave-one-subject-out
MFG	Middle frontal gyrus
MNI	Montreal Neurological Insitute
MTG	Middle temporal gyrus
mPFC	Medial prefrontal cortex
MPRAGE	Magnetization-prepared rapid acquisition gradient echo
MVPA	Multivariate pattern analysis
OFC	Orbitofrontal cortex
PCA	Principal component analysis
PCC	Posterior cingulate cortex
PFC	Prefrontal cortex
Pins	Posterior insula
PMC	Premotor cortex
pMTG	Posterior middle temporal gyrus
PO	Parietal operculum
PP	Planum polare
PPI	Psychophysiological interaction
PreCG	Pre-central gyrus

PSE	point of subjective equality
pSTG	Posterior superior tempral gyrus
pSTS	Posterior superior temporal sulcus
PT	Planum temporale
ROC	Receiver operating characteristic
ROI	Region of interest
RS	Resting-state
RSA	Representational similarity analysis
RT	Response time
S1	Primary somatosensory cortex
SEM	Structural equation modelling
SF	Superficial nucleus
SFG	Superior frontal gyrus
SMA	Supplementary motor area
SMG	Supramarginal gyrus
SN	Salience network
SPL	Superior parietal lobe
STG	Superior temporal gyrus
STL	Short-term loudness
STPI	State-trait personality inventory
STS	Superior temporal sulcus
SVM	Support vector machine
TE	Echo time
TP	Temporal pole
TPJ	Temporoparietal junction
TR	Repetition time
TTG	Transverse temporal gyrus
VAN	Ventral attention network
vmPFC	Ventromedial prefrontal cortex
VN	Visual network
VTA	Vocal temporal area

List of figures

Chapter 2: Singing in the brain: Neural representation of music and voice as revealed by fMRI

Chapter 3: Intra-individual reliability of voice- and music-elicited responses and their modulation by expertise

Chapter 4: Multivariate fMRI pattern of analysis of fear perception across modalities

Chapter 5: Neural correlates of socio-emotional ambiguity and threat processing across modalities

List of tables

Chapter 2: Singing in the brain: Neural representation of music and voice as revealed by fMRI

Chapter 3: Intra-individual reliability of voice- and music-elicited responses and their modulation by expertise

Chapter 4: Multivariate fMRI pattern of analysis of fear perception across modalities

Chapter 5: Neural correlates of socio-emotional ambiguity and threat processing across modalities

Chapter 1. General introduction

1.1 Socio-emotional processing in the 21st century

Emotional processing, particularly threat, was acknowledged as early as the 19th century, as Darwin demonstrated its contribution to human survival through one's innate ability to avoid harm by detecting danger signals within the environment (Darwin, 1872/1998). In current day, emotions provide valuable information about the affective state of others and their intentions (Britton et al., 2006) through body language (see review, de Gelder et al., 2015), facial expressions, and vocal intonations (see review, Schirmer & Adolphs, 2017), as well as more abstract mediums such as the fine arts (see review, Melcher & Bacci, 2013) or music (Koelsch et al., 2006; Brattico et al., 2011). As one might anticipate, the perception of emotion is closely linked to emotions felt and experienced (Zadra & Clore, 2011), albeit neuroimaging literature has illustrated differential regions of peak neural activity in response to the two (see for review, Wager et al., 2008). Using functional magnetic resonance imaging (fMRI), Tabei (2005) instructed subjects to *perceive* an emotion in a piece of music, in doing so, a blood-oxygen-level-dependent (BOLD) signal increase was observed in the inferior frontal gyrus. When then told to *feel* the emotion, the signal increased in the precuneus. Nevertheless, the two are not mutually exclusive, and an emotion perceived may also be experienced by the observer. Markedly, this socio-emotional contagion can result in large groups adopting uniform emotion, strengthened by the emotional embodiment of neighbouring group members (Dezecache et al., 2013). Notwithstanding, individuals are particularly variable as to how they perceive an emotion, depending on one's biological sex, age, genotype, personality, or affective disposition (Hamann, & Canli, 2004; Sergerie, Chochol & Armony, 2007). Emotion perception is also often manipulated by one's cultural background (see for review, Engelmann & Pogosyan, 2013), creating challenges for social communication in an ever-growing globalized community. Additionally, exhibiting extreme or irregular perceptions of emotion is often indicative of major psychiatric disorders (Phillips et al., 2003). Illnesses exhibiting irregular emotion perception may include psychopathy (Hastings et al., 2008; Decety et al., 2014), bipolar disorders (see review, Kohler et al., 2011), schizophrenia (see meta-analysis, Kohler et al., 2010), anxiety disorders (see review, Cisler and Koster, 2010), post-traumatic stress disorder (Williams et al., 2018; Nazarov, 2015; Passardi et al., 2018; Castro-Vale et al., 2020) and major depressive disorder (Demenscu et al., 2011; Kohler et al., 2011). It has been suggested that most every mental health disorder hinges on emotion processing as to both diagnose and treat (Kring, 2008). Through methodically deconstructing socio-emotional processes of the healthy mind, we aim to improve the foundation for investigating perceptual deviancies. Thus, the current dissertation is intended to add clarity to neuroscientific findings under debate in Socio-affective Neuroscience and demonstrate the complexity of socio-emotional perception by **(1)** providing evidence for the functional heterogeneity of neural regions and distributive processing of socio-emotional information through functional MRI. Simultaneously, I explore how individual differences may modulate these factors through **(2)** understanding the breadth of behavioural and neural variability in healthy populations—a reference point for clinical populations. Finally, I intend to **(3)** demonstrate the importance of advanced, tailored methodologies in achieving the above-mentioned objectives, supported by outcomes of strong validity and reliability.

Objective 1: Evidence for functional heterogeneity and distributive processing

To achieve the first objective and demonstrate the multifunctionality of structural regions in the brain (Price & Friston, 2005), I focus on two topical issues of discussion in socio-emotion perception. First, adding definition to the debated relationship between music and voice (Peretz et al., 2015); two major forms of auditory socio-emotional communication. Although the two act comparatively at a behavioural and neurological level, they appear to functionally diverge at an unspecified point. Second, we explore the extent in which socioemotional information converges across sensory modalities, to form a distribution of supramodal structures responsive to threat-related expressions. With this we investigate how these regions may modulate as a factor of task-difficulty during perceptual decision-making.

Compared to the visual domain, auditory socio-emotional perception has received less attention within the neuroscientific community, where only within the last decade or two, has there been a particular surge of interest in exploring music's role within society. Due to the overlapping function of music and voice, investigation has ensued as to whether the neural pathways underlying such processes are common or distinct (Norman-Haignere et al., 2015; Armony et al., 2015; Paquette et al., 2018; Angulo-Perkins and Concha, 2019). In *Study 1*, we present an intermediary singing condition to subjects, alongside music and speech excerpts, allowing for investigation into the neural relationship across music and voice, through a condition that shares attributes with both domains. *Study 2* builds on these findings and implements a series of measures to assess the internal reliability of regions exhibiting "music-preferred" activity. Collectively, the work illustrates that a simple classification of neural processing across domains as either ‘distinct’ or ‘shared’ is not a sufficient deduction, and that rather, each region likely responds to a particular preferred weighted composition of acoustic parameters.

As *Studies 1 and 2* demonstrate a specialized tuning of neurons, sensitive to a distinct configuration of acoustic parameters, *Studies 3 and 4*, conversely, aim to rationalize the multifunctionality of a single neural region. I call into question how a region such as the amygdala may be capable of processing a gamut of information (e.g., affect perception across modality; explicit or ambiguous emotion) alongside other distributed supramodal structures. For years, Affective Neuroscientists have contested the theoretical system by which the brain processes emotional information. A locationist model assumes that each discrete basic emotion (Ekman, 1992; Panksepp, 1982; 1998) is consistently and independently processed within a specified structure. This theory is actively challenged by emerging constructionist views (i.e., the *conceptual act theory* or *theory of constructed emotion*) where emotions are purported to be a composite of many dimensions that collectively form a concept (Barrett, 2006, 2017a, Barrett, 2017b). I draw this connection to acoustic parameters, where in a particular configuration, form a song or sentence (Bigand et al., 2011). For emotion, these ingredients are then processed ubiquitously across neural networks in the brain (Lindquist et al., 2012). The *neural context hypothesis* suggests that when and how a single structure processes information is dependent, in part, on when and how connecting regions respond and act (McIntosh, 2004). Thus, *Study 3* implements a multivariate pattern analysis (MVPA; Haxby et al., 2001) to investigate emotion processing across modality (see review, Kragel and LaBar, 2014). If an exclusively perceptual task of perceiving auditory and visual emotion recruits a collection of dispersed supramodal structures, one could infer that, at some point, the signal is deconstructed into modality-independent constituents to be transmitted throughout large-scale processing networks. *Study 4* then forces an emotion-cognition interaction to expose the complexity of multilevel processing across a series of high-order, modality-general neural regions (Okon-Singer et al., 2015). This study implements a perceptual decision-making task exploring subject-specific perception of ambiguous threat-related expressions, and the balance of emotion, attention and cognitive control systems that are visibly influenced by individual differences.

Objective 2: Identifying individual differences that shape perception and modulate underlying neural correlates

Genetics, experience, and personality can each shape how socio-emotional information is processed in our environment (Hamann & Canli, 2004). Through studying intra- and inter-individual variability we can better grasp how specific traits and behaviours have evolved and adapted to benefit the social nature of humans (Boogert et al., 2018). Investigating individuals of a healthy population, allows for identifying how subtle differences may manifest as an enhancement or impairment to one's neural functioning, with respect to social cognition and affect. The current thesis provides evidence for both relatively stable individual differences (i.e., music and language expertise), as well as one that is, by definition, transitory (i.e., state anxiety). Moreover, when interpreting neural recruitment, we consider that expertise reflects enhanced performance, while anxiety impedes healthy functioning, including attention/control, memory, sensory-perceptual processes and notably, executive functions such as decision-making (Robinson et al., 2013).

When exploring the unique attributes of experts, we focus on quantifying test-retest reliability, to ensure identification of consistent and genuine differences. The necessity of reliability (Elliot et al., 2020) and reproducibility (Poldrack et al., 2017) in neuroimaging has been emphasized in recent years, particularly as fMRI has seen a shift from identifying how the average brain processes information, to detecting how individual brains vary (Elliot et al., 2020). Such efforts within a healthy population may, in due course, translate to clinical applications, directing attention towards potential neural biomarkers that could help define disorders of the brain (Woo et al., 2017).

In contrast, when examining functional differences of state anxiety, we direct efforts to assessing nomological validity through *a priori* assumptions based on associative evidence of reaction times (Kalanthroff et al., 2016; Choi et al., 2012) and activity of the amygdala (Bishop et al., 2004; 2007; Somerville et al., 2004) particularly in response to threat. Moreover, we look to tasks that have historically demonstrated influences of state anxiety, such as those involving threat perception (Bishop et al., 2004; Choi et al., 2012), decision-making (Soshi et al., 2019; Panno et al., 2018) and/or variable task-difficulty (O'Neil et al., 1969; Bishop et al., 2007).

Objective 3: Adapting methodological advancements to optimize research inquiry

It is apparent that our understanding of the brain is only as good as the tools we have. Thus, throughout the entirety of this thesis, we examine the efficiency of our tools and those used by the scientific community, to ensure validity and reliability of our findings. Our primary approach to stimulus selection was using a subset of intermediary stimuli to gauge the neural response of a full stimulus spectrum. We use singing stimuli to study the relationship of music and voice, and morphed threat-related expressions along a fear-anger continuum to explore the perception of ambiguous threat-related expressions. These novel approaches were complemented using fast multiband acquisition (TR=528ms), alongside critical evaluations of construct validity (e.g., comparing continuous versus sparse sampling) and test-retest reliability—a necessary foundation to explore individual differences. Moreover, we aim to maintain continuity and ensure that the analysis approach agrees with design and acquisition protocols. For instance, when considering group effects of emotion perception, a multivariate approach can offer sensitivity to voxel-level variability and insensitivity to subject-level variability, deeming it more suitable for data with potentially noisy between-subject effects (Davis et al., 2014). Each study builds upon the last to best optimize study design, execution, and analysis, and to provide insight for the development of future experiments.

1.2 Modality-specific socio-emotional processing

In its simplest form, socio-emotional processing occurs in several dynamic and interdependent parallel stages; detecting various relevant sensory cue within ones' environment, categorizing cues, and collectively attributing them to a cause based on past experiences and knowledge (see for review, Freeman et al., 2012). Sensory cues that convey socio-emotional information in the environment may present in auditory, visual, tactile, or even chemical modes (i.e., gustation and olfaction), although the latter may involve a more challenging interpretation. Nevertheless, multisensory integration with chemical cues is common (e.g., pairing olfaction and visual emotive signals; Novak et al., 2015).

Much of our understanding has come from research conducted with emotional facial expressions, then with body expressions and a relationship of the two. For example, work by Poyo-Solanas et al. (2018) explored how congruent versus incongruent body postures can influence the perception of facial emotions, identifying neural correlates of emotional ambiguity in the visual system. Notably, contextual information of the body can also bias interpretation of an ambiguous facial expression (i.e., scream of joy versus fear; Aviezer, Trope, Todorov, 2012); an effect already observed at 8-months of age (Rajhans et al., 2016). Cross-domain investigations can provide critical understanding as to the interdependence of processing multi-source socio-emotional information.

Research in the auditory modality has moved comparatively slower; however, in recent years particular interest has grown in understanding music perception, its relationship to voice and its proposed role in emotion regulation (see for review, Hou et al., 2017). Evolution has shaped (primarily) two distinct sensory routes of social communication to assume opposing, yet complementary profiles. For example, the auditory signal can be heard at a distance, while visual information must be viewed in proximity to achieve a comparable acuity. In line with previous face-body comparisons, exploring across modality perception can provide insight into recruitment of higher-order supramodal stuctures. Given the evident distinctions across modalities, comparing the two is often challenging and requires a strategic approach to best match stimuli and protocols, as well as to identify an analysis that can equate trends across modality-specific behaviour and neural processing.

1.2.1 Visual expressions

Visual socio-emotional information is processed inherently fast, specifically, within the prefrontal cortex (120-160ms; Kawasaki et al., 2001) and amygdala (74 ms; Méndez-Bértolo et al., 2016). Partly due to the ecological saliency of a close-range threat and the ensuing need for a timely response; evolution has conserved this rapid neurobehavioral response. In its simplest terms, the brain responds to visual emotional information through two routes. The first, a fast non-conscious subcortical route that bypasses the cortex, travelling from the retina to the superior colliculus and pulvinar nucleus of the thalamus, then arriving at the amygdala for automatic processing (Öhman et al., 2007). Differentially, the conscious, slow route first moves through sensory and higher-order cortical processing regions, prior to reaching the amygdala (LeDoux, 1996). Nonetheless, as with most theories, the two-route hypothesis has its critics. Pessoa and colleagues challenge this position, proposing a *multiple waves* model that, unlike the former, integrates the existence of feedback pathways and other complexities, not previously accounted for (Pessoa & Adolphs, 2010; Pessoa, 2015). I alike, argue against reducing functioning of the brain to discrete and constrained entities, but rather suggesting that these neural processes are likely inclusive, dense, and distributed.

1.2.1.1 Facial expressions

Facial expressions of emotion have shown to elicit activity within a specialized region of the fusiform gyrus— the fusiform face area (FFA; Kanwisher et al., 1997). Curiously, this region is also reported for processing visual expertise (Bilalić, 2016), where individuals specializing in a particular class of stimuli, e.g., birds or cars (Gauthier et al., 2000), present greater response of the FFA. This may translate to a specialization of humans to detect and categorize other human facial expressions, a critical component of communication and cohabitation in a social society.

Certain universal features of facial expressions can be attributed to movements of facial musculature, of which developed from non-human primates (De Waal, 2003). For instance, the smile illustrates happiness, wide eyes and an open mouth reflect a fearful scream, and a furrowed brow with pursed lips is often associated to anger (Darwin, 1872/1998). Most emotions require attention to be directed to specific facial features (e.g., fear requires greater attention to the eyes; Gamer & Büchel, 2009). Nevertheless, the greatest FFA recruitment occurs in response to a whole face presentation—in correct configuration (Zhang et al., 2012). At a more complex level, specific regions of the FFA may respond to the whole face, while others are more sensitive to certain features (Harris & Aguirre, 2010).

1.2.1.2 Body language

Body expressions differ substantially, as they can be perceived from a sizeable distance and are often associated with biological motion (Atkinson et al., 2004). The extrastriate body area (EBA), located in the lateral occipitotemporal cortex, along with the fusiform body area (FBA), partially overlapping with the FFA, are the two principally reported body-sensitive regions (Downing et al., 2001; Peelen & Downing 2005a; 2005b; Schwarzlose et al., 2005). Albeit, unlike faces, bodies provide added information about actor's actions, requiring processing from greater connecting networks (Bachmann et al., 2018). Such includes the action observation network (AON); covering the inferior frontal gyrus (IFG), inferior and superior parietal lobes (IPL; SPL), premotor cortex (PMC), supplementary motor area (SMA), posterior middle temporal gyrus (pMTG), and the primary somatosensory cortex (S1; see metanalysis, Caspers et al., 2010). Additionally, for comprehending the actions of others— the mentalizing network (i.e., theory of mind), recruiting regions of the medial prefrontal cortex (mPFC), posterior superior temporal sulcus (pSTS) and temporal poles (TP; Frith & Frith, 2006). These networks have shown to modulate as a factor of emotion (de Gelder et al., 2015), and conclusively, differentiate perception of emotional faces versus bodies. For example, in perceiving threat-related (fear, anger) versus neutral expressions, viewing dynamic faces versus bodies illicited greater recruitment of the amygdala, while bodies versus faces demonstrated a stronger response of the EBA, cuneus, fusiform gyrus (FG), tempoparietal junction (TPJ), SPL, S1 and thalamus. Although stimuli were processed differentially, both bodies and faces were identified as expressing the correct emotion (previously validated with >80% accuracy rate; Kret et al., 2011).

1.2.2 Auditory expressions

Auditory socio-emotional information, rather, can be detected at long range and is pervious to environmental barriers. Although the field of auditory affective neuroscience is in its infancy when likened to visual perception research, it has still made considerable stride in addressing particularly complex problems of evolution and human cognition. Alike faces and body expressions, the brain processes acoustic socio-emotional information from different channels— most often music or voice. Determining the evolutionary origin of music as compared to voice, is of notable interest and weighs heavily on establishing the degree of neural overlap versus sharing observed in the brain (Peretz et al., 2015).

Auditory emotion signals are received via the cochlear receptors. They move through the brainstem, converge at the inferior colliculus, and arrive at the medial geniculate body within the auditory thalamus (LeDoux, 1996). From here the information travels to the primary auditory cortex (AC), where the signal is decoded into smaller discrete properties. In the secondary AC, these components are grouped from different inputs and integrated to form a neural representation of the emotional signal. The signal then moves to higher-order regions, such as the IFG and/or the orbitofrontal cortex (OFC; see for review, Schirmer & Kotz, 2006).

Located on the superior temporal gyrus (STG), the AC is generally concentric, comprised of the core (primary AC), containing the transverse temporal gyrus (TTG; i.e., Heschl’s gyrus; HG). The belt (secondary AC) is located posterolateral to the core and contains the planum polare (PP) and planum temporale (PT; Sweet et al., 2005), while the parabelt (tertiary AC) is anterolateral to the core, along the lateral parts of the middle and posterior STS and STG (Kaas and Hackett, 2000). Notably, the non-primary AC is the reported location for distinguishing music and the voice (Armony et al., 2015; Leaver & Rauschecker, 2010; Norman-Haignere et al., 2015).

1.2.2.1 Vocalizations

The voice conveys socio-emotional information through modulating a series of physical acoustic attributes (see for review, Belin et al., 2004). Like facial expressions, the voice can be manipulated by underlying musculature. For example, to express high arousal emotions such as anxious or happy, the glottis (i.e., the voice box) closes quickly due to laryngeal muscle tension or elevated vocal efforts (Johnstone & Scherer, 2000). Ensuing physiological adjustments are triggered by signals from the somatic and autonomic nervous systems that function alike our evolutionary predecessors (see for review, Scherer, 1986). These innate responses produce universal vocal sounds that can convey an emotional state, whether in the form of non-linguistic vocalizations (e.g., laughs, screams, yawns) or speech prosody. In humans this occurs through adjustments of vocal quality, resonance, intensity, and frequency by controlling these laryngeal muscles (see for review, Zhang, 2016). Comparable to the FFA, the dorsal region of the central STS has presented a specialization in responding to these vocal sounds (vocalizations and/or speech; Belin et al., 2000). Moreover, auditory emotional information reportedly takes both a fast and slow route to the amygdala, alike the visual system. Vocalizations seemingly use the fast route, bypassing non-primary auditory cortices that instead process the slower spectral information of prosody (see for review, Liebenthal et al., 2016). Not only are emotions recognized more accurately via vocalizations (Hawk et al., 2009), but it has been suggested that greater sensitivity of the amygdala to vocalizations (Fecteau et al., 2007; Sander and Scheich, 2001; Sander, Brechmann, et al., 2003) versus prosody (Wildgruber et al., 2005; Wiethoff et al., 2008), may reflect a heightened emotional salience partnered to the innate or primitive attribute of the expressions (see for review, Liebenthal et al., 2016). In agreement with this theory, individuals with high levels of non-clinical anxiety have demonstrated an earlier and greater response to emotional vocalizations as compared to prosody (Pell et al., 2015). Notably, in the foundational fear-conditioning literature, anxious behaviour across species is paired with heightened activity of the amygdala and an overall greater response to threat (see for review, Davis, 1992; Adolphs, 2013).

1.2.2.2 Prosody

Beyond semantics, emotion in speech is detected through temporal attributes, such as the pace of speech or the duration of a pause (reviewed in Banse & Scherer, 1996). Additional adjustments of pitch (fundamental frequency (F0), F0 contour, Jitter), intensity (intensity, attack) and quality of the voice (articulation precision, glottal waveform; high-frequency energy, formant frequencies) collectively create relatively unique emotional profiles. For instance, perceived fear in speech can be identified through attributes such as fast speech rate, low intensity, high F0/pitch or a rising pitch contour (Juslin & Laukka, 2003). Markedly, the modulation of these attributes is common to emotional and linguistic prosody, albeit the two forms function and are processed differentially in the brain. Linguistic prosody provides information about pragmatics and syntax, as expressed through stressing polysyllabic words (i.e., elevated intensity or pitch; Gay, 1978) or sentences (Ladd & Morton, 1997). It also defines whether a sentence is interrogative, imperative, declarative, or exclamatory. Belyk & Brown (2014) conducted a meta-analysis of neuroimaging studies, reporting overlap of activity in response to emotional versus linguistic prosody, primarily in the right posterior STG. Regions of the inferior frontal gyrus, including Broca's area, were more responsive to linguistic prosody, while the orbital IFG (i.e., OFC) responded more to affect (Belyk & Brown, 2014; Schirmer & Kotz, 2006). Unlike vocalizations, prosody offers multiple functions. Notably, this level of high complexity/high functionality can be mirrored in the music domain.

1.2.2.3 Music

Arguably, as a more unique channel of socio-emotional communication, music presents a supplementary (and prominent) role of emotion regulation (see for review, Hou et al., 2017). Emotions, as conveyed through music, are attributed to the acoustic components that can be controlled and manipulated by the musician. This includes tempo and/or loudness, as opposed harmony, tonality, or melody (Patel and Peretz, 1997). Although music's evolutionary origins remain under heavy scrutiny (as it relates to language perception), the ability of an instrument to mimic the human voice is undeniable (Juslin & Laukka, 2003). Nonetheless, current evidence suggests that regions of the auditory belt (i.e., planum polare and planum temporale) likely respond preferentially to music as opposed to voice (Angulo-Perkins et al., 2014; 2019; Norman- Haignere et al., 2015; Leaver & Rauschecker, 2010; Aubé et al., 2015**;** Rogalsky et al., 2011; Armony et al., 2015). For example, recruitment of the planum tempolare has been reported during sound sequence processing (Mustovic et al., 2003; Griffiths et al., 1998; 1999; Penhune et al., 1998), comparable to that of the supplementary motor area (SMA; see for review, Cona & Semenza, 2017). The SMA is often involved in motor control and preparation (Nachev et al., 2008); however, is also known to be music-sensitive, irrespective of subjects' degree of musicianship (Angulo-Perkins et al., 2019; Chen et al., 2008). Alike visual perception, the given evidence indicates neural differentiation across auditory channels; a relatively convoluted relationship that this thesis aims to elucidate.

1.3 Supramodal socio-emotional processing

Although auditory and visual channels provide unique input to emotional processing systems, the output is often a single unified percept. This could suggest that signals from different channels converge, potentially, within higher-order neural structures. Markedly, patients that present neuropsychological impairments after brain injury, have shown to exhibit deficits in emotion recognition from the face and voice alike (see for review, Young et al., 2020). Thus, the dissertation aims to identify these supramodal regions, likening modality-specific processes during passive perception, as well as at the junction of emotion and cognitive processing during perceptual decision-making.

1.3.1 The amygdala

The amygdala is often at the epicenter of emotion processing. A small, almond-shaped structure, lying deep within the medial temporal lobe, parcellated further into 13 unique nuclei. Within each nucleus lie smaller subnuclei, carefully characterized by cytoarchitecture and histochemistry. The three main functional groups include the basolateral (BLA; i.e., basal, lateral and accessory basal nuclei), superficial (SF; i.e., cortical nuclei and nucleus of lateral olfactory tract) and centromedial nuclei (CM; i.e. central and medial nuclei), while intercalated nuclei, and nuclei in the anterior amygdala, and amygdalo-hippocampal comprise the rest (see for review, Sah et al., 2003). Each amygdaloid nuclei makes distinct connections to subcortical structures, while complex circuitry transmits signals within the amygdala, aggregating critical information at the CM. Efferent connections exit the amygdala from the CM, moving through the brainstem and hypothalamus to initiate a behavioural response towards the external stimulus (Pitkänen et al., 1997). Notably, the amygdala receives input from all sensory modalities (LeDoux, 2007); however, identifying whether these neural populations are truly modality-independent, remains a question of inquiry. Recent evidence for multisensory amygdala neurons has been reported via intracranial recordings of non-human primates (Morrow et al., 2019; Domínguez-Borràs et al., 2019; Kuraoka & Nakamura; 2007); nevertheless, findings from human subjects remains limited.

Peelen et al. (2010) employed a multivariate pattern analysis to classify the neural responses made to different emotions (anger, disgust, fear, happiness, and sadness) expressed by either faces, bodies or voices. Remarkably, the model could be trained and tested using different channels to achieve above-chance emotion classification, with contributions from the mPFC and the left STS—both regions involved in emotion recognition (Kim et al., 2017; Peelen et al., 2010; Said, Moore, Engell, et al., 2010; Said, Moore, Norman, et al., 2010; Sievers et al., 2018; Lin et al., 2020) and meta-cognitive processing (mentalizing and theory-of-mind; see for review, Carrington and Bailey, 2009; Frith and Frith, 2012; Decety and Grèzes, 2006). In agreement, cross-modal adaptation effects were observed in the pSTS, where subjects adapted to repeated exposure of an emotion, irrespective of the modality (i.e., faces or voices; Watson et al., 2014). It is evident that modality-independent emotion perception occurs; however, as to whether the amygdala is imperative to the process is of relevance to this thesis.

1.3.2 Challenging modularity in the brain

Historically, the amygdala processes fear, as evident from decades of fear-conditioning literature **(**Öhman, 2005; LeDoux, 2014**)**. By the same token, the insula is reported to be the exclusive region in processing disgust (Wicker et al., 2003), the OFC—anger, the subgenual cingulate— sadness and happiness (Murphy et al., 2003). Modular theorists argue that each of six basic emotions (i.e., happiness, anger, sadness, fear, disgust, surprise; Ekman, 1992) have a unique neural bases (see for review, Celeghin et al., 2017); however, evidence for amygdala responsivity to negative and positive emotions alike (Fitzgerald et al., 2006) is only one of many disconfirming observations (see for review, Barrett, 2017a). Dimensional theorists instead argue for multidimensionality of emotions, as predominately, a linear composite of arousal and valence (Russell, 1980). Given this framework, the amygdala was since associated with modulations along the arousal dimension (see for review, Zald, 2003). However, as complementary dimensions were incorporated (e.g., lust and sociality; Kassam et al., 2013), models of greater complexity became more conceivable (e.g., Barrett, 2017a; 2017b; Posner et al., 2008; Sander et al., 2018).

Lindquist & Barrett (2012) proposed that emotions are constructs of numerous components; combining external sensations with knowledge, referred to as “situated conceptualization” (Barsalou, 2009). In agreement, Wager et al. (2015) suggested that the recruitment of 7 networks (Limbic, Salience, Default, Frontoparietal, Dorsal Attention, Somatomotor and Visual networks;

Yeo et al., 2011) work collectively during an emotional experience. Authors' brain-based Bayesian model demonstrated how specific subcortical-cortical interactions could occur in response to each independent emotion (i.e., fear, anger, disgust, sadness, and happiness). Note that although this work directs us towards a multiple-system view of emotion processing, it still defines how emotions are conceptualized, and therefore, restricts how they are investigated. This meta-analysis showed common patterns of recruitment for anger and fear (i.e., the Dorsal Attention, Frontoparietal, and Default Mode Networks), and separately, for happiness, sadness, and disgust (i.e., Ventral Attention/Salience and Somatomotor Networks). Considering these networks, authors inferred that fear/anger require a redirection of attention to the external world (often goal-driven), in accordance with the "threat-superiority effect" (Hansen & Hansen, 1988), while happiness, sadness and disgust likely orient towards internal homeostasis (Wager et al., 2015).

1.3.3 Our inherent sensitivity to threat

Threat-related socio-emotional information is the best at capturing attention and elicits the most prominent neural response (Mattavelli et al., 2014; Fusar-Poli et al., 2009; Costafreda et al., 2008; Fitzgerald et al., 2005; although see Sergerie et al., 2008). If lesioned, the most pronounced deficits in both facial (Calder, 1996) and vocal (Scott et al. 1997) emotion recognition, are to expressions of fear and anger alike. Markedly, this apparent negativity bias is dependent on subjective perception, versus significant physical attributes of the stimulus (Wang et al., 2014). For instance, in viewing facial expressions of surprise (ambiguous), greater amygdala recruitment occurred when the perception was negative versus positive. Moreover, because low-spatial features were perceived first, authors inferred that this negativity bias is a default human state (Kim et al., 2003). At the behavioural level, mouse-tracking data demonstrated that prior to selecting a positive response (to emotionally ambiguous faces), a negative option was initially approached (Neta et al., 2020). Markedly, this *initial negativity bias* has also been reported in vocal emotion recognition (Pell & Kotz, 2011). An evolutionary and psychophysiological perspective of survival would agree that a default reaction to uncertainty is likely a stress response. The reaction remains under prefrontal inhibition until an environmental change signifies unpredictability, at which it is rapidly disinhibited (see for review, Brosschot et al., 2016). Consequently, for survival it is best assumed that uncertainty of threat—is a threat. When comparing fear versus anger, fearful faces are often considered more ambiguous, as arguably, they lack information about the details or location of the approaching threat (Whalen, 1998; Whalen et al., 2001; Adams et al., 2003).

1.3.4 Multifunctionality of the amygdala

Pessoa (2008) emphasizes that neural signals are not merely "emotional" nor "cognitive" alone, but instead, likely involve an interaction of the two. As we shift towards a larger network-based model of emotion perception, we must consider that attention and neural resources are finite, and inevitably must be divided across emotion and cognitive demands—at the least. Pessoa (2015) articulates this as the "dual-competition theory", whereby competition between the two domains is particularly dependent on the salience of environmental information, as well as motivational demands. For example, in a task of competing emotion and goal-directed processing, activity of the anterior cingulate cortex (ACC) and dorsolateral prefrontal cortex (dlPFC) increased in response to greater cognitive (goal-directed) load, while activity of the amygdala decreased, negatively correlating with the ACC and dlPFC (Blair et al., 2007). Furthermore, in an emotional Stroop task, where subjects were required to resolve emotional conflict (i.e., ambiguity), authors reported that with greater conflict came greater rostral ACC recruitment, and a negative correlation with the amygdala (Etkin et al., 2006). Nonetheless, others suggest that emotion and cognition can work synergistically during task, as observed in those requiring working memory. This is conveyed through improved response accuracies and faster reaction times (Lindström & Bohlin, 2010; Jackson et al., 2009). In *Study 4*, we explore the relationship of emotion and perceptual cognition during a decision-making task, where subjects are required to discriminate emotionally ambiguous expressions. Beyond anticipating widespread recruitment of cognitive and attention-based networks, we predict a multifunctional role of the amygdala based on evidence from single-neuron recordings demonstrating differential functions of amygdala subnuclei; some responding to intensity and others to perceived emotional ambiguity (Wang et al., 2017). McIntosh (2000) suggests that the network in which a region is associated with at a given time is, in part, dependent on the functional context—in this case, the demand of the task. This is also known as the *flexible hub theory* (Cole et al., 2013), and particularly relevant given the argument for three large-scale corticolimbic networks (i.e., perception, affiliation, and aversion networks) anchored in the amygdala (Brickart et al., 2014). Thus, beyond its multidimensional and multimodal properties, the amygdala is arguably multifunctional (see for review, Gothard, 2020), recruited within social and non-social contexts, such as economic choice (Averbeck & Costa, 2017; Grabenhorst et al., 2012; Hernádi et al., 2015) and decision-making (Seymour & Dolan, 2008; Phelps et al., 2014).

1.4 Individual differences modulating perception

Measuring individual differences of emotion and cognitive processes can be challenging, both logistically and analytically. The design must allow for subjects to be tested under standardized conditions, to expose variability dependent on only specified factors (e.g., subject sex, past experiences, current mood) and moreover, for these differences to be quantifiable (Boogert et al., 2018). This dissertation employs two approaches to better understand variability across healthy subjects. Such includes exploring differences developed from years of training, as well as the transient inter-subject differences of mood, often influenced by a specific context or situation.

Isolating neural regions or systems that behave variably in a healthy population, could provide insight for the clinical community when exploring outliers and identifying pre-clinical targets.

1.4.1 Influence of past experiences: expertise and culture

Whether a genetic predisposition (Park et al., 2012; Theusch & Gitschier, 2011), years of musical practise (Gärtner et al., 2013), or a combination of the two, structural differences have been reported in musicians, not only in regions of fine motor processing, but also in areas of the AC, primarily the PT (see for review Münte et al., 2002). Moreover, musicians' enhanced performance has translated to functional differences in the brain; however, the consistency of reported distinctions has varied. *Study 2* identifies several reasons that may help explain these inconsistencies and employs a test-retest reliability assessment to address general concerns of scientific rigor (Elliot et al., 2020). Curiously, when musicians listen to sung language, EEG findings report that they learn both musical and linguistic structures better than non-musicians (Francois & Schon, 2011). This reported transfer of musical expertise to speech perception (see for review, Coffey et al., 2017) revives the ongoing debate of the intimate neural relationship between music and voice processing (Peretz et al., 2015). Moreover, with evidence for a reorganization of language processing in the brains of bilinguals and multilinguals (see for review, Del Maschio & Abutalebi, 2019), we question whether hypothesized cross-domain effects may be bidirectional. Social culture not only sculpts our brain through language, but also shapes cognitive and emotion processes (see for review, Park & Huang, 2010). For example, in a series of studies, Caucasian Americans and Japanese exhibited a comparable perception of positive affect; however, Americans were more accurate in perceiving negative emotions (e.g., anger, fear; Ekman et al., 1987; Matsumoto, 1990; Matsumoto, 1992). Effects were reportedly a reflection of Japanese society, where displays of negative affect are discouraged, marking a stark contrast against American society (Matsumoto, 1992).

1.4.2 Influence of personality, mood, and anxiety

Researchers have shown that beyond differences of sex (see meta-analysis, Filowski et al., 2017) or age (Murry & Isaacowitz, 2016; Keightley et al., 2007), specific personality traits (e.g., extraversion and neuroticism; Canli et al., 2001) or even current mood (Trilla et al., 2020) could alter ones' perception of affect and influence underlying neural correlates. For instance, activity of the MTG in response to emotional faces, has shown to positively correlate with neuroticism (Klamer et al., 2007), one of the popularized "Big Five" personality traits (McCrae and John, 1992) where individuals are susceptible to experiencing negative affect, worry and anxiety. Moreover, those with high scores of non-clinical anxiety demonstrate faster reaction times and greater basolateral amygdala activity when unconsciously (via a masked face) viewing facial expressions of fear. Notably, the dorsal amygdala, rather, responded to conscious fear perception, independent of trait anxiety (Etkin et al., 2004). Beck et al.'s (1985) cognitive model of anxiety, suggests that anxious individuals present a bias towards threat, where the initial stage of processing includes an involuntary redirection of attention to threat (i.e., orienting mode; Beck & Clark, 1997). As one could assume, these effects are likely to interact with ongoing cognitive processes. For example, during an emotional decision-making task, a hyperactive response of emotion-based systems was reported alongside an under-responsive analytic system in anxious individuals (Xu et al., 2013). Comparable effects have been reported in temporary states of anxiety (Bishop et al., 2004; Somerville et al., 2004). Pacheco-Ungetti et al., (2010) suggest that trait anxiety is associated with deficits of the executive control network, while state anxiety is instead related to overactivity of alerting and orienting systems. Investigating state anxiety is particularly pertinent for conducting fMRI studies of socio-emotion perception, as some subjects report feelings of fear in anticipation to scanning, for reasons such as claustrophobia, magnetophobia or disconcerting MRI sound exposure (Munn and Jordan 2011). Thus, *Study 4* allows for an in-depth exploration of these assumptions, focusing on state anxiety in relation to cognitive and emotion processes.

1.5 Methodology: functional Magnetic Resonance Imaging (fMRI)

Notable efforts have been put on employing the most appropriate and sound methodology in *Studies 1* through *4*, having two primary areas of focus. The first, an emphasis on using stimuli that present a full “continuous” spectrum of low-level attributes. Singing, as an intermediary stimulus between music and voice, and morph stimuli of fear-anger expressions, offer novel ways to investigate, respectively, the music-voice relationship and ambiguous emotion perception. Subsequently, to achieve validity and consistency in the findings, we account for potential confounds in design, optimize fMRI acquisition, avoid overinflation of statistics through multiple test corrections, and assess reliability and reproducibility of results. Amongst other added efforts, these steps provide assurance for the observations made and offer, at times, a unique approach to exploring socio-emotional processing across modalities.

1.5.1 Stimuli

Some of the major limitations of the current socio-emotional perception literature, reflect the stimuli used. Oftentimes, researchers use naturalistic stimuli of high ecological validity, such as long musical excerpts (Koelsch et al., 2006; Brattico et al., 2011) or clips of audiovisual films (Wagner et al., 2016; Kim et al., 2016; Pichon et al., 2015) to present a more genuine representation of the expressed emotion (see for review, Jääskeläinen et al., 2021). Nevertheless, there is a trade-off in using these stimuli, as high ecological validity often decreases experimental control and in doing so, introduces noise and reduces statistical power. This bears relevance, as low-level features, such as the brightness of an image (valence; Lakens et al., 2013) or the combined tempo, texture, sharpness, loudness, pitch level and contour of a musical piece (Coutinho & Cangelosi, 2011), may already be suffice for identifying affective states. Possible confounds can be minimized through including covariates (e.g., stimulus features, or subject behaviour over time); however, as subjects listen to, or view longer clips of emotional content with dynamic variations, they are more likely to induce an emotional response (Krumhansl, 1997; Westermann et al., 1995; Ellard et al., 2012). This felt emotion can then be challenging to dissociate from the perceived. Moreover, exposure to familiar stimuli, such as well-known musical excerpts (Pereira et al., 2011; Ali & Peynircioğğlu, 2010) or faces (Dubois et al., 1999; Schwartz et al., 2003) can also confound the observed neural responses based on previous associations stored in long-term memory. Finally, longer stimuli may attenuate the neural response through expectation or repetition suppression effects (Todorovic & Lange, 2012). A strategic approach taken in *Studies 1* through *4* was to use shorter, more diverse stimuli to maintain subjects' attention and avoid emotion induction. Brief stimuli also allow for easier association of behaviour/function to specific features, while greater stimulus diversity encourages the identification of neuronal sub-populations that may be tuned to unique dimensions (Norman-Haignere et al., 2015; Leaver and Rauschecker, 2010).

Selecting appropriate controls is critical in exploring socio-emotion perception. For example, to achieve a more reliable response of subcortical structures, such as the amygdala, it is imperative that emotion conditions are contrasted against a neutral or low arousing stimulus, as the amygdala is knowingly more responsive to both positive and negative emotions (see for review, Armony, 2013). Moreover, the control condition can be used advantageously. *Study 1* strategically contrasts vocal music (i.e., singing) versus vocal speech to minimize acoustic differences and effectively target the essential features that make voice and music unique. Using so-called "intermediary" stimuli, is helpful to observe responses across a full spectrum of attributes. *Study 4*, comparably, uses morphed emotional stimuli to determine subject-specific perceptual discrimination thresholds. This can be achieved using a two-alternative forced-choice task, where subjects are required to report their perception when presented with two unique stimuli that exist along a continuum (e.g., intensity, valence, emotion-to-emotion). Averaging responses to each stimulus identifies a point of subjective equality (PSE; Kingdom and Prins 2016). One can then observe whether the subject was sensitive to a given signal (a multiplied signal) or rather, a bias (an additive signal; Macmillan and Creelman, 2004; Green and Swets, 1966). Potential perceptual sensitivities or biases can then be related to results of personality assessments and/or demographic information to identify any plausible associations.

1.5.2 Experimental design

Block designs present challenges, particularly in perceptual studies where adaptation to stimuli is prevalent. For example, the use of block-designs in studies of auditory perception (Angulo- Perkins et al., 2014; Dick et al., 2011; Ohnishi et al., 2001; Boebinger et al., 2020; Schmithorst & Holland, 2003) raises the issues of stimulus expectancy and habituation (Liu et al., 2001) given less flexibility to randomize stimulus presentations. Thus, adaptation effects have been reported in the AC regardless of field strength (1.5 T and 3 T; Rabe et al., 2006). Nevertheless, block designs can be particularly robust, as BOLD signal changes are compared to resting baseline (see for review, Johnstone et al., 2009) demonstrating high statistical power (Bandettini & Cox, 2000). Continuous event-related (ER) designs, instead, allow for a pseudorandomization of stimuli, counterbalancing by transition orders, and jittering inter-stimulus intervals (ISI); all of which minimize subject's expectancy, prevent habituation and maintains subjects' attention (see for review, Amaro and Barker, 2006). Moreover, ER designs allow for analysis of neural responses to single trials, which as mentioned, is critical for stimulus-specific analyses (e.g., acoustic parameters) and may also help identify activity associated with behaviour (e.g., task- based responses). Conversely, a caveat of continuous ER fMRI in auditory perception is that MRI scanner noise (e.g., gradient coils, radiofrequency pulses, helium pump) may confound the acquired signal (see for review, Pelle, 2014). Together, the proceeding four chapters aim to exemplify how rigorous planning and piloting was required at each stage of experimental design to reduce noise and improve the overall signal.

1.5.3 Acquisition

An effective acquisition protocol considers the unique attributes of stimuli used, experimental design and ensuing analysis approaches. In reference to potential confounds of scanner noise during continuous-sampling, sparse-sampling may also reduce statistical power (Nebel et al., 2005) and cause high Type-I error rates (false positives; Manno et al., 2019b). A study imitating the 120dB noise of fMRI acquisition (Mansfield et al., 1998), alternated this presentation with silence/rest periods, and in doing so observed that "scanner noise" elicited activity within the bilateral AC (TTG, PP, PT, MTG and STS; Ulmer et al, 1998). To better control for scanner noise, Schwarzbauer et al. (2006) developed an interleaved silent steady state (ISSS) sparse acquisition, where auditory stimuli are presented during silent periods, and volumes are acquired when the hemodynamic response function (HRF) peaks. In a collective attempt to maintain high temporal resolution, *Study 1* compares the ISSS to continuous sampling, to confirm whether scanner noise is truly impeding on the acquired signal.

It is evident that high spatial acuity is required when exploring subtle differences in localized neural regions (e.g., distinguishing neural response to music versus voice in the AC). Or rather, when investigating deep sub-cortical structures, as during tasks of emotion perception. The development of multiband (i.e., simultaneous multi-slice) acquisition for functional MRI has allowed for notable increase in temporal resolution. The pulses, acquired in parallel, means that a multiband acceleration factor of 12, for example, can acquire 12 slices concurrently, dividing the repetition time (TR) by 12 and ultimately, reducing the time required to conduct whole-brain imaging (Moeller et al., 2008; 2010). Specifically, using an optimal TR of below 1500ms has shown to provide 12% more statistical power (see for review, Amaro and Barker, 2006), although, these high acceleration rates may come at the cost of reduced signal-to-noise ratio. As a result, it becomes particularly important to regress out movement artefacts (Chen et al., 2015). Additionally, in fast-ER designs, the ISI must be shorter than the duration of the HRF, as well as jittered. By reducing the ISI to a minimum of 4000ms, the HRF can be convolved and issues of linearity versus nonlinearity overlapping with the HRF, can be avoided (Glover, 1999). Finally, runs in general should be kept short to maintain subjects' attention, particularly during studies of passive perception. Using several runs separated in time can also allow for assessments of test-retest reliability. This is most advantageous over several days; however, may pose as a logistical problem for participant recruitment or attrition rates.

1.5.4 Analysis

Univariate and multivariate analyses each exhibit unique objectives due to their differing nature, and should thus, be applied accordingly. For instance, univariate analyses produce findings with directionality, contrasting BOLD activity between conditions, while multivariate does not have direction and instead exploits discriminability of conditions (Hebart & Baker, 2018). Jimura & Poldrack (2012) reported that although uni- and multi-variate findings may be consistent, they are likely to reflect variable sensitivity based on differential functional organization of cortical and subcortical regions. For example, authors reported that multivariate analysis provided greater sensitivity in frontal and parietal regions, while univariate was more sensitive to activity of the thalamus.

Multivariate analyses can be used in several different applications, including but not limited to dimensionality reduction, multivariate regression analyses, cluster analyses, and classification and discrimination analyses (Smith, 2018). For example, in *Study 1,* a principal component analysis (PCA) was first used to reduce dimensionality in a stimulus-based multivariate approach, proceeded by an independent component analysis (ICA); a form of blind signal separation used to identify common features (McKeown et al., 2003).

Additionally, *Studies 3 and 4*, use linear support vector machines (SVMs) to classify and discriminate conditions, which are easily interpretable, commonly used, and comparable to, if not better than, nonlinear techniques (Misaki et al., 2010). It relies on identifying the appropriate feature classification approach, the algorithm used to separate the two classes, how the data is to be divided across training and testing, and finally, how the performance of the classifier is to be measured (Kragel & LaBar, 2014). This approach can even be implemented to predict a condition, providing an accuracy rating of how well the prediction was made, or rather, how well the two (or more) conditions were discriminated (Pereira et al., 2009). Pattern classification provides information on signal detection theory measures (Swets, 1988), including area under the receiver operating characteristic (ROC) curve, as well as sensitivity and specificity; often used to classify diseased versus healthy patients in medical diagnostic test evaluations (Hajian-Tilaki, 2013). Sensitivity (specificity) identifies the likelihood in which the positive (negative) condition label is true, while the ROC reports on the relationship between sensitivity and specificity (Sokolova et al., 2006). Other comparable multivariate approaches may be used, such as a representational similarity analysis (RSA) where multiple channels of neural activity are related to one another through a similarity metric. Conclusions may then be deduced by the outcome association patterns. This approach is useful for "condition-rich" designs (Kriegeskorte et al., 2008); however, may not benefit designs with limited stimuli classes. Collectively, all aspects of design, acquisition and analysis must be considered synchronously to optimize output and best answer the questions at hand.

1.6 Conclusions

The four included chapters collectively work to meet the objectives of the current thesis, each building upon the last to ensure a thorough exploration of the proposed inquires. Through a comprehensive understanding of the current literature, we were able to identify openings for discovery that would advance our progress in achieving scientific veracity. We proposed three major areas viable for growth in the field. First, to validate the complexity of socio-emotional processing, providing evidence for regional heterogeneity and extensive whole-brain, dynamic neural processing. Secondly, we emphasized inter-subject variability, exploring behavioural and neural between-subject differences that either strengthen or weaken one's functional capacity. These findings could provide valuable insight for those developing therapeutic tools to sub-clinical and clinical populations. Finally, we emphasized several facets of methodology that we believe require greater attention in socio-emotional research through fMRI, particularly, stimulus selection, design and acquisition optimization, reliability evaluation, and lastly, analysis applicability. To satisfy these objectives we investigate two topics of a comparable nature, both of which have received less attention over the past few decades: auditory and across-modality socio-emotional perception in the brain. *Studies 1 and 2* focus on the comparison of music and voice perception, while *Studies 3 and 4* explore emotion perception across auditory and visual modalities. The general discussion bridges these themes together through satisfying each of the three outlined objectives across the two levels (i.e., across-domain; across-modality) of socio-emotional perception. Provided are two compelling examples as to the complexity and variability of perceptual processes, as well as the means of which to effectively study them.

Chapter 2. Singing in the brain: Neural representation of music and voice as revealed by fMRI

(Study 1)

Jocelyne C. Whitehead[a,b,c] & Jorge L. Armony[a,b,d]

[a]Douglas Mental Health University Institute, Verdun, Canada;
[b]BRAMS Laboratory, Centre for Research on Brain, Language and Music, Montreal, Canada;
[c]Integrated Program in Neuroscience, McGill University, Montreal, Canada;
[d]Department of Psychiatry, McGill University, Montreal, Canada

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, 4913-4924.

2.1 Abstract

The ubiquity of music across cultures as a means of emotional expression, and its proposed evolutionary relation to speech, motivated researchers to attempt a characterization of its neural representation. Several neuroimaging studies have reported that specific regions in the anterior temporal lobe respond more strongly to music than to other auditory stimuli, including spoken voice. Nonetheless, because most studies have employed instrumental music, which has important acoustic distinctions from human voice, questions still exist as to the specificity of the observed "music-preferred" areas. Here, we sought to address this issue by testing 24 healthy young adults with fast, high-resolution fMRI, to record neural responses to a large and varied set of musical stimuli, which, critically, included *a capella* singing, as well as purely instrumental excerpts. Our results confirmed that music; vocal or instrumental, preferentially engaged regions in the superior STG, particularly in the anterior planum polare, bilaterally. In contrast, human voice, either spoken or sung, activated more strongly a large area along the superior temporal sulcus. Findings were consistent between univariate and multivariate analyses, as well as with the use of a "silent" sparse acquisition sequence that minimizes any potential influence of scanner noise on the resulting activations. Activity in music-preferred regions could not be accounted for by any basic acoustic parameter tested, suggesting these areas integrate, likely in a nonlinear fashion, a combination of acoustic attributes that, together, result in the perceived musicality of the stimuli, consistent with proposed hierarchical processing of complex auditory information within the temporal lobes.

Keywords: music; speech; singing, fMRI, pulse clarity, neural selectivity, neural overlap

2.2 Introduction

The syntactic parallels that music has with speech and its comparable use for communicating emotional states have contributed to a longstanding debate over a possible common evolutionary origin (Besson & Schön, 2001). Studies highlighting their similarities, at behavioral and neural levels, have encouraged the development of several theories attempting to make sense of the close relationship that music has to speech (for a recent review, see Peretz, Vuvan, Lagrois, & Armony, 2015). For example, Brown (2000) proposed the "musilanguage" hypothesis, stating that music and language have evolved from the same origin and over time diverged, adopting their own unique domain-specific attributes. Others hypothesized an invasion of music into the language module (i.e., a "functionally specialized cognitive system"; Fodor, 1983) now acting as an adapted by-product (Pinker, 1997) that has since stabilized across cultures (Sperber & Hirschfield, 2004). In contrast, others argue that the similarities between music and speech are not unique, as they are also shared with other cognitive mechanisms (Jackendoff, 2009). Attempts at reconciling these opposing views propose that music and language processing occur across a number of discrete modules, some of which overlap, while others remain distinct (e.g., Peretz & Coltheart, 2003).

The surge in neuroimaging studies conducted over the last decade that examined the neural correlates of speech and music processing, has rekindled this debate, particularly focusing the question on whether speech and music activate distinct or overlapping regions in the brain, especially within the auditory cortex. As shown by a recent meta-analysis (Schirmer, Fox, & Grandjean, 2012), these studies have provided substantial evidence for overlapping regions of activation, in response to both music and voice, within the superior temporal gyrus (STG), superior temporal sulcus (STS), and medial temporal gyrus (MTG). However, a small but growing number of experiments, some using newly developed, more sensitive acquisition and/or analytical approaches, have reported some degree of functional separability of responses, with voice (including, but not limited, to speech) engaging mainly an area along the STS (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Belin, Zatorre, & Ahad, 2002; Belin & Zatorre, 2003; Fecteau, Armony, Joanette, & Belin, 2004; Kriegstein & Giraud, 2004; Pernet et al., 2015), and music eliciting stronger responses in a smaller cluster in the anterior STG (planum polare), often bilaterally, but more pronounced on the right hemisphere (Leaver & Rauschecker, 2010; Fedorenko, McDermott, Norman-Haignere, & Kanwisher, 2012; Angulo-Perkins et al., 2014; Aubé, Angulo-Perkins, Peretz, Concha, & Armony, 2015). Importantly, these findings were obtained with a variety of stimuli (e.g., Music: unfamiliar pop/rock music, instrumental excerpts of piano, strings, woodwind, or brass; Voice: syllables, words, connected speech, nonlinguistic vocalizations, varying languages; Controls: scrambled music or voice, songbirds, animal sounds, nonvocal human sounds, white noise, environmental, and mechanical sounds) and paradigms (e.g., block and event-related designs). Moreover, these results obtained using category-based univariate analyses were confirmed by a few others employing data-driven classification techniques based on multivariate statistics (Norman-Haignere, Kanwisher, & McDermott, 2015; Rogalsky, Rong, Saberi, & Hickok, 2011), as well as adaptation fMRI (Armony, Aubé, Angulo- Perkins, Peretz, & Concha, 2015).

Although most of the studies previously described attempted to control for the possible nonspecific effects of general acoustic characteristics of the stimuli employed (e.g., duration, intensity, and frequency), there are still important qualitative and quantitative differences between instrumental music and voice, which could, in principle, introduce confounds in the results obtained.

While it is impossible, and indeed undesirable (Leaver & Rauschecker, 2010), to remove all possible acoustic differences between music and speech (the same way it is not possible to do so for vocal vs. nonvocal sounds, or face to nonface visual stimuli), it is important to minimize them, leaving only those features that are thought to be essential to each stimulus class. In this sense, lyrical song as produced by the human voice in the absence of instruments, or a capella, may constitute an ideal candidate as an intermediary between music and speech (Schön et al., 2010). Indeed, while singing is undoubtedly a form of musical expression, its basic acoustic profile is highly similar to that of the spoken voice. In fact, a “super-expressive voice” theory of music has been put forward, suggesting that music originated simply as an exaggeration of speech, accentuating vocal speed, intensity, and timbre, as a method of enhancing communication and to ensure effective bonding (Juslin, 2001). The few studies that directly compared brain responses to speech and singing support, to some extent, this hypothesis. Schön et al. (2010) presented subjects with French tri-syllabic nouns either spoken or sung, and observed that both conditions activated, as compared with pink noise, similar clusters in the middle and superior temporal gyrus bilaterally. The comparison of Singing > Speech revealed only small clusters in those regions, leading the authors to conclude that very similar networks are engaged when listening to spoken or sung words. Callan et al. (2006) compared six well- known songs in spoken and sung form and also found very similar activation patterns for both categories. They also reported greater activity for the singing than speech condition in the right planum temporale. However, these studies did not include an instrumental music condition, so the question remains as to whether there are brain regions that respond preferentially to music, regardless of how it is expressed, either through voice or instruments.

Another potential concern when conducting fMRI studies using acoustic stimuli is the possible influence of scanner noise in the observed responses. Although a large literature exists consistently showing that auditory perception studies can be successfully conducted using standard continuous acquisition sequences, it is still generally acknowledged that the use of sparse sampling protocols, or “silent fMRI”—in which the sounds are presented during a silent period, with volume acquisitions following the silence when the hemodynamic response function is at its peak (Hall et al., 2014)—does present advantages (as well as drawbacks, particularly in terms of reduced statistical power; Nebel et al., 2005). For instance, studies that have compared the two approaches have shown the recruitment of larger networks using sparse sampling (Adank, 2012), as well as greater activation in auditory regions (Gaab, Gabrieli, & Glover, 2006), and a higher MR signal-to-noise ratio (Hall et al., 1999). Furthermore, it has been suggested that speech perception in the presence of background noise requires the recruitment of additional cognitive resources as to successfully understand what is being spoken (Manan, Yusoff, Franz, & Mukari, 2013) and that it can impair other cognitive processes, such as memory recall (Rabbit, 1968; Murphy, Craik, Li, & Schneider, 2000). Moreover, noisy speech has been shown to elicit stronger responses in several brain regions, including middle and superior temporal gyrus (Davis & Johnsrude, 2003). Because the majority of fMRI studies of music perception employed continuous acquisition, it remains unknown to what extent, if at all, scanner noise may have affected the results obtained.

