Categories, classification, and cortical processing streams:
the multiple pathways of music perception

Mike E. Klein

Doctor of Philosophy

Psychology

McGill University

Montréal, Québec

September, 2014

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

Table of contents

Acknowledgements

First and foremost, myriad thanks are due to my super-advisor Robert Zatorre. Several years ago, Robert took a chance on someone without any background in human neuroscience, an already-fading memory of patch clamps/pipettes, and inkstains from the biotech newswriting I was soon to leave behind. Throughout the several years I have been his student, Robert has empowered me with a truly unique blend of advice and trust, something I think that everyone searches for in a mentor (and supervisor!). Thank you, Robert: it's been an honour to work for you!

Various members of the Z-Lab have also helped me along the way. Much technical know how (Linux commands, scripting, and the like) was provided by Patrick Bermudez and Nick Foster. You both have the rare ability to convey very technical knowledge without making the receiver feel dumb. Other helpful MRInsights have been provided along the way by Jamila Andoh, Sibylle Herholz, Niels Disbergen, Martha Shiell, and Boris Kleber. Early friendship and encouragement in the lab was provided by Marc Schoenwiesner and the three J's: Jean Zarate, J-K Kim, and Joyce Chen, and more recently by Melanie Segado, Emily Coffey, Kuwook Cha and postdoc power due of Patrice Voss and Krystyna Grabski. Krystyna I also special thanks to for translating/transducing this thesis into la belle langue de Français. The third study that comprises this thesis would never have been imaginable without Avrum Hollinger, designer extraordinaire of metal-free musical instruments!

Various others have given me technical guidance, both on how to be a McGill student and how to scan a brain! I would particularly like to thank Annie Le Bire and Giovanna Locascio for all of their administrative help at the MNI and in the psychology department, respectively. Marc Bouffard provided the initial MRI analysis code for the first study, as well as various technical advice over the course of my studies. From the McConnell Brain Imaging Centre, I would like to thank Michael Ferreira, Ilana Leppert, André Cormier, David Costa, Ron Lopez, and Louise Marcotte, for their help in sequence design and data collection. Thanks are also due to the developers of a few pieces of software that were extremely helpful in keeping me organized and focus, particularly as the word count began to rise: Papers2 (PDF library and reference management), Scrivener (for organizing and editing the actual manuscript), WorkFlowy (for outlining and brainstorming), and Freedom (for those times that my computer needed to be just a computer).

To Mary E. Sutherland: friend, colleague, harpist and (occasional) vocalist. Perhaps the years of training provided by your countless cousins is what has given you such an ability to make your friends feel like family? To Peter Finnie, Tim Wideman, Megan Bradley, Natasha Lekes, Philip Johnson, and Lara Pierce: friends and researchers who, though not in the field of music cognition, have always provided me with *sound* advice.

To the formerly shaggy mule-men, formerly of 20 Barnes Street in Providence, Rhode Island, who have continually made their presence felt across the border. To Winston G., a prodigal (or perhaps prodigious) Providencian, who has inspired in myself a reverence for the ways in which knowledge may span the page as well as the plage. And to Mimi and Dobby for taking us places and showing us things.

Many thanks to my family for their enduring support along the way. To my brother and sister who have been a warm presence from the other side of the Green Mountains. And to my father and mother the Biology and English teachers, respectively: this thesis contains much of both!

Last thanks go to Andrea Chen, the girl with many nicknames, all of them much deserved. Your support throughout the last few years has been immeasurable and I can't thank you enough for it. You help me to see the world in a different and better way.

Contribution of authors

I am the primary author of all three manuscripts presented in this dissertation. I was the primary designer of the studies (including study protocol and data analytics), and was responsible for participant recruitment/enrollment, data collection and analysis, and manuscript authorship. My Ph.D. supervisor, Dr. Robert Zatorre, provided support and feedback throughout the course of my studies, particularly with regard to brainstorming, vetting of experimental protocols, and editing of the manuscripts. Dr. Avrum Hollinger assisted with the third study by setting up hardware for the MRI-compatible piano keyboard and interfacing its control software with my Python scripts (which in turn ran the experimental protocol). All work presented in this thesis constitutes original scholarship and distinct contributions to the scientific literature.