The goal of the present study was thus to provide a comprehensive assessment of the brain responses to music, including both instrumental and vocal (singing) stimuli. We employed a large and diverse set of unfamiliar short stimuli and controlled, either in the stimulus selection or analysis, many of the basic acoustic parameters. Analysis was conducted using complementary uni- and multivariate approaches. We employed a multiband echo-planar imaging sequence, in which the acceleration of data acquisition allowed us to achieve both high spatial resolution and sampling rate (thus maximizing statistical power). In addition, we conducted, in the same subjects, a short experiment using a subset of the stimuli using the Interleaved Silent Steady State (ISSS) sparse imaging acquisition protocol (Schwarzbauer, Davis, Rodd, & Johnsrude, 2006), to investigate the possible confounding effects of scanner noise on the results.

We expected to replicate previous studies showing that instrumental music, when compared with speech, activates a bilateral region in the anterior STG, particularly in the planum polare (PP) (Armony et al., 2015; Leaver & Rauschecker, 2010; Angulo-Perkins et al., 2014; Fedorenko et al., 2012; Rogalsky et al., 2011), whereas speech would elicit responses along the STS (Belin et al., 2000, 2002; Belin & Zatorre, 2003; Fecteau et al., 2004; Kriegstein & Giraud, 2004; Pernet et al., 2015). Critically, we hypothesized that vocal music (i.e., singing) would represent an intermediate condition between these two. Namely, when compared with music, singing should activate STS, but when compared with speech, it should yield activations overlapping with those associated with instrumental music within the PP.

2.3 Methods

2.3.1 Participants and procedure

Twenty-four healthy volunteers (11 females, mean age = 25.5) with a range of musical expertise (years of training: *M* = 4.2, *SD* = 4.73) participated in the study. Participants had normal hearing and were right-handed. All subjects were fluent in English. Eleven of them also spoke another language, and nine spoke three languages. Overall, languages understood by the participants included English (24), Finish, French (11), German, Greek, Hindi, Italian (7), Malayalam, Mandarin (3), Persian, Spanish, and Turkish.

The experiment consisted of three 8-min runs, two using a continuous multi-band sequence and one with an interleaved silent steady state (ISSS) sequence (Schwarzbauer et al., 2006), described below. Participants passively listened to auditory stimuli while watching nature scenes. Stimuli were presented using E-Prime 2.0 (Psychology Software Tools) and delivered binaurally from MRI-compatible headphones (Model S14, Sensimetrics). A sound test was conducted prior to each testing session to confirm that the acoustic stimuli were audible in the presence of the background scanner noise for the continuous acquisitions and not too loud for the sparse sampling one. Functional images were acquired on a 3T Siemens TIM TRIO MRI scanner with a 32-channel head coil. In addition to the functional runs, a high-resolution 3D T_1-weighted image (voxel size = 1 × 1 × 1 mm^3) was acquired using a magnetization-prepared rapid acquisition gradient echo

(MPRAGE) sequence (TR = 2.3 s; TE = 3 ms, 192 slices) for anatomical co- registration and normalization.

2.3.2 Stimuli

Auditory stimuli belonged to three categories:

2.3.2.1 Instrumental music

An assortment of instrumental pieces were cut to produce 60 different musical excerpts (duration: *M* = 1.49 s; *SD* = 0.13 s). The clips consisted of strings, woodwinds, or percussion instruments (40 unique instruments), each obtained from online database sources and from Vieillard et al. (2008) and Aubé et al. (2015).

2.3.2.2 Speech

A total of 60 different phrases spoken in 45 languages (ranging from English, Spanish, and French to Baatonum, Gujarati, Mongolian, and Yiddish) and one stimulus with no words ("baby talk") produced by speakers including children (*n* = 2) and adults (33 male), were obtained from various online databases (duration: *M* = 1.51 s; *SD* = 0.22 s).

2.3.2.3 Singing

Stimuli consisted of 60 different singing excerpt (duration: *M* = 1.51 s; *SD* = 0.23 s), sung by one or several individuals of varying ages, including male (*n* = 28) and female (*n* = 32), without instrumental accompaniment ("a cappella"), sung in 19 different languages (e.g., English, German, Arabic, Ilocano, Doabi, Hebrew) or without words (*n* = 6), including song excerpts produced by amateur and professional singers, lullabies, and religious chanting (e.g., Church choir and Torah reading). About 61% of these were monophonic, 37% homophonic, and 2% polyphonic.

All stimuli were monaural, but presented binaurally. The sounds were resampled to 32 bits, at a sample rate of 44,100 Hz, and adjusted for loudness by normalizing to the short-term loudness (STL) maximum using the Moore and Glasberg Loudness model (Glasberg & Moore, 2002), as implemented in the Loudness Toolbox on MATLAB. Basic acoustic parameters for each of the categories, computed using the MIRtoolbox (Lartillot, Toiviainen, & Eeorla, 2008), MATLAB scripts (Ewender, Hoffmann, & Pfister, 2009) and the Praat Vocal Toolkit (Boersma, 2002), are summarized in Table 2.1.

2.3.3 fMRI acquisition and analysis

2.3.3.1 Continuous acquisition

Each run consisted of 90 stimuli; 30 speech, 30 singing, and 30 instrumental music excerpts, which were presented in a pseudo-random fully balanced order (equal number of first-order transitions between categories), to remove any possible carry-over effects. Each stimulus was presented only once and the stimulus subsets used in each run were counterbalanced across subjects. The auditory stimuli were presented in a continuous design and were jittered using a brief ISI (duration: M = 2.49 s, SD = 0.20 s).

Functional images were acquired using a multiband accelerated pulse sequence with a factor of 12 (Setsompop et al., 2012). Eight hundred volumes (72 slices per volume, interleaved acquisition; FOV = 208 × 208 mm^2, matrix = 104 × 104, voxel size = 2 × 2 × 2 mm^3; TR = 529 ms; TE = 35 ms) were acquired. The first 10 scans of the run were discarded due to T1 saturation. Image pre-processing was conducted using SPM12 (Wellcome Department of Imaging Neuroscience, London, UK; http://www.fil.ion.ucl.ac.uk/spm). Functional images were spatially realigned to the first volume and normalized to the MNI152 template. The images were then smoothed using a 6 mm FWHM isotropic Gaussian kernel.

2.3.3.2 Univariate analysis

Statistical analysis was performed for each subject using a univariate general linear model (GLM) in which the categories of interest (Instrumental Music, Singing, and Speech) were entered as boxcars of length equal to the stimulus duration, convolved with the canonical hemodynamic response function. Subject-specific contrast instrumental music versus speech, instrumental music versus singing, and singing versus speech, were then taken to a second level, repeated-measures ANOVA. Statistical significance was determined using a voxel threshold of p = .001, with a cluster-based familywise error rate (FWE) correction for multiple comparisons of $p < .05$ (k = 90) as implemented in AFNI's algorithm 3dClustSim (AFNI version 16.3.05). To identify regions commonly activated for different categories (e.g., Instruments and Singing vs. Speech), we performed conjunction analyses (minimum statistic compared with the conjunction null; Nichols, Brett, Andersson, Wager, & Poline, 2005).

In addition, we conducted a stimulus-based analysis. For each subject, each of the 180 stimuli was entered as a separate covariate in a standard GLM. The corresponding stimulus- specific parameter estimates were then averaged across subjects. These estimates were used for post-hoc regression analyses of the significant clusters including the acoustic parameters shown in Table 2.1, as well as for the multivariate analysis described in the following paragraph.

Table 2.1 Mean and standard deviation values of acoustic features for each sound category

AUDIO FEATURES	MUSIC	SINGING	SPEECH
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 (>1.5kHz)	.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})	1335 (1206)	97(517)	1137 (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)

a.u. = arbitrary units; bpm = beats per minute.

Values were calculated with MIRToolbox, except for those related to the Fundamental Frequency (http://www.tik.ee.ethz.ch/~spr/f0_detection) and the last four features (Praat).

[a] All significantly different.

[b] Significantly different from the other two.

*Significantly different from each other (*p* < .05, Bonferroni corrected).

2.3.3.3 Multivariate analysis

The categorical univariate analyses were complemented by a simple stimulus-based multivariate approach in which the parameter estimate images obtained in the stimulus-based analysis described in the previous paragraph were submitted to an Independent Component Analysis (ICA). We restricted the observations to auditory-responsive voxels as identified by an omnibus *F*-test in the univariate group analysis. Furthermore, Principal Component Analysis (PCA) was first applied on the data to reduce the dimensionality of the signal to the subspace spanned by the first four components, which explained 87% of the total variance. The contributions of each stimulus to each of the independent components obtained (weights) were then submitted to independent-sample *t*-tests (Bonferroni-corrected for multiple tests) to assess whether there were significant differences between categories (Speech, Instrumental Music, and Singing). Finally, the weights were submitted to a multiclass, *3* (ECOC) model (a generalization of support-vector machine classification for more than two classes; Dietterich & Bakiri, 1995), implemented in MATLAB, to determine if the model could classify individual stimuli as belonging to their a priori category with above-chance accuracy.

2.3.4 Sparse acquisition

Functional images were acquired using the ISSS sequence (Schwarzbauer et al., 2006) (FOV = 224 × 224 mm^2, matrix = 104 × 104, voxel size = 2 × 2 × 2 mm^3; TR = 2,383 ms; TE = 30 ms), with 25 slices, parallel to the Sylvian fissure, covering the entire of auditory cortex. Seven TRs formed a single epoch, in which three of the volumes were acquired during the silent dummy block (no data acquisition), followed by four volumes during the acquisition block. Auditory stimuli were presented during the silent periods in a short block of four stimuli belonging to the same category (Instrumental Music or Speech) with a mean duration of 7.15 s and their onset relative to the beginning of the dummy block was jittered (latency: 0.617 ± 0.403 s). Because of time limitations, due to the longer time required to acquired images, only two categories were presented, Instrumental Music and Speech. A total of 48 stimuli per category, taken from the stimulus pool described above, were presented in 12 blocks, their order within and between blocks, pseudo-randomized, and counterbalanced across participants. In addition, there were six blocks of silence, which served for baseline estimation.

Data preprocessing was carried out as in the continuous acquisition (see above) and analysis was performed using a finite impulse response (FIR) model, in the context of the general linear model, in which each of the four acquisition volumes for the two sound types was entered as a separate category (i.e., eight in total). Dummy volumes were created using replications of the mean EPI image, to create a continuous timescale in the design matrix. The dummy scans were not included as observations in the model, to avoid skewing the degrees of freedom (Peelle, 2014). Subject-specific estimates for the contrast for Instrumental Music minus Speech were calculated and taken to a second level, one-sample *t*-test. Statistical significance was determined as in the previous analysis. Analyses were also conducted using a hrf model, yielding similar results (not shown).

To compare the results between the continuous and sparse acquisitions, conjunction analyses were conducted for each contrast of interest ($p = .01$). Additionally, we tested whether there was a correlation in the magnitude of the responses between acquisitions, by entering the corresponding cluster-averaged, subject-specific contrast estimates into a linear regression analysis.

2.4 Results

2.4.1 Continuous acquisition

2.4.1.1 Univariate analysis

Coordinates, z-scores, and cluster extents for all the significant activations obtained in the univariate analysis are reported in Table 2.2. The contrasts Instruments minus Speech yielded significant clusters in the right planum temporale (PT) and bilaterally in the planum polare (PP) (Figure 2.1a). Singing minus Speech yielded significant clusters bilaterally in the PP and in the right PT. Importantly, these clusters partially overlapped with those obtained in the preceding contrast (Figure 2.1b). This common activation for musical stimuli in general was statistically confirmed through a conjunction analysis ([Instruments − Speech] and [Singing − Speech]), which yielded significant activations in the right PT and bilateral regions in the PP (Figure 2.1a,b). Interestingly, in the more anterior regions of PP, Singing elicited stronger responses than both Speech and Instruments, whereas in the more posterior areas activation for Instruments was larger than for Speech and Singing (Figure 2.1b). Moreover, responses in this latter cluster, particularly in the left hemisphere, significantly correlated with stimuli's pulse clarity values ($z = 3.57$, $p < .001$).

Table 2.2 Significant activations associated with contrasts of interest at the group level

Anatomical location	**Left**			**Right**			**Z-score (peak voxel)**	***K*E**
	x	**y**	**z**	**x**	**y**	**z**		
CONTINUOUS MULTIBAND SEQUENCE								
Instrumental Music > Speech								
STG (posterior)				66	-28	12	5.99	155
STG (anterior)				46	-6	-6	5.88	241
STG (anterior)	-48	-6	-4				5.37	162
Singing > Speech								
STG (anterior)				50	4	-8	6.29	319
STG (posterior)				66	-26	10	5.74	213
STG (anterior)	-48	0	-6				7.43	286
[Instrumental Music & Singing] > Speech								
STG (posterior)				66	-26	10	5.74	123
STG (anterior)				48	4	-8	5.74	154
STG (anterior)	-48	-6	-4				5.37	85
Speech > Instrumental Music								
STS/STG, MTG				64	-8	-4	12.63	1570
STG/STS, MTG	-62	-12	0				15.14	2008
Singing > Instrumental Music								
STS/STG, MTG				62	-20	-2	11.04	1440
STS/STG, MTG	-60	-12	2				12.31	1835
[Speech & Singing] > Instrumental Music								
STS/STG, MTG				62	-20	-2	11.04	1250
STS/STG, MTG	-60	-12	1				12.31	1612
Speech > Singing								
STS/STG, MTG				64	-8	-4	7.22	845
STS/STG, MTG	-60	-22	-2				9.36	1239
INTERLEAVED SILENT STEADY STATE (ISSS) SEQUENCE								
Instrumental Music > Speech								
STG (posterior)				48	-32	24	4.39	238
STG (anterior)				42	-12	-10	4.88	344
STG (anterior)	-48	-6	-0				4.79	253
Speech > Instrumental Music								
STS/STG, MTG				56	-28	-2	7.54	863
STG/STS, MTG	-64	-28	4				8.01	1238

STG = superior temporal gyrus; STS = superior temporal sulcus; MTG = medial temporal

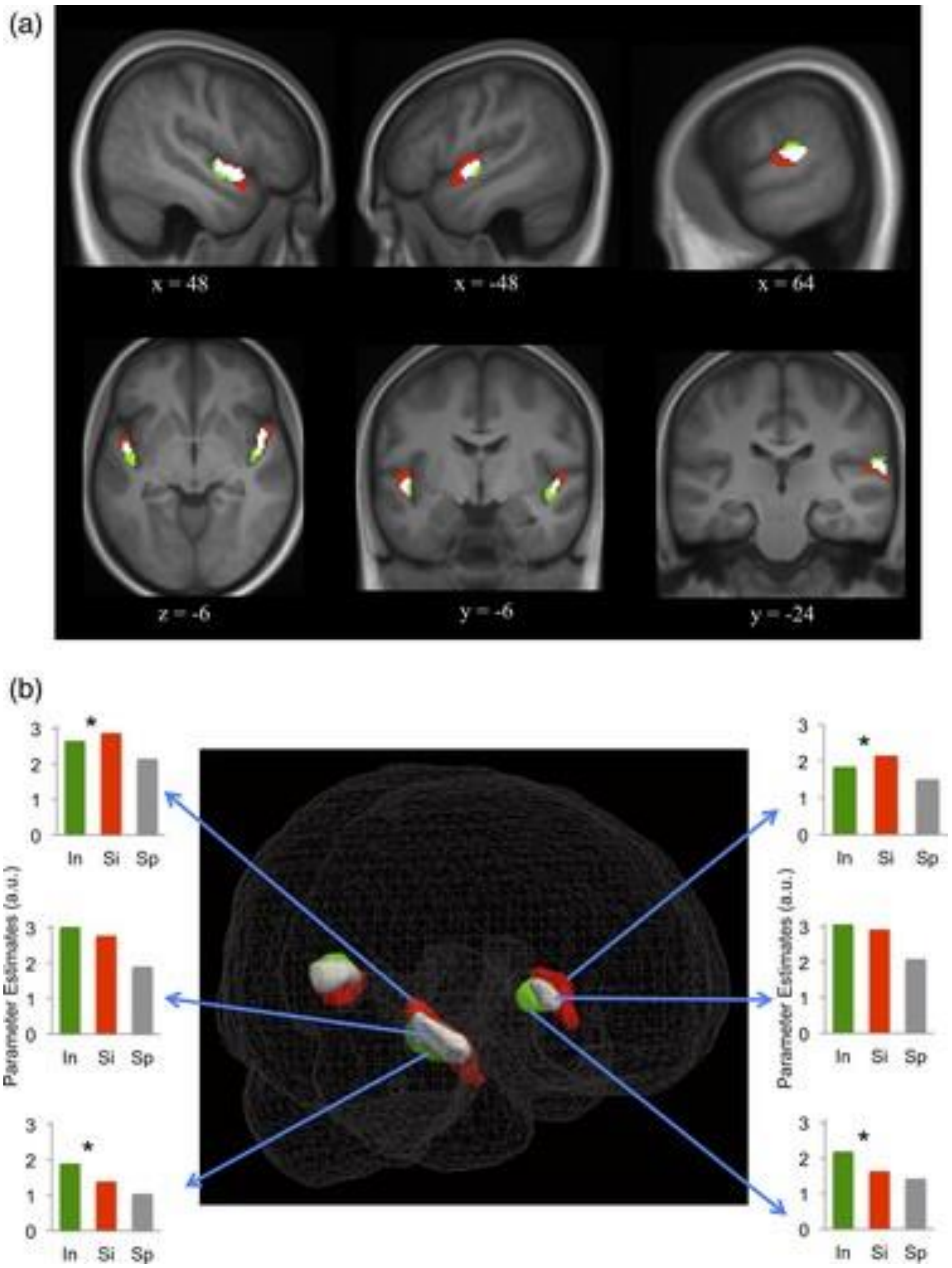

Figure 2.1. (a) 2D and (b) 3D renderings of the clusters of significant activations for the contrasts [singing − speech] (red), [instrumental music − speech] (green), as well as their conjunction (white). Threshold: *p* = .001 (corrected for multiple comparisons at the cluster level). Group average of the responses for each condition in each cluster (left and right hemispheres), using unsmoothed data. In: Instrumental music; Si: Singing; Sp: Speech; A.U.: arbitrary units. *significant difference (*p* < .001) between singing and instrumental music. In all cases, singing and instrumental music elicited significantly larger responses than speech (*p* < .001) [Color figure can be viewed at wileyonlinelibrary.com]

Speech vs. Instrumental Music revealed significant bilateral activity in in voice-preferred areas within the superior temporal sulcus (STS), superior temporal gyrus (STG), and medial temporal gyrus (MTG) (Figure 2.2, Top). Largely overlapping activations were obtained for the contrast Singing vs. Instruments, confirmed statistically using a conjunction analysis (Figure 2.2). Finally, the contrast Speech versus Singing also yielded significant clusters bilaterally in the STS, STG, and MTG (Table 2.2).

To assess whether the responses of voxels activated in the contrasts [Instruments − Speech] and [Singing − Speech] were modulated by simple acoustic parameters, we extracted the stimulus-specific parameter estimates for each of the three clusters reported in Table 2.1 and entered them (one at a time), as an additional covariate in the analysis. None of the acoustic features significantly Correlated with the BOLD parameter estimates in the music-preferred clusters (but see above for a correlation with pulse clarity in a subcluster of the contrast Instruments minus Singing and Speech).

In order to evaluate the robustness of the group-level activations of Music (Instruments and Singing) versus Speech, we tested for the presence of significant clusters in these contrasts for each subject separately, using an anatomical mask corresponding to the planum polare for each

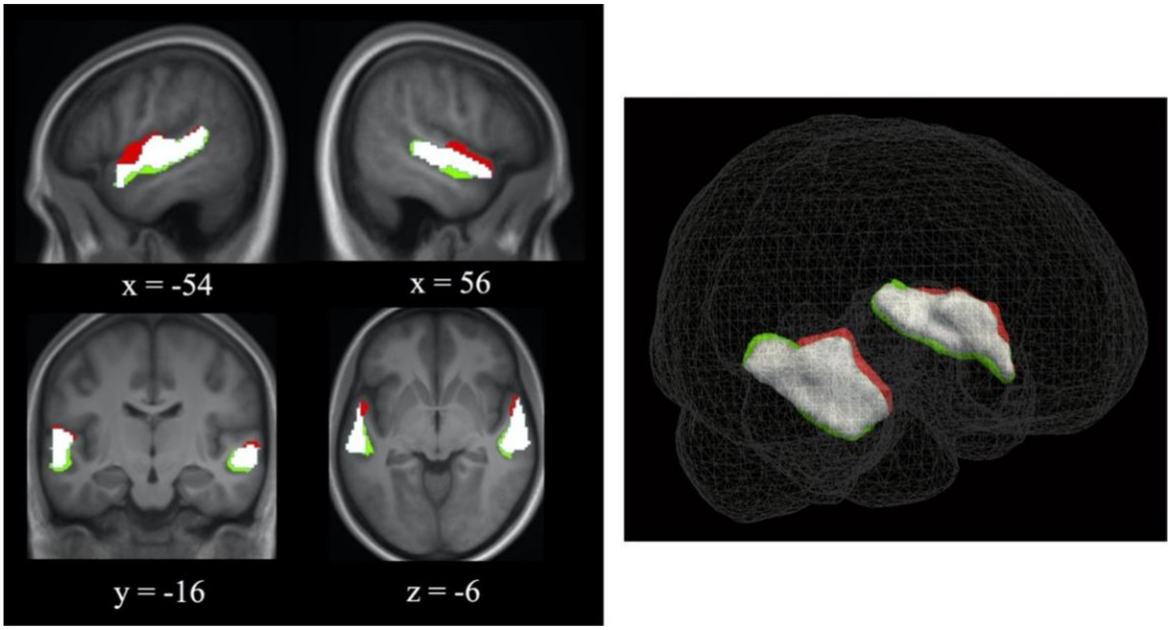

Figure 2.2 2D (Left) and 3D (Right) renderings of the clusters of significant activations for the contrasts [singing > instrumental music] (red), [speech > instrumental music] (green), as well as their conjunction (white). Threshold: *p* = .001 (corrected for multiple comparisons at the cluster level) [Color figure can be viewed at wileyonlinelibrary.com]

hemisphere, obtained from Harvard-Oxford Probabilistic Anatomical Atlas, as in our previous study (Angulo-Perkins et al., 2014). For the contrast Instruments minus Speech, 88% and 75% of subjects had significant clusters on the right and left hemispheres, respectively, using a significance threshold of $p = .01$ (uncorrected), and 75% and 63% with a more stringent threshold of $p = .001$. The proportion of subjects with significant clusters for the contrast Singing minus Speech was 83% and 75% for $p = .01$, and 63% and 58% for $p = .001$, for the right and left hemispheres, respectively. Figure 2.3 shows a prevalence map of the voxels, across the whole brain, that showed significant activation at the single-subject level ($p = .01$), for these two contrasts. Consistent with the group analysis (Figure 2.1), the individual clusters associated with Singing were slightly more anterior than those for Instrumental music.

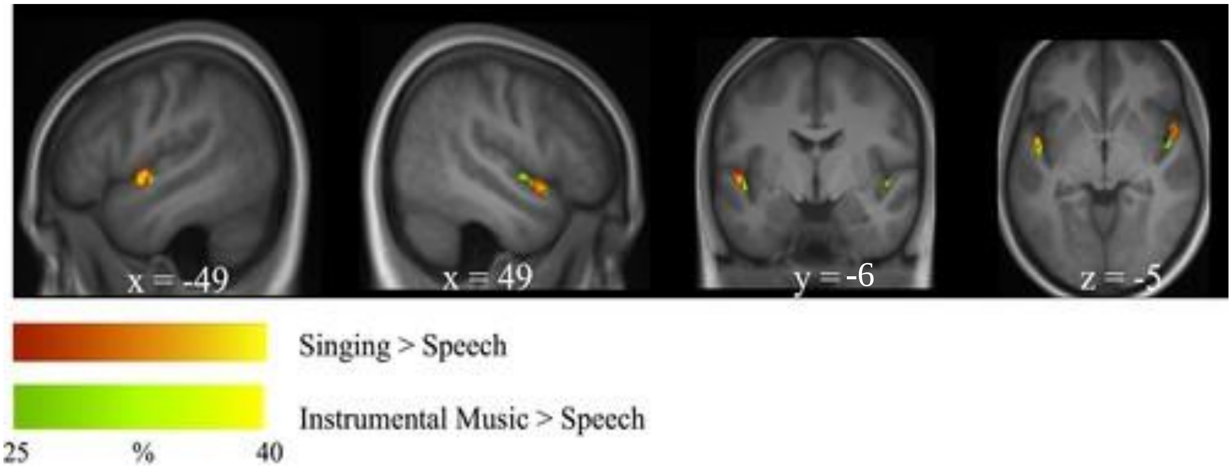

Figure 2.3 Prevalence maps showing the percentage of subject-specific significant activations at each voxel for the contrasts [singing > speech] (red scale) and [instrumental music > speech] (green scale). Clusters for singing were significantly more anterior (LH: $p = .008$; RH: $p = .02$) and lateral (LH: $p = .03$; RH: $p = .003$) than those for instrumental music [Color figure can be viewed at wileyonlinelibrary.com]

2.4.1.2 Multivariate analysis

The first ICA component (Figure 2.4a) included almost all voxels in the mask, representing, as expected, the general auditory responses elicited by all stimuli. There was a significant difference effect of category on the associated weights ($F[2,179] = 4.867$, $p = .009$), reflecting a smaller activation for Instruments compared with Speech ($p = .01$, Bonferroni corrected) and Singing ($p = .06$, Bonferroni corrected), with no difference between the two vocal sounds ($p > .9$). These results are in agreement with those from the univariate analysis.

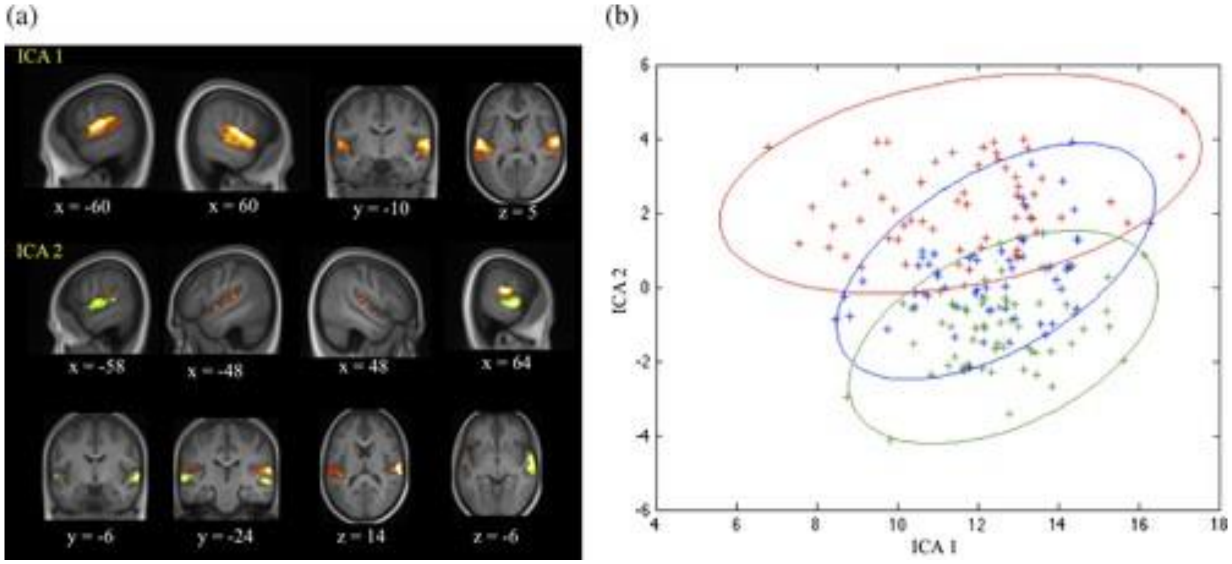

Figure 2.4 (a) First two components obtained in the stimulus-specific ICA. In the second component, red and green represent positive and negative values, respectively. (b) Scatterplots of the stimulus-specific eigenvalues corresponding to the first two ICA components. Each cross represents one stimulus: Instrumental music (red), singing (blue), and speech (green). Curves correspond to the minimum volume ellipsoid that covers all points of each category [Color figure can be viewed at wileyonlinelibrary.com]

The second ICA exhibited a bipolar pattern, with positive and negative subcomponents that largely overlapped with the music- and voice-preferred areas, respectively, obtained in the univariate conjunction analyses (Figure 2.1). Moreover, a significant category effect was observed for the corresponding weights, with all categories significantly differing from each other (all *p*'s < .001). Interestingly, the scatterplot of the weights for each stimulus (Figure 2.4b) showed almost no overlap between Instruments (positive values) and Speech (negative values), whereas singing fell in between the two, consistent with the shared activation pattern of this category with both instrumental music and spoken voice. This separation among categories was confirmed through a multiclass, ECOC model, which yielded an overall classification accuracy was 68% (leave-one-out cross-validation, chance level: 33%; *p* < .0001). Similar results were obtained when analyzing only Speech and Singing, confirming that the results were not due to simple acoustic differences between instruments and human voice.

2.4.2 Sparse Acquisition

The contrast Instrumental Music minus Speech yielded significant clusters bilaterally in the PP and the right PT. Significant clusters in the bilateral STS, STG, and MTG were obtained for the contrast Speech minus Instrumental Music (Figure 2.5 and Table 2.2). Importantly, no additional activation clusters were observed for either of the comparisons when using the “silent” sparse sampling protocol. The location of the clusters is very similar to what we found with the continuous acquisition in the same group of subjects. Furthermore, there was a significant correlation of the subject-specific, cluster-averaged parameter estimates for the contrast Instrumental Music minus Speech between both runs for each of the three main clusters ($.44 < r < .56$, p’s $< .05$).

Examination of the contrasts at single-subject level, using the same approach as described for the continuous acquisition, revealed that 79% and 88% of the subjects had significant clusters on the left and right hemispheres, respectively, using a significance threshold of $p = .01$ (uncorrected), and 58% and 75% with a threshold of $p = .001$.

2.5 Discussion

The aim of this study was to identify the brain responses to vocal and musical stimuli through the use of a high spatial- and temporal-resolution fMRI sequence. By using a stimulus set that varied widely in most of the basic acoustic measures and, critically, by including both vocal and instrumental musical excerpts, we were able to minimize potential confounding effects caused by differences in physical properties between categories. Moreover, to rule out any possible influence of scanner noise on the observed activations, we also employed a “silent” (i.e., sparse sampling) acquisition sequence with a subset of the original stimuli, in the same subjects. Finally, results obtained with a univariate categorical analysis were confirmed by a stimulus- based, multivariate approach.

2.5.1 Cortical responses to voice

When compared with musical instruments, human voice, either spoken or sung, elicited significant activations in clusters along the STS in both hemispheres. These results confirm and extend many reports in the literature showing that this region preferentially responds to the human voice (Belin et al., 2000, 2002; Belin & Zatorre, 2003; Fecteau et al., 2004; Kriegstein & Giraud, 2004; Pernet et al., 2015), with and without linguistic content. As the vocal stimuli included speech and songs

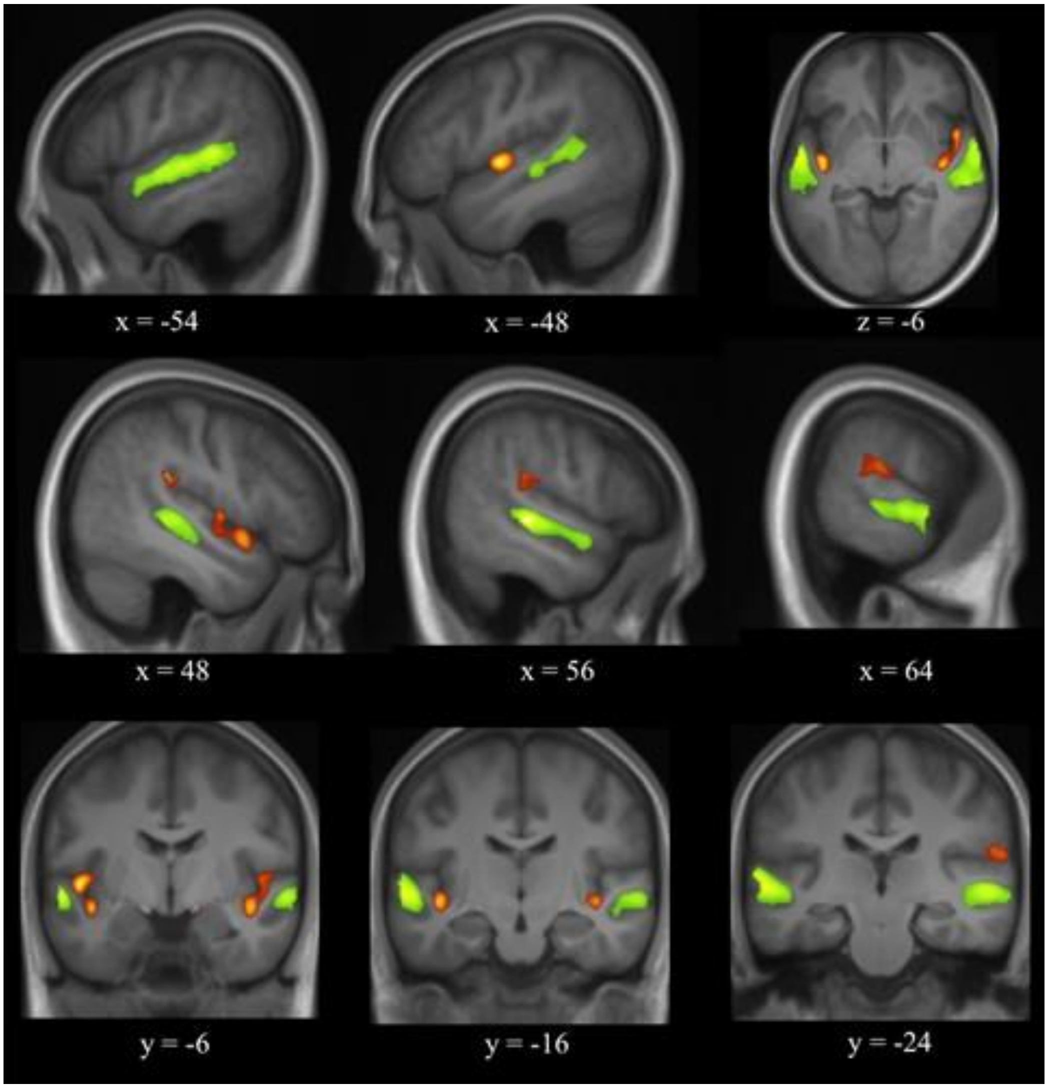

Figure 2.5 Clusters of significant activations for the contrasts [instrumental music > speech] (red) and [speech > instrumental music] (green) obtained with the sparse sampling acquisition. Threshold: *p* = .001 (corrected for multiple comparisons at the cluster level) [Color figure can be viewed at wileyonlinelibrary.com]

---

in different languages (most of which were not understood by the participants) as well as no-words singing, the responses observed in this area are likely to be related to the acoustic properties of the human voice, rather than reflecting semantic processing. This is also consistent with previous studies showing that this region responds significantly to human nonlinguisitic vocalizations (Belin et al., 2000, 2002; Fecteau et al., 2004). However, in this and previous studies, speech always elicited the strongest response, in both hemispheres. Importantly, and as reported before, while these clusters exhibited a bias, in terms of magnitude, for human voice, they also responded to nonvocal sounds, confirming that the so-called vocal temporal area (VTA) should be considered as a “voice-preferring” rather than as a “voice- selective” region (Belin et al., 2000).

2.5.2 Cortical responses to music

Conversely, contrasted to speech, music—either in instrumental or vocal form—yielded significant clusters in the anterior planum polare bilaterally and in the right planum temporale, in agreement with previous studies employing different stimulus sets and analyses approaches (Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002; Leaver & Rauschecker, 2010; Angulo-Perkins et al., 2014; Fedorenko et al., 2012; Rogalsky et al., 2011; Norman-Haignere et al., 2015; Aubé et al., 2015). In these previous studies most, when not all, musical stimuli contained an instrumental component, thus leaving open the question of whether these regions encode instrumental timbre (Leaver & Rauschecker, 2010), or music in general, including singing. Our conjunction analysis directly answers this question, confirming that clusters within these regions respond more strongly to both instrumental and vocal music than speech, with no significant differences between the first two categories. Moreover, these results also help address the often-raised concern about the possible confounding effects due to differences in acoustic parameters between instrumental music and voice. Indeed, because of the acoustic similarities between spoken and sung vocal expressions, and the substantial differences between the latter and musical instruments (ranging from drums to guitars to xylophones), it is highly unlikely that these activations simply reflect differences in basic acoustic features among categories. Instead, these areas seem to encode a higher-order feature (yet, obviously, still based on the physical characteristics of the stimuli) that is shared among different forms of musical expression, more than with other complex, social stimuli such as speech (e.g., melody vs. sentence-level intonation; Zatorre & Baum, 2012). Such a conclusion is also supported by the lack of correlation of the responses with any of the tested basic acoustic parameters, also shown previously by Leaver and Rauschecker (2010). Finally, the almost identical results obtained with the sparse-sampling sequence, rules out potential differential effects of scanner noise on voice and music (see below for further discussion of these methodological issues).

As could be expected, the overlap of the instrumental and vocal music versus speech clusters was not complete. Specifically, more posterior regions of PP responded more strongly to instrumental music than both speech and singing. Interestingly, this cluster, particularly in the left hemisphere, significantly correlated with pulse clarity, in agreement with our previous study (Angulo-Perkins et al., 2014). This acoustic parameter measures the intrinsic rhythm of a stimulus, arguably one of the defining characteristics of (instrumental) music, and appears to be involved in musical genre recognition specifically. Pulse clarity improves the ability to discriminate between genres, which differ in how audible the main pulsation is, over the texture of the base rhythm (Lartillot, Eerola, Toiviainen, & Fornari, 2008). As the key organizing structure of music, rhythm is fundamental for melody and harmony to exist (Thaut, Trimarchi, & Parsons, 2014). In contrast, the more anterior portions of PP were activated significantly more to singing than to either speech or instruments. In this case, we failed to identify one, or a linear combination of, acoustic parameters that correlated with activity in this region, including factors previously identified as differentiating singing from speech, such as duration, fundamental frequency floor, and vocal intensity (Livingstone, Peck, & Russo, 2013). One possible explanation for this null result is that the transition from speech to song involves a more complex, nonlinear weighting of several acoustic features (Saitou, Tsuji, Unoki, & Akagi, 2004; Saitou, Goto, Unoki, & Akagi, 2007; Livingstone, Peck & Russo, 2013). Overall, the brief duration of the stimuli did not allow for the computation of additional information about the acoustic features to directly explore this question. Additionally, no effects were observed based on the number of voices or melody lines on the magnitude or location of the music-related activations, again, likely due to the small variability in these features present in our stimuli. Thus, longer stimuli, with discrete categorical differences as to properly analyze the acoustic attributes, may lead to a tangible conceptualization of music, and thus a worthwhile pursuit in future studies. Another, complementary approach could be to use stimuli that have been artificially manipulated, in the line of the work of Saitou et al. (2004, 2007) to obtain the necessary independent variability of these candidate parameters to attain the statistical power required for detecting small effects, and potentially shedding light on this question.

The notion that "music-preferred" respond to a complex configuration of varying acoustic components bears some parallels with observations made in the literature regarding the processing of visual social stimuli. For example, headless bodies have been shown to elicit a greater response in body-selective areas of the brain, when presented to participants as a whole configured body, rather than as separate segregated parts appearing together, but not in full form (Brandman & Yovel, 2016). It is most likely that processing musicality reflects this pattern, in which each acoustic component is required in a particular arrangement, as to induce this response. As suggested by the development of the speech-to-singing synthesis system (Saitou et al., 2004, 2007), it is also likely that varying weights of each acoustic modification must be precise for the musical perception to be achieved. This can be related to the observed saliency- hierarchy in the fusiform face area (FFA) in response to specific facial features. Lai, Pancaroglu, Oruc, Barton, and Davies-Thomson (2014)'s fMRI-adaptation study identified that different parts of the face (e.g., nose, mouth, and eyes) contribute varying amounts to the overall neural signal in face-sensitive regions of the brain, such that greater response sensitivity is present for the upper half of the face, and more specifically, the eyes. The origin, and specificity, of category-selective, or preferred, brain regions has also been extensively studied, and debated, in the visual domain. In particular, an alternative hypothesis to the view that face selectivity, of preference, in the FFA is hard-wired (Kanwisher, McDermott, & Chun, 1997), has been put forward, suggesting that, rather than this region being face-sensitive, it may instead be better attributed as being an area of visual expertise, functioning to process and decipher highly complex visual stimuli (Gauthier, Skudlarski, Gore, & Anderson, 2000; Bilalić, 2016). According to this view, its preference for face stimuli reflects the fact that, as social individuals, we can all be considered experts of faces. Translating this idea to the auditory domain, it may be that as surrounded by music since birth, we have been able to fine-tune our perception of the particular "music algorithm," and now rely on this region of the brain to respond when needing to decipher more discrete changes within the musical framework. Some preliminary support for this idea comes from studies comparing responses to music between musicians and nonmusicians (Angulo-Perkins et al., 2014; Ohnishi et al., 2001), although further studies including musical expertise as a factor are still needed to fully test this hypothesis and better characterize the neural representation of musical and vocal stimuli in the brain.

2.5.3 Methodological considerations

The location and extent of the clusters of significant activation in the contrasts Instrumental Music versus Speech (and vice versa) obtained with the continuous and sparse acquisitions were very similar, as shown in Figures 1, 2, and 4. The concordance in results between the two sequences is in agreement with previous studies (e.g., Woods et al., 2009; Hall et al., 2014). Interestingly, we also found that the magnitudes of the responses for both runs were significantly correlated across subjects. Thus, it is very unlikely that differential effects of scanner noise on speech and music could have influenced the overall pattern of the observed activations. In turn, this provides further support for the use of continuous sampling sequences to study processing of complex auditory information, particularly when focusing on regions outside primary auditory cortex (Gaab et al., 2006). However, it should be noted that the goal of our study was not to provide a comprehensive quantitative comparisons between sequences, either in terms of how different acoustic parameters may be affected or, particularly, possible differences in their statistical power, as has been reported in some studies (Adank, 2012; Gaab et al., 2006; Hall et al., 1999).

Likewise, the activation patterns obtained with the standard univariate categorical ANOVA were very similar to those yielded by a stimulus-based multivariate ICA. This increases our confidence that the findings are not driven by a few high-leverage distinct stimuli in each category. Moreover, the distribution of the ICA stimulus-specific coefficients supports the hypothesis that the activation patterns represent the acoustic processing of the stimuli, rather than their potential categorization performed (implicitly) by participants. Indeed, the singing stimuli whose coefficients were closest to instrumental music (i.e., most positive) were chorales, whereas those with most negative values (i.e., most similar to those from speech) included amateur singing, lullabies, and a melodic Torah reading. These findings suggest the presence of a gradient from speech to music, which may be dependent on the clarity of the speech in the stimulus, irrespective of comprehension. The distribution of clusters responding preferentially to one or more of the different stimulus categories, as show in Figure 2.1b, aligns with the model proposed by Peretz and Coltheart (2003), suggesting that numerous discrete modules are involved in music and language processing, some of which overlap, while others appear independent.

Our paradigm was also designed to minimize other potential confounding effects, such as stimulus expectation, by equalizing the number of stimuli in each of the three categories and all first-order transition probabilities, as well as counterbalancing, across subjects, the specific order of stimuli within and between runs. While there was no explicit task for the participants to perform, we cannot exclude the possibility that some of the participants performed some sort of stimulus categorization (although this was not reported in the debriefing following the experiment). Nonetheless, we believe our findings are unlikely to be purely the result of such putative cognitive task, as mentioned above. Moreover, a recent meta-analysis of examining the role of attention on processing of auditory stimuli, including voice, observed that no additional areas in auditory cortex were recruited in active, compared with passive, listening conditions (Alho, Rinne, Herron, and Woods, 2014).

2.6 Conclusions

Different regions in the temporal lobe responded preferentially to vocal and musical stimuli. These included the superior temporal sulcus and gyrus for the former, and the planum polare and temporale for the latter. Consistent with its having both vocal and musical properties, singing recruited all these areas. Importantly, the results were obtained with a large and varied set of stimuli, as well as different acquisition sequences and analysis approaches. Taken together, these findings provide further support for a hierarchical processing of complex social acoustic stimuli along the temporal lobes, similar to what has been reported for the visual modality.

2.7 Acknowledgments

We are grateful to Mike Ferreira and Ilana Leppert for invaluable help implementing the multiband and sparse sequences. This study was supported by grants from the Natural Science and Engineering Research Council of Canada (NSERC, 262439-2009, 2017-05832) and the Canadian Institutes of Health Research (CIHR, MOP-130516, MOP-97967) to JLA. JCW was funded by an NSERC-Create Auditory Cognitive Neuroscience Graduate Fellowship, a CRBLM Graduate Student Stipend, a McGill's Integrated Program in Neuroscience Returning Student Award, and an NSERC - Canadian Graduate Scholarship -Masters fellowship.

2.8 Conflict of Interest

Authors have no conflicts of interest to declare.

2.9 References

Adank, P. (2012). Design choices in imaging speech comprehension: An activation likelihood estimation (ALE) meta-analysis. *Neuroimage, 8,* 360-369.

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.

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 nonmusicians. *Cortex, 59*, 126-137.

Aubé, 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,* 399-407.

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,* 35-39.

Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport, 14,* 2105-2109.

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. Brain Research. *Cognitive Brain Research, 13,* 17-26.

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

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

Bilalić, M. (2016). Revisiting the role of the fusiform face area in expertise. *Journal of Cognitive Neuroscience, 9,* 1345-1357.

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot International, 5,* 341-345.

Brandman, T., & Yovel, G. (2016). Bodies are represented as wholes rather than their sum of parts in the occipital-temporal cortex. *Cerebral Cortex, 26,* 530-543.

Brown, S. (2000). The “musilanguage” model of music evolution. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 271-300). Cambridge, MA: MIT Press.

Callan, D. E., Tsytsarev, V., Hanawaka, T., Callan, A. M., Katsuhara, M., Fuuyama, H., & Turner, R. (2006). Song and speech: Brain regions involved with perception and cover production. *Neuroimage*, *31,* 1327-1342.

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. Journal of Neuroscience, 23, 3423-3431.

Dietterich, T. G., & Bakiri, G. (1995). Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research, 2*, 263-286.

Ewender, T., Hoffmann, S., & Pfister, B. (2009). Nearly perfect detection of continuous f_0 contour and frame classification for TTS synthesis. In *Proceedings of Interspeech 2009* (pp. 100-103), Brighton, UK.

Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2004). Is voice processing species-specific in human auditory cortex? An fMRI study. *Neuroimage*, *23,* 840-848.

Fedorenko, E., McDermott, J. H., Norman-Haignere, S., & Kanwisher, N. (2012). Sensitivity to musical structure in the human brain. Journal of Neurophysiology, 108, 3289-3300.

Fodor, J. A. (1983). *The modularity of mind: An essay of faculty psychology*. Cambridge, MA: MIT Press.

Gaab, N., Gabrieli, J. D. E., & Glover, G. H. (2006). Assessing the influence of scanner background noise on auditory processing. I. An fMRI study comparing three experimental designs with varying degrees of scanner noise. *Human Brain Mapping*, *28,* 703-720.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds brain areas involved in face recognition. *Nature Neuroscience*, *3,* 191-197.

Glasberg, B. R., & Moore, B. C. J. (2002). A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society, 50,* 331-342.

Hall, A. J., Brown, T. A., Grahn, J. A., Gati, J. S., Nixon, P. L., Hughes, S. M., Menon, R.S., Lomber, S. G. (2014). There's more than one way to scan a cat: Imaging cat auditory cortex with high-field fMRI using continuous or sparse sampling. *Journal of Neuroscience Methods, 224,* 96-106.

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliot, M. R., Gurney, E.M., Bowtell, R. W. (1999). “Sparse” temporal sampling in auditory fMRI. *Human Brain Mapping, 7,* 213-223.

Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music Perception,* *26,* 195-204.

Juslin, P. N. (2001). Communication emotion in music performance: A review and a theoretical framework. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 309-397). New York, NY: Oxford University Press.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302-4311.

Kriegstein, K. V., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage, 22,* 948-955.

Lai, J., Pancaroglu, R., Oruc, I., Barton, J. J., & Davies-Thomson, J. (2014). Neuroanatomic correlates of the feature-salience hierarchy in face processing: An fMRI-adaptation study. *Neuropsychologia, 53,* 274-283.

Lartillot, O., Eerola, T., Toiviainen, P., & Fornari, J. (2008) Multi-feature modeling of pulse clarity: Design, validation and optimization. In J. P. Bello, E. Chew, & D. Turnbull (Eds.), *Proceedings of the 9th international conference on music information retrieval* (pp. 521-526). Philadelphia, PA: Drexel University.

Lartillot, O., Toiviainen, P., & Eeorla, T. (2008). A matlab toolbox for music information retrieval. In C. Preisach, et al. (Eds.), *Data analysis, machine learning and applications* (pp. 261-268). Berlin, Heidelberg: Springer.

Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: Effects of acoustic features and auditory objective category. *Journal of Neuroscience, 30,* 7604-7612.

Livingstone, S. R., Peck, K., & Russo, F. A. (2013). Acoustic differences in the speaking and singing voice. *The Journal of the Acoustical Society of America, 19,* 035080.

Manan, H. A., Yusoff, A. N., Franz, E. A., & Mukari, S. Z.-M. S. (2013). The effects of background noise on brain activity using speech stimuli on healthy young adults. Neurology, *Psychiatry and Brain Research, 19,* 207-215.

Murphy, D. R., Craik, F. I., Li, K. Z., & Schneider, B. A. (2000). Comparing the effects of aging and background noise of short-term memory performance. *Psychology and Aging,15,* 323-334.

Nebel, K., Stude, P., Wiese, H., Müller, B., Greiff, A., Forsting, M., Diener, H.-C., & Keidel, M. (2005). Sparse imaging and continuous event-related fMRI in the visual domain: A systemic comparison. *Human Brain Mapping, 24*, 130-143.

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage, 25,* 653-660.

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,* 1281-1296.

Ohnishi, T., Matsuda, H., Asada, T., Aruga, M., Hikarata, M., Nishikawa, M., Katoh, A., & Imabayashi, E. (2001). Functional anatomy of musical perception in musicians. *Cerebral Cortex, 11,* 754-760.

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,* 767-776.

Peelle, J. E. (2014). Methodological challenges and solutions in auditory functional magnetic resonance imaging. *Frontiers in Neuroscience, 8,* 1-13.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience, 6,* 688-691.

Pernet, C. R., McAleer, P., Latinus, M., Gorgolewski, K. J., Charest, I., Bestelmeyer, P. E., Watson, R.H., Fleming, D., Crabbe, F., Valdes-Sosa, M., Belin, P. (2015). The human voice areas: Spatial organization and inter-variability in temporal and extra-temporal cortices. *Neuroimage, 119*, 164-174.

Peretz, I., Vuvan, D., Lagrois, M. É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society B, 370,* 20140090.

Pinker, S. (1997). *How the mind works*. New York, NY: W.W. Norton.

Rabbit, P. M. A. (1968). Channel-capacity, intelligibility and immediate memory. *The Quarterly Journal of Experimental Psychology, 20,* 21-248.

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,* 3843-3852.

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 *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (pp. 215-218). New Paltz, NY: IEEE.

Saitou, T., Tsuji, N., Unoki, M., & Akagi, M. (2004). Analysis of acoustic features affecting “singing-ness” and its applications to singing-voice synthesis from speaking-voice. In *Proceedings of the 8th International Conference on Spoken Language Processing (ICSLP2004)* (pp. 1929-1932). Jeju Island, Korea.

Schirmer, A., Fox, P. M., & Grandjean, D. (2012). On the spatial organization of sound processing the temporal lobe: A meta-analysis. *Neuroimage, 63*, 137-147.

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*, 450-461.

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,* 774-782.

Setsompop, K., Gagoski, B. A., Polimeni, J. R., Witzel, T., Wedeen, V. J., & Wald, L.L. (2012). Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magnetic Resonance in Medicine, 67,*1210-1224.

Sperber, D., & Hirschfield, L. A. (2004). The cognitive foundations of cultural stability and diversity. *Trends in Cognitive Science, 8,* 40-46.

Thaut, M. H., Trimarchi, P. D., & Parson, L. M. (2014). Human brain basis of musical rhythm perception: Common and distinct neural substrates for meter, tempo, and pattern. *Brain Sciences, 4,* 428-452.

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognitive Emotion, 22,* 720-752.

Woods, D. L., Stecker, G. C., Rinne, T., Herron, T. J., Cate, A. D., Yund, E. W., Liao, I., Kang, X. (2009). Functional maps of human auditory cortex: Effects of acoustic features and attention. *PLoS One, 4,* e5183.

Zatorre, R. J., & Baum, S. R. (2012). Musical melody and speech intonation: Singing a different tune. *PLoS Biology, 10,* e1001372.

Connecting Chapters 2 and 3

Music has created a particular intrigue as a form of socio-emotional communication. The long evolutionary history of music, and its similarities to language and the voice (Fitch, 2006) suggests a potential overlap in their neural architecture. *Study 1* presents the first use of instrumental and vocal (singing) musical excerpts to investigate the neural relationship of music and voice. Particularly, in comparing prosody to singing, one eliminates the assumption that differences of neural activation are attributed the acoustic profile of voice alone. In doing so, an anterior-posterior gradient can be observed in the bilateral planum polare and right planum temporale, where more posterior regions present a greater sensitivity to instrumental music as compared to vocal music. This gradient was also reflected in the distribution of ICA stimulus-specific coefficients, where, based on acoustic attributes, singing overlapped with music and voice. We propose that along this gradient, lie neurons fine-tuned to their preferred composition of acoustic features, where specifically, neurons located posteriorly may be sensitive to the nonlinear composition of acoustic attributes proposed by Saitou et al. 2004; 2007) as “music”. That said, if neurons and their connections are refined over years of exposure (see for review, Münte et al., 2002), functional changes may take shape. Previous literature has shown greater recruitment of the planum polare (Angulo-Perkins et al., 2014) and planum temporale (Ohnishi et al., 2001; Angulo-Perkins et al., 2014) in musicians versus non-musicians, however, assessments of internal consistency have not been conducted to determine reliability of this across-subject variance (Boebinger et al., 2020). Markedly, *Study 2* recommends that the intrasubject reliability of the neural response to music, must be evaluated prior to any exploration into influences of expertise. To do so, we implemented a fast, high-resolution fMRI, presenting subjects with a series of sessions composed of varying design attributes (i.e., uniform vs. diverse stimuli; emotion vs. neutral stimuli). The internal consistency of neural responses to music and voice stimuli are evaluated with intraclass correlation coefficient (ICC) scores (Shrout & Fleiss, 1979) and pairwise cosine distances (Norman-Haignere et al., 2015) between sessions and subjects. Positive results for a reliable “music-preferred” response, may then provide strong conviction for an accurate detection of individual differences as a factor of musical expertise.

Chapter 3. Intra-individual reliability of voice- and music-elicited responses and their modulation by expertise

(Study 2)

Jocelyne C. Whitehead[a,b,c] & Jorge L. Armony[a,b,d]

[a]Douglas Mental Health University Institute, Verdun, Canada;
[b]BRAMS Laboratory, Centre for Research on Brain, Language and Music, Montreal, Canada;
[c]Integrated Program in Neuroscience, McGill University, Montreal, Canada;
[d]Department of Psychiatry, McGill University, Montreal, Canada

Under review in *Neuroscience* (2021)

3.1 Abstract

A growing number of functional neuroimaging studies have identified regions within the temporal lobe, particularly along the planum polare and planum temporale, that respond more strongly to music than other types of acoustic stimuli, including voice. This “music preferred” regions have been reported using a variety of stimulus sets, paradigms and analysis approaches and their consistency across studies confirmed through meta-analyses. However, the critical question of intrasubject reliability of these responses has received less attention. Here, we directly assessed this important issue by contrasting brain responses to musical vs. vocal stimuli in the same subjects across three consecutive fMRI sessions, using different types of stimuli. Moreover, we investigated whether these music- and voice-preferred responses were reliably modulated by expertise.

Results demonstrated that music-preferred activity previously reported in temporal regions, and its modulation by expertise, exhibits a high intrasubject reliability. However, we also found that activity in some extra-temporal regions, such as the precentral and middle frontal gyri, did depend on the particular stimuli employed, which may explain why these are less consistently reported in the literature. Taken together, our findings confirm and extend the notion that specific regions in the brain consistently respond more strongly to certain socially-relevant stimulus categories, such as faces, voices and music, but that some of these responses appear to depend, at least to some extent, on the specific features of the paradigm employed.