The results of my thesis have been organized into 3 manuscripts (Chapters 2, 3, and 4). Chapter 2 is published in *Neuropsychologia*, Chapter 3 is published in *Cerebral Cortex*, and Chapter 4 is in preparation to be submitted.

Chapter 2: Klein ME, Zatorre RJ. 2011. A role for the right superior temporal sulcus in categorical perception of musical chords. Neuropsychologia. 49:878–887.

Chapter 3: Klein ME, Zatorre RJ. 2014. Representations of invariant musical categories are decodable by pattern analysis of locally distributed BOLD responses in superior temporal and intraparietal sulci. Cereb Cortex. doi:10.1093/cercor/bhu003

Chapter 4: Klein ME, Hollinger AD, Zatorre RJ. (in preparation). fMRI pattern analysis of played vs. perceived piano sequences in dorsal vs. ventral cortical streams.

List of figures

List of tables

Abstract

How does the brain effortlessly recognize, classify, and transform incoming sounds? My thesis aims to tackle this question via an exploration of musical stimuli with behavioral and neurobiological data drawn from functional magnetic resonance imaging (fMRI). The categorical "this-or-that" nature of musical interval sounds, contained in the long-term memories of musically-trained participants, allowed a unique opportunity to test top-down vs. bottom-up perception via audio-motor interactivity and a phenomenon known as categorical perception (CP). Speech CP has been linked to the left cerebral hemisphere's ventral (i.e. identification) stream of information processing, but dorsal regions comprising a perceptuomotor stream have also been implicated. Music, like speech, has strong and necessary links with the motor system and certain core musical sounds are perceived categorically. However, little is known about the neural correlates of such processes. In three studies, differential roles of the left/right hemispheres and ventral/dorsal streams were examined for various musical conditions. Study 1 employed harmonic chords and contrasted brain activity in multi-category conditions against matched control sounds, via both passive adaptation and active discrimination paradigms. Study 2 examined passive perception of melodic two-tone intervals. Instead of analyzing levels of brain *activation* between conditions, multivariate pattern analysis (MVPA, brain "decoding") was performed to directly dissociate between categorical percepts. Study 3 also employed MVPA, in combination with an MR-compatible piano keyboard used for active performance and feedback. This protocol allowed for the examination of audio-motor interactivity, including the neural correlates of movement and/or sound identity. The global results highlighted a bilateral network sensitive to these musical sounds, most notably the right superior temporal sulcus (STS) and left intraparietal sulcus (IPS), which were implicated in all three studies. The third experiment, explicitly aimed at testing audio-motor interactions, additionally highlighted the ventrolateral prefrontal/premotor cortex (VLPFC / PMv). The right STS, a non-primary ventral stream region,

is well positioned to carry out identification of categorical sound units, as such processes must rely upon "upstream" extraction of individuals pitches from complex spectrotemporal scenes. The IPS finding, meanwhile, has no close analogy in the speech literature; the observed recruitment may relate to the lack of 1-to-1 relationships between music perception and production, thus requiring a layer of transformation/recoding not present for speech. It follows that the IPS, long thought to underlie spatial perception or *visuo*-motor transformations, may be considerably more flexible in the processes it subserves: across modes (auditory as well as visual/motor) and types of transformations (pitch normalization; audio-motor recoding; mental visual rotation). The VLPFC / PMv, meanwhile, sits at the junction of the ventral stream, dorsal stream, and frontal planning circuitry, and may integrate various kinds of bottom-up information with top-down cognitive processes. Overall, the results resonate with the broader literature implicating intra-modal ventral processing streams for conscious identification of perceptual objects; dorsal pathways in sensory transformation (including abstraction into inter-modal representations); and posterior frontal cortex in perceptually-"informed" planning and behavior.