Keywords: music, prosody, musical expertise, language, reliability

3.2 Introduction

There is now considerable evidence, mainly coming from functional magnetic resonance imaging (fMRI) studies that although there is substantial overlap in the response to voice and music within the temporal lobes, there is also some degree of separability. Specifically, regions of the superior temporal gyrus/superior temporal sulcus (STG/STS) and middle temporal gyrus (MTG) have been consistently shown to be particularly sensitive to the human voice (Belin et al., 2002, 2000). Critically, these voice-preferred responses have been shown to be highly consistent not only across individuals (e.g., through meta-analyses; Schirmer et al., 2012), but also within subjects, through assessments of test-retest reliability (Pernet et al., 2015), as it had been done with other social stimuli, such as fusiform and occipital face areas (Kanwisher et al., 1997; Halgren et al., 1999; Puce et al., 1996), and the extrastriate body area (Downing et al., 2001). On the other hand, activity in antero- and postero-lateral STG (planum polare (PP) and planum temporale (PT), respectively) have been shown to respond preferentially to music. Importantly, this music preference is unlikely to be due to some basic acoustic features, as highly consistent findings have been obtained using a wide variety of stimuli, paradigms and analysis techniques (Armony et al., 2015; Angulo-Perkins et al., 2014; Whitehead & Armony, 2018; Leaver & Rauschecker, 2010; Fedorenko et al., 2012; Aubé et al., 2015; Rogalsky et al., 2011; Norman-Haignere et al., 2015). Moreover, several studies have found that music-evoked activity in these regions is modulated by individuals' music experience and/or expertise (Ohnishi et al., 2001; Dick et al., 2011; Angulo-Perkins et al., 2014; Schmithorst & Holland, 2003), providing further support for a "holistic" processing of musical stimuli in these regions. Nonetheless, unlike the case of voice, the issue of within-subject consistency in music-elicited activity has not been fully addressed. This is particularly important, given that responses that are highly consistent across individuals can still show poor within-subject test-retest reliability (Elliot et al., 2020). Naturally, the latter can have a particularly negative impact when studying how individual differences may modulate these responses.

Thus, the aim of this study was to directly assess, and compare, the within-subject reliability of responses to music and voice across sessions and stimulus sets. We also investigated how individual differences in expertise modulated these responses, as well as their reliability. Given that the definition of a "music expert" (or musician) varies considerably across studies (Angulo-Perkins et al., 2014; Luo et al., 2012; Liu et al., 2018; Zhao et al., 2015; Onishi et al., 2001; Schmithorst & Holland, 2003) – and that it is generally agreed that, rather than a dichotomous approach, a parametric one is usually preferable, as it is less sensitive to arbitrary cut-offs and it typically yields higher statistical power (MacCallum et al., 2002) – we used self-reported years of formal and informal music as a quantitative measure of music expertise. As a control, we also explored the effects of voice expertise, measured as the self-reported number of languages spoken by the subject. In terms of stimuli, we used a set of highly uniform musical (piano and violin) and vocal (pseudo-sentences spoken by 4 speakers) and one composed of unique instrumental excerpts and a wide variety of spoken languages. Two sessions using the uniform set were conducted (before and after the varied one), to control for any effect of habituation or practice effects (Shahin et al., 2008). Potential within-session carry-over effects were minimized through the use of a fast event-related design with appropriate stimulus counterbalancing and randomization order.

3.3 Materials and methods

3.3.1 Participants

Thirty right-handed healthy volunteers (age: M = 24 years, SD = 2.6 years; 15 female), with no self-reported history of neurological or psychiatric illness, and with normal hearing and normal or corrected-to-normal vision, participated in the study. Subjects' experience in playing a musical instrument ranged from 0 to 23 years (mean: 9 yrs; median: 9 yrs). All subjects were fluent in English, with 13 subjects speaking one other language and 11 speaking three or more languages (for details see, *Supplementary Materials*). One subject, a monolingual and non-musician, was removed from the study as she was unable to stay awake during the experiment session. The study was approved by the McGill University Faculty of Medicine Research Ethics Office and followed the guidelines of the Declaration of Helsinki. Subjects provided written informed consent prior to and received financial compensation after their participation.

3.3.2 Procedure

Prior to scanning, subjects completed a questionnaire on demographics, languages spoken and musical experience (number of years of formal musical training received, years of practice, and instruments learned). Subjects' total years of musical experience were calculated by adding years of formal training and years of practicing thereafter. The number of languages spoken, and total years of musical experience were used as indices of language and music expertise, respectively, in the fMRI analyses described below.

Participants completed three 8-minute fMRI sessions within one testing period (i.e., acquisition was stopped and subjects received a brief break between sessions; however, subjects did leave the MRI scanner), in which they passively listened to auditory stimuli presented using E-Prime 2.0 (Psychology Software Tools) and delivered binaurally through MRI-compatible headphones (Model S14, Sensimetrics). Sound tests were conducted prior to each scanner session to confirm stimuli could be heard above the scanner noise. The first and third sessions consisted of uniform stimuli; male/female pseudospeech phrases and piano/violin excerpts, expressing either a fearful or neutral emotion. Hereafter, such sessions will be referred to as "*Uniform*" sessions, while the second session consisting of highly diverse speech and music excerpts will be identified as the "*Mixed*" sessions.

3.3.3 Stimuli

All stimuli were monaural, but presented binaurally, and resampled to 32 bits, at a sample rate of 44100Hz. The Moore and Glasberg Loudness model (Glasberg & Moore, 2002), through the Genesis Loudness Toolbox implemented in MATLAB, was used to normalize stimuli through their short-term loudness (STL) maximum.

3.3.3.1 Uniform Sessions

Music: A total of 120 instrumental excerpts (M = 1.68s; SD = 0.16s) specifically composed for research were used (Vieillard et al., 2008; Aubé et al., 2013). Half were played with a piano and the other half with violin, both by professional musicians. Half of these stimuli expressed fear, while the other half were emotionally neutral. These novel pieces were used to avoid potential confounding effects of familiarity and memory (Janata et al., 2007).

Voice: A total of 120 pseudo-speech excerpts (M = 1.62s; SD = 0.21s) spoken by 4 different speakers (2 female) with fearful or neutral vocal intonation were selected from a previously validated dataset (Pell et al., 2009; Rigoulot et al., 2015). Pseudo-sentences reflected similar morpho-syntactic and phonotactic attributes as the English language, to preserve the expressed emotion while removing semantics.

3.3.3.2 Mixed Session

Music: Sixty different musical excerpts (M = 1.49s; SD= 0.13s), each featuring a single instrument were used. These included 40 unique instruments, including strings, percussion, brass and woodwinds. Stimuli were obtained from Vieillard et al. (2008), Aubé et al. (2015) and online databases.

Voice: Sixty different phrases (M = 1.51s, SD = 0.22 s) were spoken by a variety of speakers, including adults (33 male; 25 female) and children (n=2) in 45 different languages. The languages included those understood and/or commonly encountered by most participants –such as English, French, Spanish and Italian–, as well as some heard more infrequently, such as Baatonum, Mongolian, Gujarati, or Yiddish. One stimulus had no identifiable language (i.e., "baby talk"). These stimuli were used in a previous fMRI study (Whitehead & Armony, 2018).

3.3.4 fMRI acquisition and analysis

3.3.4.1 Acquisition protocol

Functional images were acquired using a continuous multi-band accelerated pulse sequence with a factor of 12 (Setsompop et al., 2012) implemented in a 3T Siemens TIM TRIO scanner with a 32-channel head coil. One thousand volumes (72 slices per volume, interleaved acquisition; FOV = 208 x 208 mm2, matrix = 104 x 104, voxel size = 2 x 2 x 2 mm3; TR = 529 ms; TE = 35 ms) were acquired per session. The first 10 scans of the session were discarded due to T1 saturation. A high-resolution T1-weighted image (192 slices, voxel size = 1 mm3, TR = 2.3 s, TE = 3 ms) was acquired for anatomical co-registration and normalization.

In each session, 120 stimuli (60 vocal and 60 musical excerpts) were presented. Stimuli were presented in a pseudo-random, fully balanced order (equal number of first-order transitions between categories and emotions, when applicable), to remove any potential carry-over effects. Each stimulus was presented only once, and the stimuli presented in each Uniform session were counterbalanced across subjects. The stimuli were presented in a continuous design and were jittered with a short ISI (duration: M = 2.49 s, SD = 0.20 s).

3.3.4.2 Preprocessing and statistical analyses

Image pre-processing was conducted using SPM12 (Wellcome Department of Imaging Neuroscience, London, UK; http://www.fil.ion.ucl.ac.uk/spm) using standard procedures, as in our previous studies (Whitehead and Armony, 2018; 2019). Briefly, functional images were spatially realigned to the first volume of the first session, normalized to the Montreal Neurological Institute

(MNI) 152 template and smoothed with a 6 mm FWHM Gaussian kernel.

3.3.4.3 Main effect of music vs. voice

Subject-specific statistical analysis was performed in SPM12 using a general linear model (GLM) in which categories of interest were included as boxcars of a length equal to the duration of the stimulus, convolved with the canonical hemodynamic response function. For Uniform sessions, conditions were coded as a function of Stimulus *Category* (Music/Voice), *Sub-Category* (Piano/Violin and Female/Male for music and voice, respectively) and *Emotion* (Fear/Neutral), whereas for the mixed session only *Category* (Music/Voice) was used. Subject-specific parameter estimates for each condition were taken to a second level, repeated-measures ANOVA, where contrasts of interest (Music > Voice; Voice > Music) were computed. Statistical significance was determined using a voxel threshold of $p = .001$, with a cluster-based familywise error rate (FWE) correction for multiple comparisons of $p < .05$ (k=97) as implemented in AFNI's algorithm 3dClsutSim (AFNI version 19.0.17). A binary mask of the two-tailed Music versus Voice contrast was used for further orthogonal analyses. To assess similarities and differences in activation patterns across sessions, we conducted whole-brain conjunction analyses and repeated-measures ANOVAs (IBM SPSS Statistics, version 23) on the ROI (cluster-averaged) contrast estimates, respectively.

3.3.4.4 Linear regression of expertise

First-level contrasts of the Music vs. Voice main effect for each session were taken to a second level linear regression analysis with years of musical expertise and number of languages spoken as independent variables. We restricted the search space to the music- and voice-preferred regions identified in the first whole-brain analysis. Statistical significance was determined using a voxel threshold of $p = .001$, with a cluster-based familywise error rate (FWE) correction for multiple comparisons of $p < .05$ (k=20) within the mask defined by the Music vs. Voice contrast.

3.3.4.5 Assessments of within-subject reliability

For the main effects of Music vs. Voice, consistency across the 3 sessions was assessed by calculating the intraclass correlation coefficient (ICC) for each voxel having a significant difference between categories. We used the ICC (3,1) score (Shrout & Fleiss, 1979) in the context of a mixed-effects model with sessions and subjects as fixed and random factors, respectively. Statistical significance (p<0.05, corrected for multiple comparisons FWE) was determined using the corresponding F-score, which resulted in a threshold of rho=.53 (which is typically considered to represent fair-to-good intraclass correlation; Cicchetti & Sparrow, 1981).

We also computed the pairwise cosine distance between sessions and subjects for the main effect, represented as the vector sum of each session- and subject-specific contrast estimate in an orthogonal multidimensional voxel space (Norman-Haignere et al., 2015). We then compared, for each of the possible three session pairs, the distance between sessions for the same subject and to each of the other subjects. The hypothesis that same-subject distances should be smaller than between-subjects was tested through a binomial distribution ($p < .05$, corrected for multiple testing).

To determine reliability of the results of the correlation between brain activity and music and language expertise, we conducted the regression analysis for each session separately and tested the correlation using the cluster-average (independent) contrast estimates from the other sessions. Statistical significance was determined using a threshold of $p = .05$, Bonferroni-corrected for multiple testing. Additionally, we computed these correlations, for sessions 1 and 2, using only a subset of the stimuli as a function of their emotional expression (fear or neutral) or their subcategory (piano or violin for music, and female or male speaker for voice).

3.3.4.6 Acoustic parameters

Ten basic acoustic parameters were computed using the MIRtoolbox (Lartiollot, Toiviainen, Eorla, 2008), MATLAB scripts (Ewender et al., 2009) and the Praat Vocal Toolkit (Boersma, 2002). They are summarized for each Category and Emotion in Supplementary Table 1. Independent samples t-tests were conducted for temporal and spectral acoustic parameters between categories and emotions. Significance was determined with a threshold of $p = .05$ and corrected for multiple comparison using a Bonferonni correction. The acoustic parameters used were selected from the auditory and emotion processing literature (Koelsch et al., 2018; Brattico et al., 2011; Whitehead & Armony, 2018; Aubé et al., 2013; Patel et al., 201l; Wiethoff et al., 2008; Frühholz et al., 2016; Zatorre & Belin, 2001). These included three temporal (tempo, pulse clarity, jitter) and six spectral (mean fundamental frequency (F0)/pitch, variability of f0, spectral centroid, spectral spread, spectral centroid, mean harmonics-to-noise-ratio, shimmer) parameters.

3.3 Results

3.3.1 Music and voice main effects

The coordinate, z-score, and extent of the significant activation clusters obtained in the main effect of Music vs. Voice analysis including the three sessions are reported in Table 3.1. The Music minus Voice contrast revealed significant clusters within the bilateral anterior STG (planum polare, PP), temporal pole (TP), and central operculum (CO), as well as in the bilateral anterior (AIns) and posterior (PIns) insula. Activity was also observed in the right planum temporale (PT), supramarginal gyrus (SMG) and parietal operculum (PO). Activity outside of the auditory cortices and neighbouring regions included the right pre-central gyrus (preCG), middle frontal gyrus (MFG), superior frontal gyrus (SFG) and supplementary motor area (SMA). The contrast Voice minus Music yielded significant clusters within the bilateral STS/STG, MTG and PP as well as the right TP and left PT, inferior temporal gyrus (ITG), and fusiform gyrus (FG; Figure 3.1).

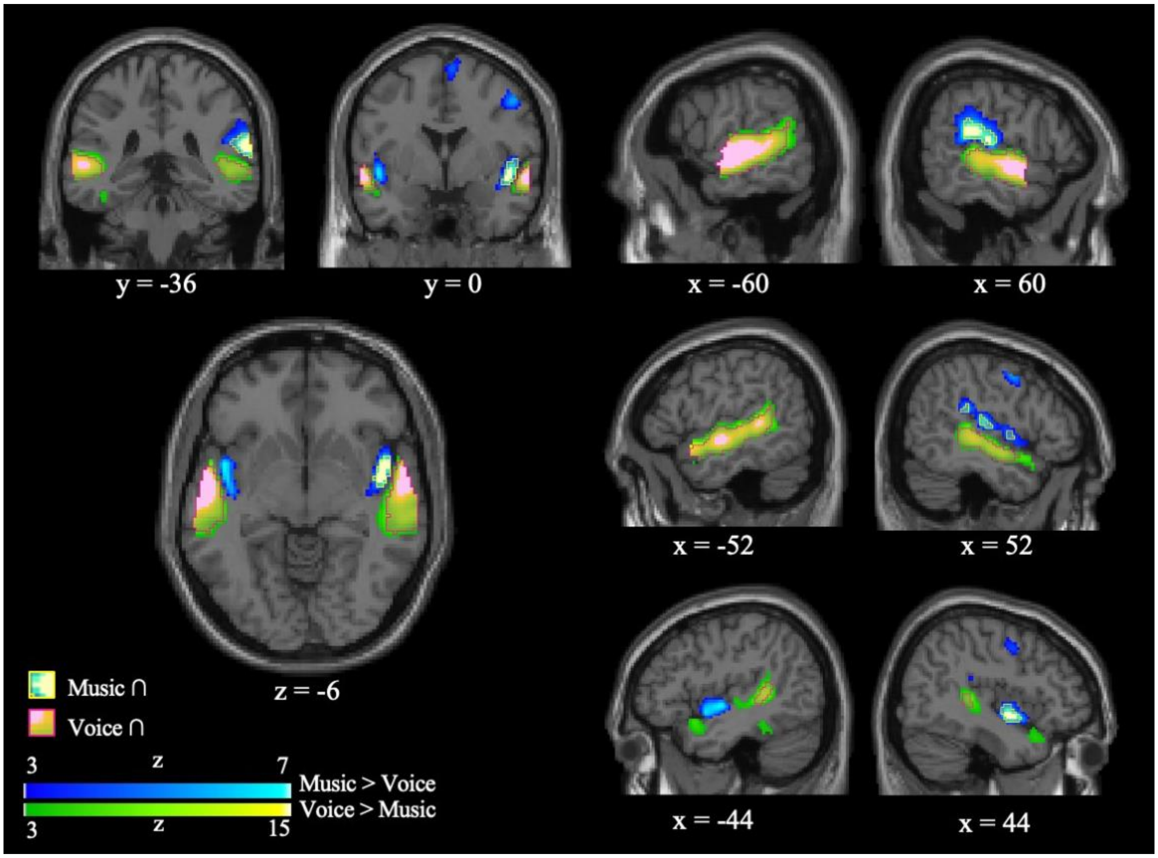

Figure 3.1 2D renderings of the clusters of significant activations for the contrasts Music > Voice (green) and Voice > Music (blue), and the conjunction of the three sessions for the same contrasts; Music > Voice (overlaid with yellow) and Voice > Music (overlaid with pink), at the group level. Threshold: p = .001 (corrected for multiple comparisons at the cluster level). A.U.: arbitrary units.

Table 3.1 Significant activations associated with main effect at the group level of Uniform and Mixed sessions, for the contrast of interest (cluster threshold at p<0.05; ke=97). Post-hoc analyses were conducted on reported clusters. Significant activations observed in the conjunction across sessions, for each contrast of interest (cluster threshold at p<0.05; ke=97).

Anatomical Regions	**Left**			**Right**				
	X	**Y**	**Z**	**X**	**Y**	**Z**	***Ke***	***Z***
MAIN EFFECT								
Music>Voice								
PP, aIns, TMP, pIns, CO				**48**	**4**	**-8**	**1525**	**10.31**
STG, SMG, PT, PO				66	-34	18		10.01
				62	-22	12		7.72
PreCG, MFG				**50**	**2**	**44**	**173**	**5.51**
R/L SMA, R SFG				**6**	**4**	**66**	**131**	**4.83**
PP, pIns, aIns, CO, aIns,	**-44**	**-6**	**-4**				**273**	**6.70**
TMP	-46	4	-8					6.69
Voice > Music								
STG/STS, MTG, PP,				**60**	**-6**	**-6**	**2005**	**20.15**
TMP				58	-28	0		13.84
				48	20	-24		5.70
STG/STS, PP, MTG, PT	**-58**	**-8**	**-4**				**2663**	**23.50**
	-62	-16	0					21.09
	-56	-36	4					17.06
FG, ITG	**-40**	**-42**	**-20**				**103**	**5.02**
CONJUNCTION								
Music>Voice								
STG, SMG, PT, CO				**64**	**-34**	**18**	**384**	**7.30**
				64	-22	12		5.14
PP, aIns, pIns, STG, CO				**48**	**2**	**-8**	**207**	**6.98**
Voice > Music								
STG/STS, PP, MTG				**60**	**-6**	**-4**	**1304**	**14.00**
				58	-26	0		10.11
STG/STS, PP, MTG,	**-58**	**-10**	**-4**				**1688**	**16.08**
SMG, PT	-56	-36	4					9.40
	-58	-48	18					3.56

For each contrast, an ROI-based (i.e., cluster-averaged ROIs; Table 3.1) post-hoc repeated measures ANOVA was conducted to identify any differences across sessions. In the case of Music minus Voice (i.e. music-preferred) contrast estimates, there was a significant ROI-by-Session interaction (F(6, 168)=6.941, p=.002, η2p=.117), whereby less activity was observed in the Mixed than Uniform 1 session in the right PreCG/MFG ROI, and in Uniform 2 than Uniform 1 in the left PP ROI; however, neither effect reached statistical significance after correcting for multiple comparisons. Post-hoc repeated measures ANOVA for voice-preferred ROIs presented a significant difference of contrast estimates across sessions (F(2, 56)=6.941, p=.002, η2p=.199), as well as a ROI-by-Session interaction (F(4, 112)=5.266, p=.001, η2p=.158). Post-hoc pairwise comparisons revealed that the Mixed session presented significantly lower activity than both Uniform 1 (t(28)=-2.34, pbonf=.038) and Uniform 2 (t(28)= -2.89, pbonf =.015) sessions, particularly in the left STG/STS (t(28)=-3.682, pbonf=.009) and bilateral STG/STS (L: t(28)=-4.023, pbonf = .004; R: t(28)=-3.030, pbonf =.045), respectively. Conjunction analyses across sessions yielded significant activity for Music > Voice within the right-lateralized PP, aIns, pIns, STG, CO, PT, and SMG, whereas Voice > Music conjunction revealed significant activity within the bilateral STG/STS, PP, MTG, and posterior into the left SMG and PT (Table 3.1; Figure 3.1).

3.3.2 Intrasubject reliability of music vs. voice activations

Figure 3.2 shows the map of voxels with a significantly positive intraclass correlation coefficient ICC (3,1) across the 3 sessions (rho≤.53, p<.05 FWE corrected). Reliable music-preferred activity (Figure 3.2A) was observed within regions of the bilateral PP and pIns, left-lateralized aIns, CO, and MTG, and right-lateralized TP and posterior STG, PT and PO. Voice-preferred voxels that presented above-threshold positive scores across sessions (Figure 3.2B) were located within the bilateral STG/STS, MTG, as well as anteriorly in the bilateral TP, and posteriorly in the left-lateralized PP, TTG and PT, and CO.

Pairwise cosine distance calculation (see Methods) between sessions 1-2, 1-3 and 2-3, revealed that for the contrast Music>Voice, 86, 83 and 66%, respectively, of subjects were closer to themselves than to every other one (chance level: 3.5%; all p’s<.001). For the contrast Voice>Music, the percentage of subjects that were closer to themselves than to every other subject were 100, 93 and 90%, respectively (all p’s < .001).

3.3.3 Influence of music- and language-expertise

A linear regression analysis including all sessions showed that subjects’ activity in the Music>Voice contrast significantly correlated with years of musical expertise in clusters within the right STG and SMG. On the other hand, language expertise (i.e., number of languages spoken) predicted activity within voice-preferred regions of the left PT, CO, STS/STG, TTG (HG), MTG, and post-central gyrus (Figure 3.3A; Table 3.2). No voxels within the music-preferred mask positively correlated with language expertise, nor any in the voice-preferred regions showed a positive correlation with music expertise.

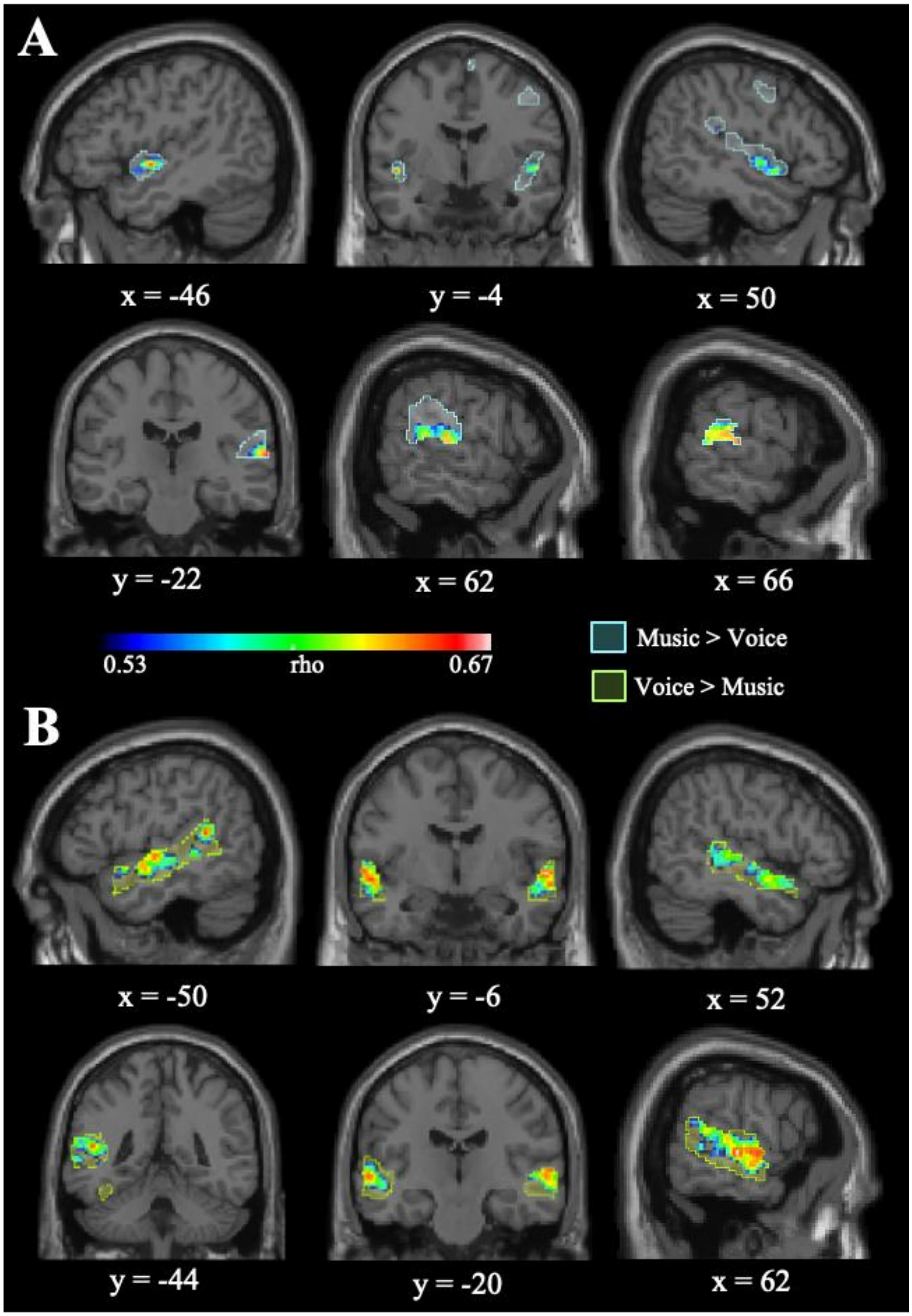

Figure 3.2 2D renderings of heat maps reflecting the positive intraclass correlation coefficients across the three sessions for A) Music > Voice and B) Voice > Music, with a significant threshold of rho=.53, at the group level. Heats maps are overlaid by the corresponding main effects of the contrasts of interest, Music > Voice (blue) and Voice > Music (green).

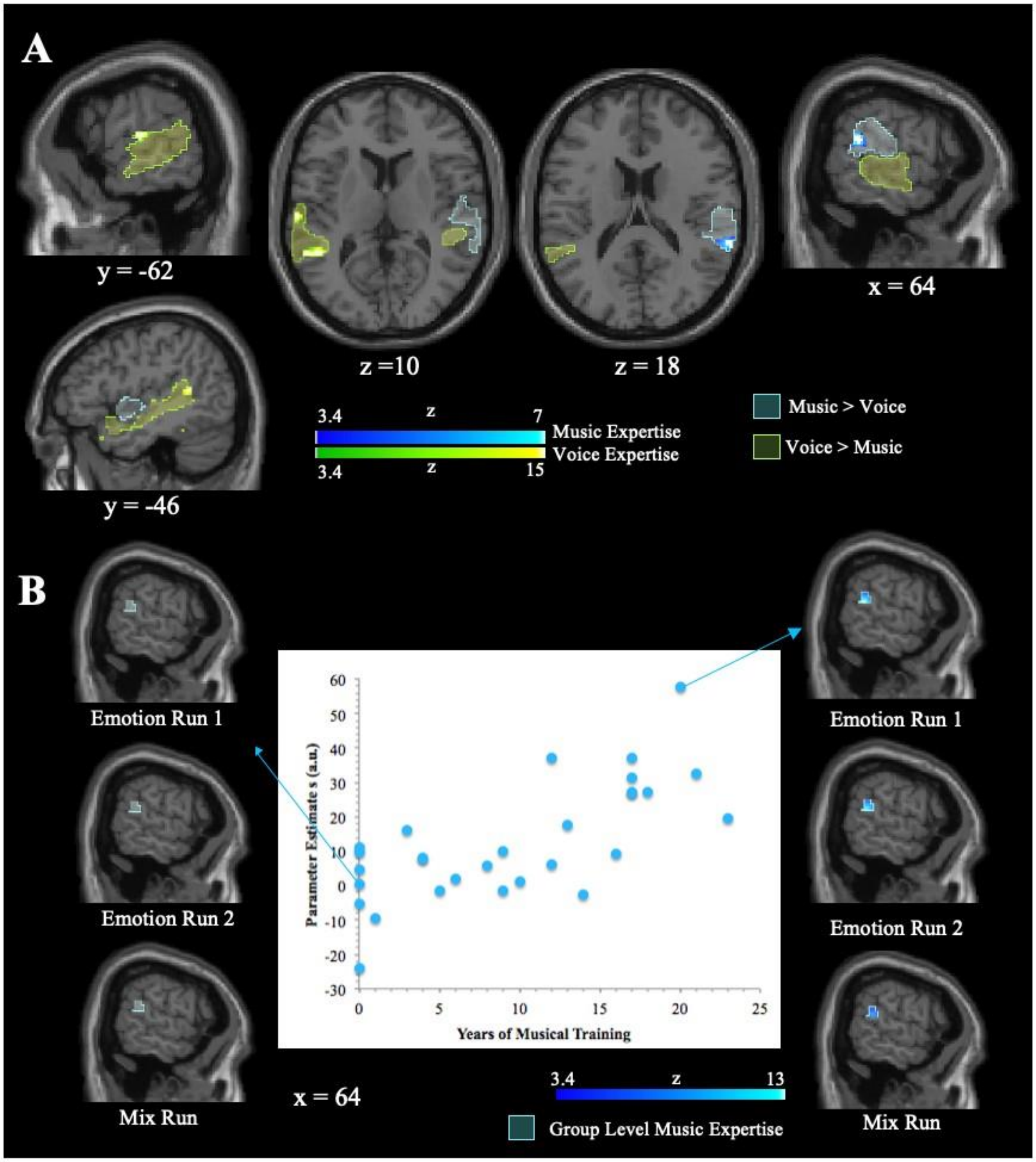

Figure 3.3 A) 2D renderings of the clusters of significant activations for linear regression of musical expertise (blue) and voice expertise (green) overlaid by the respective main effects of Music > Voice (blue overlay) and Voice> Music (green overlay) at the group level. Threshold: p = .001 (corrected for multiple comparisons at the cluster level) B) Plot of years of musical training correlated with parameter estimates in response to contrast Music > Voice, and 2D renderings of significant clusters of activity in response to Music > Voice across the three sessions (Uniform 1, Uniform 2, Mixed) for two sample subjects having a low (left) and high (right) number of years of musical training. Threshold: p = .001 uncorrected. A.U.: arbitrary unit.

Table 3.2 Significant activations associated with linear regression of contrasts (Voice > Music; Music > Voice) of Uniform and Mixed sessions collectively, with years of musical training, and number of languages spoken, as masked by music- and voice-preferred activity (cluster threshold p<0.05, ke=20).

Anatomical Regions	**Left**			**Right**				
	X	**Y**	**Z**	**X**	**Y**	**Z**	***Ke***	***Z***
Musical expertise in (Music > Voice)								
Posterior STG, SMG				64	-40	18	78	4.49
Language expertise in (Voice > Music)								
PT, CO, STS/STG, TTG (HG), PoCt	-62	-20	10				30	4.24
STS/STG, MTG	-46	-48	10				47	4.07

3.3.4 Reliability of expertise effects

To assess consistency of the correlations between brain activity and music and language expertise, we first conducted a linear regression analysis separately for each session. We then extracted the coordinates of the significant cluster(s) and computed the contrast estimate spatial average for all three sessions and calculated the correlation with the corresponding expertise scores. Results are reported in Table 3.3. Figure 3.3B illustrates the correlation of contrast estimates for Music > Voice with musical expertise, as well as across-session activity of two sample subjects who present values at the low and high end of musical expertise scale for illustration. Correlations were also conducted separately for individual emotions (Fear/Neutral) and subcategories (Piano/Violin or Male/Female) for the Uniform sessions. Positive correlations were observed across each condition for Music (.63 < r < .67, p's < .01, Bonferroni-corrected) and Voice (Cluster1 ([-62, -20,10]) .50 < r < .71, p's < .01; Cluster2 ([-46, -48, 10]) .55 <r< .65, p's < .01, Bonferroni-corrected).

Table 3.3 Significant correlations of each session (Uniform 1, 2, and Mixed), for each cluster obtained independently in session-specific linear regression of years of musical training (musical expertise) and number of languages spoken (voice expertise), in contrasts Music > Voice and Voice > Music, respectively. Masked by music- and voice-preferred activity (cluster threshold p<0.05, ke=20).

	Musical expertise in (Music > Voice)				**Data**		
	Session	**Coords**	**Z-score**	**Size**	**Uniform 1**	**Uniform 2**	**Mixed**
Cluster	**Uniform 1**	66 -40 20	3.93	51	0.66	0.66	0.73
	Uniform 2	64 -42 16	3.58	23	0.63	0.68	0.70
	Mixed	64 -42 20	4.47	70	0.65	0.66	0.75
					All p's < 0.001, Bonferroni corrected		
	Language expertise in (Voice > Music)				**Data**		
Cluster	**Uniform 1**	-44 -46 10	4.41	60	0.66	0.60	0.26
		-62 -20 8	3.99	40	0.68	0.60	0.35

Uniform 2	-62 -22 10	3.81	11	0.65	0.67	0.30
Mixed	–	–	–	–	–	–

No cluster survived session 3 alone. For sessions 1 and 2 p's<0.001 (Bonferroni-corrected), *session 3 p's>0.1*

3.4 Discussion

The aim of the current study was to test the within-subject reliability of music-preferred responses and their modulation by expertise, and compare them to those preferentially elicited by vocal stimuli. This was motivated by a growing concern in the neuroimaging literature that, despite across-study consistency in brain regions engaged in different processes, as evidenced, for instance, by meta-analyses, there is often markedly poor intrasubject reliability (Elliot et al., 2020), which can be exacerbated by differences in paradigms and stimuli (Yarkoni et al., 2010).

Results from standard group analyses confirmed the involvement of brain regions, particularly located along the temporal lobe, that were previously identified as either music- or voice- preferred (i.e., responding more strongly, though not necessarily selectively, to these stimulus types). Critically, we confirmed that the observed activations were highly consistent across sessions and stimulus sets for most subjects and for most, but not all regions. Moreover, the relation between brain activity and music or voice expertise was also very reliable across sessions, although in some cases, this relation seemed to depend on the stimulus set employed. Below, we discuss these findings in more detail.

3.4.1 Music- and voice-preferred regions

When all sessions were combined, the main contrast of interest yielded the expected voice- and music-preferred regions, identified in previous studies (Belin et al., 2000; Ethofer et al., 2009; Grandjean et al., 2005; Armony et al., 2015; Angulo-Perkins et al., 2014; Whitehead & Armony, 2018; Leaver & Rauschecker, 2010; Fedorenko et al., 2012; Aubé et al., 2011; Norman-Haignere et al., 2015), namely the bilateral STS/STG and MTG, and the bilateral PP and right PT, respectively.

Interestingly, in addition to the PP and PT, the passive perception of music also elicited activity outside the auditory cortex, in regions commonly associated with the ventral attention network (VAN; Corbetta & Shulman, 2002; Corbetta et al., 2008), such as the anterior insula, middle frontal gyrus and regions of the temporoparietal junction/IPL, as well as some belonging to the Salience Network (Goulden et al., 2014; Seeley et al. 2007), including the bilateral insulae, and right SMA and superior frontal gyrus. The VAN, often right-lateralized, is thought to function primarily in a bottom-up fashion, contributing to the assessment of the behavioural relevance or salience of the stimuli (see for review, Corbetta & Shulman, 2002), while the SN responds to the level of subjective salience, allocating necessary resources to specific cognitive functions (Sridharan et al. 2008). The recruitment of SN and premotor regions may also reflect activity of the motor planning system in response to entrainment of a musical beat, for predicting subsequent beats (Patel & Iverson, 2014). For instance, Schubotz (2007) suggests that the premotor cortex accumulates a sensorimotor representation of melodic sounds over time from the SMA, to then create a specified model for motor prediction. Moreover, anticipatory imagery of familiar music has been shown to be associated with frontal and premotor activity, which can be interpreted as evidence for sound and motor sequence learning (Leaver et al., 2009). Interestingly, pure auditory tasks involving passive listening to familiar music has shown to produce activity in secondary motor regions, while pure motor tasks, where subjects play the piano without acoustic feedback, has evoked activity in auditory processing areas. Markedly, overlapping activity of the two processes can be observed in the premotor cortex, SMA and the planum temporale (Baumann et al., 2005; Bangert, et al., 2006).

3.4.2 Intrasubject reliability of music- and voice-elicited responses

Different types (although not necessarily independent) of analyses (intra-class correlation coefficient, conjunction, cosine distance in voxel space, and repeated-measured ANOVAs) confirmed that voice- and music-preferred activations within the temporal lobe were highly consistent for all subjects across sessions, both in magnitude and location. The evident intra-subject consistency of processing high-complexity stimuli provides an optimistic outlook for test-retest reliability in neuroimaging (Kragel et al., 2021; Noble et al., 2021). These findings provide strong support for the development and use of “music localizers”; that is, a standardized tool to identify music-preferred regions in individual subjects, similar to those already employed for voice (Pernet et al. 2015) or face (Kanwisher et al., 1997) responses.

In contrast, and despite being highly significant when all sessions were combined, extra-temporal regions more responsive to music than voice, including parts of the SN and VAN, showed much less consistency across sessions, as evidenced by the ICC reliability analysis. This could explain why, unlike the regions within the temporal lobe previously mentioned, these areas are not always observed when contrasting music perception to other stimuli, such as voice. The lack of reliable activity in SMA, frontal, and premotor regions across sessions may suggest fluctuating, within-subject salience attribution to stimuli, or variability in attention allocation to stimuli, given the different stimulus sets used across sessions. Indeed, the ANOVA ROI-by-Session interaction showed that the precentral and middle frontal gyri were recruited to a lesser extent in the *Mixed* than in the *Uniform 1* session, while pairwise cosine distance calculations demonstrated a suggestive drop in similarity from *Uniform* sessions (83, 86%) to *Mixed* (66%; although chance level stands significantly lower at 3.5%). Music stimuli in the *Uniform* sessions consisted of short excerpts played with piano and violin, which had more rhythmicity than most of the stimuli presented during the *Mixed* session. The *Dynamic Attending Theory* proposes that greater perceptual resources are recruited when attention is drawn to an anticipated periodicity of a rhythm (Jones and Boltz, 1989). Thus, greater variability in the *Mixed* session could have resulted in a higher frequency of prediction errors, and reduced recruitment of attentional processes. Such findings have been observed in an fMRI passive perception study of rhythmic versus random music sequences, where greater activity was seen in SMA, preSMA and dorsal premotor cortex in response to more rhythmic sequences (Bengtsson et al., 2009). Ramnani and Passingham (2001) reported that activity in these regions increases over the course of motor-timing learning, as one improves their correct prediction rate. Additionally, musicians have shown enhanced activity of sensorimotor regions (e.g., SMA, precentral gyrus) when listening to music played by their own instrument type versus others (Margulis et al., 2009; Krishnan et al., 2018). Given that the majority of participants (66%) reported some experience in playing the piano/keyboard or a strings instrument (e.g., guitar or violin), subjects may have been more likely to attend to the salient, self-relevant stimuli (Moran et al., 2015) of *Uniform* versus *Mixed* sessions.

As with music, the voice-preferred regions typically reported (see for review, Belin et al., 2011) were consistent across sessions; however, activity seemed marginally stronger in *Uniform* than *Mixed* sessions, which may reflect subjects' familiarity with the syntactic properties of the English language. The conjunction analysis identified smaller clusters within only the bilateral STS/STG, PP, MTG, left SMG and PT (Figure 3.1). Particularly, the STS/STG, MTG and left PT were also areas of voxel-based significant positive correlations from the ICC reliability analysis

(Figure 2B). Other areas included the bilateral temporal pole, left planum polare, central operculum and transverse temporal gyrus. As part of the association cortex, the TP has been proposed as a hub for socio-emotional integration, making modulations through top-down processes (Olson et al., 2007). Moreover, this left-lateralized PT activity (also referred to as Wernicke's area; Bogen & Bogen, 1976) may reflect a site of convergence for cortical processes to access language comprehension systems (Nakada et al., 2001). ANOVA findings demonstrated that left-lateralized STG/STS activity was more prominent in *Uniform* than *Mixed* sessions. Importantly, this could not be attributed to habituation effects, as the *Mixed* session was presented between *Uniform* sessions. These findings coincided with single-subject reliability results (Table 3.3), in which *Uniform* sessions demonstrated a marginally higher chance of within-subject consistency, although *Mixed* sessions remained high, i.e., closer pairwise cosine distance to themselves than to 90% of other subjects. As mentioned earlier, this difference may be due to the fact that pseudo-utterances presented in *Uniform* sessions reflected phonotactic and morph-syntactic aspects of the English language (Pell et al., 2009). As all subjects spoke English, there may have been some familiarity bias, and a demand for greater recruitment of semantic processing regions. Consistent with this, Bestelmeyer et al. (2015) observed that English, Scottish, and American speakers had an enhanced response to their in-group accent, and neural repetition suppression to their out-groups', after repeated presentations. In our *Mixed* session, only 6% of words on average were understood by the participants, while only 4 participants (14%) could understand over 10% of the words (a maximum of 13% of all words were understood by an individual subject). Thus, differences between the *Uniform* and *Mixed* sessions likely reflect that differential processing may occur within the voice-preferred regions when a language does not sound like one's own (or a highly familiar one).

3.4.3 Influence of expertise

Developing skills or expertise, such as in music or sports, often results in neuroplasticity, inducing morphological and functional changes in the brain (see for review, Chang, 2014). Morphometric analyses have shown that absolute size of the right PT, caudal of the superior temporal plane, is the best predictor for distinguishing musical expertise groupings (i.e., musicians with and without absolute pitch, and non-musicians). A smaller right PT is seen in those with greater musical abilities (Schlaug et al., 1995; Zatorre et al., 1998; Luders et al., 2004), and suggested to reflect synaptic pruning and refining over early development (Keenan et al., 2001). Consistent with those findings, we observed a correlation between years of musical experience and music-evoked activity in the right posterior STG (neighbouring the PT) and SMG (i.e., rostral IPL; Figure 3). Upadhyay et al. (2008) identified that the pSTG acts as an extension of the PT. This heightened activity of the right PT/pSTG in musicians versus non-musicians had already been reported (Angulo-Perkins et al., 2014; Dick et al., 2011; Seung et al., 2005). Angulo-Perkins et al. (2014) suggested that this right-lateralized stronger activity in musicians may reflect the processing of sequences varying in pitch (Patterson et al., 2002). Rauschecker (2011; 2018) connected spectro-temporal sequences of music to motor movement, proposing that the dorsal auditory or “what” stream should also include the function of anticipatory regulation of sensorimotor responses. The acoustic signal is translated to a motor output, as initiated in the posterior STG/PT, moving through the dorsal stream to the IPL and premotor regions. Critically, our analyses confirmed that the modulation of music-related activity as a function of expertise is highly consistent for a given individual, across repetitions (Uniform sessions 1 and 2), stimulus types (Uniform and Mixed sessions), emotional expressions (fear and neutral) and instrument (piano and violin).

Interestingly, findings for voice-evoked activity, as modulated by number of languages spoken reflected an auditory-sensorimotor pathway comparable to that of music perception driven by musical expertise. Signfiicant correlations were observed along the STS/STG, MTG, Heschl’s gyrus and posterior in the PT, rolandic operculum and post-central gyrus (Figure 3.3). The left-lateralized activity overlapped with Wernicke’s area, part of the language network responsible for lexical and phonological recognition and understanding of words (Ardila et al., 2016). Through a series of language tasks, Parker Jones et al. (2012) demonstrated that bilinguals recruit greater left-lateralized activity in the STG, PT, dorsal precentral, and inferior frontal gyri. They suggest that bilinguals exhibit increased processing demands to manage control over multiple languages. Moreover, bilinguals have a greater ability in coupling attention and auditory processes (Krizman et al., 2014), which may explain, in part, their heightened activity.

Activity in the left postcentral gyrus has been previously reported during passive listening in response to articulatory features of speech (Correia et al., 2015; Arsenault and Buchbaum, 2015). Simmonds et al. (2011) attributed this to the close link auditory and somatosensory regions share for providing online feedback to the motor system during speech. Unlike music, reliability analyses demonstrated positive correlations with language expertise only for the Uniform sessions (pseudo-speech) and not for the Mixed session, which consisted of brief sentences spoken in many different languages, unfamiliar for most subjects. We could therefore infer that multi-linguals exhibit refined processing for familiar prosody (i.e. English pseudo-speech; Pell et al., 2009). In agreement, Gandour et al. (2007) observed greater activity in the left-lateralized posterior MTG in response to subjects' first, but not second language, in a task requiring the discrimination of sentence-initial versus sentence-final stress location.

3.4.4 Limitations and future directions

Whereas musical expertise, measured as number of years of training and practice, was fairly varied among participants (ranging from 0 to 23 years), the number of languages spoked by participants, used as an index of language expertise, was more restricted (1-5 languages). Thus, future studies would benefit by including a larger number of multi-lingual (i.e., three or more languages) participants to ensure the generalizability of the results. Additionally, differences of structural neuroplasticity have been reported based on the age at which second languages were learned (see for review, Li et al., 2014), as well as one's proficiency in the language (Mechelli et al., 2004). Moreover, differences of resting-state functional connectivity distinguished early versus late learners (Berken et al., 2016). Given this evidence, a language proficiency assessment and indication of age at which subsequent languages were learned, would prove useful in future studies. Importantly, recruiting more individuals fluent in a tonal language would improve ecological validity, as only two subjects spoke Mandarin. This is important as speech is processed differentially in tonal versus non-tonal language speakers (see for meta-analysis, Liang & Du, 2018).

Finally, it is important to keep in mind that the three sessions were conducted within the same scanning period (i.e., subjects took a brief break between sessions but did not leave scanner). Therefore, our findings cannot speak to the issue of intrasubject reliability of music-elicited brain responses across time. As previously done with voice (Pernet et al., 2015), future studies assessing test-retest reliability across days (and even months) are necessary. Moreover, including other types of vocal and musical stimuli (e.g., singing; Whitehead & Armony, 2018) could greatly confirm and extend the current findings.

3.5 Conclusion

In summary, music-preferred activity was consistently observed in the right planum polare and planum temporale, whereas activity in the right posterior STG and SMG correlated with musical expertise, indexed as number of years of training/practice. Conversely, voice-preferred activity in the left posterior STS/STG, MTG and PT was modulated by vocal expertise, measured by number of years spoken by the participant; however, these effects were substantially weaker when using highly diverse stimuli (i.e., many different instruments or languages). Overall, the study confirms a within-subject reliability of neural response to voice and music across sessions and stimulus sets. Beyond auditory perception, this type of paradigm and analysis may be applied to other modalities, stimuli, tasks, and/or paradigms, to isolate sources of intra-subject variability and, ultimately, improve understanding of the issues of reliability in fMRI highlighted in recent studies (Elliot et al., 2020; Kragel et al., 2021).

3.6 Acknowledgements

This study was supported by grants from the Natural Science and Engineering Research Council of Canada (NSERC, RGPIN-2017-05832) and the Canadian Institutes of Health Research (CIHR, MOP-130516) to JLA. JCW was funded by an NSERC-Create ACN Graduate Fellowship, a CRBLM Graduate Student Stipend, a McGill's IPN Returning Student Award, an NSERC - Canadian Graduate Scholarship –Masters (CGS-M) fellowship and an Alexander Graham Bell Canada Graduate Scholarship-Doctoral fellowship (CGS-D).

3.7 References

Angulo-Perkins, A., Aubé, W., Peretz, I., Barrios, F.A., Armony, J.K., & Concha, L. (2014). Music listening engages specific cortical regions within the temporal lobes: Differences between musicians and non-musicians. *Cortex, 59,* 126-137.

Ardila, A., Bernal, B., & Rosselli, M. (2016). The role of Wernicke's area in language comprehension. *Psychology & Neuroscience, 9,* 340-343.

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*, 35-39.

Arsenault, J.S., & Buchsbaum, B.R. (2015). Distributed neural representations of phonological features during speech perception. *Journal of Neuroscience, 35,* 634-642.

Aubé, W., Angulo-Perkins, 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*, 399-407.

Aubé, W., Peretz, I., & Armony, J.L. (2013). The effects of emotion on memory for music and vocalisations. *Memory, 21*, 981-990.

Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., Heinze, H-J., & Altenmüller, E. (2006). Shared networks for auditory and motor processing in professional pianists: Evidence for fMRI conjunction. *Neuroimage*, *30,* 917-926.

Baumann, S., Koeneke, S., Meyer, M., Lutz, K., & Janke, L. (2005). A network for sensory-motor integration: What happens in the auditory cortex during piano playing without acoustic feedback? A*nnals of the New York Academy of Sciences, 1060,* 186-188.

Belin, P., Bestelmeyer, P.E.G., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology, 102,* 711-725.

Belin, P., Zatorre, R.J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research, 13,* 17-26

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in the human auditor cortex. *Nature, 403,* 309-312.

Bengtsson, S.L., Ullén, F., Ehrsson, H.H., Hasimoto, T., Kito, T., Naito, E., Forssberg, H., & Sadato, N. (2009). Listening to rhythms activates motor and premotor cortices. *Cortex, 45,* 62-71.

Berken, J.A., Chai, X., Chen, J-K., Gracco, V.L., & Klein, D. (2016). Effects of early and late bilingualism on resting-state functional connectivity. *Journal of Neuroscience*, *36,* 1165-1172.

Bestelmeyer, P.E.G., Belin, P., & Ladd, D.R. (2015). A neural marker for social bias toward in-group accents. *Cerebral Cortex, 25,* 3953-3961.

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot International*, *5*, 341-345.

Bogen, J.E., & Bogen, G. (1976). Wernicke's region-where is it? *Annals of the New York Academy of Sciences*, *280,* 834-843.

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,* 1-16.

Chang, Y. (2014). Reorganization and plastic changes of the human brain associated with skill learning and expertise. *Frontiers in Human Neuroscience, 8,* 1-7.

Cicchetti, D.V., & Sparrow, S.A. (1981). Developing criteria for establish interrater reliability of specific items: Applications to assessment of adaptive behaviour. *American Journal of Mental Deficiency, 86*, 127-137.

Corbetta, M., Patel, G., & Shulman, G.L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron*, *58,* 306-324.

Corbetta, M., & Shulman, G.L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews*, *3,* 201-215.

Correia, J.M., Jansma, B.M., & Bonte, M. (2015). Decoding articulatory features from fMRI responses in dorsal speech regions. *Journal of Neuroscience*, *35*, 15015-15025.

Dick, F., Ling Lee, H., Nusbaum, J., & Price, C.J. (2011). Auditory-motor expertise alters "speech selectivity" in professional musicians and actors. *Cerebral Cortex, 21*, 938-948.

Downing, P.E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*, 2470-2473.

Elliot, M.L., Knodt, A.R., Ireland, D., Morris, M.L., Poulton, R., Ramrakha, S., Sison, M.L., Moffitt, T.E., Caspi, A., & Hariri, A.R. (2020). What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. *Psychological Science*, *31*, 792-806

Ethofer, T., Van de Ville, D., Scherer, K., & Vuilleumier, P. (2009). Decoding of emotional information in voice-sensitive cortices. *Current Biology, 19,* 1028-1033.

Ewender, T., Hoffmann, S., & Pfister, B. (2009). Nearly Perfect Detection of Continuous F_0 Contour and Frame Classification for TTS Synthesis. In *Tenth Annual Conference of the International Speech Communication Association*. Brighton, UK.

Fedorenko, E., McDermott, J.H., Norman-Haignere, S., & Kanwisher, N. (2012). Sensitivity to musical structure in the human brain. *Journal of Neurophysiology, 108,* 3289-3300.

Frühholz, S., van der Zwaag, W., Seanz, M., Belin, P., Schobert, A-K., Vuilleumier, P., & Grandjean G. (2016). Neural decoding of discriminative auditory object features depends on their socio-affective value. *Social Cognitive and Affective Neuroscience, 11,* 1638-1649.

Gandour, J., Tong, Y., Talavage, T., Wong, D., Dzemidzic, M., Xu, Y., Li, X., & Lowe, M. (2007). Neural basis of first and second language processing of sentence-level linguistic prosody. *Human Brain Mapping, 28,* 94-108.

Glasberg, B.R., & Moore, B.C.J. (2002). A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society, 50*, 331-342.

Goulden, N., Khusnulina, A., Davis, N.J., Bracewell, R.M., Bokde, A.L., McNulty, J.P., & Mullins P.G. (2014). The salience network is responsive for switching between the default mode and the central executive network: Replication from DCM. *Neuroimage, 99,* 180-190.

Grandjean, D., Sander, D., Pourtois, G., Schqaerz, S., Seghier, M., Scherer, K.R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience, 8,* 145-146.

Halgren, E., Dale, A.M., Sereno, M.I., Tootell, R.B., Markinkovic, K., & Rosen, B.R. (1999). Location of human face-selective cortex with respect to retinotopic areas. *Human Brain Mapping, 7,* 29-37.

IBM Corp. Released 2015. IBM SPSS Statistics for Macintosh, Version 23.0. Armonk, NY: IBM Corp.

Janata, P., Tomic, S.T., & Rakowski, S.K. (2007). Characterization of music-evoked autobiographical memories. *Memory, 15,* 845-860.

Jones, M.R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96,* 459-491.

Kanwisher, N., McDermott, J., & Chun, M.M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17,* 4302-4311.

Keenan, J.P., Thangaraj, V., Halpern, A.R., & Schlaug, G. (2001). Absolute pitch and planum temporale. *Neuroimage, 14,* 1402-1408.

Koelsch, S., Skouras, S., & Lohmann G. (2018). The auditory cortex hosts network nodes influential for emotion processing: An fMRI study on music-evoked fear and joy. *PLoS One*, *13,* e0190057.

Kragel, P.A., Han, X., Kraynak, T.E., Gianaros, P.J., & Wager, T.D. (2020). Functional MRI can be highly reliable but it depends on what you measure: A commentary on Elliot et al., (2020). *Psychological Science*, *32,* 622-626.

Krishnan, S., Lima, C.F., Evans, S., Chen, S., Guldner, S., Yeff, H., Manly, T., & Scott, S.K. (2018). Beatboxers and guitarists engage sensorimotor regions selectively when listening to the instruments they can play. *Cerebral Cortex, 28,* 4063-4079.

Krizman, J., Skoe, E., Marian, V., & Kraus N. (2014). Bilingualism increases neural response consistency and attentional control: Evidence for sensory and cognitive coupling. *Brain and Language*, *128,* 34-40.

Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: Effects of acoustic features and auditory objective category. *Journal of Neuroscience*, *30,* 7604-7612.

Leaver, A.M, Van Lare, J., Zielinski, B.A., Halpern, A., & Rauschecker, J.P. (2009). Brain activation during anticipation of sound sequences. *Journal of Neuroscience*, *29,* 2477-2485.

Liang, B., & Du, Y. (2018). The functional neuroanatomy of lexical tone perception: An activation likelihood estimation meta-analysis. *Frontiers in Neuroscience*, *12,* 1-17.

Li, P., Legault, J., & Litcofsky, K.A. (2014). Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. *Cortex, 58,* 301-324.

Liu, Y., Liu, G., Wei, D., Li, Q., Yuan, G., Qu, S., Wang, G., & Zhao, X. (2018). Effects of musical tempo on musicians' and nonmusicians' emotional experience when listening to music. *Frontiers in Psychology, 9,* 2118.

Luders, E., Gaser, C., Jancke, K., & Schlaug, G. (2004). A voxel-based approach to gray matter asymmetries. *Neuroimage*, *22,* 656-664.

Luo, C., Guo Z-w., Lai, Y-x., Liao, W., Liu, Q., Kendrick, K.M., Yao, D-z., & Li, H. (2012).

Musical training induces functional plasticity in perceptual and motor networks: Insights from resting-state fMRI. *PLoS One, 7,* e36568.

MacCallum, R.C., Zhang, S., Preacher, K.J., & Rucker, D.D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7,* 19-40.

Margulis, E.H., Mlnsa, L.M., Uppunda, A.K., Parrish, T.B., & Wong, P.C.M. (2009). Selective neurophysiological responses to music in instrumentalists with different listening biographies. *Human Brain Mapping, 30,* 267-275.

Mechelli, A., Crinion, J.T., Noppeney, Y., O'Doherty, K., Ashburner, J., Frackowiak, R.S., & Price, C.J. (2004). Structural plasticity in the bilingual brain. *Nature, 431,* 757.

Nakada, T., Fuji, Y., Yoneoka, Y., & Kwee, I.L. (2001). Planum temporale: Where spoken and written language meet. *European Neurology, 46,* 121-125.

Noble, S., Sheinost, D., & Constable R.T. (2021). A guide to the measurement and interpretation of fMRI test-retest reliability. *Current Opinion in Behavioral Sciences, 40,* 27-32.

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,* 1281-1296.

Ohnishi, T., Matsuda, H., Asada, T., Aruga, M., Hikarata, M., Nishikawa, M., Katoh, A., & Imabayashi, E. (2001). Functional anatomy of musical perception in musicians. *Cerebral Cortex, 11,* 754-760.

Olson, I.R., Plotzker, A., & Ezzyat, Y. (2007). The enigmatic temporal pole: A review of findings on social and emotional processing. *Brain, 130,* 1718-1731.

Parker Jones, 'Õ., Green, D.W., Grogan, A., Pliatsikas, C., Filippopolitis, K., Ali, N., Lee, H.L., Ramsden, S., Gazarian, K., Prejawa, S., Seghier, M.L., & Price, C.J. (2012). Where, when and why brain activation differs for bilinguals and monolinguals during picture naming and reading aloud. *Cerebral Cortex, 22,* 892-902.

Patel, A.D., & Iverson, J.R. (2014). The evolutionary neuroscience of musical beat perception: The Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience, 8,* 1-14.

Patel, S., Scherer, K.R., Björkner, E., & Sundberg, J. (2011). Mapping emotions into acoustic space: The role of production. *Biological Psychology, 87,* 93-98.

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,* 767-776.

Pell, M.D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S.A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics, 37,* 417-435.

Pernet, C.R., McAleer, P., Latinus, M., Gorgolewski, K.J., Charest, I, Bestelmeyer, P.E., Watson, R.H., Fleming, D., Crabbe, F., Valdes-Sosa, M., & Belin, P. (2015). The human voice areas: spatial organization and individual variability in temporal and extra-temporal cortices. *Neuroimage, 119,* 164-174.

Puce, A., Allison, T., Asgari, M., Gore, J.C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional magnetic resonance imaging study. *Journal of Neuroscience, 16,* 5205-5215.

Ramnani, N., & Passingham, R.E. (2001). Changes in the human brain during rhythm learning. *Journal of Cognitive Neuroscience, 13,* 952-66

Rauschecker, J.P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hearing Research, 271,* 16-25.

Rauschecker, J.P. (2018). Where, when and how: Are they all sensorimotor? Towards a unified view of the dorsal pathway in vision and audition. *Cortex, 98,* 262-268.

Rigoulot, S., Pell, M.D., & Armony, J.L. (2015). Time course of the influence of musical expertise on the processing of vocal and musical sounds. *Neuroscience, 290,* 175-184.

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,* 3843-3852.

Schmithorst, V.J., & Holland S.K. (2003). The effect of musical training on music processing: A functional magnetic resonance imaging study in humans. *Neuroscience Letters, 348,* 65-68.

Schirmer, A., Fox, P.M., & Grandjean, D. (2012). On the spatial organization of sound processing in the human temporal lobe: A meta-analysis. *Neuroimage, 62,* 137-147.

Schubotz, R.I. (2007). Prediction of external events with our motor system: Towards a new framework. *Trends in Cognitive Neuroscience, 11,* 211-218.

Seeley, W.W., Menon, V., Schatzberg, A.F., Keller, J., Glover, G.H., Kenna, H., Reiss, A.L., & Greicius, M.D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience*, *27,* 2349-2356.

Setsompop, K., Gagoski, B.A., Polimeni, J.R., Witzel, T., Wedeen, V.J., & Wald, L.L (2012). Blipped controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magnetic Resonance in Medicine*, *67,* 1210-1224.

Seung, Y., Kyong, J-S., Woo, S-H., Lee, B-T., & Lee, K-M. (2005). Brain activation during music listening in individuals with or without prior music training. *Neuroscience Research*, *52*, 323-329.

Shahin, A.J., Roberts, L.E., Chau, W, Trainor, L.J., & Miller, L.M. (2008). Music training leads to the development of timbre-specific gamma band activity. *Neuroimage*, *41*, 113-122.

Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86,* 420-428.

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,* 3341-3349.

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary and peaceful musical clips for research on emotions. *Cognition and Emotion, 2,* 720-752.

Whitehead, J.C., & Armony, J.L. (2019). Multivariate fMRI pattern analysis of fear perception across modalities. *European Journal of Neuroscience*, *49,* 1552-1563.

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,* 4913-4924.

Wiethoff, S., Wildgruber, D., Kreifelts, B., Becker, H., Herbert, C., Grodd, W., & Ethofer, T. (2008). Cerebral processing of emotional prosody-influence of acoustic parameters and arousal. *Neuroimage*, *39,* 885-893.

Yarkoni, T., Poldrack, R.A., Van Essen, D.C., & Wager, T.D. (2010). Cognitive neuroscience 2.0: building a cumulative science of human brain function. *Trends in Cognitive Sciences, 14,* 489-496.

Zatorre, R.J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 11,* 946-953.

Zatorre, R.J., Perry, D.W., Beckett, C.A., Westbury, C.F., & Evans, A.C. (1998). Functional anatomy of musical processing in listeners with absolute pitch and relative pitch.

Proceedings of the National Academy of Sciences, 95, 3172-3177.

Zhao, T.C., Lam, H.T.G., Sohi, H., & Kuhl, P.K. (2017). Neural processing of musical meter in musicians and non-musicians. *Neuropsychologia, 106*, 289-297.

3.8 Supplementary Materials

3.8.1 Participants

Instruments by played included piano/keyboard (15), guitar (9), violin (3), clarinet (3), saxophone (2), drums (2), trumpet (2), bass guitar (2), bass clarinet (1), trombone (1), flute (1), baritone (1), ukulele (1), djembe (1), and tabla (1), with one remaining unspecified. The languages understood included English (30), French (15), German (3), Spanish (4) Hindi (1), Dutch (3), Tamil (1), Italian (1), Mandarin (2), Farsi (1), Polish (1), Finish (1), Kannada (1), Kokani (1) and Urdu (1).

3.8.2 Supplementary Table

Supplementary Table 3.1 Mean and standard deviation values of spectral and temporal acoustic features for each stimulus domain of Uniform and Mixed sessions.

Audio Features	Uniform Sessions				Mixed Session	
	Music		Voice		Music	Voice
	Fear	Neutral	Fear	Neutral		
TEMPORAL						
Tempo (bpm)	119(31)	124(32)	131(29)	119(28)	125(30)	126(30)
Pulse Clarity (a.u.)	.22(.12)	.22(.10)	.26(.13)	.22(.07)	.28(.17) [a]	.23(.08)
Jitter (local) (%)	1.52(1.30)	1.47(1.41)	2.18(0.47) [a]	2.45(0.69) [a]	2.24(2.57)	1.43(1.19)
SPECTRAL						
Mean Fundamental Frequency (F0)	204(95)	261(133) [a b]	271(47) [a b]	156(40)	257(138) [a]	185(56)
Std. Dev. Fundamental Frequency (F0)	43(41)	64(44) [a b]	51(23) [b]	34(33)	47.2(37.9) [a]	29.6(15.4)
Spectral Centroid (kHz)	1.8(.09)	1.8(.08)	4.1(1.0) [a b]	3.3(1.0) [a]	2.3(1.5)	2.1(1.0)
Spectral Spread (Hz)	1.9(.7)	1.7(.6)	4.5(.5) [a b]	4.2(.6) [a]	5.8(3.3)	5.1(1.6)
Mean HNR	9.7(6.3)	13.5(7.7) [b]	13.5(2.2) [a]	13.2(2.6)	11.4(8.0)	11.4(3.1)
Shimmer (local) (%)	14.5(6.1) [a]	11.7(6.9) [a]	8.1(2.2)	8.5(2.2)	12.7(6.8) [a]	10.2(3.2)

a.u. = arbitrary units; bpm = beats per minute. Values were calculated with MIRToolbox, except for those related to the Fundamental Frequency (http://www.tik.ee.ethz.ch/~spr/f0_detection) and the last four features (Praat). [a]Significantly different from other Domain; [b]Significantly different from other Emotion (p < .05, Bonferroni corrected).

Connecting Chapters 3 and 4

Studies 1 and 2 provided substantial insight as to the differential neural processing of auditory socio-emotional information, particularly, given the good intrasubject reliability and generalizability. Specifically, *Study 2* demonstrates that irrespective of affect or stimulus, one can reliably observe a neural response within the music-preferred area of the right planum polare and planum temporale. This approach offers an alternative solution to meta-analyses that often struggle with issues of experimental heterogeneity (e.g., control conditions, image acquisition, data reduction, analyses; Schirmer et al., 2012), or identifying and filtering out low-quality studies (Sharpe et al., 1997), among other concerns (see for review, Samartsidis, et al., 2017; Wager et al., 2007; Costafreda, 2009). Conclusively, meta-analyses may best function as a preliminary scope or overview of the literature, particularly when evaluating high-level processes (e.g., emotion or social cognition; see limitations, Phan et al., 2002; Fusar-Poli et al., 2009; Van Overwalle et al., 2014) with intricate, non-uniform study designs, requiring careful interpretation. This is evident given the inconsistent reports of amygdala recruitment in response to emotion across-modality (see meta-analyses, Phan et al., 2002; Costafreda et al., 2008; Wager et al., 2015), suggestively attributed to relaxed study inclusion criteria (Costafreda et al., 2008) or reporting biases that do not appropriately consider other widespread recruitment (Phan et al., 2002). Thus, *Study 3* offers an alternative approach, employing a within-subject design to investigate emotion perception across a series of socio-emotional channels (faces, bodies, prosody, and non-linguistic vocalizations). Comparable to *Study 2*, it explores generalizability and makes *a priori* assumptions from a constructionist view; here hypothesizing extensive whole brain processing to, in part, detect the signal, sort the signal based on past experiences, and form a response based on the perception (Barrett et al., 2017a; 2017b). We thus employed a multivariate analysis (Lindquist et al., 2012; Wager et al., 2015) to compare the *passive* perception of fear to a neutral control. A neutral control was used as Peelen et al. (2010) did not report amygdala activity when comparing fear versus other emotions that were also likely to recruit the amygdala (for reviews, see Armony, 2013). We anticipated that if music and voice are comprised of the same foundational building blocks (i.e., acoustic parameters), albeit, of unique compositional arrangements, then emotion conversely, could be re-constructed from a decomposition of unique signals (modalities) to match a common "emotion" (fear) signature of unlimited dimensions (Lindquist, 2013; Lindquist & Barret, 2012; Barrett & Satpute, 2013).

Chapter 4. Multivariate fMRI pattern of analysis of fear perception across modalities

(Study 3)

Jocelyne C. Whitehead[1,2,3] & Jorge L. Armony[1,2,4]

[1] Douglas Mental Health University Institute, Verdun, Canada;
[2] BRAMS Laboratory, Centre for Research on Brain, Language and Music, Montreal, Canada;
[3] Integrated Program in Neuroscience, McGill University, Montreal, Canada;
[4] Department of Psychiatry, McGill University, Montreal, Canada

Whitehead, J.C., & Armony, J.L. (2019). Multivariate fMRI pattern analysis of fear perception across modalities. *European Journal of Neuroscience*, 49, 1552-1563.

4.1 Abstract

The emotional expression of fear can be processed through a number of modalities, and of varying forms, however, much of the functional imaging literature has centered on investigating fear as expressed through faces. Findings point to an active involvement of the amygdala, and remain fairly consistent in other studies of unimodal fear perception; however, few studies have looked at within-subject cross-modal responses to fear. Thus, we approached this inquiry by testing 30 healthy young adults with fast, high-resolution fMRI, recording the neural responses of fear perception, as expressed through faces, bodies, prosody, and vocalizations. The study was analyzed using a multivariate approach (multi-voxel pattern analysis) and yielded a significant distinction in the responses associated with the perception of fearful vs. neutral emotions. Calculated weights highlighted areas in the amygdala and surrounding subcortical structures as contributing the greatest to the discrimination; however, a whole-brain analysis was necessary to obtain above-chance classification accuracy, suggesting that processing fear across modalities likely involves a broad, distributed network. Thus, our findings support a multivariate approach to studying a highly complex construct such as emotion, as it accounts for multiple voxels simultaneously and can accommodate the high subject-level variability that oftentimes comes with studying emotion perception.

4.2 Abbreviations

AAL	automated anatomical labeling
AUC	area under the curve
fMRI	functional magnetic resonance imaging
FOV	field of view
FWHM	full width at half maximum
GLM	general linear model
ISI	interstimulus interval
LOSO	leave-one-subject-out
MVPA	multi-voxel pattern analysis
ROC	receiver operating characteristic
ROI	Region of Interest
RSA	representational similarity analysis
STL	short-term loudness
SVM	support vector machine
TE	echo time
TR	repetition time

4.3 Introduction

Functional neuroimaging studies have provided substantial evidence for a central role of the amygdala in processing fearful expressions. While most of these studies have employed faces (Aubé, Angulo-Perkins, Peretz, Concha, & Armony, 2015; Breiter et al., 1996; Morris et al., 1996; Van der Zwaag, Da Costa, Zürcher, Adams, & Hadjikhani, 2012; Vuilleumier, Armony, Driver, & Dolan, 2001), there is also empirical support for amygdala recruitment during processing of fear expressed through other visual means, such as body postures (e.g., De Gelder, Snyder, Greve, Gerard, & Hadjikhani, 2004; Grèzes, Pichon, & de Gelder, 2007; Hadjikhani & de Gelder, 2003; Peelen, Atkinson, Andersson, & Vuilleumier, 2007; Pichon, de Gelder, & Grèzes, 2009). In addition, significant amygdala responses to fearful expressions in the auditory modality, particularly non-linguistic vocalizations (e.g., Aubé et al., 2015; Fecteau, Belin, Joanette, & Armony, 2007; Herrmann et al., 2016; Phillips et al., 1998) have also been reported. In contrast, studies investigating brain responses to fear expressed through prosody have yielded inconsistent results in terms of amygdala involvement (for a review, see Liebenthal, Silbersweig, & Stern, 2016). Given the large variability among studies in terms of experimental design, image acquisition parameters and analysis methodology, it is unclear what the source of this discrepancy in the literature is. One way to minimize these potential confounding factors is to conduct studies measuring responses to fearful expressions in different modalities, using comparable stimuli and conducted within the same subjects. While a few of such studies have been performed, they all used non-linguistic vocalizations as the only form of auditory emotional expression. For instance, Phillips et al. (1998) studied perception of facial expressions and vocalizations in the same participants, observing bilateral activation in the amygdala for both modalities, when a response to fear was compared to that of mildly happy expressions. Similarly, previous work from our group examined brain responses to several basic emotions (fear, happy, sadness) as expressed through faces, non-linguistic vocalizations, and music (Aubé et al., 2015). Significant responses to fear expressions were observed in the posterior amygdala and posterior insula, regardless of the modality in which they were presented. Interestingly, fear was the only emotion that yielded significant activity in the same voxels for all three domains, as compared to neutral. A comparable cross-modal study, analyzed with a multivariate approach (multi-voxel pattern analysis, MVPA), examined responses to dynamic clips of faces, bodies, and non- linguistic vocalizations, expressing five basic emotions (Peelen, Atkinson, & Vuilleumier, 2010). Although the patterns for each emotion remained fairly consistent across modalities, as compared to the others, the amygdala did not contribute significantly to the successful classification of fear expressions. As the authors mentioned, this lack of amygdala involvement was likely due to the fact that the response to fear was classified against the average of the other emotions, which also recruit the amygdala (for reviews, see Armony, 2013; Sander, Grafman, & Zalla, 2003; Zald, 2003;), rather than to emotionally neutral stimuli. Finally, a recent study by Sachs, Habibi, Damasio, and Kaplan (2018) employed a multivariate searchlight to investigate the response of emotion perception as expressed through music (clarinet and piano) and human voice (non-linguistic vocalizations), in three emotions (fear, happy, sad). They observed a significant contribution of the auditory cortex, posterior insula, and the parietal operculum, but not from the amygdala, for the classification of emotions across categories. As with the previous study, the absence of amygdala contribution is likely a result of classifying the response to fear against the other emotional conditions (happy, sad).

Here, we conducted a high spatial- and temporal-resolution functional magnetic resonance imaging (fMRI) study to better characterize the brain responses to fearful expressions, compared to neutral ones. Critically, we employed a within-subject design and presented biologically and socially relevant cues of different sensory modalities and categories, namely faces, bodies, non-linguistic vocalizations, and prosody. Moreover, given the growing literature suggesting that emotional processing engages a network of cortical and subcortical regions (Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012; Wager et al., 2015), we performed a multivariate (MVPA) analysis on the acquired data. By maintaining the within-subject voxel-level variability and eliminating between-subject-level variability (Davis et al., 2014), MVPA is ideally suited for studying emotion perception, which is particularly sensitive to individual differences (e.g., Aleman & Swart, 2008; Canli, Sivers, Whitfield, Gotlib, & Gabrieli, 2002; Dickie & Armony, 2008). We anticipated that the accurate classification of fearful vs. neutral expressions across categories and modalities would rely on a cross-modal network, with the amygdala playing a central role, and including regions previously implicated in emotional processing, such as the insula, cingulate, and areas within the prefrontal cortex.

4.4 Materials and methods

4.4.1 Participants

Thirty right-handed healthy volunteers (age *M* = 24 years, *SD* = 2.6 years; 15 female) with no self-reported history of neurological or psychiatric illness, and normal or corrected-to-normal vision, participated in the fMRI study. Subjects provided written informed consent prior to participation. The study was approved by the McGill University Faculty of Medicine Research Ethics Office, and followed the guidelines of the Declaration of Helsinki.

4.4.2 Materials and procedure

Participants completed two 8-minute fMRI runs, using a continuous multi-band sequence (TR = 0.529 s, voxel: 2 mm isotropic) implemented in a 3T Siemens TIM TRIO scanner with a 32-channel head coil. Participants passively listened to auditory stimuli (vocalizations and pseudo-speech) or viewed static images (facial and body expressions) that expressed a fearful or neutral emotion. Stimuli were presented in a pseudo-randomized with equal probability transitions across categories and emotions.

4.4.3 Stimuli

4.4.3.1 Auditory stimuli

Vocalizations

Sixty non-linguistic vocalizations (*M* = 1.54 s; *SD* = 0.26 s), produced by 40 different speakers (20 female) were used, with half expressing fear (screams) and half being neutral (yawns) (Armony, Chochol, Fecteau, & Belin, 2007; Aubé, Peretz, & Armony, 2013; Fecteau et al., 2007).

Prosody

Sixty pseudo-speech excerpts (*M* = 1.59 s; *SD* = 0.19 s) spoken by four different speakers (two female) with fearful or neutral vocal intonation were selected from a previously validated dataset (Pell, Paulmann, Dara, Alasseri, & Kotz, 2009; Rigoulot, Pell, & Armony, 2015). The pseudo-sentences reflected similar morphosyntactic and phonotactic attributes as the English language as to preserve the emotion being expressed, but without semantic meaning attached to them.

All stimuli were monaural, but presented binaurally. The sounds were resampled to 32 bits, at a sample rate of 44,100 Hz, and adjusted for loudness by normalizing to the short-term loudness (STL) maximum using the Moore and Glasberg Loudness model (Glasberg & Moore, 2002), as implemented in the Loudness Toolbox in MATLAB.

4.4.3.2 Visual stimuli

Faces

Pictures of faces of 60 individuals (30 female) with fearful and neutral expressions were selected from the KDEF database (Lundqvist, Flykt, & Öhman, 1998). The images had the hair removed around the exterior of the faces and were altered to attain uniform face size, resolution, and contrast (Sergerie, Lepage, & Armony, 2006, 2007).

Bodies

Sixty full body pictures of individuals (half female) expressing fear or in a neutral position were used (De Gelder & Van den Stock, 2011). The faces were blurred to isolate the emotion expressed only by the body.

Visual stimuli were presented for a variable duration (*M* = 1.57 s; *SD* = 0.22 s) with a similar distribution to that of the auditory stimuli.

4.4.4 fMRI acquisition and analysis

4.4.4.1 Acquisition protocol

In each of the two runs, 120 stimuli (30 vocalizations, 30 pseudo-sentences, 30 faces, and 30 body expressions, half neutral and half fearful) were presented in a pseudo-random, fully balanced order (equal number of first-order transitions between categories and emotions) to remove any possible carry-over effects. Each stimulus was presented only once per run and the stimulus subsets used in each run were counterbalanced across subjects. The stimuli were presented in a continuous design and were jittered using a short ISI (duration: *M* = 2.49s, *SD* = 0.20 s).

Functional images were acquired using a multi-band accelerated pulse sequence with a factor of 12 (Setsompop et al., 2012). Thousand volumes (72 slices per volume, interleaved acquisition; FOV = 208 × 208 mm^2, matrix = 104 × 104, voxel size = 2 × 2 × 2 mm^3; TR = 529 ms; TE = 35 ms) were acquired per run. The first 10 scans of the run were discarded due to T1 saturation. A high-resolution T1-weighted image (voxel size = 1 mm^3) was acquired (192 slices, TR = 2.3 s; TE = 3 ms) for anatomical co-registration and normalization.

4.4.4.2 Preprocessing

Image preprocessing was conducted using SPM12 (Wellcome Department of Imaging Neuroscience, London, UK; http://www.fil.ion.ucl.ac.uk/spm). Functional images were spatially realigned to the first volume of the first run and normalized to the Montreal Neurological Institute (MNI) 152 template. The images were then moderately smoothed with a 2 mm FWHM kernel, to enhance the signal-to-noise ratio (Gardumi, Ivanov, Hausfeld, & Valente, 2016).

4.4.4.3 Analysis

Subject-specific statistical analysis was performed in SPM12 using a univariate general linear model (GLM) where categories of interest (fearful and neutral faces, bodies, vocalizations, and prosody) were included as boxcars of a length equal to the duration of the stimulus, convolved with the canonical hemodynamic response function.

Category-, subject-, and run-specific parameter estimates (*betas*) obtained from the univariate GLM were used for the pattern analysis as implemented in PRoNTo (Schrouff, Rosa, et al., 2013), in which a kernel classifier was trained to identify voxel activation patterns in the beta images with a support vector machine (SVM; Burges, 1998) using LIBSVM implementation (Chang & Lin, 2011) and a leave-one-subject-out (LOSO) cross-validation. The whole-brain feature set consisted of a single kernel that was mean-centered and used in a binary classification with SVM. The model was then cross-validated and performance was evaluated using the classification accuracy obtained (Schrouff et al., 2018). Significance of the performance of the model was set at $p = 0.05$, determined through 1000 permutations of the training labels. A mask of the cerebrum was used to define the voxel space, covering 90 regions (all but cerebellum and vermis) of the Automated Anatomical Labeling (AAL) atlas (Tzourio- Mazoyer et al., 2002) using the Wake Forest University (WFU) Pickatlas Tool (Maldjian, Laurienti, Kraft, & Burdette, 2003) in the MNI template. Additionally, masks of (a) the bilateral amygdala and (b) the limbic system (amygdala, hippocampus, parahippocampus, and cingulate gyrus) were created to compare accuracy across models with different search volumes. Weights per voxel were computed a posteriori (Schrouff, Cremers, Garraux, Baldassarre, & Mourao- Miranda, 2013). In order to identify the regions that most contributed to the classification in each model, clusters corresponding to the top 1% positive and negative weights, with a minimum extent of five contiguous voxels were extracted. The *p*-value associated with the weight for each voxel was computed through a permutation test in which prior to feature selection, the condition labels were randomized and the weight maps were obtained for the classification analysis (*N* = 1,000). Statistical significance of each cluster was obtained by computing the average of the voxel-specific *p*-values.

4.5 Results

Model performance, reported in Table 4.1 is expressed through the accuracy obtained for each classification, the area under the receiver operating characteristic (ROC) curve (AUC), and the sensitivity and specificity of the classification. The models discriminating Auditory vs. Visual stimuli, Faces vs. Bodies, and Vocalizations vs. Prosody all performed significantly above chance. In modeling Auditory vs. Visual stimuli, the clusters corresponding to the top 1% of voxels (see Materials and Methods) were located in primary and secondary auditory and visual cortices, as well as subcortical (hippocampus, amygdala, and putamen) and supramodal cortical (temporoparietal junction, middle frontal gyrus, and left prefrontal cortex) regions (Figure 4.1). Anatomical regions that contributed the most to discriminating the conditions Vocalizations vs. Prosody included auditory cortical regions, right amygdala, insula, and right supplementary motor area (Figure 4.2a). Regions that contributed the greatest weight in discriminating Faces vs. Bodies included visual cortical regions, as well as some subcortical (bilateral amygdala, right thalamus, and hippocampus) and supramodal cortical (left insula and right retrosplenial cortex) areas (Figure 4.2b). Chi-square tests revealed no significant Chi-square tests revealed no significant differences in the proportion of correct classifications of the Fear and Neutral stimuli in any of these models.

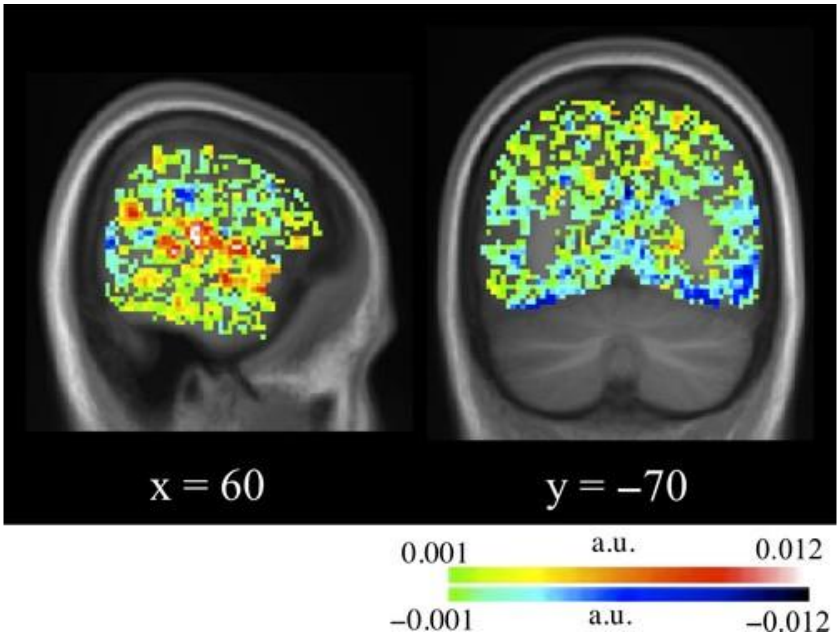

Figure 4.1 Weight map of positive (red) and negative (blue) weights depicting the clusters that correspond to the classification of Auditory vs. Visual conditions. A.U., arbitrary units.

Table 4.1 Ranking of the anatomically defined clusters (having a cluster size >5 voxels) that contributed to the top 1% of the weight contribution in the above-chance classification accuracies. Peak coordinates of cluster, percent occurrence in 1,000 permutations, and cluster size reported for each anatomical location

Auditory vs. Visual	**Accuracy** = 93.54 (p = 0.001) **AUC** = 0.98							**Sensitivity** = 92.31% **Specificity** = 94.85%					
Anatomical Location	**Positive**						**Anatomical Location**	**Negative**					
	L/R	**x**	**y**	**z**	**%**	**K_E**		**L/R**	**x**	**y**	**z**	**%**	**K_E**
Sup.Temporal Gyrus/Sulcus	R	60	-16	10	99.4	193	Fusiform Gyrus	R	32	-48	-18	98.7	23
	L	-56	-24	10	99.7	136		R	38	-60	-18	99.3	21
	R	62	-34	6	99.7	35		R	34	-70	-14	98.8	13
	R	58	-12	-6	99.5	18		R	18	-62	-6	99.7	10
Planum Polare	R	50	0	0	98.9	47		L	-22	-78	-12	100	7
	R	56	0	2	99	11		L	-36	-54	-12	99.7	6
Post. Insula	R	42	-14	0	99.1	13		R	32	-54	-12	98.7	6
Planum Temporale	R	66	-28	0	99.2	11	Cuneus	R	14	-100	14	99.5	18
Temporoparietal Junction	R	58	-52	22	99.9	10	Post. Mid. Temporal Gyrus	R	54	-64	18	99.9	11
	R	36	-30	18	99.6	7	Putamen	L	-16	14	-10	98.8	10
Heschl's Gyrus	R	44	-26	14	99.3	8		R	12	12	-10	98	8
	R	46	-24	4	99.9	6		L	-28	6	-6	98.7	6
Thalamus	R	14	-26	14	97.6	7		L	-20	-10	-12	98	6
							Hippocampus	R	24	-32	-14	96.3	8
							Ant. Orbitofrontal Cortex	L	-42	54	-14	100	7
							Lingual Gyrus	R	28	-72	-14	99.6	7
							Mid. Frontal Gyrus	R	24	52	32	100	6

							Precuneus	R	16	-66	26	100	6
							Lat. Occipital Cortex	R	50	-68	-4	99.7	6
							Amygdala	L	-22	0	-16	96.8	6
Faces vs. Bodies		**Accuracy** = 85.00 (p=0.001) **AUC** = 0.92						**Sensitivity** = 84.43% **Specificity** = 85.59%					
Lat. Occipital Cortex	R	48	-72	4	100	393	Calcarine Cortex	R	4	-90	2	99.8	370
	L	-48	-76	6	100	217		L	-2	-84	-6	99.1	10
	L	-52	-68	-8	100	6	Lingual Gyrus	R	14	-40	0	96.5	33
Sup. Occipital Gyrus	R	26	-92	18	99.6	11		R	14	-82	-12	99.5	25
Thalamus	R	18	-26	6	98.6	11	Hippocampus	R	16	-4	-16	96.2	14
Mid. Occipital Gyrus	L	-36	-88	0	100	7	Amygdala	L	-22	-10	-12	97	10
	R	34	-88	10	99.9	7		R	18	2	-16	95.7	9
	R	30	-90	0	99.8	6	Retrosplenial Cortex	R	16	-54	8	98.2	7
Post. Insula	L	-40	-10	10	97.9	7	Mid. Occipital Gyrus	R	18	-98	10	99.7	6
Mid.Temporal Pole	L	-30	0	-34	98.9	6							
Prosody vs. Vocalizations		**Accuracy** = 81.25 (p=0.001) **AUC** = 0.89						**Sensitivity** = 86.41% **Specificity** = 77.37%					
Sup.Temporal Gyrus/Sulcus	L	-60	-14	0	99.7	741	Heschl's Gyrus	R	60	-24	14	98.8	107
	R	60	-6	-2	99.4	356		L	-42	-28	10	99.3	6
	R	60	-28	-2	99.6	94	Planum Polare	R	42	-6	-6	98.6	32
	L	-64	-24	12	99	7	Insula	R	30	10	-18	94.8	28
Supplementary Motor Area	R	14	2	76	94.2	6		L	-36	10	-18	93.9	8
								R	30	2	-32	98.9	7
							Post. Sup. Temporal Sulcus	R	60	-40	-10	99.8	18

							Amygdala	R	16	2	-18	96.2	12
Fear vs. Neutral		**Accuracy** = 56.04 (p=0.017) **AUC** = 0.58						**Sensitivity** = 56.87% **Specificity** = 55.39%					
Amygdala	L	-22	-4	-14	96.7	21	Fusiform Gyrus	L	-32	-40	-16	98.8	16
	R	20	-2	-12	95.4	14	Insula	L	-44	0	-6	97.5	14
Planum Polare	R	38	12	-18	97	17		R	42	4	-8	97.9	11
Sup. Temp. Pole	R	52	4	-4	99.3	13	Putamen	L	-32	-6	2	98.1	12
	L	-46	6	-14	98.1	7	Precuneus	L	-10	-42	6	96.7	10
Mid. Temporal Pole	R	36	14	-32	97.8	9	Premotor Cortex	L	-52	8	42	99.9	9
	R	34	4	-34	97.8	9		R	10	-22	16	97.1	8
Hippocampus	L	-26	-26	12	97.3	9	Thalamus	R	18	-22	-20	94	8
Heschl’s Gyrus	L	-46	-14	8	98.8	8		R	6	-26	8	93.4	6
Thalamus	L	-22	-26	12	98.7	8	Lingual Gyrus	R	10	-40	-2	94.9	7
	R	10	-24	2	97.6	7							
Post. Cingulate Gyrus	L	-12	-46	10	99.3	7							
Insula	R	34	4	12	96.6	7							
Ventromedial Orbitofrontal Gyrus	L	-12	12	-14	97.9	6							

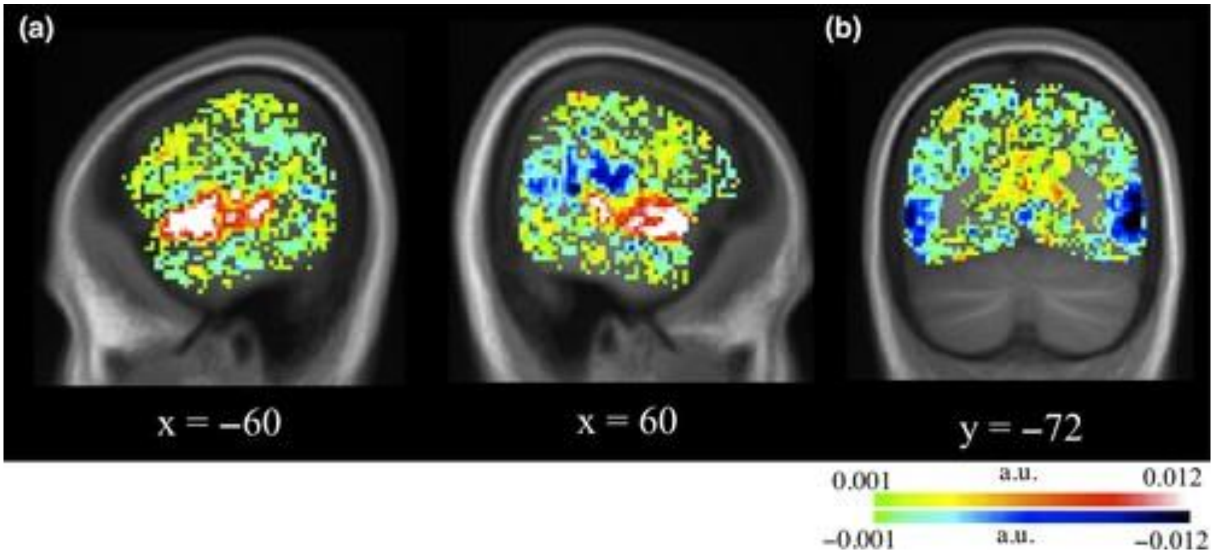

Figure 4.2 Weight map of positive (red) and negative (blue) weights depicting the clusters that correspond to the classification of (a) Prosody vs. Vocalizations conditions, (b) Faces vs. Bodies conditions. A.U, arbitrary units

Above-chance accuracy was observed in classifying Fear vs. Neutral across modalities and categories, with similar levels of sensitivity (56.9%) and specificity (55.4%). The classification of Fear and Neutral was above chance for all conditions separately except Vocalizations, which yielded a sensitivity and specificity of only 50% (Table 4.2). Chi-square tests did not reveal any significant differences in the frequency of correct classifications for Fear vs. Neutral stimuli between (a) Auditory vs. Visual information (b) Faces vs. Bodies, or (c) Vocalizations vs. Prosody. The weight map showed that the most discriminative voxels were located primarily in subcortical structures, particularly the dorsomedial amygdala (Figure 4.3), as well as the thalamus, putamen, and the left hippocampus. Cortical regions with significant weights included the insula, ventromedial orbitofrontal cortex, and posterior cingulate cortex. Additional clusters were located in visual and auditory cortical regions, as well as premotor cortex.

Table 4.2 The deconstructed classification analysis of fear vs. neutral conditions, with classification accuracies as percentages, corresponding to each of the four domains.

Emotional domains	**Sensitivity (%)**	**Specificity (%)**
Faces	63.46	60.29
Bodies	62.75	56.25
Speech	57.78	54.67
Voice	50.00	50.00

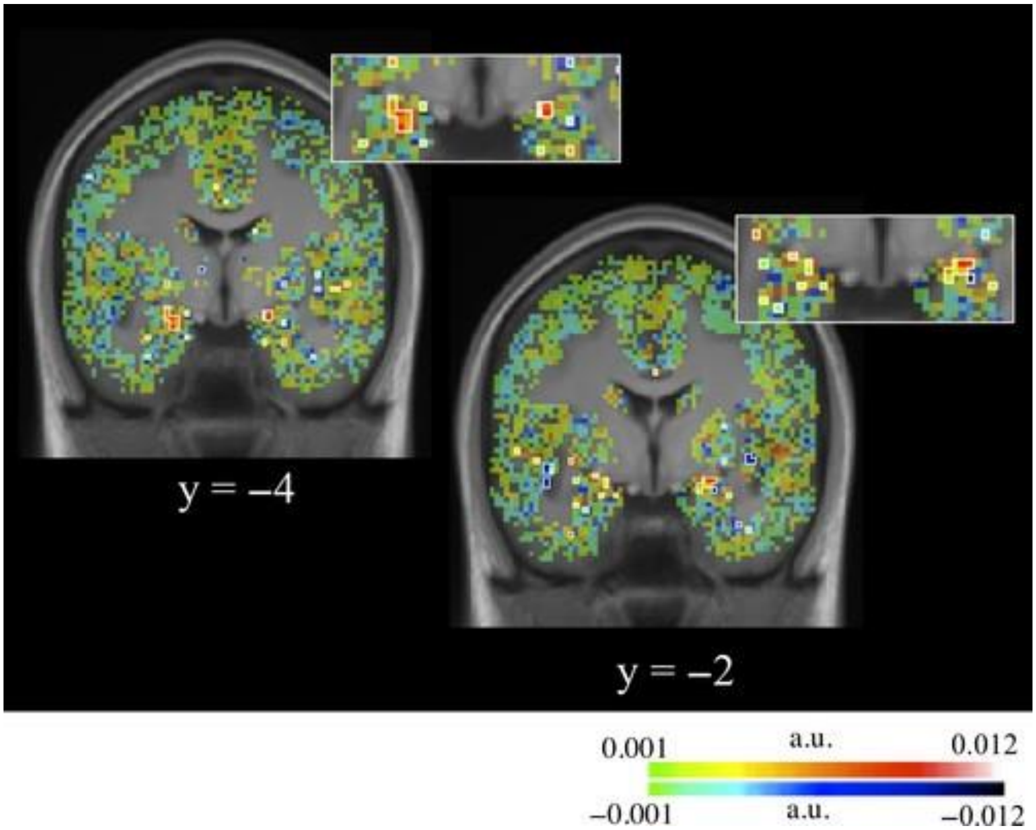

Figure 4.3 Weight map of positive (red) and negative (blue) voxel-wise weight contribution in the classification of Fear vs. Neutral conditions, across Faces, Bodies, Prosody, and Vocalizations. Highlighted region depicts contribution of the top 1% positive and negative weights, with a minimum extent of five contiguous voxels, located primarily in the bilateral amygdala. A.U., arbitrary units

Interestingly, the model performance was not significantly above chance in discriminating fearful from neutral stimuli when restricting the search space to the amygdala alone (Accuracy = 48.3%; Accuracy p-value = 0.74; AUC = 0.50; Sensitivity = 48.3%; Specificity = 48.3%) nor for an ROI of the limbic system (amygdala, hippocampus, cingulate gyrus, parahippocampal gyrus; Accuracy = 51.7%; Accuracy p-value = 0.28; AUC = 0.51; Sensitivity = 51.7% Specificity = 51.7%]). Restricting the voxel space to ROIs of equal size to that of the amygdala (Left: 211 voxels; Right: 240 voxels), but centered in other regions with high-weight value (hippocampus, insula, and ventromedial orbitofrontal gyrus) also failed to yield above-chance accuracy in the fear vs. neutral classification (*p*'s > 0.3). To further address whether this below-chance classification of the amygdala alone simply reflected fewer voxels available for classification masks of regions within the bilateral (a) superior temporal gyrus and sulcus, and (b) fusiform gyrus of equal size as the amygdala mask was used for comparison within the classification of Auditory vs. Visual modalities. Both models performed significantly above chance in discriminating Auditory from Visual within the superior temporal gyrus and sulcus alone (Accuracy = 67.5%; Accuracy p-value = 0.001; AUC = 75; Sensitivity = 66.7% Specificity = 68.4%), the fusiform alone (Accuracy = 78.5%; Accuracy p-value = 0.001; AUC = 0.86; Sensitivity = 77.7% Specificity = 79.4%), suggesting that a small ROI is capable of supporting accurate classification of a more localized stimulus representation, such as that based on sensory modality.

4.6 Discussion

The objective of the current study was to identify common patterns in the processing of fearful emotional expressions across different sensory modalities and stimulus categories. As expected, the multivariate model was very accurate for modality- and category-based classification. Moreover, the analysis yielded significant above-chance classification accuracy in distinguishing fear vs. neutral expressions, involving several cortical and subcortical regions, notably the amygdala. These results seem to point to the existence of a common, or at least overlapping, distributed network of cortical and subcortical regions involved in processing fear expressed with faces, bodies, prosody, and non-linguistic vocalizations.

4.6.1 Distinguishing sensory modalities and categories

The multivariate model yielded significantly accurate classification of subjects' responses based on stimulus modality (visual vs. auditory) and category (faces vs. bodies and prosody vs. vocalizations). The regions with the largest weights were consistent with the corresponding modality- and category-preferred areas previously reported in the literature, such as the fusiform gyrus (Kanwisher, McDermott, & Chun, 1997), lateral occipital cortex (Downing, Jiang, Shuman, & Kanwisher, 2001), and superior temporal gyrus and sulcus (Pernet et al., 2015). Interestingly, additional subcortical and supramodal cortical regions also contributed to the accurate classification of stimuli, based on their modality and category. Some of these multimodal regions were also observed in the emotion-based classification analysis (see discussion below), and could reflect differences in the perception of social information across modalities, regardless of emotional content. For instance, the temporoparietal junction, a polymodal association area, has been shown to be involved in higher-level social processing (Decety & Lamm, 2007) and in attention processing, particularly when based on stimulus saliency defined by attributes such as behavioral relevance, stimulus novelty, or frequency of its presentation (Downar, Crawley, Mikulis, & Davis, 2001, 2002). In the current study, recruitment of the right temporoparietal junction could have differed between auditory and visual modalities due to variability of stimulus saliency across modalities. Additionally, in the classification of auditory stimuli regardless of emotion, contributions of the supplementary motor area were observed. This region has been shown to respond to speech and non-linguistic vocalizations, and is suggested to play a role in connecting auditory perception and production (Lima, Krishnan, & Scott, 2016).

4.6.2 Distinguishing fear vs. neutral expressions

The multivariate analysis allowed for the accurate classification of neural response to Fear vs. Neutral expressions, regardless of stimulus category or modality, through clusters primarily located in subcortical and multimodal- and/or higher-processing regions, as well as some unimodal visual and auditory processing regions. The high weight contribution of the amygdala to the classification of fear vs. neutral (upper 1% of the entire brain) is consistent with the strong evidence in the literature for a significant role of this structure in the processing of fear-related information across different types of stimuli and modalities (Armony, 2013; Zald, 2003). It should be noted that despite the prominent contribution of the amygdala to the classification of fear vs. neutral stimuli, the classification accuracy obtained using data restricted to the amygdala or the limbic network (including the amygdala, hippocampus, cingulate gyrus, and parahippocampal gyrus) was not significantly above chance, suggesting that fear processing likely requires a distributed contribution of cortical and subcortical structures (Lindquist et al., 2012; Wager et al., 2015). This conclusion is also supported by studies of emotion induction, in which multivariate approaches were used to illustrate a distributed emotion representation across the brain. For instance, Saarmäki and colleagues demonstrated above-chance classification accuracies for 12 emotions in response to guided mental imagery. Most notably, contributions from regions in the insula, supplementary motor area, and subcortical structures were identified in discriminating four negative emotions including fear from neutral (Saarimäki et al., 2018). Additionally, activity within- and across-subjects in response to five emotions, as induced through movies was accurately classified above chance for all emotions, with contributions from parietal and frontal midline regions, and in particular, from the insula and amygdala for fear (Saarimäki et al., 2016).

Additional regions that contributed to the accurate classification of fear expressions also agree with the existing literature. For instance, a meta-analysis of neuroimaging studies of emotional perception (Wager et al., 2015) revealed a strong co-activation between the basal ganglia and the thalamus, and with the amygdala in response to fear. Moreover, the insula, an integral part of the emotion-processing network, has bilateral connections with the thalamus; in addition, its anterior portion also has reciprocal connections with the amygdala (for review, see Lévêsque, 2014).

Although usually associated with memory, the hippocampus is shown to play a role in emotional processing, particularly through its regulatory role over the amygdala. For example, the hippocampus engages in regulating the stress response (Lopez, Akil, & Watson, 1999). Moreover, Kensinger and Corkin (2004) observed strengthened connections between the hippocampus and amygdala during stimulus encoding of salient information.

Contributions in the classification of fear vs. neutral stimuli from the posterior ventromedial orbitofrontal cortex may reflect its direct connections with anterior temporal sensory association areas, as well as indirect connections with the amygdala (for review see, Barbas, 2007). Through this circuitry, the posterior orbitofrontal cortex is known to play a role in regulating the response of the amygdala (for review see, Kim et al., 2011). Moreover, work by De Gelder et al. (2004) showed that the orbitofrontal cortex acts cohesively with the amygdala and insula, collectively showing a stronger response to the presentation of fearful body expressions than neutral or happy ones. This effect was seen in parallel with elevated activity in the posterior cingulate and retrosplenial cortices, the nucleus accumbens, and most notably, the premotor cortex, a prominent player in the mirror neuron circuit (Rajmohan & Mohandas, 2007). The authors suggest that the concomitant involvement of emotion-processing circuits and those related to action representation may illustrate an action-oriented preparation in response to the fearful information (De Gelder et al., 2004). This could explain the high weight contribution of the premotor cortex we observed in this study. The posterior cingulate, specifically the retrosplenial cortex, was another supramodal structure found to contribute to the accurate classification of fear stimuli. This region receives input from the prefrontal cortex, superior temporal sulcus, precuneus, thalamus, and claustrum, and has been shown to consistently activate in response to emotionally salient information (Maddock, 1999). Activity in the precuneus is frequently associated with higher-order cognitive processing, as reported in tasks requiring subjects to self- process and incorporate concepts of empathy into judgment-making (see for review, Cavanna & Trimble, 2006). Lastly, several unimodal-processing structures also exhibited large weights. These regions included the fusiform and lingual gyrus, as well as the planum polare, a region located in the anterior superior temporal gyrus. These weights likely represent the flow of information from unimodal sensory areas to the multi- or supramodal regions, described above, involved in emotional processing. In summary, the multivariate analysis revealed a large, yet restricted network of cortical and subcortical regions necessary for the accurate classification of fearful expressions, in agreement with the literature that argues for a distributed and multifaceted model of emotional processing (e.g., Bush, Inman, Hamann, Kilts, & James, 2017; Lindquist et al., 2012; Wager et al., 2015).

4.6.3 Limitations and future directions

One limitation of the study was the use of only one emotion. This was done to maximize statistical power (in terms of number of trials per condition) while keeping the experiment short enough to maintain the subjects' attention (we chose passive perception to avoid any potential confounding effects of task-related processes). Nonetheless, it would be interesting to determine whether emotions of different valence and arousal elicit the same patterns and trends in model performance, or whether the observed pattern is specific to fear. If additional emotions were to be included in the experimental design, one could make use of other multivariate approaches, such as representational similarity analysis (RSA), which is particularly useful for condition-rich designs (Kriegeskorte, Mur, & Bandettini, 2008). Moreover, although we employed a passive perception design, it would be of interest to explore the potential influence of a task, particularly by comparing incidental and emotion-related ones, in order to study potential differences between explicit and implicit emotion perception. This would also allow for the investigation of the relation between subjective perception (e.g., through online emotion categorization or valence/intensity ratings) and brain representation of emotional information across modalities and categories.

Finally, it must be recognized that the observed results do not reflect a direct representation of how different regions of the brain respond to fearful expressions. Instead, the multivariate approach highlights areas that strongly contribute to an accurate distinction across conditions.

This distinction is dependent on the varying levels of activity across the whole brain, not just a discrete region. Therefore, statements cannot be made about a significant neural response to the stimulus within, for example, the amygdala, but, rather, that the amygdala contributes information to the classification, contingent on the varying pattern of activity across the brain, and thus involved in a subtler way (Haufe et al., 2014).

4.7 Acknowledgments

We are grateful to Dr. Jessica Schrouff for invaluable help implementing the multivariate analysis in PRoNTo. This study was supported by grants from the Natural Science and Engineering Research Council of Canada (NSERC, RGPIN-2017-05832) and the Canadian Institutes of Health Research (CIHR, MOP-130516) to JLA. JCW was funded by an NSERC- Create ACN Graduate Fellowship, a CRBLM Graduate Student Stipend, a McGill's IPN Returning Student Award, an NSERC - Canadian Graduate Scholarship -Masters (CGS-M) fellowship and an Alexander Graham Bell Canada Graduate Scholarship-Doctoral fellowship (CGS-D).

4.8 Conflict of interest

The authors declare no conflict of interest.

4.9 Data accessibility

Data are available upon request to the corresponding author.

4.10 Authors’ contributions

J.C.W and J.L.A design the research. J.C.W. and J.L.A conducted the research. J.C.W. analyzed the results. J.C.W. drafted the manuscript. J.C.W and J.L.A reviewed and commented the first draft. J.L.A. read and approved the final manuscript.

4.11 References

Adank, P. (2012). Design choices in imaging speech comprehension: An activation likelihood estimation (ALE) meta-analysis. *Neuroimage, 8,* 360-369.

Aleman, A., & Swart, M. (2008). Sex differences in neural activation to facial expressions denoting contempt and disgust. *PLoS One, 3,* e3622.

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.

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 nonmusicians. *Cortex, 59*, 126-137.

Armony, J. L. (2013). Current emotion research in behavioural neuroscience: The role(s) of the amygdala. *Emotion Review, 5,* 104-115.

Armony, J. L., Chochol, C., Fecteau, S., & Belin, P. (2007). Laugh (or cry) and you will be remembered: Influence of emotional expression on memory for vocalizations. *Psychological Science, 18,* 1027-1029.

Aubé, W., Angulo-Perkins, A., Peretz, I., Concha, L., & Armony, J. L. (2015). Fear across the senses: Brain responses to music, vocalizations and facial expression. *Social Cognitive and Affective Neuroscience, 10,* 399-407.

Aubé, W., Peretz, I., & Armony, J. L. (2013). The effects of emotion on memory for music and vocalizations. *Memory, 21*, 981-990.

Barbas, H. (2007). Flow of information for emotions through temporal and orbitofrontal pathways. *Journal of Anatomy, 211,* 237-249.

Bilalić, M. (2016). Revisiting the role of the fusiform face area in expertise. *Journal of Cognitive Neuroscience, 28,* 1345-1357.

Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., Strauss, M.M., Hyman, S.E., Rosen, B. R. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron, 17,* 875-887.

Burges, C. J. C. (1998). *A tutorial on support vector machines for pattern recognition*. Boston, MA: Kluwer Academic Publishers.

Bush, K. A., Inman, C. S., Hamann, S., Kilts, C. D., & James, G. A. (2017). Distributed neural processing predictors for multi-dimensional properties of affect. *Frontiers in Human Neuroscience, 11*, 1-14.

Canli, T., Sivers, H., Whitfield, S. L., Gotlib, I. H., & Gabrieli, J. D. (2002). Amygdala response to happy faces as a function of extraversion. *Science, 296,* 2191.

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioural correlates. *Brain, 129,* 564-583.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 27*, 1-27.

Davis, T., LaRocque, K. F., Mumford, J., Norman, K. A., Wagner, A. D., & Poldrack, R. A. (2014). What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. *Neuroimage, 15,* 271-283.

Gelder, B., Snyder, J., Greve, D., Gerard, G., & Hadjikhani, N. (2004). Fear fosters flight: A mechanism for fear contagion when perceiving emotion expressed by a whole body. *Proceedings of the National Academy of Sciences, 101*, 16701-16706.

Gelder, B., & Stock, J. (2011). The Bodily Expressive Action Stimulus Test (BEAST). Constructions and validation of a stimulus basis for measuring perception of whole body expression of emotions. *Frontiers in Psychology, 2,* 181.

Decety, J., & Lamm, C. (2007). The role of the right temporoparietal in social interaction: How low level computational processes contribute to metacognition. *Neuroscientist*, *13*, 580-593.

Dickie, E. W., & Armony, J. L. (2008). Amygdala responses to unattended fearful faces: Interaction between sex and trait anxiety. *Psychiatry Research Neuroimaging, 162,* 51-57.

Downar, J., Crawley, A. P., Mikulis, D. J., & Davis, K. D. (2001). The effect of task-relevance on the cortical response to changes in visual and auditory stimuli: An event-related fMRI study. *Neuroimage, 14,* 1256-1267.

Downar, J., Crawley, A. P., Mikulis, D. J., & Davis, K. D. (2002). A cortical network sensitive to stimulus salience in a neutral behavioural context across multiple sensory modalities. *Journal of Neurophysiology, 87*, 615-620.

Fecteau, S., Belin, P., Joanette, Y., & Armony, J. L. (2007). Amygdala responses to nonlinguistic emotional vocalizations. *Neuroimage, 36,* 480-487.

Gardumi, A., Ivanov, D., Hausfeld, L., & Valente, G. (2016). The effect of spatial resolution on decoding accuracy in fMRI multivariate pattern analysis. *Neuroimage, 132*, 32-42.

Glasberg, B. R., & Moore, B. C. J. (2002). A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society, 50*, 331-342.

Grèzes, J., Pichon, S., & Gelder, B. (2007). Perceiving fear in dynamic body expressions. *Neuroimage*, *35,* 959-967.

Hadjikhani, N., & Gelder, B. (2003). Seeing fearful body expression activates the fusiform cortex and amygdala. *Current Biology, 13,* 2201-2205.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J. D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage, 87,* 96-110.

Herrmann, M. J., Boehme, S., Becker, M. P. I., Tupak, S. V., Guhn, A., Schmidt, B., Brinkman, L., Straube, T. (2016). Phasic and sustained brain responses in the amygdala and the bed nucleus of the stria terminalis during threat anticipation. *Human Brain Mapping, 37,* 1091-1102.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302-4311.

Kensinger, E. A., & Corkin, S. (2004). Two routes to emotional memory: Distinct neural processes for valence and arousal. *Proceedings of the National Academy of Sciences, 101,* 3310-3315.

Kim, J. M., Loucks, R. A., Palmer, A. L., Brown, A. C., Solomon, K. M., Marchante, A. N., & Whalen, P. J. (2011). The structural and functional connectivity of the amygdala: From normal emotion to pathological anxiety. *Behavioral Brain Research, 223,* 403-410.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience, 2,* 1-28.

Lévêsque, M. (2014). The neuroanatomy of emotions. In M.Lévêsque (Ed.), Psychosuregery: *New techniques for brain disorders* (pp. 49-106). Basel, Switzerland: Springer International Publishing.

Liebenthal, E., Silbersweig, D. A., & Stern, E. (2016). Language, tone and prosody of emotions: Neural substrates and dynamics of spoken-word emotion perception. *Frontiers in Neuroscience, 10,* 1-13.

Lima, C. F., Krishnan, S., & Scott, S. K. (2016). Roles of supplementary motor areas in auditory processing and auditory imagery. *Trends in Neurosciences, 39,* 527-542.

Lindquist, K. A., Wager, T. D., Kober, G., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences, 35*, 121-202.

Lopez, F., Akil, H., & Watson, S. J. (1999). Neural circuits mediating stress. *Biological Psychiatry, 450*, 111-123.

Lundqvist, D., Flykt, A., & Öhman, A. (1998) *The Karolinska Directed Emotional Faces - KDEF*, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630-7164-9.

Maddock, R. J. (1999). The retrosplenial cortex and emotion: New insights from functional neuroimaging of the human brain. *Trends in Neurosciences, 22,* 310-316.

Maldjian, J. A., Laurienti, P. J., Kraft, R. A., & Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage, 19,* 1233-1239.

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., & Dolan, R. J. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature, 383,* 812-815.

Peelen, M. V., Atkinson, A. P., Andersson, F., & Vuilleumier, P. (2007). Emotional modulation of body-selective visual areas. Social Cognitive and Affective Neuroscience, 2, 273-283.

Peelen, M. V., Atkinson, A. P., & Vuilleumier, P. (2010). Supramodal representation of perceived emotions in the human brain. *Journal of Neuroscience, 30*, 10127-10134.

Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics, 37*, 417-435.