Résumé

Comment le cerveau est-t-il capable à reconnaitre, classer et transformer les sons entrants sans effort? Ma thèse a pour objectif d'aborder cette question à travers une exploration des stimuli musicaux avec des données comportementales et neurobiologiques tirées de l'imagerie par résonance magnétique fonctionnelle (IRMf). La nature catégorielle « ceci-ou-cela » des intervalles musicaux, contenus dans les mémoires à long terme des participants formés musicalement, a présenté une occasion unique pour tester la perception dans l'aspect ascendant vs descendant à travers l'interaction audiomotrice et un phénomène connu en tant que « perception catégorielle » (CP). La parole a été considéré reliée à la voie ventrale (pour l'identification) de l'hémisphère gauche, mais les régions dorsales comprenant une voie perceptuomotrice ont également été impliquées. La musique, comme la parole, entretient des liens étroits et nécessaires avec le système moteur et certains sons musicaux essentiels sont perçus de manière catégorielle. Cependant, peu est connu sur les corrélats neuraux de ces processus. Dans trois études, les rôles différents des hémisphères gauche/droite et des voies ventrale/dorsale ont été examinés en différentes conditions de musique. Dans l'Etude 1, en utilisant des accords harmoniques, nous avons contrasté l'activité cérébrale sous-jacente dans des conditions multi-catégorielles par rapport aux sons contrôles appariés. L'Etude 2 avait pour but d'examiner la perception passive des intervalles mélodiques à deux tons. Au lieu d'analyser les niveaux d'activations cérébrales entre conditions, une analyse multivariée des patterns d'activité (MVPA) a été effectuée afin de dissocier directement les percepts catégoriels. Dans l'Etude 3 nous avons également employé la méthode de MVPA, en combinaison avec un clavier de piano compatible-IRM utilisé pour des performances actives et les retours sensoriels sous-jacents. Ce protocole a permis l'examen des interactions audiomotrices, y compris les corrélats neuronaux des mouvements et/ou de l'identité sonore. Les résultats globaux ont mis en évidence un réseau bilatéral sensible à ces sons musicaux, notamment le sillon temporal supérieur (STS) et le sillon

intrapariétal gauche (IPS), qui ont été impliqués dans toutes les trois études. En outre, la troisième étude, ayant pour but explicitement d'investiguer les interactions audiomotrices, a mis en évidence l'importance du cortex ventrolatéral préfrontal/prémoteur (VLPFC / PMV) dans ce processus. Le STS droit, une région non-primaire de la voie ventrale, est bien positionnée pour mener à bien l'identification des unités sonores catégorielles, puisque ces processus doivent se reposer sur une extraction "en amont" des hauteurs individuelles depuis des scènes spectrotemporales complexes. Le résultat trouvé concernant l'IPS, quant à lui, n'a pas d'équivalent direct dans la littérature de la parole; son recrutement observé peut être lié à l'absence de relations de type un-à-un entre perception et production de la musique, ce qui nécessite un processus de transformation/recodage absent pour la parole. Il s'ensuit que l'IPS, longtemps considéré à sous-tendre la perception spatiale ou les transformations visuomotrices, pourrait être considérablement plus souple dans son rôle: à travers les modes (auditif ainsi que visuel/moteur) et les types de transformation (normalisation de la hauteur; recodage audio-moteur; rotation mentale visuelle). Enfin, le VLPFC / PMV se trouve à la jonction de la voie ventrale, la voie dorsale, et le réseau frontal de planification, et peut intégrer différents types d'informations ascendantes avec les processus cognitifs descendants. Dans l'ensemble, les résultats résonnent avec la littérature portant plus largement sur les mécanismes de traitements intra-modaux par la voie ventrale dans l'identification des objets perceptuels; la voie dorsale dans les transformations sensoriels (incluant l'abstraction en représentations intermodales); et le cortex frontal postérieur dans la planification et le comportement perceptuellement « informés ».