Pernet, C. R., McAleer, P., Latinus, M., Gorgolewski, K. J., Charest, I., Bestelmeyer, P. E., Watson, R.H., Fleming, D., Crabbe, F., Valdes-Sosa, M,. & Belin, P. (2015). The human voice areas: Spatial organization and inter-variability in temporal and extra-temporal cortices. *Neuroimage, 119,* 164-174.

Phillips, M.L., Young, A.W., Scott, S.K., Calder, A.J., Andrew, C., Giampietro V., Williams, S.C., Bullmore, E.T., Brammer, M., & Gray, J.A. (1998). Neural responses to facial and vocal expressions of fear and disgust. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 265,* 1809-1817.

Pichon, S., Gelder, B., & Grèzes, J. (2009). Two different faces of threat. Comparing the neural systems for recognizing fear and anger in dynamic body expressions. *Neuroimage, 47,* 1873-1883.

Rajmohan, V., & Mohandas, E. (2007). Mirror neurons system. *Indian Journal of Psychiatry, 49,* 66-69.

Rigoulot, S., Pell, M. D., & Armony, J. L. (2015). Time course of the influence of musical expertise on the processing of vocal and musical sounds. *Neuroscience, 290,* 175-184.

Saarimäki, H., Ejtehadian, L. F., Glerean, E., Jääskeläinen, I. P., Vuilleumier, P., Sams, M., & Nummenmaa, L. (2018). Distributed affective space respresents multiple emotion categories across the human brain. *Social Cognitive and Affective Neuroscience, 13,* 471-482.

Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I. P., Lampien, J., Vuilleumier, P., Hari, R., Sams, M., & Nummenmaa, L. (2016). Discrete neural signatures of basic emotions. *Cerebral Cortex, 26,* 2563-2573.

Sachs, M. E., Habibi, A., Damasio, A., & Kaplan, J. T. (2018). Decoding the neural signatures of emotions expressed through sound. *Neuroimage, 174,* 1-10.

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: An evolved system for relevance detection. *Reviews in Neuroscience, 14,* 303-316.

Schrouff, J., Cremers, J., Garraux, G., Baldassarre, L., & Mourao-Miranda, J. (2013). Localizing and comparing weight maps generated from linear kernel machine learning models. In *2013 3rd International Workshop on Pattern in Neuroimaging (PRNI2013)* (pp. 124-127) Philadelphia, PA: IEEE Computer Society Conference Publishing Services.

Schrouff, J., Monteiro, J. M., Portugal, L., Rosa, M. J., Phillips, C., & Mourão-Miranda, J. (2018). Embedding anatomical or functional knowledge in whole brain multiple kernel learning models. *Neuroinformatics, 16,* 117-143.

Schrouff, J., Rosa, M. J., Rodina, J. M., Marquand, A. F., Chu, C., Ashburner, J., Phillips, C., Richiardi, J., Mourao-Miranda, J. (2013). PRoNTo: Pattern recognition for neuroimaging toolbox. *Neuroinformatics, 11,* 319-337.

Sergerie, K., Lepage, M., & Armony, J. L. (2006). A process-specific functional dissociation of the amygdala in emotional memory. *Journal of Cognitive Neuroscience, 18,* 1359-1367.

Sergerie, K., Lepage, M., & Armony, J. L. (2007). Influence of emotional expression on memory recognition bias: A functional magnetic resonance imaging study. *Biological Psychiatry, 62*, 1126-1133.

Setsompop, K., Gagoski, B. A., Polimeni, J. R., Witzel, T., Wedeen, V. J., & Wald, L.L. (2012). Blipped controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magnetic Resonance in Medicine, 67,* 1210-1224.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joilioy, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage, 15,* 273-289.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the brain: An event-related fMRI study. *Neuron, 30,* 829-841.

Wager, T. D., Kang, J., Johnson, T. D., Nichols, T. E., Satpute, A. B., & Barrett, L. F. (2015). A Bayesian model of category-specific emotional brain responses. *PLoS Computational Biology, 11,* e1004066.

Zald, D. H. (2003). The human amygdala and the emotional evaluation of sensory stimuli. *Brain Research Reviews, 41*, 88-123.

Zwaag, W., Costa, S. E., Zürcher, N. R., Adams, R. B., & Hadjikhani, N. (2012). A 7 Tesla fMRI study of amygdala responses to fearful faces. *Brain Topography, 25,* 125-128.

Connecting Chapters 4 to 5

Study 3 successfully demonstrated an above-chance classification of neural activity in response to fear versus neutral expressions, across modality. Markedly, the amygdala contributed significant weight to the classification, amongst other supramodal structures, including the hippocampus, thalamus, ventromedial orbitofrontal gyrus, precuneus and posterior cingulate gyrus; regions that comprise a validated model of amygdala connectivity during affect perception (Stein et al., 2007). Alike Touroutoglou et al. (2015), we challenge the idea of *basic emotion*-specific networks, where proponents argue for whole-brain patterns to reflect discrete emotion categories (Kragel & LaBar, 2014). Wager et al. (2015) instead offers that emotion-specific patterns largely overlap cortical "resting-state" connectivity networks (Yeo et al., 2011). As in *Study 3*, Diano et al. (2017) emphasizes involvement of the amygdala, acknowledging its capacity to process multiple basic emotions (Kirby & Robinson, 2017) and its multifunctionality from varied network recruitment (Robinson et al., 2010). In investigating psychophysiological interactions (PPI) of the amygdala, authors illustrated dynamic neural signatures of perceived emotions defined by the amygdala and its extensive connections (Diano et al., 2017). Given the role of the amygdala outside of emotion perception (Robinson et al., 2010), it is of interest to identify how this proposed signature may fluctuate alongside task-based executive functioning. Thus, in *Study 4* we explore the interaction of emotion perception and perceptual cognition during an emotion-based, forced-choice decision-making task. Previous work has shown amygdala recruitment during decision-making tasks, when the decision is based on the arousal versus valence of the emotion (Pessoa & Padmala, 2005). Thus, we explore the neural response to anger and fear, two negative emotions exhibiting a comparable arousal (see for review, Mauss & Robinson, 2009) and physiological profiles (Stemmler, 2004). Moreover, we create an ambiguous condition by morphing expressions of fear-anger, to explore how the brain may discriminate ambiguity (Wang et al., 2017). Finally, we examine whether non-clinical anxiety or task-demand, may influence behavioural and neural processes of perception. For example, high anxiety subjects present higher recognition accuracies for fear (Surcinelli et al., 2006), whereby this fear-sensitivity can modulate dependent on perceptual load (Bishop et al., 2007). Exploring these interactions offers a novel approach to investigating supramodal structures, particularly, given the proposed multifunctionality of the amygdala (see review, Gothard, 2020).

Chapter 5. Neural correlates of socio-emotional ambiguity and threat-related processing across modalities

(Study 4)

Jocelyne C. Whitehead[a,b,c,] Ignacio Spiousas[a,b,d,e] and Jorge L. Armony[a,b,d]

[a] Douglas Mental Health University Institute, Verdun, Canada s
[b] BRAMS Laboratory, Centre for Research on Brain, Language and Music, Montreal, Canada
[c] Integrated Program in Neuroscience, McGill University, Montreal, Canada
[d] Department of Psychiatry, McGill University, Montreal, Canada
[e] Department of Science and Technology, National University of Quilmes, Buenos Aires, Argentina

To be submitted.

5.1 Abstract

The amygdala is actively involved in emotion perception, particularly social threat, and may modulate by factors of uncertainty, stress, and anxiety. The current study investigates neural correlates of processing auditory and visual ambiguous threat-related information during decision-making, and how state anxiety may impact these processes. Healthy subjects (*N*=29) were presented with morphed stimuli, created along a fear-anger continuum while performing a two-alternative forced-choice task (fear/anger). Subject-specific psychometric curves were built, and the morph level corresponding to maximal perceived ambiguity calculated. Visual and auditory subject-specific stimuli representing explicit anger, fear and ambiguous threat, were presented as subjects underwent fMRI with a fast (TR=529ms), high-resolution (2 mm^3 isotropic) multiband sequence. Behavioural results demonstrated decision biases for visual fear, and to a lesser degree, auditory anger, while greater task-difficulty was observed for ambiguous stimuli, and imaging results showed increased activity in regions of the salience and frontoparietal control networks and deactivation in those of the default mode network. The right amygdala responded more strongly to explicit stimuli, also modulating in response to ambiguity, favouring the decision biases. Finally, state anxiety correlated with faster response times and greater amygdala recruitment, illustrating exaggerated impulsivity and affinity towards intrinsic biases. Results provide a first look into the neural perception of emotional ambiguity and subject-specific biases of threat perception across modalities, as modulated by non-clinical anxiety.

KEYWORDS: ambiguity; emotion perception; fMRI; state-anxiety, threat

5.2 Introduction

Most often recognized for processing fearful facial expressions (Vuilleumier et al., 2001; Morris et al., 1996; Breiter et al., 1996; Van der Zwaag et al., 2012; Aubé et al., 2015), the amygdala has since shown to process positive and negative affect (for review, see Sergerie et al., 2008) across varying domains and modalities (Whitehead et al., 2018; Aubé et al., 2015; Phillips et al., 1998). Still, several outstanding questions as to the extent and generalizability of its function, remains. Specifically, little is known as to how emotional information with no clear emotional label (i.e., ambiguous affect) may be processed. Neta et al. (2013) are one of few groups that investigated the neural response to perceiving emotional ambiguity in the amygdala. Authors used facial expressions of surprise, interpreted as either positive or negative, presented in a decision-making task of perceived valence. Oftentimes, when evaluating surprise, the judgment depends on the subjective experience of the observer, as influenced by context, personality or even their current mood. To some degree, this can be controlled for by using stimuli of the same perceptual valence. As the amygdala is recognizably responsive to processing threat (Mattavelli et al., 2014), where lesions present pronounced deficits to the recognition of fear and anger in faces (Calder, 1996) and voices (Scott et al., 1997), auditory and visual morphed expressions of fear-anger are used in the current study to explore the neural response to ambiguous emotion. Moreover, as emotion perception is highly subjective, we explore influences of individual differences and the potential for interactions with task-difficulty and emotional salience.

Accounting for subject-specific profiles can help explain differences of behavior and/or underlying neural correlates of perception, particularly in decision-making tasks. For instance, individuals with high non-clinical anxiety have shown an attentional bias towards (Fox et al., 2002), or enhanced sensitivity to (Richards et al., 2002) fearful facial expressions, as well as higher emotion recognition accuracies for fear (Surcineli et al., 2006) and faster rates of responding (Richards et al., 2002). At a neural level, non-clinical anxiety has shown to increase amygdala response when viewing emotional faces (Bishop et al., 2004; Stein et al., 2007), as well as influencing attention-emotion interactions and modulating underlying functional connectivity (see for review, Dolcos et al., 2020). Notably, state and trait anxiety correlate under anxiogenic situations (Leal et al., 2017); however, they are also associated to differential recruit of neural processes (Saviola et al., 2020). Williams et al. (1998) defined trait anxiety as generating the habit to reorient towards threat, while state anxiety; the transient elevation of the autonomic nervous system (Saviola et al., 2020), elevates the value of threat linked to the specific situation or stimulus. Specifically, when subjects undergo functional MRI, they may experience uneasiness and fear in anticipation of being in the scanner, due to reasons such as claustrophobia, magnetophobia or disconcerting MRI sound exposure (Munn and Jordan, 2011). These subjects often report higher levels of state anxiety, reportedly dependent upon genetic composition (Mutschler et al., 2014). Markedly, those with high state anxiety have shown greater connectivity with regions of the Salience Network (SN; Menon, 2015), including the insula, planum polare, precentral gyrus, and posterior and Default Mode Network (DMN; e.g., precuneus, posterior cingulate gyrus; Saviola et al., 2020; Dennis et al., 2011). This is particularly important, as these networks are readily recruited during decision-making tasks alongside the central executive or frontoparietal control network (FPCN; Chand and Dhamala, 2016). Notably, during these tasks, behavioural performance has been altered by anxiety (see for review Hartley and Phelps, 2012). Thus, specific attention should be paid to modulation of "sub-networks" comprising the *task-positive* (SN and FPCN activation) and *task-negative* (DMN deactivation) networks (Fox et al., 2005; Fransson, 2005; Di & Biswal, 2014). Additionally, as bidirectional connections between the insula and amygdala have been proposed as a plausible biomarker for state anxiety (Baur et al., 2013), identifying common patterns across the regions, as mediated by state anxiety, would be imperative. Particularly given that the insula integrates information from salient stimuli to modulate the internal body state (Paulus and Stein, 2006). Thus, through a forced-choice emotion decision-making task with threatening and ambiguous stimuli (fear-anger morphs), we examine individual differences of transient state anxiety, given a particular focus on the role of the amygdala.

Neta et al. (2013) reported increased signal change in the SN and ventral amygdala when processing facial expressions of surprise. Interestingly, differences were not observed in the dorsal amygdala, similar to Kim et al. (2003), who instead used a passive-viewing versus task-based paradigm. A meta-analysis of PET and fMRI studies suggested that inhibition of the amygdala may be explained by a greater demand for cognitive processing; often associated with frontal activity apparent in task-based designs (Costafreda et al., 2008). Of a similar nature, Wyble et al.'s (2008) neural network model illustrates how limited attentional resources are modulated between the cognitive demands of a task and the emotional salience of the stimulus. Thus, we use this principle in investigating the neural response to morphed stimuli presenting graded levels of cognitive demand (i.e., ambiguity decision-making) and emotional salience. Response times are used as an indirect evaluation of task-difficulty, acting as a marker for cognitive load (Deupree and Simon, 1963), while responses to equal-morph stimuli (50% fear/50% anger) may illustrate influences of individual biases and/or sensitivities to a specific emotion (El Zein et al., 2015). The study equates and contrasts within-modality (auditory; visual) responses to two forms (anger/direct; fear/indirect) of threatening social information. Using this approach, we can identify potential neural deviations when stimuli are physically identical (i.e., subjectively *ambiguous* condition), but perceived differentially (i.e., fearful or angry), in addition to differentiating the perception of 'ambiguous' versus evidently threatening expressions (e.g., 100% fear or 100% anger).

Uniquely, we created subject-specific stimuli using psychometric curves calculated prior to functional imaging acquisition, in a two-alternative forced choice task (2AFC; fear or anger). This approach identifies the *point of subjective equality* (PSE, i.e., central point of symmetry in the curve (Kingdom and Prins, 2016)), or rather, the stimulus that is perceived as maximally ambiguous (labeled fear or anger 50% of the trials). Most studies that use this approach average PSEs across subjects. Instead, creating subject-specific stimuli maintains a uniform degree of perceived ambiguity, and accounts for individual differences. Uniquely, we explore emotion perception in different modalities, given recent findings for distributed neural processing of fear irrespective of domain (expressed by faces, bodies, non-linguistic vocalizations, pseudospeech; Whitehead and Armony, 2018). Thus, we use non-linguistic vocalizations and faces due to the reliable neural response of the amygdala when viewing threatening facial expressions (Mattavelli et al., 2004), and their close link to vocalizations (i.e., the 'auditory face' (Belin et al., 2004)). We implemented complementary univariate and data-driven multivariate pattern analyses to illustrate how the brain as a whole, may process ambiguous emotion. Finally, using high spatial-and temporal-resolution fMRI will support imaging the potentially subtle differences of graded threat perception, as well as subcortical structures, such as the amygdala.

5.3 Materials and methods

5.3.1 Participants

Twenty-nine right-handed, healthy volunteers (age: M = 25 years old, SD = 3.2; 16 female) with no self-reported history of neurological or psychiatric illness, and normal audition and normal or corrected-to-normal vision, participated in the fMRI study. One male subject was removed due to poor behavioral performance, while three subjects (2 female) were removed as they scored above normative values for their corresponding age group in measures of state anxiety (Spielberger, 1995). Subjects provided written informed consent prior to participation and received financial compensation at the end of the experiment. The study was approved by the McGill University Faculty of Medicine Research Ethics Office, in accordance with the Declaration of Helsinki.

5.3.2 Measures of non-clinical state anxiety

Subjects completed the state anxiety subscale (transitory mood; 10-item) of the Spielberger State-Trait Personality Inventory (STPI; Spielberger et al., 1995) before starting the experiment. The STPI, derived from the unidimensional State-Trait Anxiety Inventory (STAI: Spielberger et al., 1983), requires subjects to score the intensity of described feelings (e.g., “I feel nervous”) using a four-point Likert scale (1=Not at All; 2=Somewhat; 3=Moderately So; 4=Very Much So). The state subscale requires subjects to respond based on feelings experienced *right now, at this very moment.* Results were used to relate subject-specific anxiety levels to behavioral and neuroimaging findings of the current study. The reliability and validity of the STPI has been confirmed (see for review, Spielberger and Reheiser, 2009; Boyle et al., 2015), and has been previously used as a measure of anxiety in healthy subjects (e.g., Matthews et al., 2011; Smith et al., 2007; Loebach Wetherell et al., 2002).

5.3.3 Stimuli

Vocal stimuli consisted of brief emotional interjections (Duration: M = 0.68 s; SD = 0.22 s) using the French vowel *ah* (/a/) produced by one male and one female actor (Belin et al., 2008). Vocal expressions of fear and anger were used as the prototypical emotions from which 99 intermediate steps on a fear-anger continuum were obtained through morphing. Morph stimuli were created using STRAIGHT software (Kawahara et al., 1999), where intermediate morphing steps were obtained through re-synthesis determined by the linear interpolation of time-frequency landmark templates (F0, frequency, and amplitude; Figure 5.1). Stimuli were monaural, but presented binaurally, and resampled to 32 bits, at a sample rate of 44100Hz. Furthermore, stimuli were adjusted for loudness by normalizing to the short-term loudness (STL) maximum using the Moore and Glasberg Loudness model (Glasberg and Moore, 2002), as implemented in the Loudness Toolbox in MATLAB ver. R2015b.

Faces of two individuals (one female) expressing fear and anger were selected from the KDEF database (Lundqvist et al., 1998). Hair around the exterior of the face was removed from the images, which were further altered using Adobe Photoshop CS5.1 (Adobe Systems, San Jose CA) to attain uniform face size, resolution, and contrast (Sergerie et al., 2006; 2007). Similar to auditory stimuli, (Figure 5.1), 99 equally spaced intermediate morph steps along a fear-anger continuum were created using FantaMorph 5 (Abrosoft http://www.fantamorph.com/). Visual stimuli were presented for 1s (determined during pilot testing to yield >95% mean accuracy in classifying the emotion expressed by the original, 100% prototypical emotions).

5.3.4 Pre-scan experimental procedure

5.3.4.1 Subject-specific stimulus selection

Subjects initially completed two-alternative forced-choice tasks outside of the scanner (*pre-scan session*) to determine the morph level corresponding to their subjective perception of emotionally ambiguous stimuli for use during the *scan session* (see below). Participants were presented with voices (auditory run) or faces (visual run) of one identity, using Psychtoolbox-3 (Brainard, 1997; Pelli, 1997) and made a response by pressing the mouse (left or right button) in accordance with the emotion that they perceived to be expressed by the actor. The words "*Fear*" and "*Anger*" were presented on either side of the screen (pseudo-randomized) and subjects were given a maximum of 2s to respond, after which the subsequent stimulus was presented with a brief (~0.5s) delay. The task consisted of two auditory and visual runs, with an additional practice run prior to testing. An option to repeat the practice run was given. The order of runs, in terms of modality and stimulus identity was counterbalanced across subjects. Subjects received verbal and written instructions for the task through Psychtoolbox-3.

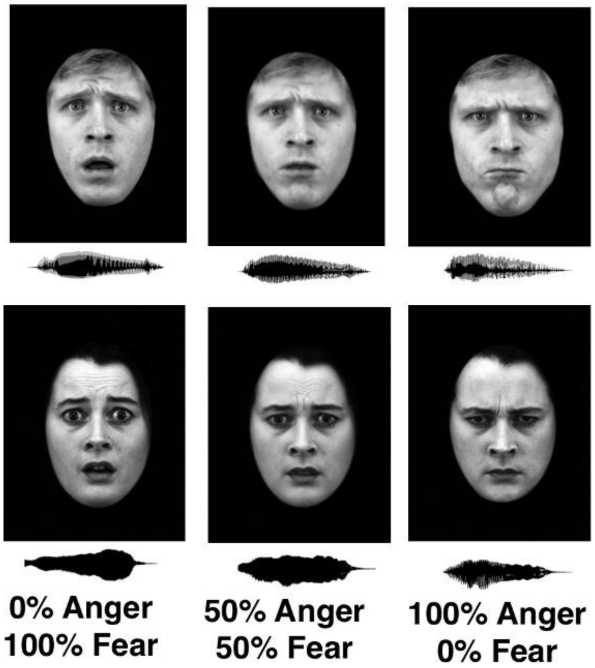

Figure 5.1 Samples of female and male morph stimuli expressing 0%, 50% and 100% anger, where auditory morphs are visualized by the shape and texture of their sound wave.

In the pre-scan session, morph steps were presented on a scale of percent-anger from 0% anger (i.e., 100% fear) to 100% anger (i.e., 0% fear). In each unimodal run, 17-morph steps (0%, 10%, 20%; 25%; 30%; 35%; 40%; 45%; 50%; 55%; 60%; 65%; 70%; 75%; 80%; 90%; 100% anger) were presented 15 times each, except for the continuum extremes (0% and 100% anger: 7 times, 10% and 90% anger: 8 times). Differences of presentation numbers for these morphs were based on pilot studies, which showed that across- and within-subject responses were highly consistent. The first trial of each run was discarded as a dummy trial for analysis. Stimulus presentation was fully balanced with equal number of first-order transitions between different morph steps (Brooks, 2012). Each run lasted on average 14 minutes, with no significant differences between modalities. Once subjects completed all runs, a psychometric function was fitted to the data of each modality with a logistic curve, using the Palamedes toolbox (Prins and Kingdom, 2018) implemented in MATLAB. The three stimuli form the 101-morph continuum closest to 45%, 50% (point of subjective equivalence, PSE) and 55% anger responses derived from the subject- and modality-specific psychometric curve were identified and used in the scan session as emotionally *ambiguous* exemplars. The extreme morphs representing prototypical expressions of fear (0%, 10%, 20% anger) and anger (80%, 90%, 100% anger) were used as emotionally *explicit* stimuli during the scan session.

5.3.5 fMRI acquisition and analysis

5.3.5.1 fMRI experimental protocol

Participants completed four 8-minute runs, which were similar to those done before scanning, with the exception that both auditory and visual stimuli were presented in each run and that the inter-stimulus interval was 4 seconds. Each run consisted of six presentations of nine morph steps of each modality, grouped as *explicit fear* (0%, 10% and 20% anger morph), *ambiguous* (subject-specific pre-scan 45%, 50% and 55% anger responses; see above) and *explicit anger* (80%, 90% and 100% anger morph). Stimuli were pseudo-randomized with equal probability transitions across morph steps and modalities. Participants responded to the stimuli using an MRI-compatible two-button mouse. Prior to the scan session, auditory and visual tests were conducted to ensure subjects could comfortably hear and see the stimuli and press the buttons.

The proportion of anger responses given to the ambiguous stimulus was averaged across the four runs and analyzed in an ANCOVA, with modality as within-subject factor and state anxiety as a covariate. The corresponding response times (RTs) were log-transformed and entered in an

ANCOVA with modality and emotion response as within-subject factors and state anxiety as a covariate. Only "correct" responses were included in the case of explicit expressions (i.e., anger responses to Anger and fear responses to Fear). To test for potential sex differences, the subject self-identified sex was entered into analyses as a between-subject factor. Correction for multiple tests was applied to p-values when appropriate the Holm-Bonferonni method (Holm, 1979) and reported as p_{HB}. Mean response times and responses from pre-scan and scan sessions are reported in Supplementary Table 5.1.

5.3.5.2 Image acquisition and preprocessing

Functional images were acquired in a 3T Siemens Magnetom Prisma Fit scanner with a 32-channel head coil using a continuous multi-band accelerated pulse sequence with a factor of 12 (Setsompop et al., 2012). One thousand volumes (72 slices per volume, interleaved acquisition; FOV = 208 x 208 mm^2, matrix = 104 x 104, voxel size = 2 x 2 x 2 mm^3; TR = 529 ms; TE = 35 ms) were acquired per run. The initial 10 scans of each run were discarded based on T1 saturation. A high-resolution T1-weighted image (voxel size = 1 x 1 x 1mm^3) was acquired using a magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence (192 slices, TR = 2.3 s; TE = 3 ms) for anatomical co-registration.

Image pre-processing was conducted using SPM12 (Wellcome Department of Imaging Neuroscience, London, UK; http://www.fil.ion.ucl.ac.uk/spm), as in our previous studies (Whitehead & Armony, 2018; 2019). Briefly, images were realigned to the first volume of the first run, normalized to the Montreal Neurological Institute (MNI) 152 template and for the univariate analysis, smoothed using a 6 mm FWHM isotropic Gaussian kernel.

5.3.5.3 Univariate Analysis

Subject-specific statistical analysis was performed in SPM12 using a univariate general linear model (GLM) with 4 categories of interest for each modality, as function of the *a priori* emotion category (Anger, Fear and Ambiguous) and subject's response (Anger or Fear): Anger/Anger, Fear/Fear, Ambiguous/Anger and Ambiguous/Fear. That is, in the case of explicit emotions, only correct responses were included (any incorrect or no-response trials were included as additional covariates of no interest). Each stimulus was represented as a boxcar of a length equal to its duration, convolved with the canonical hemodynamic response function. Subject-specific conditions were averaged across the 4 runs, and taken to a second level, repeated-measures GLM. Regions responding significantly differently to ambiguous and explicit expression obtained by computing the corresponding contrasts. As previous studies have shown that task-difficulty modulates both so-called *task-positive* (i.e., positive responses relative to baseline) and *task-negative* (i.e., negative responses relative to baseline) regions (Chand and Dhamala, 2016), we used the overall main>0 and main<0 contrasts to distinguish activations from deactivations, respectively. That is, we used both [(Ambiguous>Explicit) ∩ Main>0] and [(Ambiguous<Explicit) ∩ Main<0)] contrasts to identify voxels responding significantly more to ambiguous than explicit expressions (and vice versa). Statistical significance was determined using a voxel threshold of $p = .05$, corrected for multiple comparisons with a family-wise error rate (FWE) computed through Gaussian random field theory. Post-hoc repeated-measures MANCOVAs were conducted using the mean absolute value of parameter estimates for each contrast, with modality (Auditory vs. Visual), emotion response (Fear vs. Anger) and emotion intensity (Ambiguous vs. Explicit) and network for respective MANCOVAs, as within-subject factors, subjects' self-reported sex as a between-subject factor and state anxiety score as a covariate.

5.3.5.4 Multivariate Pattern Analysis

We conducted a complementary whole-brain multivariate pattern analysis to assess whether explicit vs. ambiguous emotional stimuli could be accurately distinguished based on the overall brain activation patterns they elicited (Whitehead & Armony, 2019). Specifically, preprocessing occurred alike the univariate analysis, however, volumes were instead moderately smoothed using a 2 mm FWHM isotropic Gaussian kernel (Gardumi et al., 2016). Condition- and subject-specific parameter estimates (betas) obtained from the univariate GLM were averaged across the 4 runs and used in a multivariate pattern analysis conducted in PRoNTo (Schrouff et al., 2013). A kernel classifier was trained to identify activity patterns across voxels, as extracted from mean beta images, using a binary classification with support vector machine (SVM; Burges, 1998) and LIBSVM implementation (Chang and Lin, 2011). A leave-one-subject-out (LOSO) cross-validation was conducted, and classification accuracy was acquired (Schrouff et al., 2018). Model performance was evaluated at a p = .05 threshold, calculated using 1000 permutations of the training labels. A mask of the cerebrum was used formed by 90 regions (excluding cerebellum and vermis) of the Automated Anatomical Labeling (AAL; Tzourio-Mazoyer et al., 2002) atlas.

Weights per voxel were computed *a posteriori* (Schrouff et al., 2013) and identification of regions contributing greatest to classifications done by identifying the top 1% positive and negative weights, with a minimum size of ten neighbouring voxels. The associated p-value for each voxel weight, was calculated through permutation testing, where the condition labels were randomized, and weight maps acquired from each classification (N= 1,000). Statistical significance of clusters was calculated by averaging the voxel-specific p-values. Model performance was evaluated based on classification accuracy, area under the receiver operating characteristic (ROC) curve (AUC) and the sensitivity and specificity of the classification.

5.4 Results

5.4.1 Pre-scan

The group-averaged PSEs for visual and auditory stimuli were 60.8% (*SD*=6.8%) and 47.5% (*SD*=6.3%) anger morph, respectively (t(25) = 6.62, p < .001, d = 1.63), with no significant correlation between modalities (r=-.22, p=.24). In contrast, the slopes of the fitted psychometric curves were very similar (Visual: 14.8, SD=4.7; Auditory: 14.6, SD=4.8, t(28)<1) and significantly correlated (r=.60, p<.001; Figure 5.2A). As expected, the group-averaged RTs exhibited an inverted U-shape pattern centered on the subjects' mean PSE (Figure 5.2B). This was confirmed statistically by a significant correlation between the individual subjects' PSE and their morph step with largest RT (Visual: r= .57, p=.002; Auditory: r=.79, p <.001).

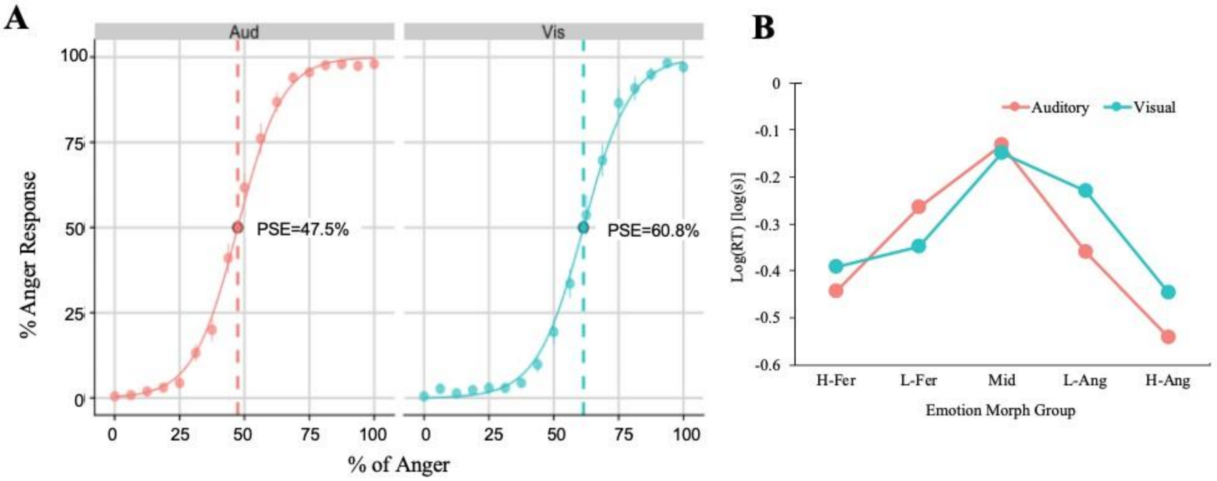

Figure 5.2 *A*) Auditory (left) and visual (right) psychometric curves of pre-scan responses averaged across 26 subjects. The curves illustrate the percent-anger of the morph stimuli, and the corresponding average response made by subjects as to the percent anger perceived. PSE=point of subjective equality. *B)* The average auditory (red) and visual (blue) pre-scan log reaction times for each emotion-by-intensity step.

5.4.2 Scan session

5.4.2.1 Behaviour

As expected, subjects correctly identified the explicit emotional expressions for auditory (Anger: M=97.2%, *SD*=2.9%; Fear: 96.9%, SD=3.9%) and visual (Anger: M=94.0%, *SD*=7.6%; Fear: 98.7%, SD=2.2%) stimuli, confirming that they could hear and see the stimuli without problem. Nonetheless, the proportion of Anger responses for the visual explicit angry condition was significantly smaller compared to the pre-scan session for the same stimuli (M=97.4%, *SD*=3.4%; Wilcoxon signed rank test: z=-2.74, p=.006/p_{HB}=.02), with no significant differences for any of the other conditions (p_{HB}>.5). Interestingly, this decrease in Anger responses for visual explicit angry stimuli between sessions was significantly correlated with subjects' state anxiety scores (r=-.508, p=.008; Figure 5.3A).

Analysis of responses to ambiguous stimuli as a function of the emotion perceived (Figure 5.3B) revealed a main effect of modality (*F*(1,23) = 9.02, *p* = .006, partial η^2 =.29). Post-hoc paired samples t-tests showed that this effect was driven by a significant difference in responses to visual ambiguous stimuli, rated more often as fearful than angry (t(1,23) = 5.65, p_{HB} =<.001, Cohen's d= .43), whereas no difference was observed in the auditory modality (F(1,23) <1). The proportion of anger responses to visual ambiguous stimuli was negatively correlated with the decrease in Anger responses for visual explicit angry stimuli between sessions mentioned above (r= -.481, p=.01). There were no main effects or interactions of subjects' sex or state anxiety scores in response to ambiguous stimuli.

Analysis of logRTs revealed a main effect of modality [F(1,23)=34.98, p<.001, partial η^2 =.60: faster responses for visual than auditory stimuli]; intensity [F(1,23)=323.36, p<.001, partial η^2 =.93: faster responses for explicit than ambiguous stimuli] and state anxiety [F(1,23)=5.78, p=.025, partial η^2 =.20: negative correlation with logRTs]. In addition, there were response-by modality [F(1,23)=18.52, p<.001, partial η^2 =.45: faster RT for fear than anger in visual and opposite in auditory] (Figure 5.3C), response-by-intensity [F(1,23)=7.38, p=.012, partial η^2 =.24: faster RTs for fear than anger in Explicit and opposite in Ambiguous] and near to significant response-by-intensity-by-anxiety (F(1,23)=4.22, p=.052, partial η^2 =.16) interactions. The latter was due to a reduction in the RT difference between explicit and ambiguous stimuli judged as fearful as a function of anxiety (r=.55, p=.003; Figure 5.3D). There were no main effect nor interactions associated with subjects' sex.

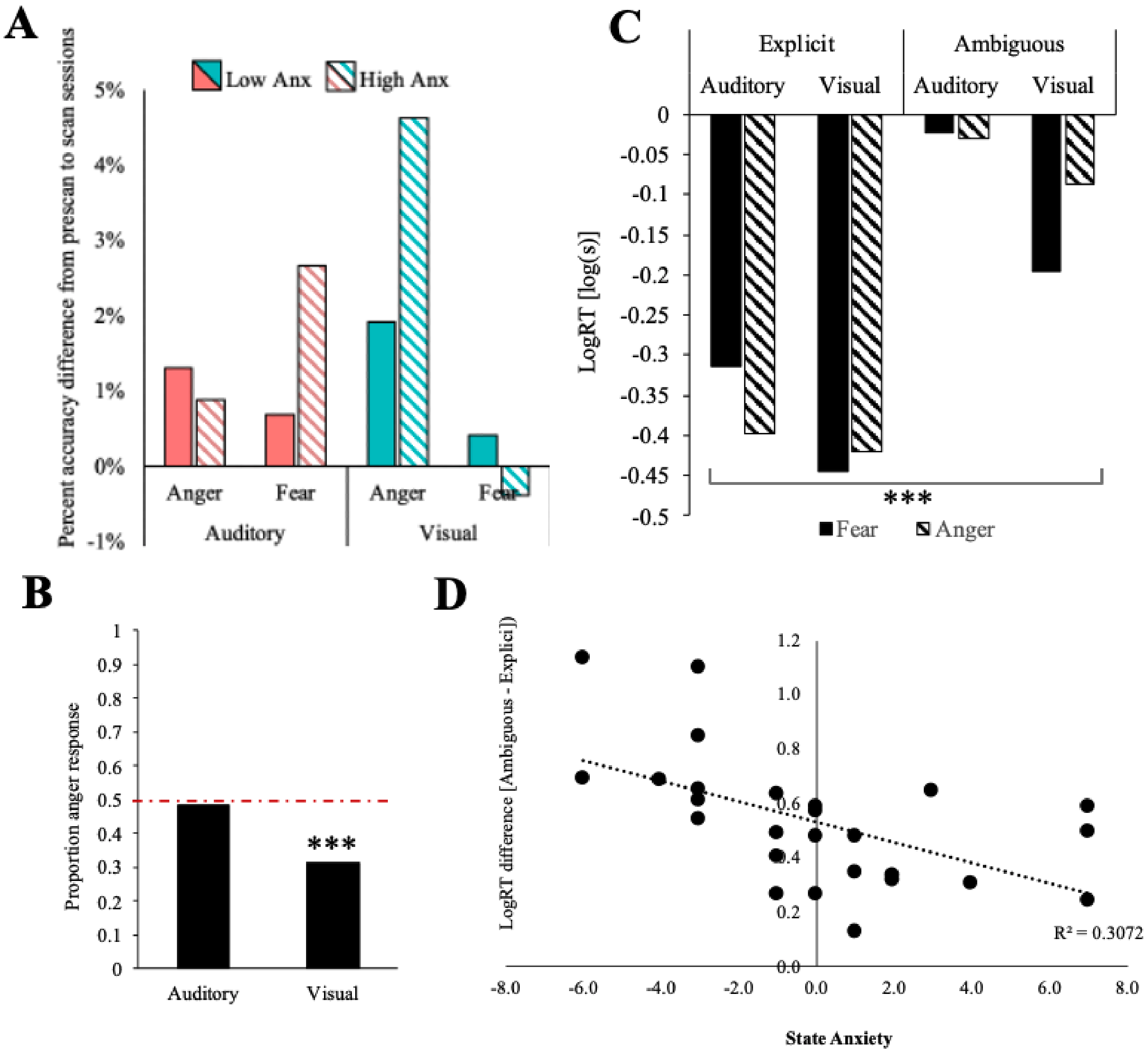

Figure 5.3 *A)* Difference of percent accuracy between pre-scan and scan session mean responses to explicit auditory (red) and visual (blue) fear and anger stimuli. For visualization purposes, subjects were median split by low (<0; solid fill) and high (≥0; striped) state anxiety scores. *B)* Proportion of anger responses made to auditory and visual ambiguous stimuli perceived as anger, showing significant difference in visual modality and *C)* mean *l*og reaction times (log(s)) to ambiguous and explicit stimuli, illustrating significant response-by-modality interaction. *D)* Correlation of state anxiety scores with difference of ambiguous versus explicit log reaction times for stimuli perceived as fearful. Asterisks indicate a significant difference (***p<.001).

5.4.3 fMRI data analysis

5.4.3.1 Univariate analysis

Significant clusters of activity for the contrasts Ambiguous vs. Explicit expressions are reported with their coordinates, z-scores, and cluster extents in Table 5.1. Greater *activation* in response to Ambiguous than Explicit Emotion was observed bilaterally in several frontal regions– including the bilateral supplementary motor area (SMA), superior frontal gyrus (SFG), anterior cingulate cortex (ACC), the dorsolateral prefrontal cortex (dlPFC) and anterior insula (AI), as well as the inferior parietal lobe (IPL) and occipital activity. The majority (81.4%) of the significant voxels were located within three of the 7-network cortical Parcellation derived by Yeo et al. (2011) from

Table 5.1 Significant activations associated with contrasts of interest at the group level.

Ambiguous vs. Explicit Emotion								
Anatomical Location	***L***			***R***			***Z-score***	***K_E***
	x	***y***	***z***	***x***	***y***	***z***	***(peak voxel)***	
Ambiguous > Emotion – Activation								
L/R Supplementary Motor Area	-2	22	44				10.35	1393
L/R sup. Frontal gyrus	-4	10	56				8.45	
L ant. Cingulate cortex	-10	26	30				6.1	
L mid. Cingulate cortex								
L anterior insula	-30	22	2				9.23	1390
L dorsolateral Prefrontal cortex	-46	18	26				8.11	
L Precentral gyrus	-36	8	30				6.17	
L dorsolateral Prefrontal cortex	-42	38	8				5.43	20
R anterior insula				34	24	-2	9.03	1890
R dorsolateral Prefrontal cortex				44	24	22	8.11	
R mid. Frontal gyrus				42	18	-4	7.88	
L inf. Parietal lobe	-32	-48	42				5.9	51
R inf. Parietal lobe	38	-48	46				5.57	33
L inf. Occipital gyrus	-18	-100	2				6.9	520
L mid. Occipital gyrus	-18	-104	-6				6.36	
	-36	-90	-2				5.74	
R inf. Occipital cortex				22	-96	4	6.43	258
R mid. Occipital cortex				18	-104	8	5.85	
R sup. Occipital cortex				30	-90	-6	5.05	
R Calcarine gyrus								
L Cerebellum	-8	-80	-28				5.14	7
L Cerebellum	-32	-58	-30				5.08	1
Ambiguous > Emotion – Deactivation								
L Angular gyrus	-44	-68	30				8.25	1538
L sup. Temporal gyrus	-38	-78	40				7.31	
L mid. Temporal gyrus	-52	-54	38				7.13	
L mid. Occipital gyrus								
R Angular gyrus				48	-68	30	7.19	673
R sup. Temporal gyrus				40	-78	38	6.46	

R mid. Temporal gyrus				56	-58	22	6.14	
R mid. Occipital gyrus								
R Supramarginal gyrus				52	-30	28	5.29	32
				58	-26	22	5.07	
L sup. Frontal gyrus	-6	52	-6				7.54	3295
L ventromedial Prefrontal cortex	-16	42	48				7.27	
L mid. Frontal gyrus	-16	48	38				7.21	
L ant. Cingulate cortex								
L ventromedial Prefrontal cortex	0	44	-24				5.78	39
L mid. Cingulate gyrus	0	-10	34				4.94	1
R sup. Frontal gyrus				28	32	44	6.97	223
R mid. Frontal gyrus								
R Ventromedial Prefrontal cortex								
R sup. Frontal gyrus				24	20	58	5.69	51
R mid. Frontal gyrus								
R ant. Cingulate cortex				4	28	12	5.38	9
R sup. Frontal gyrus				14	50	36	4.98	4
L/R precuneus	-6	-52	34				6.98	1060
L/R Calcarine gyrus	-4	-44	36				6.88	
L/R mid. Cingulate cortex				0	-40	46	6.93	
L/R post. Cingulate cortex								
L mid. Temporal gyrus	-60	-22	-12				6.92	569
L inf. Temporal gyrus	-60	-6	-22				6.76	
	-56	4	-24				6.24	
L mid. Temporal gyrus	-60	-56	-2				5.88	23
L inf. Temporal gyrus								
L Temporal pole	-40	16	-34				5.12	4
R mid. Temporal gyrus				64	-8	-14	6.02	157
				62	-24	-16	5.48	
				64	0	-16	5.2	
R Temporal pole				46	18	-32	5.83	56
R mid. Temporal gyrus				52	0	-26	5.21	13
L Fusiform gyrus	-30	-42	-10				6.4	87
R Fusiform gyrus				32	-32	-16	6.28	65
R parahippocampus								
Explicit > Ambiguous – Activation								
R Amygdala				26	0	-22	5.22	11
L sup. Temporal sulcus	-62	-22	-8				6.08	44
	-62	-8	-10				5.33	
L mid. Temporal gyrus	-50	-62	18				5.63	6
L mid. Temporal gyrus	-60	-42	0				4.98	2
R sup. Temporal sulcus				62	-6	-12	5.92	74
				64	-20	-6	5.52	
R Parietal Operculum				46	-30	20	4.9	1

resting-state data: the Frontoparietal Control (FPCN; 43.1%), Ventral Attention (26.5%; also known, and referred henceforward, as Salience, SN; Seeley et al., 2007) and Visual (VN; 11.9%) networks (Figure 5.4A; Supplementary Table 5.2).

Greater *deactivation* observed in response to Ambiguous as compared to Explicit Emotion was located within the bilateral ventromedial prefrontal cortex (vmPFC), angular gyrus, precuneus and posterior cingulate cortex (PCC), as well as in the bilateral middle temporal gyrus and fusiform gyrus (FG; Figure 5.4A). Most of these voxels (74.6%) were located within the Default Mode network (DMN).

Clusters exhibiting significantly greater *activation* in response to Explicit than Ambiguous Emotion conditions included the right amygdala (AMG) and bilateral superior temporal sulcus (STS; Figure 5.5A).

5.4.3.2 Post-hoc (M)ANCOVA interactions

The three task-positive networks that overlapped with the majority of the ambiguous-preferred clusters were used as individual masks to conduct a post-hoc MANCOVA on the subject- and network-specific, session-averaged parameter estimates with Modality, Emotion Response, Intensity and Network as within subject factors, sex as a between-subjects factor and state anxiety as a covariate. In addition to the expected significant effect of intensity ($F(3,21) = 35.87$, $p <.001$, partial $\eta^2 = .84$), there was a significant effect of modality ($F(3,21) = 38.27$, $p <.001$, partial $\eta^2 = .84$) with greater activity in response to auditory stimuli for the SN and FRPN, and the opposite for VN. There was also a modality-by-intensity interaction ($F(3,21) = 9.35$, $p <.001$, partial $\eta^2 = .57$) driven by the VN ($F(1,23) = 4.39$, $p =.047$, partial $\eta^2 = .16$) where the aforementioned difference was most prominent for ambiguous than explicit conditions. Moreover, a three-way modality-by-intensity-by-anxiety interaction ($F(3,21) = 3.64$, $p = .030$ partial $\eta^2 =.34$) was due to a decreasing difference between emotion intensities (ambiguous minus explicit) in the visual modality as anxiety scores increased and the opposite pattern for auditory stimuli, as driven by the SN ($F(1,23) = 5.32$, $p = .030$ partial $\eta^2 =.19$) and FPCN ($F(1,23) = 9.86$, $p = .005$ partial $\eta^2 =.30$; Figure 5.6A). Finally, there was a near-to-significant response-by-intensity interaction ($F(3,21) = 2.97$, $p = .055$, partial $\eta^2 =.30$) driven by a greater difference to ambiguous minus explicit for anger than fear responses for SN ($F(1,23) = 7.28$, $p = .013$, partial $\eta^2 =.24$) and a trend for FPCN ($F(1,23) = 4.05$, $p = .056$, partial $\eta^2 =.15$; Figure 5.4B).

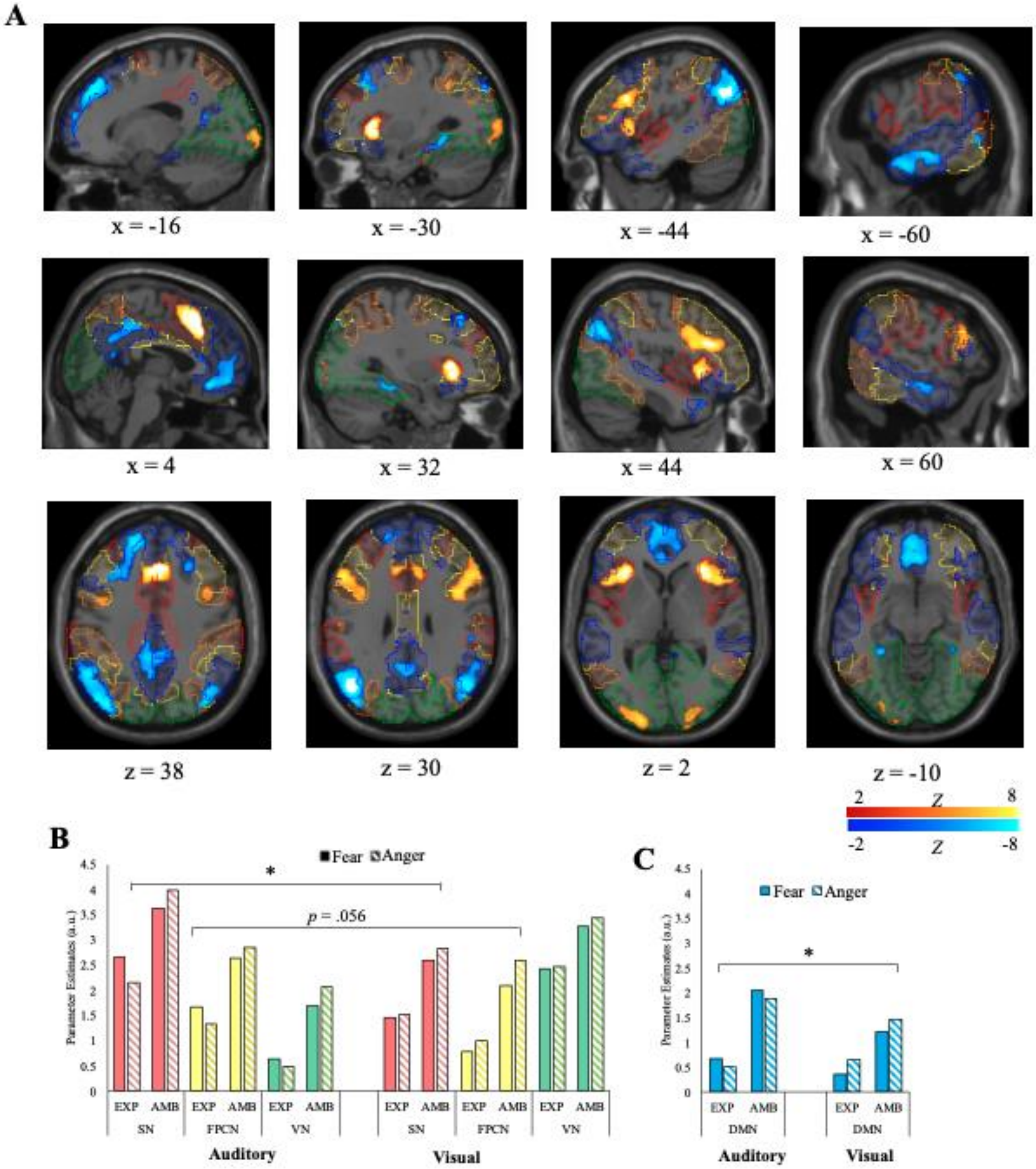

Figure 5.4 *A)* Five networks taken from Yeo et al. (2011)'s 7-network cortical parcellation are outlined, including the Salience (Ventral Attention; red), Dorsal Attention (orange), Frontoparietal (yellow), Default Mode (blue), and Visual Networks (green). 2D renderings of the clusters of significant activation (red scale) and deactivation (blue scale) in response to contrasts Ambiguous versus Explicit Emotion. Threshold of p=.05 FWE. Parameter estimates (absolute values) for regions of *B)* activation resulting from MANCOVA that presents post-hoc results of a response-by-intensity interaction for SN and FPCN, and *C)* deactivation resulting from ANCOVA, illustrating a response-by-modality interaction. Regions overlap (>10%) with Yeo et al. (2011) networks, presenting preferred response of fear (solid fill) and anger (striped) to Ambiguous versus Explicit stimuli. Asterisks indicate significant difference (*p<.05).

A post-hoc ANCOVA was conducted on mean beta estimates for clusters of significant Ambiguous vs. Explicit Emotion *deactivation,* using a mask of the overlapping Default Mode Network (>10% *deactivation*). In addition to the expected significant effect of intensity ($F(1,23) = 60.83$, $p < .001$, partial $\eta^2 = .73$), there was a significant effect of modality ($F(1,23) = 8.95$, $p = .007$, partial $\eta^2 = .31$) with a greater response to auditory than visual stimuli. There was also a significant response-by-modality interaction ($F(1,23) = 4.40$, $p = .047$, partial $\eta^2 = .16$), due to a greater response to auditory fear ($p_{HB} = .30$) and visual anger ($p_{HB} = .038$) and a modality-by-intensity interaction ($F(1,23) = 5.72$, $p = .025$, partial $\eta^2 = .20$) driven by a greater difference between emotion intensities for auditory than for visual modalities (Figure 5.4C).

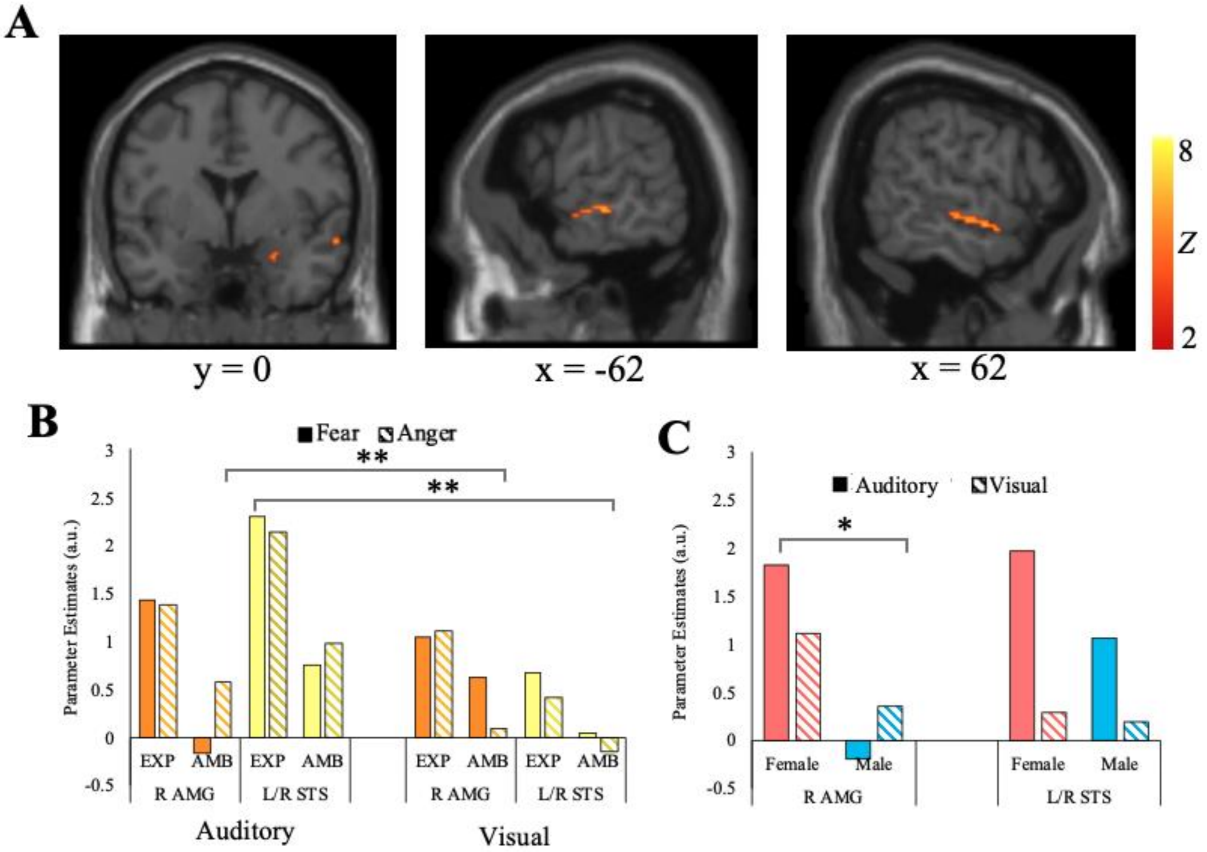

Figure 5.5 *A)* 2D renderings of the clusters of significant activation in response to contrasts Explicit minus Ambiguous Emotion. Threshold of p=.05 FWE. Results of MANCOVA of right Amygdala and bilateral STS parameter estimates as separated by *B)* intensity (ambiguous and explicit), modality (auditory and visual) and emotion response of fear (solid fill) and anger (striped), illustrating significant modality-by-intensity interaction driven by the STS, and a modality-by-intensity-by-response interaction driven by the amygdala; *C)* modality and subject-sex (female and male), illustrating a sex-by-modality interaction. Asterisks indicate significant difference (*$p < .05$, **$p < .01$).

A similar MANCOVA for the cluster in the right amygdala and bilateral STS obtained in the contrast Explicit minus Ambiguous Emotion conditions (Figure 5.5B) yielded the expected significant effect of intensity ($F(2,22) = 57.60$, $p<.001$, partial $\eta^2 = .84$), as well as an effect of modality ($F(2,22) = 15.95$, $p<.001$, partial $\eta^2 = .59$) and a modality-by-intensity interaction ($F(2,22) = 6.26$, $p=.007$, partial $\eta^2 = .36$) driven by greater difference between auditory explicit minus ambiguous conditions than visual in the bilateral STS clusters ($F(1,23) = 12.39$, $p=.002$, partial $\eta^2 = .35$). Additionally, an observed 3-way modality-by-intensity-by-response interaction ($F(2,22) = 4.72$, $p=.020$, partial $\eta^2 = .30$) was driven by activity of the amygdala ($F(1,23) = 9.81$, $p=.005$, partial $\eta^2 = .30$) demonstrating a larger magnitude for auditory ambiguous stimuli judged as angry than fearful, with the opposite trend for visual ambiguous stimuli. A sex-by modality interaction was observed ($F(2,22) = 5.60$, $p=.011$; partial $\eta^2 = .34$) where females responded more to auditory stimuli, while males responded more to visual. This finding was driven by activity of the amygdala ($F(1,23) = 7.18$, $p=.013$; partial $\eta^2 = .24$, as both females and males presented greater activity within the STS in response to auditory stimuli (Figure 5.5C). Finally, an intensity-by-

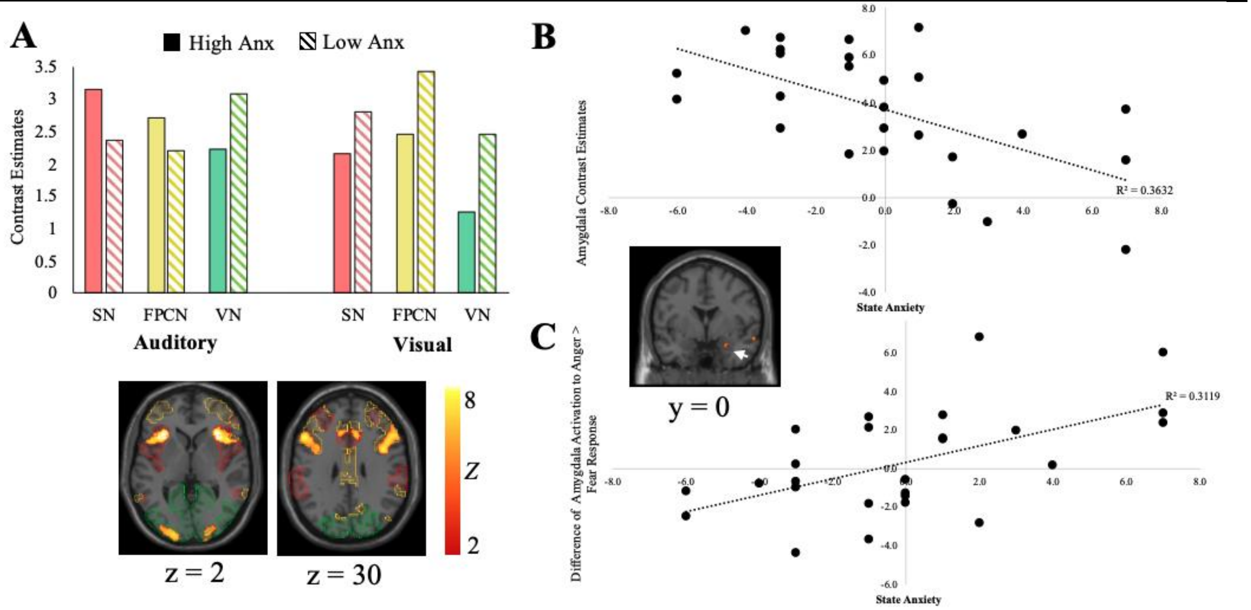

Figure 5.6 *A)* Results from MANCOVA reporting contrast estimates of Ambiguous minus Explicit *activation* that overlapped (>10%) with Yeo et al's (2011) networks, as separated by modality and state anxiety, where for visualization purposes, is median split for low (<0; solid fill) and high (≥0; striped) state anxiety scores. 2D renderings from Figure 5.4A indicate regions of *activation*, outlined by the Salience (Ventral Attention) (red), Frontoparietal (yellow) and Visual Networks (green). Correlation of subjects' state anxiety scores with *B)* Amygdala Explicit minus Ambiguous contrast estimates and *C)* Amygdala response to stimuli perceived as anger versus fear. 2D renderings from Figure 5.5A indicate amygdala activation in response to Explicit minus Ambiguous conditions.

anxiety interaction (*F*(2,22) = 5.42, p=.012, partial η^2 = .33) was observed, driven by the amygdala (*F*(1,23) = 10.14, p=.004, partial η^2 = .31) demonstrating a reduction in the explicit minus ambiguous difference with increased anxiety, (r=-.603, p=.001; Figure 5.6B), and a response-by-anxiety interaction (*F*(2,22) = 3.97, p=.034, partial η^2 = .27) driven by the amygdala (*F*(1,23) = 8.29, p=.008, partial η^2 = .27) due to a positive correlation between the Anger minus Fear difference and anxiety scores (r=.558, p=.003; Figure 5.6C).

There were no main effects of subjects' sex, nor state anxiety in any of the above analyses.

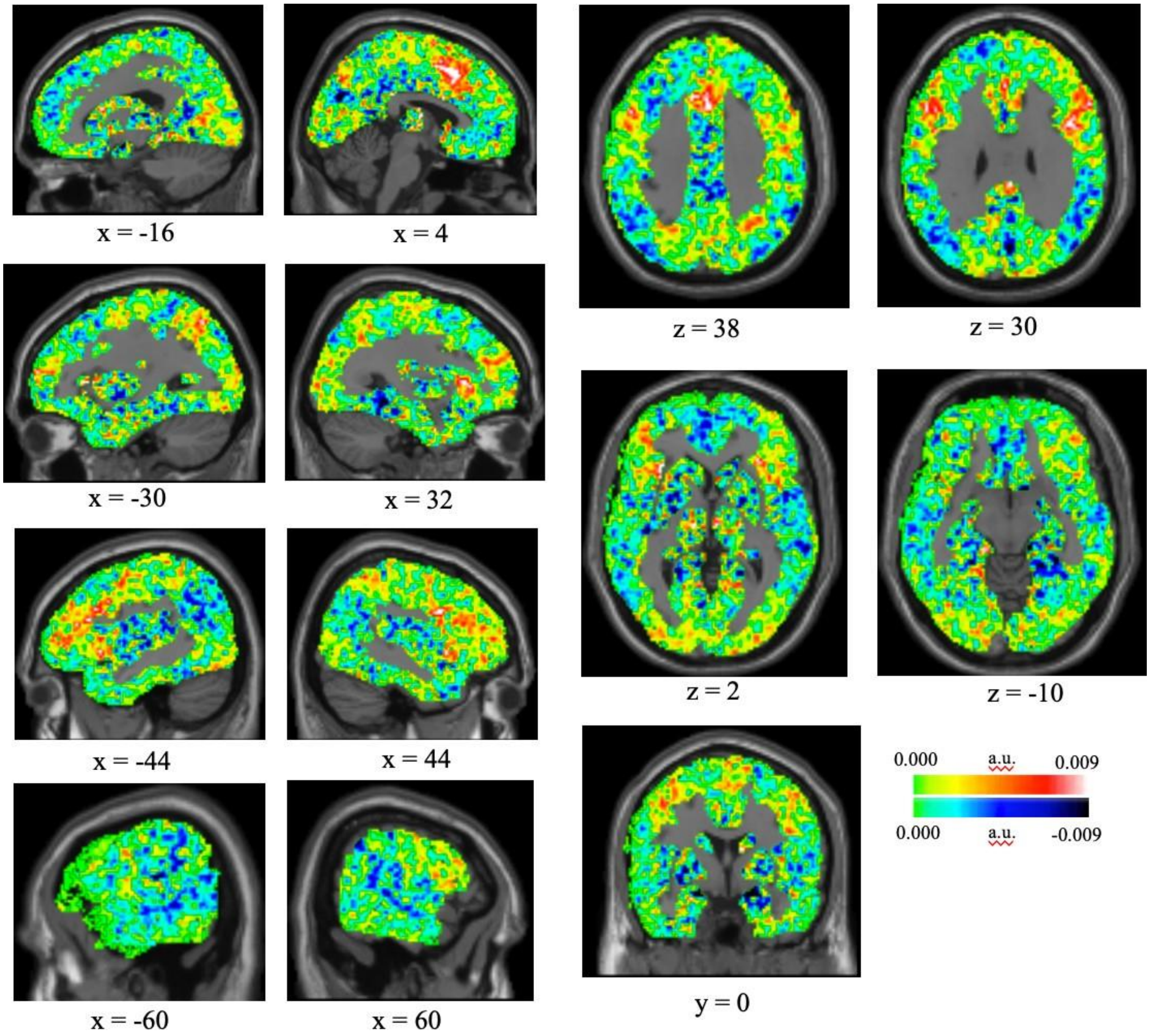

Figure 5.7 Weight map of positive (red) and negative (blue) weights presenting clusters corresponding to the classification of Ambiguous versus Explicit Emotion conditions. A.U., arbitrary units.

5.4.4 Multivariate results

The model performance of Ambiguous versus Explicit emotion perception yielded an above chance classification accuracy of 85.1%, p<.001), with an AUC of 0.90, true positives for ambiguous of 74.0%, and true positives for explicit of 96.1%. The clusters corresponding to the top 1% of voxels that contributed the greatest weight to the classification included frontal regions (e.g., SMA, dlPFC and insula), superior parietal lobe, thalamus, lingual and parahippocampal gyri. Negative weights contributions came primarily from subcortical regions, such as amygdala, cuneus, putamen, and caudate, as well as parahippocampal, fusiform, lingual and posterior superior and mid temporal gyri (Figure 5.7; Table 5.2).

5.5 Discussion

Interpreting emotional expressions in everyday life is not always clear cut, nor are the underlying neural processes to do so. The current study uses high spatial-and temporal-resolution fMRI to identify behavioural and neural properties that differ when perceiving ambiguous versus explicit socio-emotional threat-related expressions across modalities. We explore differences of auditory and visual perception through measuring within-modality stability of subject-specific behaviour as it relates to the underlying neural activity. By forcing subjects to make perceptual decisions about ambiguous stimuli, we expose systems of perception, independent of the physical attributes of the stimulus. Moreover, we provide evidence that the amygdala not only tracks the perception of threat (explicit fear and anger), but also modulates in accordance with the degree of ambiguity, or rather, the task-difficulty enforced by the stimulus (Wang et al., 2017). Finally, due to the close relationship of stress and uncertainty, we report on influences of transient state anxiety, identifying trends that span from behaviour to the brain. Specifically, attention is drawn to interactions of state-anxiety with task-difficulty, demonstrating dynamic modulations of regions within concerting networks associated with emotion, attention, and cognition.

5.5.1 Influence of task-difficulty on emotion perception processes

Involvement of the amygdala in detecting and valuating salient emotional information is apparent throughout the literature (see for review, Phelps and LeDoux, 2005). Specifically, our group and others, have observed amygdala recruitment when processing emotion across sensory modalities (Lin et al., 2020; Aubé et al., 2015; Whitehead et al., 2019). In the current study, the right amygdala

Table 5.2 Ranking of anatomically defined clusters (having a cluster size >10 voxels) that contributed to the top 1% of the weight contribution in the above-chance classification accuracy of Ambiguous versus Explicit emotion conditions. The peak coordinates of cluster, cluster sizes and percent occurrence in 1,000 permutations are reported for each anatomical location.

Anatomical location	**Positive Weights**						**Anatomical Location**	**Negative Weights**					
	L/R	***x***	***y***	***z***	K_E	**%**		**L/R**	***x***	***y***	***z***	K_E	**%**
Supplementary Motor Area **Mid. Cingulate cortex**	L/R	-4	10	58	294	99.6	Amygdala Parahippocampal gyrus	R	18	0	-12	43	95.2
Dorsolateral Prefrontal cortex **Precentral** gyrus	L	-46	16	32	95	99.4		R	24	4	-22	19	97.6
	L	-46	34	18	11	99.6		L	-18	-2	-14	30	94.6
	L	-46	26	20	10	99.9	Fusiform gyrus	R	26	-44	-14	33	96.0
	R	54	20	34	49	99.5	Cuneus	R	4	-78	28	19	98.4
	R	46	12	30	30	99.9	Sup. Temporal gyrus (posterior)	L	-56	-30	18	18	99.1
Insula	L	-32	24	4	45	99.0	Mid. Temporal gyrus (posterior)	R	62	-32	4	13	99.4
	R	32	24	6	21	99.5		L	-62	-38	-4	10	99.4
Sup. parietal lobe	L	-32	-66	54	15	99.8	Putamen	L	-32	0	-6	12	96.1
Thalamus	L	-16	-18	0	15	96.1	Lingual gyrus	R	20	-46	-6	12	95.6
	L	-12	-24	14	11	94.3		R	14	-32	-10	11	93.1
	R	8	-16	12	11	91.9	Caudate	R	10	10	-8	10	97.5
	R	4	-8	12	10	95.3							
Caudate	R	10	6	12	13	94.4							
Lingual gyrus	R	12	-56	4	11	96.1							
Parahippocampal gyrus	L	-24	-38	-6	10	96.1							
	L	-16	-34	-8	10	85.3							

and bilateral STS demonstrated sensitivity to explicit threat, confirmed by MVPA results illustrating top weight contributions from the bilateral amygdala to the above-chance classification (85%) of ambiguous versus explicit emotion perception. Markedly, the STS is recognized for processing speech, faces and audiovisual integration, as well as theory of mind (Hein and Knight, 2008; Deen et al., 2015; Beauchamp, 2015). It is plausible that here, the STS is engaging when processing the actor's intentions (anger) or empathizing with the actor (fear; see meta-analysis, Bzdok et al., 2012). Acquiring self-report measures of empathy (Empathy Quotient [EQ]; Baron-Cohen and Wheelwright, 2004) and theory of mind (Theory of Mind Inventory [ToMI]; Hutchins et al., 2010) may clarify this recruitment.

Like others before us (Wang et al., 2017; Ito et al., 2017; Bestelmeyer et al., 2014), we identified that subjects respond faster to expressions of explicit versus ambiguous affect, potentially reflecting lower task-difficulty (Deupree and Simon, 1963). In contrast, as perceptual uncertainty increases during decision-making, we observe greater recruitment of regions in *task-positive* and *task-negative* networks (Fox et al., 2005) often reported alongside greater task-difficulty (see meta-analysis, Keuken et al., 2014). These regions included those of primarily the Salience, Frontoparietal and Visual Networks (*activation*), as well as the Default Mode Network (*deactivation*). Multivariate analyses reinforced our assumptions of network recruitment (Yeo et al., 2011), given significant weight contributions to the above-chance classification dispersed across these same networks. Comparable regions have been reported in response to ambiguous affect (Thielscher & Pessoa, 2007; Bestelmeyer et al., 2013); however, have yet to be associated to these specified networks.

The SN, responding to level of subjective salience elicited by a stimulus or task (Menon, 2015; Menon & Uddin, 2010; Peters et al., 2016), is actively involved in resolving ambiguity and uncertainty (Neta et al., 2013; Lamichhane & Dhamala, 2015; Lamichhane et al., 2016a; 2016b). The FPCN, recruited during problem-solving and decision-making (Sridharan et al., 2008), functions by modifying and sustaining relevant information in working memory (see for review, Menon, 2011). These two networks present opposing activity to the DMN, which is often active at rest, and involved in self-reference, mentalizing and theory of mind (Mars et al., 2012). DMN activity then decreases when attention is goal-directed towards external stimuli (Raichle et al., 2001; Buckner et al., 2008). Interestingly, activation of the SN, FPCN and deactivation of the DMN appeared to fluctuate in contrast to the amygdala—as task-difficulty increased, amygdala activity decreased. We could infer that from explicit to ambiguous emotion perception, attentional demands may shift from the emotional salience of the stimulus, to instead, the task of resolving ambiguity. Comparable findings have been reported where decreased activity of amygdalohippocampal regions were observed when subjects had to explicitly label, versus implicitly or passively process emotion (Lange et al., 2003; Hariri et al., 2000; Critchley et al., 2000). This supports the idea that the amygdala is part of an “impulsive system” during decision-making, most often observed in the context of reward behaviour (Bechara, 2005; Kim et al., 2018; Xie et al., 2015; Ko et al., 2015).

Following the rationale that response time serves as a proxy for task-difficulty (Deupree and Simon 1963), auditory stimuli may be more challenging to classify than visual, based on the longer response times observed and greater recruitment of regions in task-based networks. Additionally, the auditory ambiguous ratings remained relatively stable at chance (50% anger) in scan sessions, while visual ambiguous stimuli were, on average, perceived as more fearful (34% anger). When subjects present a strong intuitive bias (as see in the visual modality), they purportedly make responses more rapidly (Starns & Ma, 2018; Evans & Stanovich, 2013; Dekel & Sagi, 2019). In contrast, if subjects present doubt after forming an initial prediction, they may return to information held in working memory (Ester et al., 2014) to re-evaluate the decision before responding (Sterzer, 2016), or rather, before meeting the decision threshold (refer to drift diffusion model; Smith, 2000). The greater difference of activity in regions of the SN, FPCN, and DMN (deactivation) to auditory versus visual conditions may also reflect greater recruitment of attentional resources (Posner, 1980) to maintain task performance (reflected by response times and/or continuity of within-subject responses). Based on Fox et al. (2005)’s task-related dichotomy, subject’s reduce attention to self-referential processes and instead redirect it outwards to the task-at-hand.

Sensitivity of the Visual Network to the resolution of ambiguity was distinguishable from above-mentioned networks—responding more strongly to visual stimuli. This may be a product of late-stage processing of attention-directed eye movements during decision-making (see for review, Orquin and Loose 2013). It may also reflect a sustained response of the visual cortex during visual working memory and/or visual attention (see for review, Sreenivasan et al., 2014; faces specifically, Nelissen et al., 2013) from using strategies of visual imagery (Keogh and Pearson, 2011).

5.5.2 Exploring intrinsic decision biases

A response bias observed at both the behavioural and neural level was apparent for visual stimuli, and to a lesser degree, auditory. In exploring the bias for visual fear, we noted that subjects were faster at identifying faces expressing fear versus anger, while the opposite was true for auditory. Remarkably, this behavioural pattern was mirrored by the amygdala, driven by ambiguous stimuli (in the contrast explicit versus ambiguous). Greater activity was observed in response to visual stimuli perceived as fearful, and auditory perceived as angry, while in the opposing contrast (ambiguous versus explicit), we reported less deactivation of regions in the DMN to visual fear and auditory anger.

Markedly, the bias for visual fear was augmented during the scan session, reflecting—in signal detection theory terms (MacMillan and Creelman, 2004; Green and Swets, 1966)— a shift of the decision criteria. This occurred even after exhibiting an initial tendency (possibly an *a priori* bias) to respond “fear” during pre-scan sessions. This *re-calibration phenomenon* cannot be explained by aftereffects (Skuk and Schweinberger, 2013) as stimulus presentation order was counter-balanced across modality and morph-level. These measures minimize neuronal adaptation to attributes of a stimulus after repeated exposure (e.g., visual fear preceded by visual fear) that, otherwise, would have enhanced sensitivity to a change of stimulus (e.g., visual fear preceded by visual anger).

These findings would suggest that greater attentional resources were directed towards the more *challenging* modality-emotion pairs (visual anger, auditory fear). In a comparable fMRI study of facial emotion recognition, Sreevinas et al. (2012) attributed greater deactivation of the DMN to a lack of automaticity when perceiving sad and angry versus happy faces. They suggest that in viewing these upsetting emotions, a disruption to cognitive processes occurred while decision-making and attentional resources were reallocated, resulting in longer response times.

Notably, the neural response of the amygdala and regions of the DMN reflect differences of perception only, as the physical attributes of the stimulus remain unchanged. This would suggest that activity of the amygdala may not only modulate in accordance with the emotional salience of the stimulus, but appears sensitive to the degree of stimulus ambiguity (i.e., the level of certainty in accurately perceiving the stimulus). Wang and colleagues (2014; 2017) drew similar conclusions through a series of single-cell recording studies where subjects were presented with face morphs, ranging from happy to fearful expressions. They identified two sets of neuronal populations in the right amygdala: one responding positively to primarily, graded emotions of fear or happy (i.e., basolateral nuclei; BLA), while the other mostly responded to the degree of categorical ambiguity (i.e., centromedial nuclei; CM) or rather, the certainty or confidence in emotional categorization.

5.5.3 Influences of state-anxiety on brain and behaviour

This study demonstrates several means by which higher state anxiety scores may influence ones' processing of threat-related stimuli and ambiguity. Firstly, subjects with higher scores of state anxiety made decisions more quickly, with faster response times towards, particularly, visual ambiguous stimuli perceived as fearful, i.e., the proposed *visual fear* response bias. Additionally, they more often mistook visual explicit stimuli expressing anger, for fear, during scan versus pre-scan sessions. At the neural level, anxious subjects exhibited greater recruitment of the amygdala when perceiving ambiguous stimuli, and decreased activity of *task-positive* regions to visual ambiguous versus explicit conditions, while the opposite was true for auditory.

We deduce that mistakes made to explicit visual stimuli may reflect, to some degree, faster motor responses (as in subjects with high non-clinical anxiety; Ciucurel, 2012; Shrooten et al., 2012), but moreover, signs of enhanced impulsivity pertinent to both clinical (see for review, Jakuszkowiak-Wojten et al., 2015) and non-clinical anxiety (Xia et al., 2017). This evidence supports findings from Brown et al. (2017), who propose that an automatic emotional response is prioritized over the intended decision when under acute stress—as may occur in the confines of an MRI scanner. Conducting a control task outside of the scanner would be necessitated to validate assumptions of scanner-induced stress resulting in accentuated perceptual biases and increased impulsivity.

Clear evidence has shown that activity of the amygdala modulates as a factor of anxiety (Bishop et al., 2004; Somerville et al., 2004). Greater amygdala recruitment reported here could suggest an enhanced emotional salience of ambiguous stimuli, more easily classifying them as one explicit emotion versus the other. Eysenck et al.'s (2007) *attentional control theory* proposes that when (predominantly nonclinical) anxiety does not impair performance (i.e., continuity of responses to ambiguous stimuli) then compensatory strategies are likely engaged, and the *efficiency* of the system is reduced compared to non-anxious subjects. Efficiency equates to the relationship between performance effectiveness (measured by behaviour, e.g., response accuracy) and the effort exerted (e.g., quantified by psychophysiological measures or self-reports) or resources spent during the task (reported using neuroimaging techniques, e.g., reduced recruitment of executive-control system; Bishop et al., 2004; 2009). Additionally, individuals with state anxiety often experience worry and self-preoccupation, which may further consume attentional resources, or even engage self-regulatory processes to reduce negative thoughts (Eysenck et al., 2007). Thus, as performance was not impaired as a factor of state anxiety, we could infer that recruitment of regions in the FPCN may reflect engagement of compensatory mechanisms (Balderston et al., 2020; Basten et al., 2011; Ansari and Derakshan, 2011a; Fales et al., 2008, although see Bishop, 2009; Ansari and Derakshan, 2011b). This is evident in response to auditory ambiguous stimuli, as compensatory recruitment is most prominent under high task-load (Balderston et al., 2020). Therefore, when an *a priori* bias exists (and thus, the task is easier—i.e., visual ambiguous stimuli), the response will be made faster, and recruitment of high-order executive functions will be less.

Curiously, high-anxiety subjects demonstrated a heightened response of the amygdala to stimuli perceived as angry versus fear, in both modalities. Markedly, the opposite was true for low-anxious subjects. As anxiety can heighten vigilance to imposing threat (Kastner-Dorn et al., 2018; see review, Grupe & Nitschke, 2013), we propose that high-anxious subjects may present an "impulsive" amygdala response to expressions of direct threat. Anger demonstrates a clear source of threat and may increase the propensity for a fight or flight reaction. This is evident through, for example, heightened startle responses (Springer et al., 2007) or increased systolic blood pressure (Garfinkel et al., 2015). For these same reasons, anxious subjects are more sensitive (El Zein et al., 2015) and present greater amygdala activity (Ewbank et al., 2010) when facial expressions of anger exhibit a direct gaze, compared to fear.

5.5.4 Sex differences in the amygdala

Interestingly, the right amygdala in female subjects was most responsive to auditory stimuli; however, males instead responded more to visual. To our knowledge, no study has yet to report sex effects when investigating the perception of both auditory and visual threat-related stimuli. The amygdala is one of few neural structures that presents sexual dimophism, which is purportedly due to it sensitivity to sex steroids during critical periods of neural development (Goldstein et al., 2001). Further exploration is required using a larger sample size to confirm and understand this prospective three-way interaction of modality, sex, and amygdala recruitment.

5.5.5 Future directions and limitations

Several limitations became apparent, including the restrictive set of stimuli (1 male, 1 female identify per modality) that prevented a generalization of findings across modality. It is possible that stimulus identity may bias emotion perception, as evident in variability reported during emotion recognition (e.g., KDEF validation study; Goeleven, et al., 2008). Albeit, presentations of additional identities increases experiment duration, and could reduce subjects' attention during the task. Here, the study objective was to, instead, explore the within-subject brain-behaviour relationship, supporting the use of auditory and visual subject-specific stimuli.

In line with investigating subject-specific responses, the current assessment of state anxiety was limited to a single self-report measure. In addition, one could conduct MRI-specific fear assessments to indicate likelihood of adverse psychological responses while scanning (e.g., MRI-Fear Survey Schedule (FSS); Harris et al., 2004) and/or to include concurrent recordings of in-time physiological measures, such as electrodermal activity (Strohmaier et al., 2020) or heart rate (Kantor et al., 2001). These measures of transient stress could track a continual state of anxiety during testing and detection of temporal fluctuations. Moreover, manipulating the design to incorporate an assessment of test-retest intrasubject reliability would be critical to ensure reliability of individual differences (Elliot et al., 2020).

5.6 Conclusions

The study illustrates a link between brain and behaviour and provides a clear demonstration of how task-difficulty may influence emotion processing pathways when perceiving an emotionally salient stimulus. Findings illustrated that regions of the SN, FPCN and DMN are critical to emotional decision-making, recruited during evaluation of ambiguous stimuli. The amygdala, in contrast plays a prominent role in evaluating explicit emotion; however, also appears to modulate in response to ambiguity. Moreover, results illustrated a link between amygdala and the *task positive* network with state anxiety, as well as behavioural changes reflected by faster response times and greater mistakes made to explicit emotion. The experiment was designed to minimize confounds, and capture responses specific to each subject. These findings provide neural targets for further investigation into how emotion and cognition processes may interact, particularly for those who may be subject to heightened emotional perception, such those with anxiety disorders, post-traumatic stress disorder (PTSD) or panic disorders (Holzschneider and Mulert, 2011).

5.7 Acknowledgements

This study was supported by grants from the Natural Science and Engineering Research Council of Canada (NSERC, RGPIN-2017-05832), the Canadian Institutes of Health Research (CIHR, MOP-130516) and Global Affairs Canada to JLA. JCW was funded by a CRBLM Graduate Student Stipend, a McGill IPN Returning Student Award, an Alexander Graham Bell Canada Graduate Scholarship-Doctoral fellowship (CGS-D), and the Graduate Mobility Award funded by the Government of Quebec. IS was funded by Canadian Institutes of Health Research (CIHR, MOP-130516).

5.8 References

Ansari, T.L., & Derakshan, N. (2011a). The neural correlates of cognitive effort in anxiety: Effects on processing efficiency. *Biological Psychology, 86,* 337-348.

Ansari, T.L., & Derakshan, N. (2011b). The neural correlates of impaired inhibitory control in anxiety. *Neuropsychologia, 49,* 1146-1153.

Aubé, W., Angulo-Perkins, A., Peretz, I., Concha, L., & Armony, J.L. (2015). Fear across the senses: brain responses to music, vocalizations and facial expression. *Social Cognitive and Affective Neuroscience*, *10,* 399-407.

Balderston, N.L., Flook, E., Hsiung, A., Liu, J., Thongarong, A., Stahl, S., Makhoul, W., Sheline, Y., Ernst, M., & Grillon C. (2020). Patients with anxiety disorders rely on bilateral dlPFC activation during verbal working memory. *Social Cognitive and Affective Neuroscience, 15*, 1288-1298.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: AN investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders, 34*, 163-175.

Basten, U., Stelzel, C., & Fiebach, C.J. (2011). Trait anxiety modulates the neural efficiency of inhibitory control. *Journal of Cognitive Neuroscience*, *23,* 3132-3145.

Bechara, A. (2005) Decision making, impulse control and loss of willpower to resist drugs: A neurocognitive perspective. *Nature Neuroscience*, *8,* 1458-1463.

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Science*, *8,* 129-135.

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behaviour Research Methods, 40,* 531-539.

Bestelmeyer, P.E.G., Maurage, P., Rouger, J., Latinus, M., & Belin, P. (2014) Adaptation to vocal expressions reveals multistep perception of auditory emotion. *Journal of Neuroscience, 34,* 8098-8105.

Bishop, S.J. (2009). Trait anxiety and impoverished prefrontal control of attention. *Nature Neuroscience, 12,* 92-98.

Bishop, S.J., Duncan, J., & Lawrence, A. D. (2004). State anxiety modulation of the amygdala response to unattended threat-related stimuli. *Journal of Neuroscience, 24,* 10364-10368.

Boyle, G.J., Helmes, E., Matthews, G., & Izard, C.E. (2015). Measures of Affect Dimensions. In Boyle, G.J., Saklofske, D.H., Matthews, G. (Eds), *Measures of Personality and Social Psychological Constructs* (pp.190-224). Academic Press: Elsevier.

Brainard, D.H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433-436.

Breiter, H.C., Etcoff, N.L., Whalen, P.J., Kennedy, W.A., Rauch, S.L., Buckner, R.L., Strauss, M.M., Hyman, S.E., & Rosen, B.R. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron, 17,* 875-887.

Brooks, J.L. (2012). Counterbalancing for serial carryover effects in experimental condition orders. *Psychological, 17,* 600-614.

Brown, C.C., Raio, C.M., & Neta, M. (2017). Cortisol responses enhance negative valence perception for ambiguous facial expressions. *Scientific Reports, 7,* 1-8

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. In A. Kingstone & M. B. Miller (Eds.), *The year in cognitive neuroscience* (pp. 1–38). Blackwell Publishing.

Burges, C.J.C. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery 2*, 121-167.

Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A.R., Langner, R., & Eickhoff, S.B. (2012). Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Structure and Function*, *217,* 783-796.

Calder, A.J. (1996). Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear. *Cognitive Neuropsychology, 13*, 699-745.

Chand, G.B., & Dhamala, M. (2016). Interactions among the brain default-mode, salience, and central-executive networks during perceptual decision-making of moving dots. *Brain Connectivity*, *6*, 15254-15262.

Chang, C-C., Lin, C-J. & (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2,* 1-27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

Ciucurel, M.M. (2012). The relation between anxiety, reaction time and performance before and after sport competitions. *Procedia Social and Behavioral Sciences, 33,* 885-889.

Costafreda, S.G., Brammer, M.J., David, A.S., & Fu, C.H. (2008). Predictors of amygdala activation during the processing of emotional stimuli: A meta-analysis of 385 PET and fMRI studies. *Brain Research Reviews, 58,* 57-70.

Critchley, H., Daly, E., Phillips, M., Brammer, M., Bullmore, E., Williams, S., Van Amelsvoort, T., Robertson, D., David, A., & Murphy D. (2000). Explicit and implicit neural mechanisms for processing of social information from facial expressions: A functional magnetic resonance imaging study. *Human Brain Mapping*, 9, 93-105.

Deen, B., Koldewyn, K., Kanwisher, N., & Saxe, R. (2015). Functional organization of social perception and cognition in the superior temporal sulcus. *Cerebral Cortex, 25,* 4596-4609.

Dekel, R., & Sagi, D. (2019). Seeing, fast and slow: Effects of processing time on perceptual bias. *bioRxiv,* 1-25.

Dennis, E.L., Gotlib, I.H., Thompson, P.M., & Thomason, M.E. (2011). Anxiety modulates insula recruitment resting-state functional magnetic resonance imaging in youth and adults. *Brain Connectivity, 1,* 245-254.

Deupree, R.H., & Simon, J.R. (1963). Reaction time and movement time as a function of age, stimulus duration and task difficulty. *Ergonomics, 6,* 403-411.

Di, X., & Biswal, B.B. (2014). Modulatory interactions between the default mode network and task positive networks in resting-state. *PeerJ, 2*, e367.

Dolcos, F., Katsumi, Y., Moore, M., Berggren, N., de Gelder, B., Derakshan, N., Hamm, A.O., Koster, E.H.W., Ladouceur, C.D., Okon-Singer, H., Pegna, A.J., Richter, T., Schquizer, S., Van den Stock, J., Ventura-Bort, C., Weymar, M., & Dolcos, S. (2020). Neural correlates of emotion-attention interactions: From perception, learning and memory to social cognition, individual differences, and training interventions. *Neuroscience and Biobehavioural Reviews, 108,* 559-601.

Elliot, M.L., Knodt, A.R., Ireland, D., Morris, M.L., Poulton, R., Ramrakha, S., Sison, M.L., Moffit, T.E., Caspi, A., & Hariri, A.R. (2020). What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. *Psychological Science, 7,* 792-806.

El Zein, M., Wyart, V., & Grèzes, J. (2015). Anxiety dissociates the adaptive functions of sensory and motor response enhancements to social threats. *Elife, 4,* e10274.

Ester, E.F., Ho, T.C., Brown, S.D., & Serences, J.T. (2014). Variability in visual working memory ability limits the efficiency of perceptual decision making. *Journal of Vision, 14,* 1-12.

Evans, J.S.B.T, & Stanovich, K.E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8,* 223-241.

Ewbank, M.P., Fox, E., & Calder, A.J. (2010). The interaction between gaze and facial expression in the amygdala and extended amygdala is modulated by anxiety. *Frontiers in Human Neuroscience, 4,* 1-11.

Eysenck, M.W., Derakshan, N., Santos, R., & Calvo, M.G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion, 7,* 336-353.

Fales, C.L., Barch, D.M., Burgess, G.C., Schaefer, A., Gray, J.R., & Braver, T.S. (2008). Anxiety and cognitive efficiency: Differential modulation of transient and sustained neural activity during a working memory task. *Cognitive, Affective, & Behavioral Neuroscience, 8,* 239-253.

Fox E. (2002). Processing emotional facial expressions: The role of anxiety and awareness. *Cognitive and Affective Behaviour, 2,* 52-63.

Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., & Raichle, M.E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings in the National Academy of Sciences, 102,* 9673-9678.

Fransson, P. (2005). Spontaneous low-frequency BOLD signal fluctuations: An fMRI investigation of the state default mode of brain function hypothesis. *Human Brain Mapping, 26,* 15-29.

Gardumi, A., Ivanov, D., Hausfeld, L., & Valente, G. (2016). The effect of spatial resolution on decoding accuracy in fMRI multivariate pattern analysis. *Neuroimage, 132,* 32-42.

Garfinkel, S.N., Zorab, E., Navaratnam, N., Engels, M., Mallorquí-Bague, N., Minati, L., Dowell, N.G., Brosschot, J.F., Thayer, J.F., & Critchley, H.D. (2016). Anger in brain and body: The neural and physiological perturbation of decision-making by emotion. *Social Cognitive and Affective Neuroscience, 11,* 150-158.

Glasberg, B.R., Moore, B.C.J. (2002). A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society, 50,* 331-342.

Goeleven, E., De Raedt, R., Leyman, L., Verschuere, B. (2008). The Karolinska Directed Emotional Faces: A validation study. *Cognition and Emotion, 22,* 1094-1118.

Goldstein, K., Seidman, N.J., Horton, N., Makris, D.N., Kennedy, V.S., Caviness, V.S., Jr. Faraone, S.V., & Tsuang, M.T. (2001). Normal sexual dimorphism of the adult human brain assessed by in vivo MRI. *Cerebral Cortex, 11,* 490-497.

Green, D.M., & Swets, J.A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley and Sons.

Grupe, D.W., & Nitschke, J.B. (2013). Uncertainty and anticipation in Anxiety: An integrated neurobiological and psychological perspective. *Nature Reviews Neuroscience, 14,* 488-501.

Hariri, A.R., Bookheiner, S., & Mazziotta, J.C. (2000). Modulating emotion responses: Effects of neocortical network on the limbic system. *Neuroreport*, *11,* 43-48.

Harris, L.M., Cumming, S.R., & Menzies, R.G. (2004) Predicting anxiety in magnetic resonance imaging scans. *International Journal of Behavioural Medicine*, 11:1-7.

Hartley, C.A., & Phelps., E.A. (2012) Anxiety and decision-making. *Biological Psychiatry, 72,* 113-118.

Hein, G., & Knight, R.T. (2008). Superior temporal sulcus—It's my area: or is it? *Journal of Cognitive Neuroscience*, *20,* 2125-2136.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics, 6,* 65–70.

Holzschneider, K., & Mulert, C. (2011). Neuroimaging in anxiety disorders. *Dialogues in Clinical Neuroscience, 13,* 453-461.

Hutchins, T.L., Prelock, P.A., & Bonazinga, L. (2010) *The theory of mind inventory (ToMI). Available from:* http://www.theoryofmindinventory.com/download-tests-test-materials/.

Ito, T., Yokokawa, K., Yahata, N., Isato, A., Suhara, T., & Yamada, M. (2017). Neural basis of negativity bias in the perception of ambiguous facial expression. *Scientific Reports.*, *7,* 1-9.

Jakuszkowiak-Wojten, K., Landowski, J., Wiglusz, M.S., Cubala, W.J. (2015). Impulsivity in anxiety disorders. A critical review. *Psychiatry Danubina*, *27,* 452-455.

Kantor, L., Endler, N.S., Heslegrave, R.J., Kocovski, N.L. (2001). Validating self-report measures of state and trait anxiety against a physiological measure. *Current Psychology*, *20,* 207-215.

Kastner-Dorn, A.K., Andreatta, M., Pauli, P., & Wieser, M.J. (2018). Hypervigilence during anxiety and selective attention during fear: Using steady-state visual evoked potentials (ssVEPs) to disentangle attention mechanisms during predictable and unpredictable threat. *Cortex*, *106,* 120-131.

Kawahara, H., Masuda-Katsuse, I., de Cheveigné, A. (1999). Restructuring speech representations using pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction. *Speech Community, 27,* 187-207.

Keogh, R., & Pearson, J. (2011). Mental imagery and visual working memory. *PLoS One, 6,* e29221.

Keuken, M.C., Müller-Axt, C., Langner, R., Eickhoff, S.B., Forstmann, B.U., & Neumann, J. (2014). Brain networks of perceptual decision-making: An fMRI ALE meta-analysis. *Frontiers in Human Neuroscience, 8,* 1-14.

Kim, B., Yoon, S., Nakajima, R., Lee, R.J., Lim, H.J, Lee, Y-K., Choi, J-S., Yoon, B-J., Augustine, G.J. & Baik, J-H. (2018). Dopamine D2 receptor-mediated circuit from the central amygdala to the bed nucleus of the stria terminalis regulates impulsive behaviour. *Proceedings of the National Academy of* Sciences, *115,* E10730-E10739.

Kim, H., Somerville, L.H., Johnstone, T., Alexander, A.L., & Whalen, P.J. (2003). Inverse amygdala and medial prefrontal cortex responses to surprised faces. *Neuroreport, 14,* 2317-2322

Kingdom, F., & Prins, N. (2016). *Psychophysics: A practical introduction.* London: Elsevier Academic Press.

Ko, C.H., Hsieh, T-J., Wang, P-W., Lin, W-C., Yen, C-F., Chen, S-S., & Yen, J-Y. (2015). Altered gray matter density and disrupted functional connectivity of the amygdala in adults with internet gaming disorder. *Progress in Neuro-psychopharmacology and Biological Psychiatry, 57*, 185-192.

Lamichhane, B., Adhikari, B.M., & Dhamala, M. (2016a). The activity in the anterior insulae is modulated by perceptual decision-making difficulty. *Neuroscience, 327,* 79-94.

Lamichhane, B., Adhikari, B.M., & Dhamala, M. (2016b). Salience network activity in perceptual decisions. *Brain Connectivity, 6,* 1-14.

Lamichhane, B., & Dhamala, M. (2015). The salience network and its functional architecture in a perceptual decision: An effective connectivity study. *Brain Connectivity, 5,* 1-9.

Lange, K., Williams, L.M., Young, A.W., Bullmore, E.T., Brammer, M.J., Williams, S.C.R., Gray, J.A., & Phillips, M.L. (2003). Task instructions modulate neural responses to fearful facial expressions. *Biological Psychiatry, 53,* 226-232.

Leal, C.P., Goes, T.C., Ferreira da Silva, L.C., Teixeira-Silva, F. (2017). Trait and state anxiety in different threatening situations. *Trends in Psychiatry and Psychotherapy, 39,* 147-157.

Liebman, S.E., & Allen, G.J. (1995). Anxiety sensitivity, state anxiety and perceptions of facial emotions. *Journal of Anxiety Disorders, 9,* 257-267.

Lin, H., Müller-Bardorff, M., Gathmann, B., Brieke, J., Mothes-Lasch, M., Bruchmann, M., Miltner, W.H.R., & Straube, T. (2020) Stimulus arousal drives amygdalar responses to emotional expressions across sensory modalities. *Scientific Reports, 10,* 1-12.

Loebach Wetherell, J., Reynolds, C.A., Gatz, M., & Pedersen, N.L. (2002). Anxiety, cognitive performance, and cognitive decline in normal aging. *The Journal of Gereontology: Series B, 57,* 246-255.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska Directed Emotional Faces – KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630-7164-9.

Macmillan N.A., & Creelman, C.D. (2004). *Detection Theory: A user's guide.* New York: Psychology Press.

Mars, R.B., Neubert, F-X., Noonan, M.P., Sallet, J., Toni, I., & Rushworth, M.F.S. (2012). On the relationship between the "default mode network" and the "social brain". *Frontiers in Human Neuroscience, 6,* 1-9.

Mattavelli, G., Sormaz, M., Flack, T., Asghar, A.U.R., Fan, S., Frey, J., Mansseur, L., Usten, D., Young, A.W., & Andrews, T.J. (2014). Neural responses to facial expressions support the role of the amygdala in processing threat. *Social Cognitive and Affective Neuroscience, 9,* 1684-1689.

Matthews, G., Panganiban, A.R., Hudlicka, E. (2011). Anxiety and selective attention to threat in tactical decision-making. *Personality and Individual Differences, 50,* 949-954.

Menon, V. (2011). Large-scale brain networks and psychopathology: A unifying triple network model. *Trends in Cognitive Science, 15,* 483-506.

Menon, V. (2015). Salience Network. In Toga AW (Ed.), *Brain Mapping: An Encyclopedic Reference Vol. 2.* (p. 597-611). Academic Press: Elsevier.

Menon, V., & Uddin, L.Q. (2010). Saliency, switching, attention and control: A network model of insula function. *Brain Structure and Function, 214*, 655-667.

Mogg, K., Bradley, B.P. (2018). Anxiety and threat-related attention: Cognitive-motivational framework and treatment. *Trends in Cognitive Sciences*, *22*, 225-240.

Morris, J.S., Frith, C.D., Perrett, D.I., Rowland, D., Young, A.W., Calder, A.J., & Dolan, R.J. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature, 383,* 812-815.

Munn, Z., & Jordan, Z. (2011). The patient experience of high technology medical imaging: A systematic review of the qualitative evidence. *Radiography, 17*, 323-331.

Mutschler, I., Wieckhorst, B., Meyer, A.H., Schweizer, T., Klarhöfer, F.H., Wihelm, F.H., Seifritz, E., & Ball, T. (2014). Who gets afraid in the MRI-scanner? Neurogenetics of state-anxiety changes during an fMRI experiment. *Neuroscience Letters, 583,* 81-86.

Nelissen, N., Stokes, M., Nobre, A.C., & Rushworth, M.F.S. (2013). Frontal and parietal cortical interactions with distributed visual representations during selective attention and action selection. *Journal of Neuroscience, 33,* 16443-16458.

Neta, M., Kelley, W.M., & Whalen, P.J. (2013). Neural responses to ambiguity involve domain-general and domain-specific emotion processing systems. *Journal of Cognitive Neuroscience, 25,* 547-557.

Orquin, J.L., & Loose, S.M. (2013). Attention and choice: A review on eye movements in decision-making. *Acta Psychologica,* 144, 190-206.

Paulus, M.P., & Stein, M.B. (2006). *Biological Psychiatry, 60*, 383-387.

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437-442.

Peters, S.K., Dunlop, K., & Downar, J. (2016). Cortico-striatal-thalamic loop circuits of th salience network: A central pathway in psychiatric disease and treatment. *Frontiers in Systems Neuroscience, 10*, 1-23.

Phelps, E.A., & LeDoux, J.E. (2005). Contributions of the amygdala to emotion processing: From animal models to human behaviour. *Neuron, 48,* 175-187.

Phillips, M.L., Young, A.W., Scott, S.K., Calder, A.J., Andrew, C., Giampietro V., Williams, S.C., Bullmore, E.T., Brammer, M., & Gray, J.A. (1998). Neural response to facial and vocal expressions of fear and disgust. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 265*, 1809-1817.

Posner, M.I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32,* 3-25.

Prins, N., Kingdom, F.A.A. (2018). Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamades toolbox. *Frontiers in Psychology, 9,* 1-14.

Raichle, M.E., MacLeod, A.M., Snyder, A.Z. Power, W.J., Gusnard, D.A., & Shulman, G.L. (2001) A default mode of brain function. *Proceedings of the National Academy of Sciences, 98,* 676-682.

Richards, A., Holmes, A., Pell, P.J., & Bethell, E.J. (2013). Adapting effects of emotional expression in anxiety: Evidence of late positive potential. *Society for Neuroscience*, 8, 650-664.

Saviola, F., Pappaianni, E., Monti, A., Grecucci, A., Jovicich, J., & De Pisapia, N. (2020). Trait and state anxiety are mapped differently in the human brain. *Scientific Reports, 10,* 1-11.

Schrouff, J., Monteiro, J.M., Portugal, L., Rosa, M.J., Phillips, C., & Mourão-Miranda, J. (2018). Embedding anatomical or functional knowledge in whole brain multiple kernel learning models. *Neuroinformatics, 16*, 117–143.

Schrouff, J., Rosa, M.J., Rodina, J.M., Marquand, A.F., Chu, C., Ashburner, J., Phillips, C., Richiardi, J., & Mourao-Miranda, J. (2013). PRoNTo: Pattern recognition for neuroimaging toolbox. *Neuroinformatics, 11,* 319–337.

Scott, S.K., Young, A.W., Calder, A.J., Hellawell, D.J., Aggleton, J.P., & Johnson, M. (1997). Impaired auditory recognition of fear and anger following bilateral amygdala lesions. *Nature, 385*, 254-257.

Seeley, W.W., Menon, V., Schatzberg, A.F., Keller, J., Glover, G.H., Kenna, H., Reiss, A.L., & Greicius, M.D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience, 27*, 2349-2356.

Sergerie, K., Chocol, C., & Armony, J.L. (2008). The role of the amygdala in emotional processing: A quantitative meta-analysis of functional neuroimaging studies. *Neuroscience & Biobehavioural Reviews,* 32, 811-830.

Sergerie, K., Lepage, M., & Armony, J.L. (2006). A process-specific functional dissociation of the amygdala in emotional memory. *Journal of Cognitive Neuroscience, 18,* 1359-1367.

Sergerie, K., Lepage, M., & Armony, J.L. (2007). Influence of emotional expression on memory recognition bias: A functional magnetic resonance imaging study. *Biological Psychiatry, 62,* 1126-1133.

Setsompop, K., Gagoski, B. A., Polimeni, J. R., Witzel, T., Wedeen, V. J., & Wald, L.

L. (2012). Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magnetic Resonance in Medicine, 67,*1210-1224.

Shrooten, M.G., Smulders, F.T.Y., Mogg, K., Bradley, B.P. (2012). Anxiety-related biases in visual orienting and spatial motor response selection independently assessed by a probe-classification task. *Journal of Experimental Psychopathology, 3,* 393-408.

Skuk, V.G., & Schweinberger, S.R. (2013). Adaptation aftereffects in vocal emotion perception by expressive faces and voices. *PLoS One, 8,* e81691.

Somerville, L.H., Kim, H., Johnstone, T., Alexander, A.L., & Whalen, P.J. (2004). Human amygdala responses during presentation of happy and neutral faces: Correlations with state anxiety. *Biological Psychiatry, 55,* 897-903.

Smith, C., Hancock, H., Blake-Mortimer, J., & Eckert, K. (2007). A randomized comparative trial of yoga and relaxation to reduce stress and anxiety. *Complementary Therapies in Medicine, 15*, 77-83.

Smith, P.L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of mathematical psychology*, *44,* 408-463.

Spielberger, C.D. (1983). Manual for the State-Trait Anxiety Inventory STAI (form Y) ("Self-Evaluation Questionnaire"). Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C.D. (1995). *State-Trait Personality Inventory Research Manual Sampler Set*. Menlo Park, CA: Mind Garden Inc.

Spielberger, C.D., & Reheiser, E.C. (2009). Assessment of emotions: Anxiety, anger, depression and curiosity. *Applied Psychology: Health and Wellbeing*, *1,* 271-302.

Spielberger, C.D., Ritterband, L.M., Sydeman, S.J., Reheiser, E.C., & Unger, K.K. (1995). Assessment of emotional states and personality traits: Measuring psychological vital signs. In J.N. Butcher (Ed.) *Clinical personality assessment: Practical approaches* (pp. 42-58). New York: Oxford University Press.

Springer, U.S., Rosas, A., McGetrick, H., & Bowers, D. (2007). Differences in startle reactivity during the perception of angry and fearful faces. *Emotion*, *7,* 516-525.

Sreenivas, S., Boehm, S.G., & Linden, D.E.J. (2012). Emotional faces and the default-mode network. *Neuroscience Letters, 506*, 229-234.

Sreenivasan, K.K., Curtis, C.E., & D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Sciences*, *18,* 82:89.

Sridharan, D., Levitin, D.J., Menon, V. (2008). A critical role for the right fronto-insular cortex in switching between central-executive and default mode networks. *Proceedings of the National Academy of Sciences*, *105,* 12569-12574.

Starns, J.J., & Ma, Q. (2018). Response biases in simple decision making: Faster decision making, faster repsonse execution, or both? *Psychonomic Bulletin & Review, 25,* 1535-1541.

Stein, M.B., Simmons, A.N., Feinstein, J.S., & Paulus, M.P. (2007). Increased amygdala and insula activation during emotion processing in anxiety prone subject. *The American Journal of Psychiatry, 164,* 318-327.

Sterzer, P. (2016). Moving forward in perceptual decision making. *Proceedings of the National Academy of Sciences of the United States, 113*, 5771-5773.

Strohnmaier, A.R., Schiepe-Tiska, A., & Reiss, K.M. (2020). A comparison of self-reports and electrodermal activity as indicators of mathematics state anxiety. An application of the control-value theory. *Frontline Learning Research*, *8*, 16-32.

Surcineli, P., Codispoti, M., Montebarocci, O., Rossi, N., & Baldaro, B. (2006). Facial emotion recognition in trait anxiety. *Journal of Anxiety Disorders, 20,* 110-117.

Sussman, T.J., Jin, J., Mohanty, A. (2016). Top-down and bottom-up factors in threat-related perception and attention in anxiety. *Biological Psychology, 121,* 160-172.

Thielscher, A., & Pessoa, L. (2007). Neural correlates of perceptual choice and decision making during fear-disgust discrimination. *Journal of Neuroscience*, *27*, 2908-2917.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joilioy, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, *15,* 273-289.

Van der Zwaag, W., Da Costa, S.E., Zürcher, N.R., Adams, R.B., & Hadjikhani, N. (2012). A 7 Tesla fMRI study of amygdala responses to fearful faces. *Brain Topography, 25*, 125-128.

Vuilleumier P., Armony, J.L., Driver, J., & Dolan, R.J. (2001). Effects of attention and emotion on face processing in the brain: An event-related fMRI study. *Neuron, 30,* 829-841.

Wang, S., Tudusciuc, O., Mamelak, A.N., Ross, I.B., Adolphs, R., & Rutishauser, U. (2014). Neurons in the human amygdala selective for perceived emotion. *Proceedings of the National Academy of Sciences*, *111,* E3110-E1119.

Wang, S., Yu, R., Tyszka, J.M., Zhen, S., Kovach, C., Sun, S., Huang, Y., Hurlemann, R., Ross, I.B., Chung, J.M., Mamelak, A.N., Adolphs, R., & Rutishauser, U. (2017). The human amygdala parametrically encodes the intensity of specific facial emotions and their categorical ambiguity. *Nature Communications, 8,* 14821.

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,* 4913-4924.

Whitehead, J.C., & Armony, J.L. (2019). Multivariate fMRI pattern analysis of fear perception across modalities. *European Journal of Neuroscience, 49,* 1552-1563.

Williams, J.M.G., Watts, F.N., Macleod, C., & Matthews, A. (1988). *Cognitive psychology and emotional disorders, 2nd Edition*. Chichester, England: Wiley.

Wyble, B., Sharma, D., & Bowman, H. (2008). Strategic regulation of cognitive control by emotional salience: A neural network model. *Cognition and Emotion, 22,* 1019-1051.

Xia, L., Gu, R., Zhang, D., & Luo, T. (2017). Anxious individuals are impulsive decision-makers in the delay discount dark: An ERP study. *Frontiers in Behavioral Neuroscience, 11, 1-11.*

Xie, C., Shao, Y., Fu, L., Goveas, J., Ye, E., Li, W, Cohen, A.D., Chen, G., Zhang, Z., & Yang, Z. (2011). Identification of hyperactive intrinsic amygdala network connectivity associated with impulsivity in abstinent heroin addicts. *Behavioural Brain Research, 216,* 639-646.

Yeo, B.T.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., Fischl, B., Liu, H., & Buckner, R.L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology, 106,* 1125-1165.

5.9 Supplementary Materials

5.9.1 Supplementary Tables

Supplementary Table 5.1 Pre-scan and scan session mean response times and responses for each perceived emotion and intensity.

Session	*Auditory* Mean Anger Responses *(SD)*	*Auditory* Mean logRTs *(SD)*	*Visual* Mean Anger Responses *(SD)*	*Visual* Mean logRTs *(SD)*
Pre-Scan				
High Fear	.01(.03)	-.47(.22)	.01(.03)	-.45(.15)
Low Fear	.11(.08)	-.28(.28)	.04(.03)	-.36(.26)
Mid-morph	**.59(.21)**	**-.17(.32)**	**.21(.16)**	**-.17 (.29)**
Low Anger	.93(.07)	-.39(.30)	.749 (.21)	-.25 (.23)
High Anger	.98(.02)	-.56(.25)	.97(.03)	-.46(.23)
Scan				
Fear	.03(.04)	-.32(.20)	.01(.02)	-.44(.18)
Ambiguous	**.49(.21)**	**-.03(.22)**	.32(.16)	**-.15(.24)**
Response Fear	**.51(.21)**	**-.03(.24)**	**.68(.16)**	**-.20 (.24)**
Response Anger	**.49(.21)**	**-.04(.20)**	**.32(.16)**	**-.09(.24)**
Anger	.97(.03)	-.40(.21)	.94(.07)	-.43(.21)

Supplementary Table 5.2 Percent overlap of activity clusters (K_E>10) in response to Ambiguous vs. Explicit emotion, with 7-network cortical parcellation (Yeo et al. 2011). Red text highlights regions that overlap 20% or greater with the corresponding network. VN: Visual Network; SMN: Somatomotor Network DAN: Dorsal Attention Network; SN: Salience Network; LN: Limbic Network; FPCN: Frontoparietal Control Network; DMN: Default Mode Network.

Anatomical Location	**Yeo et al. (2011) 7-Network Parcellation (% overlap)**							**K_E**
	VN	**SMN**	**DAN**	**SN**	**LN**	**FPCN**	**DMN**	
L/R SMA/ACC	0%	0%	0%	52%	0%	55%	2%	1392
L Insula/dlPFC	0%	0%	2%	23%	0%	57%	13%	1410
R Insula/dlPFC	0%	0%	13%	23%	0%	38%	3 %	1890
L IPL	0%	0%	84%	0%	0%	16%	0%	51
R IPL	0%	0%	42%	0%	0%	58%	0%	33
L Occipital cortex	90%	0%	0%	0%	0%	0%	0%	523
R Occipital cortex	76%	0%	0%	0%	0%	0%	0%	257
Total (92.3%)	**11.9%**	**0%**	**6.0%**	**26.5%**	**0%**	**43.1%**	**4.8%**	**5556**
L Angular gyrus	0%	0%	8%	0%	0%	6%	80%	1532
R Angular gyrus	1%	0%	16%	0%	1%	9%	68%	733
L/R vmPFC	0%	0%	0%	0%	6%	4%	76%	3315
R vmPFC	0%	1%	0%	11%	0%	11%	76%	255
L/R precuneus/PCC	0%	2%	2%	6%	0%	14%	69%	1057
L MTG	0%	0%	0%	0%	6%	0%	89%	605
R MTG	0%	0%	0%	0%	0%	25%	74%	221
L Fusiform gyrus	86%	0%	0%	0%	0%	0%	0%	65
R Fusiform gyrus	86%	0%	1%	0%	0%	0%	13%	87
Total (91.1%)	**1.9%**	**0.3%**	**3.4%**	**1.3%**	**3.0%**	**6.7%**	**74.6%**	**7870**

Chapter 6. General Discussion

The intent of this thesis was to expose complexities of socio-emotional perception, where through advanced methodologies, we reported recruitment of widespread multifunctional structures during passive and task-based studies of perception. Evidence suggests influences of individual differences, task demands, and an interaction of the two; insight that may collectively contribute to the ongoing isolation of dispersed neural targets/systems of dysfunction. Notably, we manipulate experimental design, acquisition, and analysis throughout, to provide assurance of validity and reliability. Through exploring the neural response to full-spectrum auditory (i.e., music, singing, speech) and across-modality (anger-fear morphs of faces and voices) socio-emotional stimuli we improve specificity for detecting subtle differences across conditions.