Chapter 1 - General introduction

1.1 Categorical perception

"*Perception thus differs from sensation by the consciousness of farther facts associated with the object of the sensation." —William James (1891)*

1.1.1 Perception and categorical perception

While there are many ways to define perception, this strong yet simple early definition from William James highlights consciousness and, more specifically, perception as the consciousness of *objects*. Whereas James saw pure sensation as "an abstraction never realized in adult life," perception, on the other hand, was "the consciousness of particular material things present to sense." Put differently, perceptions are by definition —organized— and, in effect, the only media through which we have access to sensory information.

At the heart of perception lies the interface between what are termed "bottom-up" and "top-down" processes. Perception as a primarily "data-driven" (i.e. bottom-up) phenomenon was argued for by Gibson (1950; 1966), who believed that visual stimulation, for example, was so information-rich, it need not be "interpreted" and instead could be subject to what Gibson termed "direct perception." This one-way theory has largely been disregarded in favor of two-way models of perception, in which bottom-up and experience-driven top-down processes interact heavily with one another. (An influential counter-argument to Gibson was made by Gregory (1970), although the roots of this two-way model can be traced back to at least Helmholtz (1867).) Whereas sensation concerns primarily passive receipt by the nervous system of physical or chemical signals (via the sensory organs responsible for transduction), perception, on the other

hand, is highly influenced by both the senses as well as pre-existing (i.e. top-down) factors, notably memory and attentional state.

Turning to the auditory system, a particularly vivid example of the interactions of bottom-up and top-down perceptual processes can be seen in what has become known as *Categorical Perception* (CP). Discovered via speech research and now recognized to be a global and multi-modal phenomenon, CP concerns the processes by which continuously-varying physical signals (which may consist of an infinite number of states) are perceived as members of one of a small number of discrete categories (Liberman, Harris, Hoffman, & Griffith, 1957). This phenomenon of "within-category compression and between-category separation in similarity space" causes "members of the same category to look more alike and members of different categories look more different" (Harnad, Hanson, & Lubin, 1995), and can be demonstrated behaviorally via a combination of identification and discrimination tasks. Such testing will, for categorically-perceived stimuli, show (a) labeling (identification) plateaus separated by sharp boundaries and (b) discrimination accuracy peaks straddling such boundaries, with concurrent troughs located near the center of identification plateaus (Liberman et al., 1957). The resulting phenomenon is such that physical continua "out there" are converted into sets of discrete meaningful units "in here," with the conversion happening so quickly and effortlessly that it feels as if none has taken place. Stimuli grouped into the same category appear similar, are easily identified as same-category members, and are relatively difficult to tell apart. Stimuli grouped in different categories appear un-alike, are easily identified as having membership in separate categories, and are relatively easy to differentiate. The representations of these complex physical signals have thus become dramatically simplified and, importantly, *meaningful*: James' "farther facts" from above. Early studies such as that by Goto (1971), which demonstrated that CP of speech phonemes is strongest in an individual's first language, indicated that CP's role is experience-driven, adaptive, and linked to meaningful inputs in one's environment.

Categorical perception, thus, is well positioned as a prime example of the distinction between perception and "pure" sensation. Whereas the nervous system can *sense* the infinite states of physical signals (or at least a down-sampled version thereof, according to the physiological limits of the sensory organs, e.g. number of photoreceptors or inner hair cells, etc.), perceptual experience is more limited than this. According to Liberman's initial (and "strong") definition of CP, the ability to discriminate between two sounds should be limited by the categories to which these sounds are assigned. Stated differently, two sounds should be easily discriminable if belonging to two separate categories and, alternatively, essentially non-discriminable (chance-level accuracy) if belonging to the same perceptual category. The latter conclusion implies that the conscious mind loses access to —or explicitly disregards— non/pre-categorical information (the "continuously varying physical signals" from above), leaving only the discrete categorical information to operate upon. (Separately, the strong form of CP relates to the famous "motor theory" of perception, also put forward by Liberman's group (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Motor theory states that the perception of speech sounds is limited by the knowledge of how to *make* such sounds: an idea I will return to later on when discussing the dorsal cortical stream of information processing.) Later findings, such as those by Studdert-Kennedy (1963) which demonstrated that reaction times varied with distance from a category center, suggested that categorical perception does not imply a *complete* loss of continuous information and engendered an updated "weak" form of the theory. The revised theory allowed for context-dependent access to non-categorical information, information which is both less robust and more subject to degradation/interference than the categorical variety (Pisoni, 1975).