6.1 Summary of the findings

Findings are reported in three-fold, as directed by the three objectives of the thesis, First, we provide evidence of functional heterogeneity and distributive processing to demonstrate the complexity of socio-emotional perception. We explore how signals of a similar nature (i.e., music versus prosody) diverge at the voxel level—potentially reflecting differential patterns of fine-grained spectral and temporal tuning of neurons. Nevertheless, we argue against domain-*specificity*, where regions are reported to independently process either music or voice (Peretz & Coltheart, 2003). Instead, we offer evidence for domain-*preferred* activity, where domain-general structures exhibit some degree of partiality via variable response magnitudes. We strengthen this argument of neural heterogeneity, by demonstrating that across-modality, socio-emotional signals may reach the same higher-order cortical and subcortical processing regions to output a consistent categorical percept. Moreover, during both passive and active (i.e., decision-making) emotion perception, we observe widespread recruitment of multifunctional regions that collectively comprise validated functional networks (Yeo et al., 2011). Notably, through perceptual decision-making, we report fluctuations of said networks dependent upon bottom-up (emotional salience) versus top-down (task-difficulty) attentional demands.

Second, in complement to the above-mentioned findings, we explored how this extensive neural recruitment may differ dependent upon individual variability within a healthy population. Investigating expert proficiencies in music and language have shown, that even during passive perception, one can identify reliable, within-domain neural distinctions modulated by expertise.

Additionally, we use measures of state anxiety to investigate deficits of affect perception in healthy subjects, adding experimental depth through manipulating aspects of threat and uncertainty. We report that individuals with high levels of state anxiety present signs of impulsivity and accentuated perceptual biases. These conclusions are evident through observed differences of reaction times, increased amygdala recruitment, as well as regions in the Task-Positive Network (Salience and Frontoparietal Control) that are largely responsible for directing attention, problem solving and decision-making (see for review, Boyatzis et al., 2014).

Finally, by using methods systematically designed to reach each objective we were able to validate the use of intermediary stimuli, report that continuous sampling is sufficient for studies of auditory perception, and that conducting test-retest reliability analyses are invaluable and ought to be a normative practice within the context of fMRI research. Finally, we provided evidence for the complementary use of univariate and multivariate analyses to approach unique questions and promote a holistic understanding of the brain.

6.2 Functional heterogeneity and distributive processing in the brain

The terms *degeneracy* (many-to-one mapping; Edelman & Gally, 2001) and *neural reuse* (one-to-many mapping; Anderson, 2010; 2014) have been used to highlight two means by which the brain exemplifies complexity, specifically in socio-emotion research (Barrett, 2017a). The current section extends upon this framework, providing evidence for multifunctional (*neural reuse*), dispersed neural recruitment in processing distinct signals as to form a single uniform perception (*degeneracy*). This evidence is reported in both an auditory *(Studies 1 and 2)* and across-modality (auditory and visual; *Studies 3 and 4*) context.

6.2.1 Diverging perception: music and voice

6.2.1.1 A stored musical representation

Disentangling signals of music and voice in the brain to ultimately isolate perceived musicality, often involves identifying, controlling, and manipulating acoustic attributes related to duration, intensity, or frequency of the sound. Reports of neural separation in response to speech and singing (Callan et al., 2006; Schön et al., 2010) validated our methodological rational for *Study 1*—controlling the common acoustic vocal traits to isolate musicality. Using stimuli of instrumental music, vocal music (singing) and speech, we observed regions preferentially responsive to either voice, instrumental tones, or musicality (of both domains). We propose that regions sensitive to domain-general musicality are specially tuned to a unique (musicality) acoustic profile. I also offer the idea that these regions are likely to interact with functional systems/networks related to emotion, salience, and sequencing.

Animal models, particularly, non-human primates have provided us with insight into the flow of information throughout the AC. The AC receives most input from the medial geniculate body of the thalamus, where the signal then travels in a feedforward lateral direction (as well as feedback; de la Mothe et al., 2006) from the core to primarily the belt, with dense connections to the surrounding parabelt (Hackett et al., 1998). In this direction, neurons become increasingly sensitive to sound complexity. For instance, the PP and neighbouring lateral HG and STS, respond more to complex versus fixed pitch manipulations (Hall et al., 2000; Patterson et al., 2002). Congruently, greater activity is reported in mid- and anterolateral-HG, and PP, versus posteromedial-HG and PT when perceiving rich harmonic tones, (Norman-Haignere et al., 2013).

In *Studies 1 and 2*, a sensitivity to instrumental music (as compared to speech prosody) was reported in rostral (bilateral PP) and caudal belt regions (right PT; Angulo-Perkins et al., 2014; Norman-Haignere et al., 2015; Leaver & Rauschecker, 2010; Angulo-Perkins & Concha, 2019). Moreover, the PP was further segregated, by preference for vocal music anteriorly, and instrumental music posteriorly. Demarcating this distinction was a novel contribution to the field and moreover, challenged the idea that the PP best responds to timbre alone (Leaver & Rauschecker, 2010). In fact, findings suggest that no single acoustic parameter can define activity of music-preferred regions, and rather, the PP is likely most responsive to a pattern of acoustic attributes unique to musicality. Alike the necessary configural processing of emotion expressed by the body or face (see Chapter 2; Brandman & Yovel, 2016; Lai et al., 2014), we suggest *(Study 1)* that a weighted model of parameters may collectively define musicality, and transition perceived speech into song (Saitou et al., 2004; 2007; Livingstone et al., 2013).

Peretz & Coltheart (2003) propositioned the idea of a "musical lexicon"—a stored representation of music in the brain, based off a lifetime of experiences and exposure to musical sequences. This definition is congruous with Friston & Friston (2013)'s free energy principle and active inference to music, where one attempts to minimize prediction error of the outside world (e.g., musical sequences) by maximizing their stored knowledge and experiences. Thus, we could infer that activity observed in non-primary AC in response to music, reflects the matched prediction of subjects' own stored "musical representation". Like sensitivities to face configuration in the FFA (Zhang et al., 2012), this may explain why activity in the anterior STG, PP and posterior STG/PT more prominently responded to sequential versus scrambled music (Fedorenko et al., 2012; Angulo-Perkins & Concha, 2019). Based on the location of identified music-preferred regions, we propose that tuning and specialization of neurons may reflect nearby, connecting structures, such as those involved in sequence processing.

6.2.1.2 Music processing: emotion, saliency, and sequencing

From the primate AC, information moves rostrally to temporal poles, caudally to the TPJ and posterior parietal regions (see for review, Hackett, 2011), or along the opposing axis—laterally towards the STS, or medially to the insula (de la Mothe et al., 2006; Hackett et al., 1998). The direction in which the signal travels is dependent upon the type of information transmitted. This is determined by a combined input from thalamic nuclei, of which each transmit a specific element of the signal (Hackett et al., 1998). Neural signals travelling along the rostral or *what* pathway tend to have slow latencies and sustained responses, with heavy feedforward and feedback connectivity, mediating processes of signal recognition, integration, and perception. Neurons along the caudal or *where/how* path, instead, respond quickly and detect modulation of sound sequences more readily. In doing so, caudal neurons are more prepared to direct ensuing actions, especially given the connection to motor system processes (see review, Jasmin et al., 2019).

Notably, musicians with absolute pitch have shown increased resting-state functional connectivity medially, between the right PP and insula (Kim & Knösche, 2017); a region involved in auditory object recognition (Binder et al., 2004) and a critical node of the salience network (Menon & Uddin, 2010; see *Study 4*). Markedly, this increased connectivity is also apparent as individuals listen to emotional (especially fearful) versus neutral music (Koelsch et al., 2018). The insula, reported in *Study 2*, exhibited a preferred response to music over prosody, alongside the PP *(Studies 1 and 2).* Interestingly, both regions contributed to an above-chance classification of vocalizations versus prosody in *Study 3*, as well as in distinguishing the neural response to fear versus neutral conditions (*Study 3*). Based on our collective findings, we question whether the salience of an auditory signal may, to some degree, determine combined recruitment of the PP and insula, particularly as it relates to emotional salience. For instance, activity of the insula in *Study 2* versus *1*, may reflect the two emotion-based runs that were uniquely incorporated into *Study 2*.

An event-related potential (ERP) study demonstrated greater P2 amplitude in response to emotional music versus vocalizations or speech (Paquette et al., 2020)—an ERP related to attentional changes that occur based on the salience of the stimulus (Iredale et al., 2013; Paulmann et al., 2013). Moreover, according to Faber & Fiveash (2014), emotion in music is perceived more intensely and consistently than in prosody, and is purportedly more memorable (Haiduk et al., 2020). Friston & Friston (2013) suggest that as the anterior insula and OFC activate in response to unexpected musical chord presentations (Koelsch et al., 2005), that emotion may therefore reflect violations of musical predictions (Gold et al., 2019). Given this insight, I agree that the PP may function as an intermediary processing site between primary AC and association areas (Angulo-Perkins & Concha, 2019), relaying relevant, salient acoustic information that ultimately distinguishes music, prosody, and vocalizations.

Comparable to the proposed "intermediary" function of the PP, the PT is commonly recognized as a computational hub, processing information of high spectro-temporal complexity, segregating sound objects and isolating their source location (as required in auditory scene analyses; Griffiths & Warren, 2002). The music-preferred response in the PT (*Studies 1 and* 2) is reportedly due to its involvement in processing sequences of complex sounds, comparable to those of motor movements (Rauschecker, 2011). Interestingly, Schön et al. (2008) observed that in learning a new language, individuals were more successful in learning speech through sung language as compared to spoken, assumingly due to the added pitch and tonal changes between syllables that aid in phonological segmentation (based on gestalt principles; Deutsch, 1999). In line with sound segmentation, deviancies to a consistent sound pattern elicited increased activity in the right PT and TPJ. To modulate temporal integration of auditory sequences, these regions would require the formation of short-term auditory memory traces, to associate past and present auditory information, particularly as these signals are received both in parallel and serially (Mustovic et al., 2003). Like others, their findings reflect regions also known to process auditory motion and space (Baumgart et al., 1999; Warren et al., 2002). Thus, we propose that the PT (like the PP) may act as a midway processor, with a particular role in sequencing auditory information.

Interestingly, *Study 2* also demonstrated music-preferred activity within the SMA; a heterogenous structure, where rostral regions share significant connections with PFC, and caudal activity is associated more to activity of the PMC (Nachev et al., 2008). Thus, it appears that the SMA plays both a role in sequence processing (functioning within domains such as linguistic, motor, numerical, temporal; see for review, Cona & Semenza, 2017) as well as in salience evaluation (see *Study 4*). Notably, regions of the salience network (SN) have shown greater functional connectivity after individuals receive musical training (Zamorano et al., 2017; Luo et al., 2014). With this framework, we can conceive as to why music may be processed differentially. Although music and voice share similar spectro-temporal cues (Juslin & Laukka, 2003) and perform comparable communication functions; the "music code" is likely to provide unique feedback to the internal model via specialized music-preferred neurons. Through the standpoint of a *distributed*-versus *sparse-coding model* (Bizley & Cohen, 2014), I propose that model predictions likely incorporate stored knowledge of emotion-, salience- or sequence-related information, acquired from associated structures and systems. Finally, I do not deny that these regions respond to the "voice-code", only that the model prediction (i.e., pattern of neural tuning) is best matched to the "music-code".

6.2.2 Converging perceptions: emotion across modality

Studies 1 and 2 support the hypothesis that regional specialization may, in part, be influenced by connecting structures and that, likely, unique patterns of activity signal "musicality". Here, we draw similar conclusions, recognizing functional links between distributed structures of emotion perception through multivoxel pattern analyses, and acknowledging the collective overlap with relevant functional networks. We suggest that the observed multifunctionality of the amygdala, may too, reflect its network of far-reaching connections (see for review, LeDoux, 2007; Freese & Amaral, 2009).

6.2.2.1 The multisensory, multifunctional amygdala

The amygdala is arguably, best known for its role as a detector, whether *threat* (Öhman, 2005), *intensity* (Bonnet et al., 2015), *motivational* (Pessoa & Adolphs, 2010) or *behavioural relevance* (see for review Sander, Grafman, et al., 2003; Murray et al., 2014). In some of the earliest stimulation studies of the cat amygdala, a reliable attention response occurred when probed, where cats immediately stopped behaviour (e.g., walking, licking) and became alert to the environment (i.e., raised head, inquisitive; Ursin & Kaada, 1960). This observed *vigilance* of the amygdala is, arguably, defining of its function (Whalen, 1998; Pessoa, 2010; Davis & Whalen, 2001; Terburg et al., 2012) and its relation to species survival (see for review, Mobbs et al., 2015). It rationalizes why highly salient information may elicit a neural response, even as the strong inhibitory circuitry of the amygdala maintains low spontaneous activity. This inhibition is necessary to thwart inappropriate action potential firing and to habituate to repeated stimulation (see for review, LeDoux, 2007). From *Studies 3 and 4*, it is evident that the amygdala is not only involved in the passive perception of explicit threat-related expressions, but also to the active classification of ambiguous threat-related expressions, fluctuating in accord with networks involved in allostasis (i.e., SN and DMN; Kleckner et al., 2017). Due to the amygdala's reported involvement in responding to ambiguity and uncertainty (Freeman et al., 2014; Wang et al., 2017; Whalen, 1998; Hsu et al., 2005; Herry et al., 2007), I conclude that the amygdala may be sensitive to the *anticipation* of threat, in addition to its well-documented *reactive* response to threat-related information (see for review, Adolphs, 2008). Previous literature has suggested that these unique functions likely occur in different subnuclei of the amygdala (Wang et al., 2017). Where, divergently, these subnuclei are also reported to be heterogenous; involved in processing multisensory information (Domínguez-Borràs et al., 2019; Morrow et al., 2019). Due to the limited spatial resolution of fMRI, we cannot confirm that the same "multisensory" neurons are involved, however, *Studies 3 and 4* provide evidence that strongly supports the possibility.

The structural and functional heterogeneity of the amygdala (Ball et al., 2007) is emphasized through these unique subdivisions. The basolateral complex is, in large part, the receiving site of the amygdala; sending excitatory glutamatergic projections either directly to the CM, or indirectly to GABAergic intercalated mass cells, which then send feedforward inhibitory signals to the CM, or back to the BL. In turn inhibitory projections from the CM become disinhibited, signaling to the brainstem and hypothalamus to initiate processes responsible for producing emotional and associated physiological responses (see for review, LeDoux, 2007; Davis, 1992).

The amygdala receives input from multiple sensory systems, where, specifically, auditory, and visual signals are transmitted either via the thalamus or cortically through high-level association areas, such as the PFC or STS (LeDoux, 2007; Sah et al., 2003). It is in these association areas, where multiple modalities may integrate into a single signal (Sah et al., 2003). Whether this occurs at the level of the amygdala has remained under scrutiny. In support of multisensory amygdala processing reported in *Studies 3 and 4*, it has been shown, that approximately 40% of neurons in the primate amygdala respond to a single modality, while about 60% are multisensory and dispersed across various subnuclei (Morrow et al., 2019). Equivalently, neurons sensitive to both auditory and visual information, comprise the majority of reported sensory-receptive amygdala neurons in humans (Domínguez-Borràs et al., 2019). The large number of multisensory neurons may be explained by a convergence of inputs, or rather, a capacity to receive already integrated multisensory signals (Morrow et al., 2019). In agreement with direct intracranial recordings, and our findings in *Studies 3 and 4,* other fMRI studies have presented comparable recruitment of the amygdala in response to arousing (versus neutral) expressions of faces and voices alike (Lin et al., 2020; Aubé et al., 2015).

6.2.2.2 Processing threat-related information: from regions to networks

Distributed contribution from regions across the brain in distinguishing the perception of threat-related versus neutral or ambiguous expressions across modalities (*Studies 3 and 4),* is indicative of common processing networks comprised of cortical and subcortical regions, including, but not limited to the amygdala. In *Studies 3 and 4,* multivariate findings reported high-weighted supramodal subcortical regions that contributed to emotion classification consistently across studies. These included the amygdala, putamen, and thalamus, while cortical regions extended to those of the prefrontal cortex, superior parietal lobe, and the insula; a region directly involved in emotion awareness (Simmons et al., 2013). In agreement, other cross-modal (auditory and visual) emotion perception studies also reported involvement of the mPFC (Peelen et al., 2010; Kim et al., 2017), PCC (Kim et al., 2017; Lin et al., 2020), MFG (Kim et al., 2017; Lin et al., 2020), insula (Lin et al., 2020; Aubé et al., 2015), amygdala (Lin et al., 2020; Aubé et al., 2015) and the posterior STS/STG (Peelen et al., 2010; Sievers et al., 2018; Kim et al., 2017; Lin et al., 2020) in distinguishing emotions, valence or arousal, across at least two modalities. Curiously, even with comparable multivariate approaches, *Study 3* did not exhibit involvement of the posterior STS; a recognized site of multisensory integration (Beauchamp, Argall et al., 2004; Beauchamp, Lee et al., 2004; Beauchamp et al., 2008). Dricu & Fruhholz (2016) suggest that the pSTS may be more involved during explicit versus passive emotion perception tasks. Markedly, most of the reported studies integrated a task that explicitly evaluated an emotional component (i.e., intensity, valence, category). In agreement, the pSTS was observed during the perceptual decision-making task in *Study 4,* which may be explained by its recognized role in social cognition (Hein and Knight, 2008; Deen et al., 2015).

Interestingly, as observed across *Studies 3 and 4,* the PCC, ventral and dorsal precuneus, vmPFC, ACC and IPL, may act as functional connector hubs, directing and integrating the flow of information in the brain (see for review, Van den Heuvel & Sporns, 2013a). Between-network connectivity in rs-fMRI demonstrated that these connector hubs not only comprise key regions of the DMN (i.e., lateral parietal, posterior cingulate, MTG), but also the Salience and Frontoparietal Control (Central Executive) networks (i.e., primarily the insula, dorsal ACC, and MFG; Bagarinao et al., 2020). In humans, data-driven analysis of functional heterogeneity ("diversity") reported that the mPFC, anterior insula, lateral PFC, lateral parietal cortex and subcortically, the thalamus and left putamen were "high diversity" regions (top 10%), while the lateral temporal cortex, right inferior PFC, vmPFC, and notably, the amygdala, were "low diversity" (bottom 10%; Anderson et al., 2013). Authors attributed low diversity of the amygdala, to an unaccounted-for confirmation bias in the literature, where recruitment was primarily associated with emotion, opposed to the larger scope of its functions, such as its involvement in decision-making, valuation, attention, awareness, and reward (see for review, Pessoa, 2010).

Although the amygdala is not a recognized central hub (although see Bickart et al., 2014), nor part of an "emotion network" per-se, modulation of the functional connectome during an emotion perception task has demonstrated that the amygdala is central to changes observed across the connectome (Markett et al., 2020). Evidence from both neuroimaging (Lindquist et al., 2012; Wager et al., 2015) and intracranial recordings (see for review, Guillory & Bujarski, 2014) support widespread processing of emotion. Comparable to our interpretation of a music-preferred pattern of recruitment across the AC and associated structures in *Studies 1 and 2*, Barrett and colleagues suggest that widespread neural patterns of multifunctional connecter nodes represent an internal model of emotions formed by past experiences. The model relays top-down predictions of the external world through active inference and receives bottom-up feedback which either corroborates with, or rejects the prediction (i.e., error; Barrett et al., 2016; Barrett, 2017a). This theory builds upon Friston et al.'s (2006) free energy principle, whereby the brain acts as an *inference engine*—it has a statistical model of the environment in which it is immersed (patterns of neural networks) and functions by minimizing the free energy of internal states through maximizing Bayesian model evidence (sense and perception), resulting in a change that will maintain a non-equilibrium steady state. Barrett (2017b) refers to this as allostasis, and reports involvement of the SN and DMN in integrating information (Kleckner et al., 2017)—two networks with the highest proportion of hubs and that share connections with all other motor and sensory networks (Van den Heuvel & Sporns, 2013b). Authors also indicate that visceromotor regions involved in allostasis (i.e., amygdala, striatum, insula, dorsal ACC, OFC and mPFC) are those most often reported in emotion processing circuits (Barrett & Satpute, 2013; Yeo et al., 2011; Barrett, 2017a; Kleckner et al., 2017). Alike the suggested role of the insula in the SN (see below; Sridharan et al., 2008; Menon & Uddin, 2010; Goulden et al., 2014), the CM nuclei of the amygdala is speculated to be involved in allostasis (Bohus et al., 1996; Ghashghaei et al., 2007; Barrett et al., 2016; Kleckner et al., 2017), as well as switching between cognitive and emotion processes and associated networks. The three subnuclei reportedly work together in a push-pull fashion, each exhibiting preference for a different neural network (e.g., DMN, FPCN, DAN, Sylvester et al., 2020).

6.2.2.3 Interacting systems of emotion and cognition

As a major part of the SN, the insula is a site for evaluating salience of homeostatic, emotional, or cognitive information in the environment (Uddin, 2015). It is recruited when sensory information poses a challenge and necessitates greater attentional resources to resolve novelty, ambiguity, uncertainty, or peculiarity (Sterzer & Kleinschmidt, 2010). As proposed above, this definition could also encompass violations of musical predictions (Friston & Friston, 2013). Markedly, the insula was involved in both passive perception and active categorization of emotion in *Studies 3 and 4,* respectively. During perceptual decision-making, the insula is reportedly, the primary initiation site of processing; responsible for comparing and integrating sensory information and directing downstream activity of the dorsal ACC (Lamichhane & Dhamala, 2015; Lamichhane et al., 2016a; 2016b). Markedly, the fronto-insular cortex and ACC are the only neural regions with Von Economo neurons; specialized projection cells that rapidly conduct. As a result, these areas reportedly detect errors in perceptual prediction, initiating a signal cascade that provides prompt negative feedback and an adaptive response (see for review, Allman et al., 2011). Menon and Colleagues (Menon & Uddin, 2010; Sridharan et al., 2008) have suggested that the insula functions as a switch in the SN (i.e., fronto-insular cortex and ACC; Seeley et al., 2007), modulating the opposing nature (Fox et al., 2005) of the FPCN (structurally overlapping with DAN) and DMN during attention-demanding cognitive tasks. An interaction of comparable nature is evident in *Study 4* by assessing the dynamics of emotion perception and decision-making using ambiguous presentations of threat-related stimuli. Subjects spontaneously shifted their inference across two mutually exclusive competing perceptions (fear versus anger), as the physical stimulus remained unchanged (Sterzer & Kleinschmidt, 2007). In doing so, we observed greater recruitment of regions within primarily, the SN, FPCN, and DMN. Auditory stimuli created during pre-scan sessions to reflect subject-specific perceived ambiguity, were then perceived more often as ambiguous during scanning sessions, compared to visual “ambiguous” stimuli. As reported in Chapter 5, this was reflected by a greater difference of activation between ambiguous and explicit processing within the task-positive network (SN and FPCN), as well as a greater difference of deactivation in the DMN. Most remarkably, we reported that the DMN exhibits greater deactivation in response to, the arguably more ambiguous modality-response pairs (e.g., auditory-fear, visual-anger). Here, decreasing activity of the DMN reduces interfering self-thought that may otherwise impede on cognitive processes (see for review, Anticevic et al., 2013), such as those associated with activity of the FPCN (Dosenbach et al., 2008).

We propose that this dynamic interaction may reflect an allocation of limited resources within interconnected networks of emotion perception and cognitive control (Pessoa, 2008; 2009). We suggest greater recruitment of these networks in *Study 4* versus *3*, due to the minimal demands of the passive perception task in *Study 3*. Notably, similar findings have been reported in perceptual decision-making tasks of emotional ambiguity (i.e., fear-neutral-disgust facial morphs; Thielscher & Pessoa, 2007; fear-anger vocalization morphs; Bestelmeyer et al., 2014). Specifically, both groups reported the same inverted U-shape correlation of morph steps with reaction times and recruitment of regions within Task-Positive and Default Mode Networks.

6.3 Modulating neural correlates through individual differences

Even in apparently homogenous, healthy populations, we can detect sources of inter-subject variability relating to neural processes and associated behaviour. We can clarify whether inconsistencies in the literature (particularly in socio-emotion perception; Hamann & Canli, 2004), may be explained by these differences (see for example, Eugène et al., 2003), particularly where overlap of recruitment is likely (i.e., music perception – music expertise; threat perception – state anxiety). Thus, in *Study 2* we addressed reservations of *music-preferred* processing regions, and influences of the relatively stable musical expertise through providing evidence for internal reliability. In *Study 4* we change course, investigating transient between-subject differences more likely to fluctuate. By investigating influences of state anxiety on perceptual decision-making and threat-related perception, we build a brain and behaviour profile that can be corroborated with the existing literature. Although the nature of inter-subject variability differs, *Studies 2* and *4* demonstrate added value to our understanding of healthy music perception (*Study 1*) and threat-related social perception (*Study 3*).

6.3.1 Musical expertise and music perception

Experienced musicians can put in over ten thousand hours of practice; that is over a decade of highly involved training, and often years of practice thereafter (Krampe & Ericsson, 1996). Thus, they are suitable subjects to observe structural and functional neural differences over the lifespan. Previous work has demonstrated that greater exposure and skill in music (Ohnishi et al., 2001; Dick et al., 2011; Angulo-Perkins et al., 2014; Schmithorst & Holland, 2003) and language (see for review, Costa and Sebastián-Gallés, 2015) leads to neurofunctional differences. In agreement, individuals with more musical experience showed greater activity in the right posterior STG and SMG (anterior region of the IPL), close to and surrounding a music-preferred region, the PT (*Studies 1* and *2).* Notably, the PT is a recognized site for neuroplasticity in musicians (see for review, Meyer et al., 2012) and often involved in auditory-sensorimotor integration and sequencing of events. When performing, a musician controls the order of pitch events (pitch sequencing) and intervals between successive events (temporal sequencing; Warren et al., 2005; Rauschecker, 2011; 2018). Coordinating auditory and motor systems is also required to perform, relying on feedforward/feedback connections (see for review, Zatorre et al., 2007) and increased coupling (Chen et al., 2008).

Even when deciphering subtle pitch variations in a foreign language, musicians can better detect distinctions, and respond faster (difference of 300msec) than non-musicians (Marques et al., 2007). Moreover, they can better detect the offset and onset of subsequent speech cues; exhibiting greater posterior left PT activity than non-musicians (Elmer et al., 2012). Nevertheless, no crossover effects were reported in *Study 2*, although language expertise was associated with activity along the left dorsal stream in primarily the posterior STS/STG, MTG, HG, and PT. Evidence for left-lateralized PT activity responsive to language expertise, and the opposite to musical expertise (*Study 2*), agrees with known behaviour of decoding fine-grained phonetic information preferably in the left PT, and speech rhythm or melody in the right (Meyer et al., 2012; Poeppel, 2003).

6.3.2 State anxiety and threat-related social perception

Individuals with high anxiety, have selective attention to threat, encouraging the development and maintenance of anxiety disorders (Singer et al., 2012; MacLeod and Matthews, 2012; Barr-Haim et al., 2007; Mogg and Bradley, 1998). Choi and colleagues (2012) observed that monitoring threat impairs those with high state anxiety when resolving conflict; explained by greater recruitment of the anterior insula, which in part, is mediated by impairment of the mPFC. In general, those with anxiety present weaker cognitive control (Derryberry & Reed, 2002) and often exhibit reduced performance *effectiveness* (accuracy) or *efficiency* (response times; Wong et al., 2013; Eysenck et al., 2007), particularly in tasks of attention, perception, executive function, and memory (see for review, Robinson et al., 2013). In *Studies 3* and *4* we observe that healthy subjects recruit the amygdala when passively or explicitly perceiving fear. Nonetheless, *Study 4* demonstrates that it is when perceiving ambiguous stimuli that state anxiety influences amygdala recruitment. The increase of amygdala activity may reflect a binarization of stimuli (more easily classified as one explicit emotion versus the other), particularly given the shorter reaction times. As highly anxious subjects made greater errors when perceiving visual explicit anger, as opposed to fear, we propose that these individuals may exhibit hypervigilance (Eysenck, 1997), impulsivity (Jakuszowiak-Wojten et al., 2015; Xia et al., 2017) and heightened attention (Williams et al., 1997) towards their perceptual bias (i.e., visual fear). Notably, regions of activity responsible for control and attention (FPCN and SN) were most prominent in anxious subjects, specifically in response to auditory ambiguous versus explicit stimuli, the arguably more challenging task. As reported in Chapter 5, these findings support Eysenck et al.'s (2007) *attentional control theory* where anxious subjects recruit compensatory strategies to maintain performance, such as modulating recruitment of the FPCN. This adjustment, therefore, results in a reduced efficiency of the system. This non-clinical behaviour is relatively consistent with clinical findings—effecting functionality of the amygdala and frontal attention and control systems. In clinically anxious subjects, both the amygdala and PFC are mediated by genetics, specifically through monoamine neurotransmission (see for review, Millan, 2003).

6.4 Methodological Advancements

The complexities of socio-emotional communication are demonstrated at the behavioural and neural level, through careful consideration of design, implementation, and analysis for each given experiment. The selected approaches were intended to demonstrate the value of methodological rigor in fMRI studies of socio-emotional perception.

6.4.1 Experimental design

The careful selection of stimuli and controls proved essential in the current thesis, particularly for *Studies 1 and 4*. Developing two studies that used intermediary stimuli contrasted against extremes to provide insight into full-spectrum perception. Moreover, using highly diverse stimuli in *Study 2,* allowed for a generalization of responses observed in music- and voice-preferred regions, with comparable conclusions drawn in *Study 3*, using faces, bodies, vocalizations, and prosody to demonstrate across-modality processing of the amygdala. Finally, subject-specific stimuli in *Study 4* allowed for a relative comparison of auditory and visual threat-related perception, and a closer look at between-subject differences.

6.4.2 Acquisition

Highly similar activity across continuous and sparse acquisition (*Study 1*), suggests that scanner noise does not impose differential effects on speech and music, encouraging further use of continuous sampling in exploring auditory socio-emotional perception. Notably, the sparse-sampling paradigm had to be reduced from 72 (continuous sampling) to 25 slices along the Sylvian fissure (covering the AC), given the limited scan time. This therefore may have prevented observed activity of extra-temporal processing regions as reported in *Study 2*. Beyond assurance of face validity, *Study 2* argues for standardized assessments of intrasubject reliability in fMRI studies (Elliot et al., 2020) and determines test-retest reliability over multiple runs to refine the precise location of music- and voice-preferred regions.

6.4.3 Analysis

Consistency across univariate categorical analyses and a stimulus-based multivariate ICA in *Study 1* enhanced the reliability of music-preferred regions, in addition to illustrating a conceivable distribution of coefficients reflecting acoustic perception along a gradient. For example, chorales were singing stimuli closely located to music, while lullabies or melodic Torah readings were near to speech. Multivariate analyses offer increased specificity by analyzing voxels collectively, exploiting the covariance between them, and likely presenting a more direct means of measuring multidimensional stimuli (Davis & Poldrack, 2013). In *Study 4*, the high percentage of voxels (univariate) that overlapped with specific networks (i.e., SN, FPCN, VN, DMN) in Yeo et al. (2011) network parcellations, combined with observing similar patterns in the MVPA, indicated plausible network recruitment. As discussed in detail below, functional connectivity is a suitable next step in furthering our argument for widespread processing of social information.

6.5 Limitations and future directions

Opportunity for improvement can be identified within each study, to enhance the generalizability and reliability of said findings. Nonetheless, design decisions were made, weighing in constraints of time, expense, and subjects' attention span. For instance, *Studies 3 and 4* would benefit from using a greater number of actors, however, other aspects of design would have to be sacrificed as to not overstretch available resources. Below we address feasible modifications that would enhance existing findings, or rather, progress the inquiry into supplementary questions.

6.5.1 Multidimensionality of the Amygdala

As we outline the multisensory and multifunctional qualities of the amygdala, we recognize the limitation of exploring the perception of only fearful, angry, and ambiguous social expressions. The substantial evidence for differential neural recruitment when processing emotions along a given dimension (e.g., valence, arousal, internally-versus externally-directed affect; Posner et al., 2009; Bush et al., 2017; Wager et al., 2015) justifies further exploration using paradigms from *Studies 3 and 4*. Markedly, we anticipate that perceptual decision-making behaviour and underlying neural correlates may vary in accordance with the chosen dimension in which stimuli are morphed.

6.5.2 Individual differences

From our findings we deduce that the inclusion of additional questions/assessments on background, proficiencies and demographics would add depth to our exploration of individual differences. For example, in *Study 2*, conducting a simple language proficiency assessment or acquiring greater detail about one's level of expertise (e.g., age at which subjects began musical training or second-language learning; Ohnishi et al., 2001; Li et al., 2014) may add value to profiling influences of expertise. With access to greater resources, using larger sample sizes or quantifying test-retest reliability across longer time scales (e.g., days, weeks, months) would help in validating our findings. Particularly when exploring deficits in healthy populations, i.e., non-clinical state anxiety, the objective is to match consistency to the diagnostic-level of clinical populations (Kraemer, 2014).

As previously stated in Chapter 5, including real-time physiological measures, such as electrodermal activity (Strohmaier et al., 2020), heart rate (Kantor et al., 2001) or simply manual feedback on ones' fluctuating mood, may help in modelling intrasubject variability. Using self-reports alongside unbiased measures of anxiety (e.g., cortisol levels; Brown et al., 2017) can improve the validity of anxiety scores, and moreover, of the neural correlates associated with anxious mood. The value in acquiring a true representation of non-clinical anxiety in the brain is important as these assessments could be developed for screening use in clinical anxiety (e.g., Wiglusz et al., 2019); a complex disorder with high comorbidities (e.g., substance abuse, major depressive disorder, personality disorders; see for review, Antony & Stein, 2009).

6.5.3 Task-based functional connectivity

Advancing current findings through task-based functional connectivity would prove useful in confirming network-based hypotheses formed throughout the dissertation. Network analyses can emphasize how malleable and dynamic the brain is, particularly relevant for illustrating interactions of emotion and cognition (Kinnison et al., 2012, Uddin et al., 2014). To expose directional effects of networks, study designs can be modified to best accommodate analyses of, for example, psychophysiological interactions (PPI; Friston et al., 1997) or structural equation modelling (SEM; McIntosh & Gonzalez-Lim, 1994). The challenge of these approaches is to develop a model (as it relates to anatomical connectivity) with strong construct, face, and predictive validity, which in its simplest form explains how the BOLD signal change of one voxel reflects a weighted sum of changes in other neural regions (Friston, 1994).

Extending on *Studies 1* and *2,* current work with collaborators (Herholz et al., *in preparation*) explores functional connectivity of voice, singing and music perception in the AC through voxel-by-voxel correlations and diffusion-based network embedding (Margulies et al., 2016). However, given above-mentioned assumptions of extra-temporal connectivity with music-preferred regions (PP, PT), supplementary intra-regional analyses are of interest. The same applies to *Studies 3* and *4*, where assumptions of large network recruitment should be confirmed, such as through PPI, where Prado et al. (2011) demonstrated that response times fluctuate with the strength of connectivity between the dlPFC and posterior parietal cortex (FPCN), with the ACC (SN). This exploration is valuable given that even disorders seemingly attributed to early sensory processes (e.g., conduction aphasia—a disconnection between Wernicke's and Broca's areas; Weems & Reggia, 2006), may indeed be regulated by feedback from other network nodes (Rowe, 2010). Moreover, abnormal functional network connectivity has been suggested as a plausible biomarker for anxiety (e.g., Sylvester et al., 2012; Zhao et al., 2007; Qiao et al., 2017; Makovac et al., 2018). Evidently, changes of connectivity in cortical and subcortical networks can manifest as cognitive dysfunction, akin to abnormalities of isolated regions.

6.6 Implications

This wide scope of the thesis offers several diverse opportunities for ongoing research and application. At the clinical level, our findings could contribute to the validation of rehabilitation protocols, i.e., cross-sensory (e.g., Streim-Amit et al., 2012) or cross-domain therapeutics (e.g., Torppa et al., 2018; Good et al., 2017) reliant on functional transference to mediate deficits. For instance, evidence suggests that musical training, either instrumental or singing, can improve speech perception in individuals with hearing loss, because of similar top-down and hierarchical processing in the AC (see for review, Lerousseau et al., 2020).

Alternatively, our findings could benefit the development of deep neural networks for human-computer interaction. Through modelling emotion recognition in faces, speech, or multisensory information (see for review, Ko, 2018; Khalil et al., 2019; Abdullah et al., 2021, respectively), our work could provide knowledge of perceptual variability in healthy populations. Understanding the scope of perceptual variability in healthy populations is also helpful for the detection of outliers. Potentially, combined behaviour and neuroimaging data could improve the sensitivity of diagnostics, particularly when examining the response to titrated socio-emotional stimuli. Finally, findings from the methodological inquiries pursued are intended to benefit future research in the field of socio-emotion perception.

General Reference List

Abdullah, S.M.S., Ameen, S.Y., Sadeeq, M.A.M., & Zeebaree, S.R.M. (2021). Multimodal emotion recognition using deep learning. *Journal of Applied Science and Technology Trends, 2,* 52-58.

Adams, R.B., Jr., Gordon, H.L., Baird, A.A., Amady, N., & Kleck, R.E. (2003). Effects of gaze on amygdala sensitivity to anger and fear faces. *Science, 300*, 1536.

Adolphs, R. (2008). Fear, faces, and the human amygdala. *Current Opinion in Neurobiology, 18*, 166-172.

Adolphs, R. (2013). The biology of fear. *Current Biology, 23*, R79-R93.

Ali, S.O., & Peynircioğğlu, Z.F. (2010). Intensity of emotions conveyed and elicited by familiar and unfamiliar music. *Music Perception, 27,* 177-182.

Allman, J.M., Tetreault, N.A., Hakeem, A.Y., Manaye, K.F., Semendeferi, K., Erwin, J.M., Park, S., Goubert, V., & Hof, P.R. (2011). The von Economo neurons in frontoinsular and anterior cingulate cortex. *Annals of the New York Academy of Sciences*, *1225,* 59-71.

Amaro, E Jr., & Barker, G.J. (2006). Study design in fMRI: Basic principles. *Brain and Cognition, 60,* 220-232.

Anderson, M.L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioural and Brain Sciences, 33,* 245-266.

Anderson, M.L. (2014). *After phrenology: Neural reuse and the interactive brain*. MIT press.

Anderson, M.L., Kinnison, J., & Pessoa, L. (2013). Describing functional diversity of brain regions and brain networks. *Neuroimage, 73,* 50-58.

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, 59,* 126-137.

Angulo-Perkins, A., & Concha, L. (2019). Discerning the functional networks behind processing of music and speech through human vocalizations. *PLoS One, 14*, e0222796.

Anticevic, A., Cole, M.W., Murray, J.D., Corlett, P.R., Wang, X-J., & Krystal, J.H. (2012). The role of the default network deactivation in cognition and disease. *Trends in Cognitive Science, 16,* 584:592.

Antony, M.M., & Stein, M.B. (Eds.). (2009). *Oxford Handbook of Anxiety and Related Disorders*. Oxford University Press.

Armony, J.L. (2013). Current emotion research in behavioural neuroscience: The role(s) of the amygdala. *Emotion Review, 5,* 104-115.

Armony, J.L., Aubé, W., Angulo-Pekins, A., Peretz, I., & Concha, L. (2015). The specific of neural responses to music and their relation to voice processing: An fMRI-adaptation study. *Neuroscience Letters*, 593, 35-39.

Atkinson, A.P., Dittrich, W.H., Gemmell, A.J., & Young, A.W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception, 33,* 717-746.

Aubé, W., Angulo-Perkins, A., Peretz, I., Concha, L., & Armony, J. L. (2015). Fear across the senses: Brain responses to music, vocalizations and facial expression. *Social Cognitive and Affective Neuroscience, 10,* 399-407.

Averbeck, B.B., Costa, V.D. (2017). Motivational neural circuits underlying reinforcement learning. *Nature Neuroscience, 20,* 505-512.

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between positive and negative emotions. *Science, 338,* 1225-1229.

Bachmann, J., Munzert, J., & Krüger, B. (2018). Neural underpinnings of the perception of emotional states derived from biological human motion: A review of neuroimaging research. *Frontiers in Psychology, 9,* 1763.

Bagarinao, E., Watanbe, H., Maesawa, S., Mori, D., Hara, K., Kawabata, K., Ohdake, R., Masuda, M., Ogura, A., Kato, T., Koyama, S., Katsuno, M., Wakabayashi, T., Kuzuya, M., Hoshiyama, M., Isoda, H., Naganawa, S., Ozaki, N., & Sobue, G. (2020). Identifying the brain's connector hubs at the voxel level using functional connectivity overlap ration. *Neuroimage, 222,* 1-11.

Ball, T., Rahm, B., Eickoff, S.B., Schulze-Bonhage, A., Speck, O., & Mutschler, I. (2007). Response properties of human amygdala subregions: Evidence based on functional MRI combine with probabilistic anatomical maps. *PLoS One, 2,* e307.

Bandettini, P.A., & Cox, R.W. (2000). Event-related fMRI contrast when using constant intersimulus interval: Theory and experiment. *Magnetic Resonance in Medicine, 43,* 540-548.

Banse, R., & Scherer, K.R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70,* 614-636.

Barr-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M.J., & van IJzendoorn, M.H. (2007). Threat-related attentional bias in anxious and non-anxious individuals: A meta-analytic study. *Psychological Bulletin, 133,* 1-24.

Barrett, L.F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review, 10,* 20-46.

Barrett, L.F. (2017a). The theory of constructed emotions: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience, 12,* 1-13.

Barrett, L.F. (2017b). *How Emotions Are Made: The Secret Life of 3the Brain*. New York, NY: Houghton-Mifflin-Harcourt.

Barrett, L.F., Quigley, K.S., & Hamilton P. (2016). An active inference theory of allostasis and interception in depression. *Philosophical Transactions of the Royal Society B, 371,* 1-17.

Barrett, L.F., & Satpute, A.B. (2013). Large-scale brain networks in affective and social neuroscience: Towards an integrative functional architecture of the brain. *Current Opinion in Neurobiology, 23,* 361-372.

Barsalou, L.W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society: Biological Sciences, 364,* 1281-1289.

Baumgart, F., Gaschler-Markefski, B., Woldorff, M.G., Heinze, H-J., & Scheich, H. (1999). A movement-sensitive area in auditory cortex. *Nature, 400,* 724-726.

Beauchamp, M.S. (2015). The social mysteries of the superior temporal sulcus. *Trends in Cognitive Science, 19*, 489-490.

Beauchamp, M.S., Argall, B.D., Bodruka, J., Duyn, J.H., & Martin, A. (2004). Unraveling multisensory integration: Patchy organization within human STS multisensory cortex. *Nature Neuroscience, 7,* 1190-1192.

Beauchamp, M.S., Lee, K.E., Argall, B.D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron, 41,* 809-823.

Beauchamp, M.S., Yasar, N.E., Frye, R., & Ro, T. (2008). Touch, sound and vision in human superior temporal sulcus. *Neuroimage, 41,* 1011-1020.

Beck, A.T., & Clark, D.A (1997). An information processing model of anxiety: Automatic and strategic process. *Behaviour Research and Therapy, 35,* 49-58

Beck, A.T., Emery, G., & Greenberg, R. (1985). *Anxiety disorders and phobias: A cognitive perspective*. New York: Basic Books.

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences, 8,* 129-135.

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

Belyk, M., & Brown, S. (2014). Perception of affective and linguistic prosody: An ALE meta-analysis of neuroimaging studies. *Social Cognitive and Affective Neuroscience, 9,* 1395-1403.

Bestelmeyer, P.E.G., Maurage, P., Rouger, J., Latinus, M., & Belin, P. (2014). Adaptation to vocal expressions reveals multistep perception of auditory emotion. *Journal of Neuroscience, 34,* 8098-8105.

Bigand, E., Delbé, C., Gérard, Y., & Tillmann, B. (2011). Categorization of extremely brief auditory stimuli: Domain-specific or domain-general process? *PLoS One, 6,* e27024.

Binder, J.R., Liebenthal, E., Possing, E.T., Medler, D.A., & Ward, B.D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience, 7,* 295-301.

Bishop, S.J., Duncan, J., & Lawrence, A.D. (2004). State anxiety modulation of the amygdala response to unattended threat-related stimuli. *Journal of Neuroscience, 24,* 10364-10368.

Bishop, S.J., Jenkins, R., & Lawrence, A.D. (2007). Neural processing of fearful faces: Effects of anxiety are gated by perceptual capacity limitations. *Cerebral Cortex, 17*, 1595-1603.

Bizley, J.K., & Cohen, Y.E. (2014). The what, where and how of auditory-object perception. *Nature Reviews in Neuroscience, 14*, 693-707.

Blair, K., Smith, B., Mitchell, D., Morton, J., Vythilingam, M., Pessoa, L., Fridberg, D., Zametkin, A., Sturman, D., Nelson, E.E., Drevets, W.C., Pine, D.S., Martin, A., & Blair, R.J.R. (2007). Modulation of emotion by cognition and cognition by emotion. *Neuroimage, 35*, 430-440.

Boebinger, D., Norman-Haignere, S., McDermott, J., & Kanwisher, N. (2020). Cortical music selectivity does not require musical training. *bioRxiv,* 1-46.

Bohus, B., Koolhas, J.M., Luiten, P.H., Korte, S.M., Roozendaal, B., & Wiersma, A. (1996). The neurobiology of the central nucleus of the amygdala in relation to neuroendocrine and autonomic outflow. *Progress in Brain Research. 107*, 447-460.

Bonnet, L., Comte, A., Tatu, L., Millot, J-L., Moulin, T., & Medeiros de Bustos, E. (2015). The role of the amygdala in the perception of positive emotions: An “intensity detector”. *Frontiers in Behavioural Neuroscience, 9,* 1-12.

Boogert, N.J., Madden, J.R., Morand-Ferron, J., & Thornton, A. (2018). Measuring and understanding individual differences in cognition. *Philosophical Transactions of the Royal Society B, 373,* 1-10.

Boyatzis, R.E., Rochford, K., & Jack, A.I. (2014). Antagonistic neural networks underlying differentiated leadership roles. *Frontiers in Human Neuroscience, 8,* 1-15.

Brandman, T., & Yovel, G. (2016). Bodies are represented as wholes rather than their sum of parts in the occipital-temporal cortex. *Cerebral Cortex, 26,* 530-543.

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,* 1-16.

Brickart, K.C., Dickerson, B.C., & Barrett, L.F. (2014). The amygdala as a hub in brain networks that support social life. *Neuropsychologia, 63,* 235-248.

Britton, J.C., Phan, K.L., Taylor, S.F., Welsh, R.C., Berridge, K.C., & Liberzon, I. (2006). Neural correlates of social and nonsocial emotions: An fMRI study. *Neuroimage, 31,* 397-409.

Brosschot, J. F., Verkuil, B., & Thayer, J. F. (2016). The default response to uncertainty and the importance of perceived safety in anxiety and stress: An evolution-theoretical perspective. *Journal of Anxiety Disorders, 41*, 22-34.

Brown, C.C., Raio, C.M., & Neta, M. (2017). Cortisol responses enhance negative valence perception for ambiguous facial expressions. *Scientific Reports, 7,* 1-8.

Bush, K. A., Inman, C. S., Hamann, S., Kilts, C. D., & James, G. A. (2017). Distributed neural processing predictors for multi-dimensional properties of affect. *Frontiers in Human Neuroscience, 11*, 1-14.

Calder, A.J. (1996). Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear. *Cognitive Neuropsychology, 13,* 699-745.

Callan, D. E., Tsytsarev, V., Hanawaka, T., Callan, A. M., Katsuhara, M., Fukuyama, H., & Turner, R. (2006). Song and speech: Brain regions involved with perception and cover production. *Neuroimage, 31*, 1327-1342.

Canli, T., Zhao, Z., Desmond, J.E., Kang, E., Gross, J., & Gabrieli, J.D. (2001). An fMRI study of personality influences on brain reactivity to emotional stimuli. *Behavioural Neuroscience, 115,* 33-42.

Carrington, S.J., & Bailey, A.J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping, 30,* 2313-2335.

Caspers, S., Zilles, K., Laird, A.R., & Eickhoff, S.B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *Neuroimage, 50,* 1148-1167.

Castro-Vale, I., Severo, M., & Carvalho, D. (2020). Lifetime PTSD is associated with impaired emotion recognition in veterans and their offspring. *Psychiatry Research, 284*, 1-8.

Celeghin, A., Diano, M., Bagnis, A., Viola, M., & Tamietto, M. (2017). Basic emotions in human neuroscience: Neuroimaging and beyond. *Frontiers in Psychology, 8,* 1-13.

Chen, J.L, Penhune, V.B., & Zatorre, R.J. (2008). Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex, 18,* 2844-2854.

Chen, L., T.Vu, A., Xu, J., Moeller, S., Ugurbil, K., Yacoub, E., & Feinberg, D.A. (2015). Evaluation of highly accelerated simultaneous multi-slice EPI for fMRI. *Neuroimage, 104,* 452-459.

Choi, J.M., Padmala, S., & Pessoa, L. (2012). Impact of state anxiety on the interaction between threat monitoring and cognition. *Neuroimage, 59,* 1912-1923.

Cisler, J.M., & Koster, E.H.W. (2010). Mechanisms of attentional biases towards threat in anxiety disorders: An integrative review. *Clinical Psychology Review, 30*, 203-216.

Coffey, E.B.J., Mogilever, N.B., & Zatorre, R.J. (2017). Speech-in-noise perception in musicians: A review. *Hearing Research, 352,* 49-69.

Cole, M.W., J.R., Reynolds, J.D. Power., Repovs, G., Anticevic, A., & Braver, T.S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature Neuroscience, 16,* 1348-1355.

Cona, G., & Semenza, C. (2017). Supplementary motor area as key structure for domain-general processing: A unified account. Neuroscience *Biobehavioural Reviews, 72,* 28-42.

Costa, A., & Sebastián-Gallés, N. (2015). How does the bilingual experience sculpt the brain. *Nature Review Neuroscience, 15,* 336-345.

Costafreda, S.G. (2009). Pooling fmri data: meta-analysis, mega-analysis and multi-center studies. *Frontiers in Neuroinformatics, 3,* 1-8.

Costafreda, S.G., Brammer, M.J., David, A.S., & Fu, C.H.Y. (2008). Predictors of amygdala activation during the processing of emotional stimuli: A meta-analysis of 385 PET and fMRI studies. *Brain Research Reviews, 58*, 57-70.

Coutinho, E., & Cangelosi, A. (2011). Musical emotions: predicting second-by-second subjective feelings of emotion from low-level psychoacoustic features and physiological measurements. *Emotion, 11*, 921-37.

Darwin, C. (1872/1998). *The expression of emotion in man and animals*. New York: Oxford University Press.

Davis, M. (1992). The role of the amygdala in fear and anxiety. *Annual Review of Neuroscience, 15*, 353-375.

Davis, M., & Whalen, P.J. (2001). The amygdala: Vigilance and emotion. *Molecular Psychiatry, 6*, 13-34.

Davis T, LaRocque KF, Mumford J, Norman KA, Wagner AD, & Poldrack RA. (2014). 'What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis'. *Neuroimage, 15,* 271-283.

Davis, T., & Poldrack, R.A. (2013). Measuring neural representations with fMRI: Practices and pitfalls. *Annals of the New York Academy of Sciences, 1296,* 108-134.

Decety, J., & Grèzes, J. (2006). The power of simulation: imagining one's own and other's behaviour. *Brain Research, 1079,* 4-14.

Decety, J., Skelly, L., Yoder, J.K., & Kiehl, K.A. (2014). Neural processing of dynamic emotional facial expressions in psychopaths. *Social Neuroscience, 9,* 36-49.

Deen, B., Koldewyn, K., Kanwisher, N., & Saxe, R. (2015). Functional organization of social perception and cognition in the superior temporal sulcus. *Cerebral Cortex, 25,* 4596-4609.

De Gelder, B., de Borst, A.W., & Watson R. (2015). The perception of emotion in body expressions. *WIREs Cognitive Science, 6,* 149-158.

De la Mothe, L.A., Blumell, S., Kjikawa, Y., & Hackett, T.A. (2006). Cortical connections of the auditory cortex in marmoset monkeys: Core and medial belt regions. *Journal of Comparative Neurology, 496*, 27-71.

Del Maschio, N., & Abutalebi, J. (2019). Language organization in the bilingual and multilingual brain. In J.W. Schwieter, & M. Paradis (Eds.), *The handbook of the neuroscience of multilingualism* (pp.197-213). John Wiley & Sons Ltd.

Demenescu, L.R., Renken, R., Kortekaas, R., van Tol, M.J., Marsman, J.B., van Buchem, van der

Wee, N.J.A., Veltman, D.J., & den Boer, J.A., Aleman, A. (2011). Neural correlates of perception of emotional facial expressions in out-patients with mild-to-moderate depression and anxiety. A multicenter fMRI study. *Psychological Medicine, 41*, 2253-2264.

Derryberry, D., & Reed, M.A. (2002). Anxiety-related attentional biases and their regulation by attentional control. *Journal of Abnormal Psychology, 111*, 225-236.

Deutsch, D. (1999). Grouping mechanisms in music. In D. Deutsch (Ed.), *The psychology of music* (pp. 299-348). Academic Press.

De Waal, F.B.M. (2003). Darwin's legacy and the study of primate visual communication. In P. Ekman, J. Campos, J. Davidson & F.B.M. De Waal (Eds.), *Emotions inside out: 130 years after Darwin's The Expression of Emotion in Man and Animals*, (pp. 7-31). New York: New York Academy of Sciences.

Dezecache, G., Conty, L., Chadwick, M., Philip, L., Soussignan, R., Serber, D., & Grèzes, J. (2013). Evidence for unintentional emotional contagion beyond dyads. *PLoS One*, *8*, e67371.

Diano, M., Tamietto, M., Celeghin, A., Weiskrantz, L., Tatu, M.K., Bagnis, A., Duca, S., Germiniani, G., Cauda, F., & Costa, T. (2017). Dynamic changes in amygdala psychophysiological connectivity reveals distinct neural networks for facial expressions of basic emotions. *Scientific Reports*, *7,* 45620.

Dick, F., Ling Lee, H., Nusbaum, J., & Price, C.J. (2011). Auditory-motor expertise alters "speech selectivity" in professional musicians and actors. *Cerebral Cortex, 21*, 938-948.

Domínguez-Borràs, J., Guex, R., Méndez-Bértolo, C., Legendre, G., Spinelli, L., Moratti, S., Frühholz, S., Mégevand, P., Arnal, L., Strange, B., Seeck, M., & Vuillemuier, P. (2019). Human amygdala response to unisensory and multisensory emotion input: No evidence for superadditivity from intracranial recordings. *Neuropsychologia*, *131*, 9-24.

Dosenbach, N.U., Fair, D.A., Cohen, A.L., Schlaggar, B.L., & Peterson, S.E. (2009). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences, 12*, 99-105.

Downing, P.E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science, 293*, 2470-2473.

Dricu, M., & Frühholz, S. (2016). Perceiving emotional expressions in others: Activation likelihood estimation meta-analyses of explicit evaluation, passive perception and incidental perception of emotions. *Neuroscience & Biobehavioural Reviews, 71,* 810-828.

Dubois, S., Rossion, B., Schiltz, C., Bodart, J.M., Michel, C., Buyer, R., & Crommelinck, M. (1999). Effect of familiarity on the processing of human faces. *Frontiers in Psychology, 9,* 278-289.

Edelman, G.M., & Gally, J.A. (2001). Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences, 98,* 13763-13768.

Ekman, P. (1992). An argument for basic emotions. Cognition and Emotion, 6, 169-200.

Ekman, P., Friesen, W.V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W.A., Pitcairn, T., Ricci-Bitti, P.E., Scherer, K., Tomita, M., Tzavaras, A. (1987). Universals and cultural differences in the judgements of facial expressions of emotion. *Journal of Personality and Social Psychology, 53,* 712-717.

Ellard, K.K., Farchione, T.J., & Barlow, D.H. (2012). Relative effectiveness of emotion induction procedures and the role of personal relevance in a clinical sample: A comparison of film, images, and music. *Journal of Psychopathology and Behavioural Assessment, 34,* 232-243.

Elliot, M.L., Knodt, A.R., Ireland, D., Morris, M.L., Poulton, R., Ramrakha, S., Sison, M.L., Moffitt, T.E., Caspi, A., & Hariri, A.R. (2020). What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. *Psychological Science, 31*, 792-806.

Elmer, S., Meyer, M., & Jäncke, L. (2012). Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. *Cerebral Cortex, 22,* 650-658.

Engelmann, J.B., & Pogosyan, M (2013). Emotion perception across cultures: The role of cognitive mechanisms. *Frontiers in Psychology, 4,* 1-10.

Etkin, A., Egner, T., Peraza, D.M., Kandel, E.R., & Hirsch, J. (2006). Resolving emotional conflict: A role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron, 51,* 871-882.