Returning again to the idea of bottom-up vs. top-down processes, CP can be now seen to represent a particularly good example of the interaction of the two. Whereas pre-learned and discrete categories are clearly a demonstration of the influence of top-down processes on perception, the presence of pre-categorical/continuous information (which makes itself known via within-category discrimination and the reaction time effects mentioned above) suggests

bottom-up data-driven processes at work. In auditory psychology, this dichotomy came to be known as the "auditory" and "categorical" stages of information processing (Fujisaki & Kawashima, 1968; 1969). Whereas a categorical memory is robust and stable over time, the auditory stage, representing sensory memory, rapidly degrades and may be more easily subject to perturbations such as retroactive interference (Pisoni, 1975). Thus, it seems that the role of "raw" sensory processes is multifaceted, as they trigger category percepts while also seeming to "fill out" the perceptual experience. Differential conscious access to high- vs. low-level information is well addressed by reverse hierarchical theory (RHT) (Ahissar, Nahum, Nelken, & Hochstein, 2008; Hochstein & Ahissar, 2002), which proposes that initial/fast perception is mediated entirely by categorical memory ("forest before trees"). RHT details a mechanism by which implicit low-level information may be consciously accessed, a process requiring attentional processes and a reverse hierarchical (i.e. *top-down*) search.

Interaction between various processing stages is a topic I will return to when covering information processing streams in cortex (*section 1.2*). For now I would like to note that CP relies critically upon both bottom-up sensory and top-down (long-term memory) processes and represents a vivid example of their interaction. RHT, meanwhile, augments Atkinson and Shiffrin's classic multimodal model for memory (Atkinson & Shiffrin, 1968) to include a dynamic two-way interplay between sensory and post-sensory processes.

1.1.2 Categorical perception in speech and non-speech

As mentioned previously, categorical perception was discovered by speech researchers and initially believed to be a speech-only phenomenon (Liberman et al., 1957; Mattingly, Liberman, Syrdal, & Halwes, 1971). However, an early complication to the story arose when it was demonstrated that clear CP effects, which could reliably be observed using stop consonant stimuli (such as phonemes like /da/ and /ta/ that differ primarily according to voice onset time

(VOT)), were not seen very clearly with vowels (Fry, Abramson, Eimas, & Liberman, 1962; Mattingly et al., 1971). Similarly, robust CP was not elicited using lexical tone contrasts (Abramson, 1977). Pisoni (1975) and others argued that rapid consonant-vowel transitions and/or release bursts in stop consonants are what drives the CP effects, in opposition to the steady formant structures found in vowels; a possible underlying explanation being that consonant structures are more constrained by the general physiology of aspiration (and thus less sensitive to large amounts of multi-dimensional variability, as compared to vowels).