Etkin, A., Klemenhagen, K. C., Dudman, J. T., Rogan, M. T., Hen, R., Kandel, E. R., & Hirsch, J. (2004). Individual differences in trait anxiety predict the response of the basolateral amygdala to unconsciously processed fearful faces. *Neuron, 44,* 1043-1055.

Eugène, F., Lévesque, J., Mensour, B., Leroux, J-M., Beaudoin, G., Bourgouin, P., & Beauregard M. (2003). The impact of individual differences on the neural circuitry underlying sadness.

Neuroimage, 2, 354-364.

Eysenck, M.W. (1997). *Anxiety and cognition: A unified theory*. Hove, England: Psychology Press.

Eysenck, M.W., Derakshan, N., Santos, R., & Calvo, M.G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion, 7,* 336-353.

Faber, S., & Fiveash, A. (2014). Emotion without words: A comparison study of music and speech prosody. *Canadian Journal of Music Therapy, 20*, 86-101.

Fecteau, S., Belin, P., Joanette, Y., & Armony, J.L. (2007). Amygdala responses to nonlinguistic emotional vocalizations. *Neuroimage, 36*, 480-487.

Fedorenko, E., McDermott, J. H., Norman-Haignere, S., & Kanwisher, N. (2012). Sensitivity to musical structure in the human brain. *Journal of Neurophysiology, 108,* 3289-3300.

Filkowski, M.M., Olsen, R.M., Duda, B., Wanger, T.J., & Sabatinelli, D. (2017). Sex differences in emotional perception: Meta analysis of divergent activation. *Neuroimage, 15,* 925-933.

Fitch, W.T. (2006). The biology and evolution of music: A comparative perspective. *Cognition, 100,* 173-215.

Fitzgerald, D.A., Angstadt, M., Jelsone, L.M., Nathan, P.J., & Phan, K.L. (2005). Beyond threat: Amygdala reactivity across multiple expressions of facial affect. *Neuroimage, 30,* 1441-1448.

Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., & Raichle, M.E. (2005). The human brain is intrinsically organized into dynamic anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the USA*, *102*, 9673-9678.

Francois, C. & Schon, D (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex, 21,* 2357-2365.

Freeman, J.B., Johnson, K.L., Adams, R.B. Jr., & Ambady, N. (2012). The social-sensory interface: Category interactions in person perception. *Frontiers in Integrative Neuroscience, 6,* 1-13.

Freeman, J.B., Stolier, R.M., Ingretsen, Z.A., & Hehman, E.A. (2014). Amygdala responsivity to high level social information from unseen faces. *Journal of Neuroscience, 34,* 10573-10581.

Freese, J.L., Amaral, D.G. (2009). Neuroanatomy of the primate amygdala. In P.J. Whalen & E.A. Phelps (Eds.), *The human amygdala* (pp. 3-42). The Guilford Press.

Friston, K.J. (1994). Functional and effective connectivity in neuroimaging: A synthesis. *Human*

Brain Mapping, 2, 56-78.

Friston, K.J., Buchel, C., Fink, G.R., Morris, J., Rolls, E., & Dolan, R. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage, 6,* 218-229.

Friston, K.J., & Friston, D.A. (2013). A free energy formulation of music generation and perception: Helmholtz revisited. In R. Bader (Ed.) *Sound—Perception—Performance, Current Research in Systematic Musicology 1*. Springer International Publishing: Switzerland.

Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. Journal of *Physiology, 100,* 70-87.

Frith, C.D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology, 63,* 287-313.

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., Benedetti, F., Abbamonte, M., Gasparotti, R., Baral, F., Perez, J., McGuire, P., & Politi, P. (2009). Functional atlas of emotional faces processing: A voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *Journal of Psychiatry and Neuroscience, 34,* 418-432.

Gamer, M., & Büchel, C. (2009). Amygdala activation predicts gaze towards fearful eyes. *Journal of Neuroscience, 29*, 9123-9126.

Gärtner, H., Minnerop, M., Pieperhoff, P., Schleicher, A., Zilles, K., Altenmüller, E., & Amunts, K. (2013). Brain morphometry shows effects of long-term practice in middle-aged keyboard players. *Frontiers in Psychology, 4*, 1-13.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds brain areas involved in face recognition. *Nature Neuroscience, 3,* 191-197.

Gay, T. (1978). Physiological and acoustic correlates of perceived stress. *Language and Speech, 21,* 8620874.

Ghashghaei, H., Hilgetag, C., & Barbas. H. (2007). Sequence of information processing for emotions based on the anatomic dialogue between prefrontal cortex and amygdala. *Neuroimage, 34,* 905-923.

Glover, G.H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage, 9,* 416-429.

Gold, B.P., Mas-Herrero, E., Zeighami, Y., Benovoy, M., Dagher, A., & Zatorre, R.J. (2019).

Musical reward prediction error engage the nucleus accumbens and motivate learning. *Proceedings of the National Academy of Sciences, 116*, 3310-3315.

Good, A., Gordon, K.A., Papsin, B.C., Nespoli, G., Hopyan, T., Peretz, I., & Russo, F.A. (2017). Benefits of music training for perception of emotion speech prosody in deaf children with cochlear implants. *Ear and Hearing, 38*, 455-464.

Gothard, K.M. (2020). Multidimensional processing in the amygdala. *Nature Reviews Neuroscience, 21*, 565-575.

Goulden, N., Khusnulina, A., Davis, N.J., Bracewell, R.M., Bokde, A.L., McNulty, J.P., & Mullins, P.G. (2014). The salience network is responsible for switching between the default mode network and the central executive network: Replications for DCM. *Neuroimage, 99,* 180-190.

Grabenhorst, F., Hernádi, I., Schultz, W. (2012). Prediction of economic choice by primate amygdala neurons. *Proceedings of the National Academy of Sciences, 109,* 18950-18955.

Green, D.M., & Swets, J.A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley and Sons.

Griffiths, T.D., Büchel, C., Frackowiak, R.S.J., & Patterson, R.D. (1998). Analysis of temporal structure in sound by the human brain. *Nature Neuroscience, 1,* 422-427.

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

Griffiths, T.D., & Warren, J.D. (2002). The planum temporale as a computational hub. *Trends in Neuroscience, 25,* 348-353.

Guillory, S.A., & Bujarski, K.A. (2014). Exploring emotions using invasive methods: Reviews of 60 years of human intracranial electrophysiology. *Social, Cognitive and Affective Neuroscience, 9,* 1880-1889.

Hackett, T.A. (2011). Information flow in the auditory cortical network. *Hearing Research, 271,*133-146.

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. *Journal of Comparative Neurology, 294,* 475-495.

Haiduk, F., Quigley, C., & Fitch, W.T. (2020). Song is more memorable than speech prosody:

Discrete pitches aid auditory working memory. *Frontiers in Psychology, 11*, 1-22.

Hajian-Tilaki, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine, 4,* 627:635.

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

Hamann, S., & Canli, T. (2004). Individual differences in emotion processing. *Current Opinion in Neurobiology, 14*, 233-238.

Hansen, C.H., & Hansen, R.D. (1988). Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology, 54,* 917-924.

Harris, A., & Aguirre, G.K. (2010). Neural tuning for face wholes and parts in human fusiform gyrus revealed by fMRI adaptation. *Journal of Neurophysiology, 104*, 336-345.

Hastings, M.E., Tangney, J.P., & Stuewig, J. (2008). Psychopathy and identification of facial expressions of emotion. *Personality and Individual Differences, 44,* 1474-1483.

Hawk, S.T., van Kleef, G.A., Fischer, A.J., & van den Schalk, J. (2009). Worth a thousand words: absolute and relative decoding of nonlinguistic vocalizations. *Emotion, 9*, 293-305.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Shouten, J.L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*, 2425-2430.

Hebart, M.N., & Baker, C.I. (2018). Deconstructing multivariate decoding for the study of brain function. *Neuroimage 180*, 4-18.

Hein, G., & Knight, R.T. (2008). Superior temporal sulcus—It's my area: or is it? *Journal of Cognitive Neuroscience, 20,* 2125-2136.

Hernádi, I., Grabenhorst, F., & Schultz, W. (2015). Planning activity for internally generated reward goals in monkey amygdala neurons. *Nature Neuroscience, 18,* 461-469.

Herry, C., Bach, D.R., Esposito, F., Di Salle, F., Perrig, W.J., Scheffler, K., Lüthi, A., & Seiftiz, E. (2007). Processing of temporal unpredictability in human and animal amygdala. *Journal of Neuroscience, 27,* 5958-5966.

Hou, J., Song, B., Chen, A.C.N., Sun, C., Zhou, J., Zhu, H., & Beauchaine, T.P. (2017). Review on neural correlates of emotion regulation and music: Implications for emotion dysregulation. *Frontiers in Psychology, 8,* 501.

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., Camerer, C.F. (2005). Neural systems responding to degrees of uncertainty in human-decision making. *Science, 310*, 1680-1683.

Iredale, J.M., Rushby, J.A., McDonald, S., Dimoska-Di Marco, A., & Swift, J. (2013). Emotion in voice matters: Neural correlates of emotional prosody perception. *International Journal of Psychophysiology, 89*, 483-490.

Jääskeläinen, I.P., Sams, M., Glearean, E., & Ahveninen, J. (2021). Movies and narratives as naturalistic stimuli in neuroimaging. *Neuroimage, 224*, 1-14.

Jackson, M.C., Wu, C-Y., Linden D.E.J., & Raymond, J.E. (2009). Enhanced visual short-term memory for angry faces. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 363-374.

Jakuszkowiak-Wojten, K., Landowski, J., Wiglusz, M.S., & Cubala, W.J. (2015). Impulsivity in anxiety disorders. A critical review. *Psychiatry Danubina, 27*, 452-455.

Jasmin, K., Lima, C.F., & Scott, S.K. (2019). Understanding rostral-caudal auditory cortex contributions to auditory perception. Nature Reviews Neuroscience, 20, 425-434.

Jimura, K., & Poldrack, R.A. (2012). Analyses of regional-average activation and multivoxel pattern information tell complementary stories. *Neuropsychologia, 50*, 544-552.

Johnstone, T., Kim, M.J., & Whalen, P.J. (2009). Functional Magnetic Resonance Imaging in the Affective and Social Sciences. In Harmon-Jones, E., & Beer, J.S., (Eds.), *Methods in Social Neuroscience*, (pp. 313-336). The Guilford Press, USA.

Johnstone, T., & Scherer, K.R. (2000). The effects of emotions on voice quality. In J.J. Ohala, Y. Hasegawa, M., Ohala, D., Granville, & A.C. Bailey (Eds.) *Proceedings of the XIVth International Congress of Phonetic Sciences* [CD-ROM]. Berkeley: University of California, Department of Linguistics.

Juslin, P.N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*, 770-814.

Kaas, J.H., & Hackett, T.A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences, 97*, 11793-11799.

Kalanthroff, E., Henrik, A., Derakshan, N., & Usher, M. (2016). Anxiety, emotional distraction, and attentional control in the Stroop task. *Emotion, 16*, 293-300.

Kantor, L., Endler, N.S., Heslegrave, R.J., & Kocovski, N.L. (2001). Validating self-report measures of state and trait anxiety against a physiological measure. *Current Psychology*, *20,* 207-215

Kanwisher, N., McDermott, J., & Chun, M.M. (1997). The Fusiform Face Area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302-4311.

Kassam, K.S. Markey, A.R., Cherkassky, V.L., Loewentstein, G., & Just, M.A. (2013). Identifying emotions on the basis of neural activation. *PLoS One, 8,* e66032.

Kawasaki, H., Kaufman, O., Damasio, H., Damasio, A.R., Granner, M., Bakken, H., Hori, T., Howard, M.A., III, & Adolphs, R. (2001). Single-neuron response to emotional visual stimuli recorded in human ventral prefrontal cortex. *Nature Neuroscience, 4,* 15-16.

Keightley, M.L., Chiew, K.S., Winocur, G., & Grady, C.L. (2007). Age-related differences in brain activity underlying identification of emotional expressions in faces. *Social Cognition in Affective Neuroscience, 2,* 292-302.

Khalil, R.A., Jones, E., Baber, M.I., Jan, T., Zafar, M.H., & Alhussain, T. (2019). Speech emotion recognition using deep learning techniques: A review: *IEEE Access, 7,* 117327-117345.

Kim, H., Somerville, L.H., Johnston T., Alexander, A.L., & Whalen, P.J. (2003). Inverse amygdala and medial prefrontal cortex responses to surprised faces. *Neuroreport, 13,* 2317-2322.

Kim, J., Shinkareva, S.V., & Wedell, D.H. (2017). Representations of modality-general valence for videos and music derived from fMRI data. *Neuroimage, 148,* 42-54.

Kim, J., Wang, J., Wedell, D.H., & Shinkareva, S.V. (2016). Identifying core affect in indviduals from fMRI responses to dynamic naturalistic audiovisual stimuli. *PLoS One, 11,* e0161589.

Kim, S-G., & Knösche, T.R. (2017). Resting state functional connectivity of the ventral auditory pathway in musicians with absolute pitch. *Human Brain Mapping, 38,* 3899-3916.

Kingdom F, & Prins N. (2016). *Psychophysics: A practical introduction*. London: Elsevier Academic Press.

Kinnison, J., Padmala, S., Choi, J.M., & Pessoa, L (2012). Network analysis reveals increased integration during emotional and motivational processing. *Journal of Neuroscience, 32,* 8361-8372.

Kirby, L.A.J, & Robinson, J.L. (2015). Affective mapping: An activation likelihood estimate (ALE) meta-analysis. *Brain and Cognition, 118,* 137-148.

Klamer, S., Schwarz, L., Krüger, O., Koch, K., Erb, M., Scheffler, K., & Ethofer, T. (2017). Association between neuroticism and emotional face processing. *Scientific Reports, 7,* 1-

8.

Kleckner, I.R., Zhang, J., Touroutoglou, A., Chanes, L., Xia, C., Simmons, W.K., Quigley, K.S., Dickerson, B.C., & Barrett, L.F. (2017). Evidence for a large-scale brain system supporting allostasis and interoception in humans. *Nature Human Behaviour, 1,* 1-14.

Ko, B.C. (2018). A brief review of facial emotion recognition based on visual information. *Sensors, 18*, 1-20.

Koelsch, S., Fritz, T., Cramon, D.Y.V., Müller, K., & Friederici, A.D. (2006). Investigation emotion with music: An fMRI study. *Human Brain Mapping, 27,* 239-250.

Koelsch, S., Fritz, T., Schulze, K., Alsop, D., & Schlaug, G. (2005). Adults and children processing music: An fMRI study. *Neuroimage, 25,* 1068-1076.

Koelsch, S., Skouras, S., & Lohmann, G. (2018). The auditory cortex hosts network nodes influential for emotion processing: An fMRI study on music-evoked fear and joy. *PLoS One, 31*, e0190057.

Kohler, C.G., Hoffman, L.J., Eastman, L.B., Healey, K., & Moberg, P.J. (2011). Facial emotion perception in depression and bipolar disorder: A quantitative review. *Psychiatry Research, 188,* 303-309.

Kohler, C.G., Walker, J.B., Martin, E.A., Healey, K.M., & Moberg, P.J. (2010). Facial emotion perception in Schizpohrenia: A maeta-analytic review. *Schizophrenia Bulletin, 36,* 1009-1019.

Kraemer, H.C. (2014). The reliability of clinical diagnoses: State of the art. *Annual Review of Clinical Psychology, 10,* 111-130.

Kragel, P.A., & LaBar, K.S. (2014). Advancing emotion theory with multivariate pattern classification. *Emotion Review, 6*, 160-174.

Krampe, R.T., & Ericsson, K.A. (1996). Maintaining excellence: Deliberate practice and elite performance in young and older pianists. *Journal of Experimental Psychology: General, 125,* 331-359.

Kret, M.E., Pichon, S., Grèzes, J., & de Gelder, B. (2011). Similarities and differences in perceiving threat from dynamic faces and bodies. An fMRI study. *Neuroimage, 54,* 1755-1762.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience, 2,* 1-28.

Kring, A.M. (2008). Emotion disturbances as transdiagnostic processes in psychopathology. In M. Lewis, J.M., Haviland-Jones, & L.F. Barrett (Eds.), *Handbook of emotions* (pp. 691-705). The Guilford Press.

Krumhansl, C.L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology/Revue Canadienne de psychologie expérimentale, 51*, 336-353.

Kuraoka, K., & Nakamura, K. (2007). Responses of single neurons in monkey amygdala to facial and vocal emotions. *Journal of Neurophysiology, 97*, 1379-1387.

Ladd, D.R., & Morton, R. (1997). The perception of intonational emphasis: Continuousu or categorical? *Journal of phonetics, 25*, 313-342.

Lai, J., Pancaroglu, R., Oruc, I., Barton, J. J., & Davies-Thomson, J. (2014). Neuroanatomic correlates of the feature-salience hierarchy in face processing: An fMRI-adaptation study. *Neuropsychologia, 53*, 274-283.

Lakens, D., Fockenberg, D.A., Lemmens, K.P.H., Ham, J., & Midden, C.J.H. (2013). The evaluation of affective pictures depends on their brightness. *Cognition and Emotion, 27,* 1225-1246.

Lamichhane, B., Adhikari, B.M., & Dhamala, M. (2016a). The activity in the anterior insulae is modulated by perceptual decision-making difficulty. *Neuroscience, 327,* 79-94.

Lamichhane, B., Adhikari, B.M., & Dhamala, M. (2016b). Salience network activity in perceptual decisions. *Brain Connectivity, 6,* 558-572.

Lamichhane, B., & Dhamala, M. (2015). The salience network and its functional architecture in a perceptual decision: An effective connectivity study. *Brain Connectivity, 5*, 1-9.

Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: Effects of acoustic features and auditory objective category. *Journal of Neuroscience, 30,* 7604-7612.

LeDoux, J.E. (1996). *The emotional brain: The mysterious underpinnings of emotional life*. Simon and Schuster, New York.

LeDoux, J.E. (2007). The amygdala. *Current Biology, 17,* R868-874.

LeDoux, J.E. (2014). Coming to terms with fear. *Proceedings of the National Academy of Sciences of the Unites States of America, 111,* 2871-2878.

Lerousseau, J.P., Hidalgo, C., & Schön, D. (2020). Musical training for auditory rehabilitation in hearing loss. *Journal of Clinical Medicine, 9,* 1-15.

Li, P., Legault, J., & Litcofsky, K.A. (2014). Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. *Cortex, 58*, 301-324.

Liebenthal, E., Silbersweig, D.A., & Stern, E. (2016). The language, tone and prosody of emotions: Neural substrates on dynamics of spoken-word emotion perception. *Frontiers in Neuroscience, 10*, 1-13.

Lin, H., Müller-Bardorff, M., Gathmann, B., Brieke, J., Mothes-Lasch, M., Bruchmann, M., Miltner, W.H.R., & Straube, T. (2020). Stimulus arousal drives amygdalar responses to emotional expressions across sensory modalities. *Scientific Reports, 10*, 1-12.

Lindquist, K.A. (2013). Emotions emerge from more basic psychological ingredients: A modern psychological constructionist model. *Emotion Review, 5,* 356-368.

Lindquist, K.A., & Barrett, L.F. (2012). A functional architecture of the human brain: Emerging insights from the science of emotion. *Trends in Cognitive Science, 16*, 533-540.

Lindquist, K. A., Wager, T. D., Kober, G., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences, 35*, 121-202.

Lindström, B.R., & Bohlin, G. (2010). Emotion processing facilitates working memory performance. *Cognition and Emotion, 25*, 1196-1204.

Livingstone, S.R., Peck, K., & Russo, F.A. (2013). Acoustic differences in the speaking and singing voice. *The Journal of the Acoustical Society of America, 19,* 035080.

Luo, C., Tu, S., Peng, Y., Gao, S., Li, J., Dong, L., Lai, Y., Li, H., & Yao, D. (2014). Long-term effects of musical training and functional plasticity in salience system. *Neural Plasticity, 2014,* 1-13.

MacLeod, C., & Matthews, A. (2012). Cognitive bias modification approaches to anxiety. *Annual Review of Clinical Psychology, 8,* 189-217.

Macmillan N.A., & Creelman, C.D. (2004). *Detection Theory: A user's guide*. New York: Psychology Press.

Makovac, E., Mancini, M., Fagioli, S., Watson, D.R., Meetan, F., Rae, C.L., Critchley, H.D., & Ottaviani, C. (2018). Network abnormalities in generalized anxiety pervade beyond the amygdala-pre-frontal cortex circuity: Insights from graph theory. *Psychiatry Research: Neuroimaging, 281*, 107-116.

Manno, F.A.M., Fernandez-Ruiz, J., Manno, S.H.C., Cheng, S.H., Lau, C., & Barrios, F.A. (2019). Sparse sampling of silence type I errors with an emphasis on primary auditory cortex. *Frontiers in Neuroscience, 13,* 1-11.

Mansfield, P., Glover, P.M., & Beaumont, J. (1998). Sound generation in gradient coil structures for MRI. *Magnetic Resonance in Medicine. 39,* 539-550.

Margulies, D.S., Gosh, S.S., Goulas, A., Falkiewicz, M., Huntenburg, J.M., Langs, G., Bezgin, G., Eickhoff, S.B., Castellanos, F.X., Petrides, M., Jefferies, E., & Smallwood, J. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences, 113,* 12574-12579.

Markett, S., Jawinski, P., Kirsch, P., & Gerchen, M.F. (2020). Specific and segregated changes to the functional connectome evoked by the processing of emotional faces: A task-based connectome study. *Scientific reports, 10,* 1-14.

Marques, C., Moreno, S., Catro, S.L., & Besson, M. (2007). Musicians detect pitch violations in a foreign language better than nonmusicians: Behavioural and electrophysiological evidence. *Journal of Cognitive Neuroscience, 19,* 1453-1463.

Matsumoto, D. (1990). Cultural similarities and differences in display rules. *Motivational Emotion, 14,* 195-214.

Matsumoto, D. (1992) American-Japanese cultural differences in the recognition of universal facial expressions. *Journal of Cross-Cultural Psychology, 23,* 72-84.

Mattavelli, G., Sormaz, M., Flack, T., Asghar, A.U.R., Fan, S., Frey, J., Manssuer, L., Usten, D., Young, A.W., & Andrews, T.J. (2014). Neural responses to facial expressions support the role of the amygdala in processing threat. *Social, Cognitive and Affective Neuroscience, 9,* 1684-1689.

Mauss, I.B., & Robinson, M.D. (2009). Measures of emotion: A review. *Cognition & Emotion, 23,* 209-237.

McCrae, R.R., & John, O.P. (1002). An introduction to the five-factor model and its applications. *Journal of Personality, 60,* 175-215.

McIntosh, A.R. (2000). Towards a network theory of cognition. *Neural Networks, 13,* 861-870.

McIntosh, A.R. (2004). Contexts and Catalysts: A resolution of the localization and integration of function in the brain. *Neuroinformatics, 2,* 175-181.

McIntosh, A.R., & Gonzalez-Lim, F. (1994). Structural equation modeling and its application to network analysis in functional brain mapping. *Human Brain Mapping, 2*, 2-22.

McKeown, M.J., Hansen, L.K., Sejnowski, T.J. (2003). Independent component analysis of functional MRI: What is signal and what is noise? *Current Opinion in Neurobiology, 13,* 620-629.

Melcher, D., & Bacci, F. (2013). Perception of emotion in abstract artworks: A multidisciplinary approach. *Progress in Brain Research, 204,* 191-216.

Méndez-Bértolo, C., Moratti, S., Toledano, R., Lopez-Sosa, F., Martínez-Alvarez, R., Mah, Y.H., Vuilleumier, P., Gil-Nagel, A., & Strange, B.A. (2016). A fast pathway for fear in human amygdala. *Nature Neuroscience, 19,* 1041-1049.

Menon, V., & Uddin, L.Q. (2010). Saliency, switching, attention and control: A network model of insula function. *Brain Structure and Function, 214,* 655-667.

Meyer, M., Elmer, S., & Jäncke, L. (2012). Musical expertise induces neuroplasticity of the planum temporale. Annals of the New York Academy of Sciences, 1252, 116-123.

Millan, M.J. (2003). The neurobiology of control of anxious states. *Progress in Neurobiology, 70,* 83-244.

Misaki, M., Kim, Y., Bandettini, P.A., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage, 53,* 103-118.

Mobbs, D., Hagan, C.C., Dalgleish, T., Silston, B., & Prévost, C. (2015). The ecology of human fear: Survival optimization and the nervous system. Frontiers in Neuroscience, 9, 1-22.

Moeller, S., Auerbach, E., Van de Moortee, P.F., Adriany, G., & & Ugurbil, K. (2008). fMRI with 16 fold reduction using multibanded multislice sampling. *Proceedings of the International Society for Magnetic Resonance Medicine, 16*, 2366.

Moeller, S., Yacoub, E., Olmna, C.A., Auerbach, E., Strupp, J., Harel, N., & Urbil, K (2010). Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magnetic Resonance in Medicine, 63,* 1144-1153.

Mogg, K., & Bradley, B.P. (1998). A cognitive-motivational analysis of anxiety. *Behaviour Research and Therapy, 36,* 809-848.

Morrow, J., Mosher, C., & Gothard, K. (2019). Multisensory neurons in the primate amygdala. *Journal of Neuroscience, 39,* 3663-3675.

Munn, Z., & Jordan, Z. (2011). The patient experience of high technology medical imaging: A systematic review of the qualitative evidence. *Radiography, 17,* 323-331.

Münte, T.F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. Nature Reviews *Neuroscience, 3*, 473-478.

Murphy, F.C., Nimmo-Smith, I., & Lawrence, A.D. (2003). Functional neuroanatomy of emotions: A meta-analysis. *Cognitive, Affective, & Behavioural Neuroscience, 3,* 2007-233.

Murray, R.J., Brosch, T., & Sander, D. (2014). The functional profile of the human amygdala in affective processing: Insights from intracranial recordings. Cortex, 60, 10-33.

Murry, M.W.E., & Isaacowitz, D.M. (2016). Age differences in emotion perception: The effects of the social environment. *International Journal of Behavioural Development, 41*, 597-604.

Mustovic, H., K. Scheffler, F. Di Salle, Esposito, F., Neuhoff, J.G., Henning, J., & Seifritz, E. (2003). Temporal integration of sequential auditory events: silent period in sound pattern activates human planum temporale. *Neuroimage, 20,* 429-434.

Nachev, P., Kennard, C., & Husain, M. (2008). Functional role of the supplementary and pre-supplementary motor areas. *Nature Reviews Neuroscience, 9,* 856-869.

Nazarov, A., Frewen, P., Oremus, C., Schellenberg, E.G., McKinnon, M.C., & Lanius, R. (2015). Comprehension of affective prosody in women with post-traumatic stress related to childhood abuse. *Acta Psychiatrica Scandinavica, 131*, 342-349.

Nebel, K., Stude, P., Wiese, H., Müller, B., Greiff, A., Forsting, M., Diener, H-C., & Keidel, M. (2005). Sparse imaging and continuous event-related fMRI in the visual domain: A systemic comparison. *Human Brain Mapping, 24,* 130-143.

Neta, M., Berkebile, M.., & Freeman, J.B. (2020). The dynamic process of ambiguous emotion perception. *Cognition and Emotion,* 1-8

Norman-Haignere S.V., Kanwisher, N., & McDermott, J.H. (2013). Cortical pitch regions in human respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. *Journal of Neuroscience, 33,* 19451-19469.

Norman-Haignere, S.V., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron, 88,* 1281-1296.

Novak, L.R., Gitelman, D.R., Schuyler, B., & Li, W. (2015). Olfactory-visual integration facilitates perception of subthreshold negative emotion. *Neuropsychologia, 77,* 288-297.

Öhman, A. (2005). The role of the amygdala in human fear: Automatic detection of threat. *Psychoneuroendocrinology, 30,* 953-958.

Öhman, A., Carlsson, K., Lundqvist, D., & Ingvar, M. (2007). On the unconscious subcortical origin of human fear. *Physiological Behavior, 92,* 180-185.

Ohnishi, T., Matsuda, H., Asada, T., Aruga, M., Hikarata, M., Nishikawa, M., Katoh, A., & Imabayashi, E. (2001). Functional anatomy of musical perception in musicians. *Cerebral Cortex, 11,* 754-760.

Okon-Singer, H., Hendler, T., Pessoa, L., & Shackman, A.J. (2015). The neurobiology of emotion-cognition interactions: Fundamental questions and strategies for future research. *Frontiers in Human Neuroscience, 9,* 1-14

O'Neil, H. F., Jr., Spielberger, C. D., & Hansen, D. N. (1969). Effects of state anxiety and task difficulty on computer-assisted learning. *Journal of Educational Psychology, 60,* 343-350.

Pacheco-Unguetti, A.P., Acosta, A., Callejas, A., & Lupiáñez, J. (2010). Attention and anxiety: Different attentional functioning under state and trait anxiety. *Psychological Science, 21,* 298-304.

Panksepp, J. (1982). Toward a general psychobiological theory of emotions. *Behavioral and Brain sciences, 5,* 407-422

Panksepp J. (1998). *Affective neuroscience: The foundations of human and animal emotions.* Oxford University Press; New York.

Panno, A., Donati, M.A., Milioni, M., Chiesi, F., & Primi, C. (2017). Why women take fewer risk than men: The mediating role of state anxiety. *Sex Roles, 78,* 286-294.

Paquette, S., Rigoulot, S., Grunewald, K., & Lehmann, A. (2020). Temporal decoding of vocal and musical emotions: Same code, different timecourse? *Brain Research, 1741,* 1-6.

Paquette, S., Takerkart, S., Saget, S., Peretz, I., & Belin, P. (2018). Cross-classification of musical and vocal emotions in the auditory cortex. *Annals of the New York Academy of Sciences, 1423*, 329-337.

Park, D.C., & Huang, C-M. (2010). Culture wires the brain: A cognitive neuroscience perspective. *Perspectives on Psychological Science, 5,* 391-400.

Park, H., Lee, S., Kim, H.J., Ju, Y.S., Shin, J.Y., von Grotthuss, M., Lee, D-S., Park C., Kim, J.H., Kim, B., Yoo, Y.J., Cho, S-I., Sung, J., Lee, C., Kim, J-I., & Seo, J-S. (2012).

Comprehensive genomic analyses associate UGT8 variants with musical ability in a Mongolian population. *Journal of Medical Genetics, 49*, 747-752.

Passardi, S., Peyk, P., Ryfer, M., Plichta, M., Mueller-Pfeiffer, C., Wingenbach, T.S.H., Hassanpour, K., Schnyder, U., & Pfaltz, M.C. (2018). Impaired recognition of positive emotions in individuals with posttraumatic stress disorder, cumulative traumatic exposure, and dissociation. *Psychotherapy and Psychosomatics, 87*, 118-120.

Patel, A., & Peretz, I. (1997). Is music autonomous from language. A neuropsychological appraisal. In I. Deliège & J. Sloboda (Eds.), *Perception and cognition of music* (pp. 191-215). Hove, England: Psychology Press/Erbaum (UK). Taylor & Francis.

Patterson, R.D., Uppenkamp, Johnsrude, I.S., & Griffiths, T.D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron, 36*, 767-776.

Paulmann, S., Bleichner, M., & Kotz, S.A. (2013). Valence, arousal, and task effects in emotion prosody processing. *Frontiers in Psychology, 4*, 1-10.

Peelen, M.V., Atkinson, A.P., & Vuilleumier, P. (2010). Supramodal representations of perceived emotions in the human brain. *Journal of Neuroscience, 30,* 10127-10134.

Peelen, M.V., & Downing, P.E. (2005a). Within-subject reproducibility of category-specific visual activation with functional MRI. *Human Brain Mapping, 25*, 402-408.

Peelen, M.V., & Downing, P.E. (2005b). Selectivity for the human body in the fusiform gyrus. *Journal of Neurophysiology, 93*, 603-608.

Pell, M.D., & Kotz, S.A. (2011). On the time course of vocal emotion recognition. *PLoS One, 6*, e27256.

Pell, M.D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology, 111*, 14-25.

Pelle, J.E. (2014). Methodological challenges and solutions in auditory functional magnetic resonance imaging. *Frontiers in Neuroscience, 8,* 1-13.

Penhune, V.B., Zattore, R.J., & Evans, A.C. (1998). Cerebellar contributions to motor timing: A PET study of auditory and visual rhythm reproduction. *Journal of Cognitive Neuroscience, 10,* 752-765.

Pereira, C.S., Teixeria, J., Figueiredo, P., Xavier, J., Castro S.L., & Brattico, E. (2011). Music and emotions in the brain: Familiarity Matters. *PLoS One, 6*, e27241.

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifies and fMRI: A tutorial overview. *Neuroimage, 45,* S199-S209.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience, 6*, 688-691.

Peretz, I., Vuvan, D., Lagrois, M. É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society B, 370*, 1-8.

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews Neuroscience, 9,* 148-158.

Pessoa, L. (2009). How do emotion and motivation direct executive control? *Trends in Cognitive Sciences, 13,* 160-166.

Pessoa, L. (2010). Emotion and cognition and the amygdala: From "what is it?" to "what's to be done?" *Neuropsychologia, 48,* 3416-3429.

Pessoa, L. (2015). Précis on the cognitive-emotional brain. *Behavioural and Brain Sciences, 38,* 1-66.

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: From a 'low road' to 'many roads' of evaluating biological significance. *Nature Reviews Neuroscience, 11*, 773-783.

Pessoa, L., & Padmala, S. (2005). Quantitative prediction of perceptual decisions during near-threshold fear detection. *Proceedings of the National Academy of Sciences, 102*, 5612-5617.

Pichon, S., Miendlarzewska, E.A., Eryilmaz, H., & Vuilleumier, P. (2015). Cumulative activation during positive and negative events and state anxiety predicts subsequent inertia of amygdala reactivity. *Social Cognitive Affective Neuroscience, 10*, 180-190.

Phan, K.L., Wager, T., Taylor, S.F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage, 16*, 331-348.

Phelps, E.A., Lempert, K.M., Sokol-Hessner, P. (2014). Emotion and decision making: Multiple modulatory neural circuits. *Annual Review of Neuroscience*, 37, 263-287.

Phillips, M.L., Drevets, W.C., Rauch, S.L., & Lane, R. (2003). Neurobiology of emotion perception II: Implications for major psychiatric disorders. *Biological Psychiatry, 54*, 515-528.

Pitkänen, A., Savander, V., & LeDoux, J.E. (1997). Organization of intra-amygdaloid circuitries in the rat: An emerging framework for understanding functions of the amygdala. *Trends in Neuroscience, 20,* 517-523.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication, 41*, 245-255.

Poldrack, R.A., Baker, C.I., Durnez, J., Gorgolewski, K.J., Matthews, P.M., Munafò, M.R., Nichols, T.E., Poline, J-B., Vul, E., & Yakoni, T. (2017). Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience, 18,* 115-126.

Posner, J., Russell, J.A., & Peterson, B.S. (2008). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology, 17,* 715-734.

Posner, J., Russell, J.A., Gerber, A., Gorman, D., Colibazzi, T., Yu, S., Wang, Z., Kangarlu, A., Zhu, H., & Peterson, B.S. (2009). The neurophysiological bases of emotion: An fMRI study of the affective circumplex using emotion-denoting words. *Human Brain Mapping, 30*, 883-895.

Poyo-Solanas, M., Zhan, M., Vaessan, M., Hortensius, R., Engelen T., & de Gelder, B (2018). Looking at the face and seeing the whole body. Neural basis of combined face and body expressions. *Social Cognitive and Affective Neuroscience, 13,* 135-144.

Prado, J., Carp, J., & Weissman, D.H. (2011). Variations of response time in a selective attention task are linked to variations of functional connectivity in the attentional network. *Neuroimage, 54,* 541-549.

Price, C.J., & Friston, K. (2005). Functional ontologies for cognition: The systematic definition of structure and function. *Cognitive Neuropsychology, 22,* 262-275.

Qiao, J., Li, A., Cao, C., Wang, Z., Sun, J., & Xu, G. (2017). Aberrant functional network connectivity as a biomarker of generalized anxiety disorders. *Frontiers in Human Neuroscience, 11,* 1-10.

Rabe, K., Michael, N., Kugel, H., Heindel, W., & Pfleiderer, B. (2006). FMRI studies of sensitivity and habituation effects within the auditory cortex at 1.5 T and 3 T. *Journal of Magnetic Resonance Imaging, 23,* 454-458.

Rajhans, P., Jessen, S., Missana, M., & Grossman, T. (2016). Putting the face in context: Body expressions impact facial emotion processing in human infants. *Developmental Cognitive*

Neuroscience, 19, 115-121.

Rauschecker, J.P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hearing Research, 271*, 16-25.

Rauschecker, J.P. (2018). Where, when and how: Are they all sensorimotor? Towards a unified view of the dorsal pathway in vision and audition. *Cortex, 98,* 262-268.

Robinson, J.L., Laird, A.R., Glahn, D.C., Lovallo, W.R., & Fox, P.T. (2010). Metaanalytic connectivity modeling: Delineating the functional connectivity of the human amygdala. *Human Brain Mapping, 31*, 173-184.

Robinson, O.J., Vytal, K., Cornwell, B.R., & Grillon, C. (2013). The impact of anxiety upon cognition: Perspectives from human threat of shock studies. *Frontiers in Human Neuroscience, 7,* 1-21.

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*, 3843-3852.

Rowe, J.B. (2010). Connectivity analysis is essential to understand neurological disorders. *Frontiers in Systems in Neuroscience, 4*, 1-13.

Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161-1178.

Sah, P., Faber, E.S.L., Lopez de Armentia, M., & Power, J. (2003). The amygdaloid complex: Anatomy and physiology. *Physiological Reviews, 83*, 803-834.

Said, C.P., Moore, C.D., Engell, A.D., Todorov, A., & Haxby, J.V. (2010). Distributed representations of dynamic facial expression in the superior temporal sulcus. *Journal of Vision, 10*, 1-12.

Said, C.P., Moore, C.D., Norman, K.A., Haxby, J.V., & Todorov, A. (2010). Graded representations of emotional expressions in the left superior temporal sulcus. *Frontiers in Systems Neuroscience, 4,* 1-8.

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 *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (pp. 215-218). New Paltz, NY: IEEE.

Saitou, T., Tsuji, N., Unoki, M., & Akagi, M. (2004). Analysis of acoustic features affecting "singing-ness" and its applications to singing-voice synthesis from speaking-voice. In *Proceedings of the 8th International Conference on Spoken Language Processing (ICSLP2004)* (pp. 1929-1932). Jeju Island, Korea.

Samartsidis, P., Montagna, S., Nichols, T.W., & Johnson, T.D. (2017). The coordinate-based meta-analysis of neuroimaging. *Statistical Science, 32*, 580-599.

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: An evolved system for relevance detection. *Reviews in the Neurosciences, 14,* 303-316.

Sander, D., Grandjean D., & Scherer, K.R. (2018). An appraisal-driven componential approach to the emotional brain. *Emotion Review, 10*, 219-231.

Sander, K., Brechmann, A., & Scheich, H. (2003). Audition of laughing and crying leads to right amygdala activation in low-noise fMRI setting. *Brain Research Protocols, 11*, 81-91.

Sander, K., & Scheich, H. (2001). Auditory perception of laughing and crying activates human amygdala regardless of attentional state. *Cognitive Brain Research, 12,* 181-198.

Scherer, K.R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99,* 143-165.

Schirmer, A., & Adolphs, R. (2017). Emotion perception from face, voice and touch: Comparisons and convergence. *Trends in Cognitive Sciences, 21*, 216-228.

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*, 137-147.

Schirmer, A., & Kotz, S.A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotion processing. *Trends in Cognitive Sciences, 10*, 24-30.

Schmithorst, V.J., & Holland S.K. (2003). The effect of musical training on music processing: A functional magnetic resonance imaging study in humans. *Neuroscience Letters, 348,* 65-68.

Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Song as an aid for language acquisition. *Cognition, 106,* 975-983

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,* 450-461.

Schwartz, C.E., Wright, C.I., Shin, L.M., Kagan, J., Whalen, P.J., McMullin, K.G., & Rauch, S.L. (2003). Differential amygalar response to novel versus newly familiar neutral faces:

A functional MRI probe developed for studying inhibited temperament. *Biological Psychiatry, 53,* 853-862.

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*, 774-782.

Schwarzlose, R.F., Baker, C.I., & Kanwisher, N. (2005). Separate face and body selectivity on the fusiform gyrus. *Journal of Neuroscience, 25,* 11055-11059.

Scott, S.K., Young, A.W., Calder, A.J., Hellawell, D.J., Aggleton, J.P., & Johnson, M. (1997). Impaired auditory recognition of fear and anger following bilateral amygdala lesions. *Nature, 385*, 254-257.

Seeley, W.W., Menon, V., Schatzberg, A.F., Keller J., Glover, G.H., Kenna, H., Reis, A.L., & Greicius, M.D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience, 27,* 2349-2356.

Sergerie, K., Chocol, C., & Armony, J.L. (2008). The role of the amygdala in emotional processing: A quantitative meta-analysis of functional neuroimaging studies. *Neuroscience and Biobehavioural Reviews, 32*, 811-830.

Seymour, B., & Dolan, R. (2008). Emotion, decision making, and the amygdala. *Neuron*, *58,* 662-671.

Sharpe, D. (1997). Of apples and oranges, file drawers, and garbage: Why validity issues in meta-analysis will not go away. *Clinical Psychology Review, 17,* 881-901.

Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420-428.

Sievers, B., Parkinson, C., Kohler, P.K., Hughes, J., Fogelson, S.V., & Wheatley, T. (2018). Visual and auditory brain areas share a neural code for perceived emotion. *bioRxiv,* 1-23.

Simmons, W.K., Avery, J.A., Barcalow, J.C., Bodurka, J., Drevets, W.C., & Bellgowan, P. (2013). Keeping the body in mind: Insula functional organization and functional connectivity integrate interoceptive, exteroceptive, and emotional awareness. *Human Brain Mapping, 34,* 2944-2958.

Singer, N., Eapen, M., Grillon, C., Ungerleider, L.G., & Hendler, T. (2012). Through the eyes of anxiety: Dissecting threat bias via emotional-binocular rivalry. *Emotion, 12,* 960-969.

Smith, P.F. (2018). On the application of multivariate statistical and data mining analyses to data in neuroscience. *Journal of Undergraduate Neuroscience Education, 16*, R20-R32.

Sokolova, M., Japkoqicz, N., & Szpakowicz, S. (2006). Beyond accuracy, f-score, and ROC: A family of discriminant measures for performance evaluation. In Satter, A., & Kang, B. (Eds.), *Australasian joint conference on artificial intelligence* (pp. 1015-1021). Springer, Berlin, Heidelberg.

Somerville, L.H., Kim, H., Johnstone, T., Alexander, A.L., & Whalen, P.J. (2004). Human amygdala responses during presentation of happy and neutral faces: Correlations with state anxiety. *Biological Psychiatry, 55*, 897-903.

Soshi, T., Nagamine, M., Fukuda, E., & Takeuchi, A. (2019). Pre-specified anxiety predicts future decision-making performances under different temporally constrained conditions. *Frontiers in Psychology, 10*, 1-17.

Sridharan, D., Levitin, D.J., & Menon, V. (2008). A critical role of the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proceedings of the National Academy of Sciences, 105*, 12569-12574.

Stein, J.L., Wiedholz, L.M., Bassett, D.S., Weinberger, D.R., Zink, C.F., Mattay, V.S., & Meyer-Lindenberg, A. (2007). A validated network of effective amygdala connectivity. *Neuroimage, 36,* 736-745

Stemmler, G. (2004). Physiological processes during emotion. In P. Philippot & R.S. Feldman (Eds.), *The regulation of emotion* (pp. 33-70). Lawrence Erlbaum Associates Publishers.

Sterzer, P., & Kleinschmidt, A. (2007). A neural basis for inference in perceptual ambiguity. *Proceedings of the National Academy of Sciences, 104,* 323-328.

Sterzer, P., & Kleinschmidt, A. (2010). Anterior insula activations in perceptual paradigms: Often observed but barely understood. *Brain Structure and Function, 214*, 611-622.

Streim-Amit, E., Guendelman, M., & Amedi, A. (2012). ‘Visual’ acuity of the congenitally blind using visual-to-auditory sensory substitution. *PLoS One,* 7, e33136.

Strohmaier, A.R., Schiepe-Tiska, A., & Reiss, K.M. (2020). A comparison of self-reports and electrodermal activity as indicators of mathematics state anxiety. An application of the control-value theory. *Frontline Learning Research, 8,* 16-32.

Surcinelli, P., Codispoti, M., Montebarocci, O., Rossi, N., & Baldaro, B. (2006). Facial emotion recognition in trait anxiety. *Journal of Anxiety Disorders, 20,* 110-117.

Sweet, R.A., Dorph-Peterson, 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*, 270-289.

Swets, JA. (1988). Measuring the accuracy of diagnostic systems. Science. 240, 1285-1293.

Sylvester, C.M., Corbetta, M., Raichle, M.E., Rodebaugh, T., Schlagger, B.L., Sheline, Y.I., Zorumski, C.F., & Lenze, E.J. (2012). Functional network dysfunction in anxiety and anxiety disorders. *Trends in Neuroscience, 35,* 527-535.

Sylvester, C.M., Yu, Q., Srivastava, A.B., Marek, S., Zheng, A., Alexopoulos, D., Smyser, C.D., Shimony, H.S., Ortega, M., Dierker, D.L., Patel, G.H., Nelson, S.M., Gilmore, A.W., McDermott, K.B., Berg, J.J., Drysdale, A.T., Perino, M.T., Snyder, A., Raut, R.V., … Dosenbach, N.U.F. (2020). Individual-specific functional connectivity of the amygdala: A substrate for precision psychiatry. *Proceedings of the National Academy of Sciences, 117,* 3808-3818.

Tabei, K-I. (2015). Inferior frontal gyrus activation underlies the perception of emotions, while precuneus activation underlies the feeling of emotions during music listening. *Behavioural Neurology, 2015,* 1-6.

Theusch, E., & Gitschier, J. (2011). Absolute pitch twin study and segregation analysis. *Twin Research and Human Genetics, 14,* 173-178.

Thielscher, A., & Pessoa, L. (2007). Neural correlates of perceptual choice and decision making during fear-disgust discrimination. *Journal of Neuroscience, 27,* 2908-2917.

Todorovic, A., & de Lange, F.P. (2012). Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *Journal of Neuroscience, 32,* 389-395.

Torppa, R., Faulkner, A., Kujala, T., Huotilainen, M, & Lipsanen, J. (2018). Developmental links between speech perception in noise, singing, and cortical processing of music in children with cochlear implants. *Music Perception: An Interdisciplinary Journal, 36,* 156-174.

Touroutoglou, A., Lindquist, K.A., Dickerson, B.C., & Barrett, L.F. (2015). Intrinsic connectivity in the human brain does not reveal networks for ‘basic’ emotions. *Social Cognitive and Affective Neuroscience, 10,* 1257-1265.

Trilla, I., Weigand, A., & Dziobek, I. (2020). Affective states influence emotion perception: Evidence for emotional egocentricity. *Psychological Research, 85,* 1005-1015.

Uddin, L.Q. (2015). Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscience, 15,* 55-61.

Uddin, L.Q., Kinnison, J., Pessoa, L., & Anderson, M.L. (2014). Beyond the tripartite cognition-emotion-interoception model of the human insular cortex. *Journal of Cognitive Neuroscience, 26,* 16-27.

Ulmer, J.L., Biswal, B.B., Yekin, F.Z., Mark, L.P., Matthews, V.P., Prost, R.W., Estkowski, L.D., McAuliffe, T.L., Haughton, V.M., & Daniels, D.L. (1998). Cortical activation response to acoustic echoplaner scanner noise. *Journal of Computer Assisted Tomography, 22,* 111-119.

Ursin, H., & Kaada, B.R. (1960). Functional localization within the amygdaloid complex in the cat. *Electroencephalography and Clinical Neurophysiology, 12,* 1-20.

Van den Heuvel, M.P., & Sporns, O. (2013a). Network hubs in the human brain. *Trends in Cognitive Science, 7,* 683-696.

Van den Heuvel, M.P., & Sporns, O. (2013b). An anatomical substrate for integration among functional networks in human cortex. *Journal of Neuroscience, 33,* 14489-14500.

Van Overwalle, F., Baetens, K., Mariën, P., & Vandekerckhove, M. (2014). Social cognition and the cerebellum: A meta-analysis of over 350 fmri studies. *Neuroimage, 86,* 554-572.

Wager, T. D., Barrett, L. F., Bliss-Moreau, E., Lindquist, K. A., Duncan, S., Kober, H., Joseph, J., Davidson, M., & Mize, J. (2008). The neuroimaging of emotion. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (p. 249-271). The Guilford Press.

Wager, T., Kang, J., Johnson, T.D., Nichols, T.E., Satpute, A.B., & Barrett, L.F. (2015). A Bayesian model of category-specific emotional brain responses. *PLoS Computational Biology, 11,* 1-27.

Wager, T.D., Lindquist, M., & Kaplan, L. (2007). Meta-analysis of functional neuroimaging data: Current and future directions. *Social Cognitive and Affective Neuroscience, 2,* 150-158.

Wagner, D.D., Kelley, W.M., Haxby, J.V., & Heatherton, T.F. (2016). The dorsal medial prefrontal cortex responds preferentially to social interactions during natural viewing. *Journal of Neuroscience, 36,* 6917, 6925.

Wang, S., Rongjun, Y., Tyszka, J.M., Zhen, S., Kovach, C., Sun, S., Huang, T., Hurlemann, R., Ross, I.B., Chung, J.M., Mamelak, A.N., Adolphs, R., & Rutishauser, U. (2017). The human amygdala parametrically encodes the intensity of specific facial emotions and their categorical ambiguity. *Nature Communications, 8,* 1-13.

Wang, S., Tudusciuc, O., Mamelak, A.N., Ross, I.B., Adolphs, R., & Rutishauser, U. (2014). Neurons in the human amygdala selective for perceived emotion. *Proceedings of the National Academy of Science, 111,* E3110-E3119.

Wang, Y., Zhou, W., Cheng, Y., & Bian, X. (2017). Gaze patterns in auditory-visual perception of emotion by children with hearing aids and hearing children. *Frontiers in Psychology, 8,* 1-9.

Warren, J.D., Zielinski, B.A., Green, G.G.R., Rauschecker, J.P., & Griffiths, T.D. (2002). Perception of sound-source motion by the human brain. *Neuron, 34,* 139-148.

Warren, J.E., Wise, R.J.S., Warren, J.D. (2005) Sounds do-able: Auditory-motor transformations and the posterior temporal plane. *Trends in Neuroscience*, 28, 636-643.

Watson, R., Latinus, M., Noguchi, T., Garrod, O., Crabbe, F., & Belin, P. (2014). Crossmodal adaptation in right posterior temporal sulcus during face-voice emotional integration. *Journal of Neuroscience, 34,* 6813-6821.

Weems, S.A., & Reggia, J.A. (2006). Simulating single word processing in the classic aphasia syndromes based on the Wernicke-Lichtheim-Geschwind theory. *Brain and Language, 98,* 291-309.

Westermann, R., Spies, K., Stahl, G., & Hesse, F.W. (1995). Relative effectiveness and validity of mood induction procedures: A meta-analysis. *European Journal of Social Psychology, 26,* 557-580.

Whalen, P.J. (1998). Fear, vigilance, and ambiguity: Initial neuroimaging studies of the human amygdala. *Current Directions in Psychological Science, 7,* 117-188.

Whalen, P.J., Shin, L.M., McInerney, S.C., Fischer, H., Wright, C.I., & Rauch, S.L. (2001). A functional MRI study of human amygdala responses to facial expression of fear versus anger. *Emotion, 1*, 70-83.

Wicker, B., Keysers, C., Plailly, J., Royet, J-P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron, 40,* 655-664.

Wiethoff, S., Wildgruber, D., Kreifelts, B., Becker, H., Herbert, C., Grodd, W., & Ethofer, T. (2008). Cerebral processing of emotional prosody—influence of acoustic parameters and arousal. *Neuroimage, 39,* 885-893.

Wildgruber, D., Riecker, A., Hertirch, I., Erb, M., Grodd, W., Ethofer, T., & Ackermann, H.

(2005). Identification of emotional intonation evaluated by fMRI. *Neuroimage, 24,* 1233-1241.

Wiglusz, M.S., Landowski, J., & Cubala, W.J. (2019). Psychometric properties and diagnostic utility of the State-Trait Anxiety Inventory in epilepsy with and without comorbid anxiety disorder. *Epilepsy Behaviour, 92,* 221-225.

Williams, C.L., Milanak, M.E., Judah, M.R., & Berenbaum, H. (2018). The association between PTSD and facial affect recognition. *Psychiatry Research, 265,* 298-302.

Williams, J.M.G., Watts, F.N., MacLeod, C., & Mathews, A. (1997). *Cognitive psychology and emotional disorders* (2nd ed.). Chichester, England: Wiley

Wong, I.Y., Mahar, D.P., Titchener, K., & Freeman, J. (2013). The impact of anxiety on processing efficiency: Implications for the attentional control theory. *Open Behavioural Science Journal, 6,* 7-15.

Woo, C-W., Chang, L.J., Lindquist, M.A., & Wager, T.D., (2017). Building better biomarkers: Brain models in translational neuroimaging. *Nature Neuroscience, 20,* 365-377.

Xia, L., Gu, R., Zhang, D., & Luo, T. (2017). Anxious individuals are impulsive decision-makers in the delay discount dark: An ERP study. *Frontiers in Behavioral Neuroscience, 11*, 1-11.

Xu, P., Gu, R., Broster, L.S., Wu., R., Van Damn, N.T., Jiang, Y., Fan, J., & Luo, Y-J. (2013). Neural basis of emotional decision making in trait anxiety. *Journal of Neuroscience, 33,* 18641-18653.

Yeo, B.T.T., Krienen, F.M., Sepulcre, J., Sabunca, M.R., Laskari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., Fischl, B., Liu, H., & Buckner, R.L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology, 106*, 1125-1165.

Young, A.W., Frühholz, S., & Schweinberger, S.R. (2020). Face and voice perception: Understanding commonalities and differences. *Trends in Cognitive Sciences, 24*, 398-410.

Zadra, J.R., & Clore, G.L. (2011). Emotion and perception: The role of affective information. *Wiley Interdisciplinary Review of Cognitive Science, 2*, 676-685.

Zald, D.H., & Zatorre, R.J. (2011). Chapter 19. Music. In: J.A. Gottfried (Ed.). *Neurobiology of Sensation and Reward. Boca Raton*: CRC Press/Taylor & Francis. p. 1-98.

Zamorano, A.M., Cifre, I., Montoya, P., Riquelme, I., & Kleber, B. (2017). Insula-based networks in professional musicians: Evidence for increased functional connectivity during resting state. *Human Brain Mapping, 38*, 4834-4849.

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

Zhang, J., Li, X., Song, Y., & Liu, J. (2012). The fusiform face area is engaged in holistic, not parts-based, representation of faces. *PLoS One, 7,* e40390.

Zhang, Z. (2016). Mechanisms of human voice production and control. *The Journal of the Acoustical Society of America, 140*, 2614-2635.

Zhao, X-H., Wang, P-J., Li, C-B., Hu, Z-H, Xi, W., Wu, W-Y., & Tang, X-W. (2007). Altered default mode network activity in patient with anxiety disorders: An fMRI study. *European Journal of Radiology, 63,* 373-378.

Appendix 1: List of Doctoral Degree Publications

Kondo, F., **Whitehead, J.C.,** Corbalán, F., Beauileu, S., & Armony, J.L. (*in preparation*). Emotion regulation in bipolar disorder type-I: MVPA analysis of fMRI data.

Herholz, P., **Whitehead, J.C.**, Poline, J-B., & Armony, J.L. (*in preparation*). Investing principles of the auditory cortex via multivariate, connectivity and meta-analyses.

Whitehead, J.C., Spiousas, I., & Armony, J.L. (*pre-submission*). Neural correlates of socio-emotional ambiguity and threat-related processing across modality.

Whitehead, J.C., & Armony, J.L. (*under review*). Intra-individual reliability of voice- and music-elicited repsonses and their modulation by expertise.

Whitehead, J.C., & Armony, J.L. (2019). Amygdala Processing of Vocal Emotions, in S. Frühholz and P. Belin (eds.) The Oxford Handbook of Voice Perception. Oxford University Press, pp.473-494.

Whitehead, J.C., & Armony, J.L. (2018). Multivariate fMRI pattern analysis of fear perception across modalities. Eur. J. Neurosci, doi:10.1111/ejn.14322

Whitehead, J.C., & Armony, J.L. (2018) Singing in the brain: Neural representation of music and voice as revealed by fMRI. Hum. Brain Mapp. doi:10.1002/hbm.24333

Whitehead, J.C., & Armony, J.L. (2015). Neural representation of emotion across modalities. In: Abstracts of the 2015 CSBBCS Annual Meeting: Résumés du congrès annuel 2015 de la SCCCSC. Can J Exp Psychol. 69(4): 327-376.