This last point harkens back to Liberman's motor theory of perception (Liberman et al., 1967), as the categorical nature of certain percepts was theorized as being driven by the perceiver's ability to *produce* certain sounds. In other words, researchers like Fry et al. (1962) postulated that the amount of "articulatory discontinuity" between sounds —high for stop consonants, but low for vowels— was correlated (and perhaps causal) with the sharpness of their perceptual CP functions. This line of reasoning ties into a larger point about early CP research: it was framed more-or-less exclusively within the speech system and was thought to be a special property of spoken language. However, a major blow was dealt to both this "speech is special" position as well as the strong form of motor theory of categorical perception when it was discovered that chinchillas showed behavioral functions essentially indistinguishable from humans when responding to speech consonants (Kuhl, 1978), given that chinchillas don't speak. Non-verbal human infants were found to exhibit a similar effect (Eimas, Siqueland, Jusczyk, & Vigorito, 1971), raising additional problems. More recently, however, it has been clearly shown that speech-relevant perturbations to the somatosensory (Ito, Tiede, & Ostry, 2009) and motor (Nasir & Ostry, 2009) systems can alter the perceptual boundaries (and thus classification) of speech sounds. Thus, there is certainly a role played by somatomotor processes in auditory speech perception (i.e. support for a "weak" version of motor theory), a topic to which I will return in the neuroscience section later in this chapter.

While it is not my intent to argue against the major role that CP surely plays in the perception and understanding of speech, the above literature indicates that CP is a more complex and global phenomenon than the early speech literature would suggest. On the one hand, the chinchilla and infant studies seem to suggest that CP is *not* driven by learned cognitive categories (and instead a function of low-level psychoacoustical phenomena). On the other hand, non-speech auditory research (including studies of music (Locke & Kellar, 1973; Zatorre & Halpern, 1979), which will be expanded upon next), as well as the learning studies mentioned above, show that learning and experience *can* play a definitive role in the observation of CP.

As mentioned, one need not look outside the human auditory system to observe a particularly rich source of *non*-speech categorical perception: music. Similar to speech, where stop consonants may vary along a single dimensions such as voice onset time, musical intervals (such as minor thirds, perfect fifths, etc.) also vary uni-dimensionally and are defined according to the ratio of frequencies between the high and low tones of the interval. And similar to recognition of speech sounds, trained musicians can quickly and effortlessly identify a musical interval as belonging to one of a relatively small number of learned categories. CP of musical intervals by trained musicians was first observed by Locke & Kellar (1973) using simultaneous (i.e. harmonic) 3-note chords, and later by Siegel & Siegel (1977) using sequential (i.e. melodic) 2-note intervals. Neither of these studies found categorical effects in *non*-musicians, indicating that CP was strongly experience-dependent. Burns & Ward (1978) showed that (a) CP in musical intervals could be observed even when the tones were roved in absolute pitch space and (b) discrimination functions followed very closely what would be predicted from identification data. The former effect demonstrated that participants had to have been using abstracted *interval* information in order to make their judgements, as opposed to cueing in on particular pitches of single tones. Zatorre & Halpern (1979) extended Burns & Ward's work to harmonic intervals, while Zatorre (1983) showed that discrimination accuracy could be negatively impacted by selectively interfering with auditory sensory memory. Those later data provide support for dual

stage auditory vs. categorical memory processing, as the manipulation did not affect the overall shape of the categorical discrimination function (i.e. within- and across-category comparisons were equally affected), echoing earlier speech CP research such as that by Pisoni (1975).

In addition to music, there have been multiple demonstrations of CP in the auditory system using non-speech stimuli, including via the use of noise-buzz sequences (J. D. Miller, 1976) and rhythmic units (Raz, 1977). Categorical perception has also been demonstrated as a robust phenomenon in non-auditory modalities, particularly in vision. Beale & Keil (1995) and Kikutani, Roberson, & Hanley (2010), among others, have shown categorical processing of human face stimuli and Bornstein & Korda (1984) demonstrated a strong CP effect for perception of color/hue. Interestingly, Gilbert, Regier, Kay, & Ivry (2006) later showed that this categorical effect for color was only evident for stimuli presented in the right visual field, and thus more directly available to the language-dominant left hemisphere. This result, combined with those of Winawer et al. (2007) that showed that CP of color was mediated by linguistic boundaries specific to particular languages, relates to what is known as the "principle of linguistic relativity" (also known as the Sapir-Whorf (or simply "Whorf") hypothesis)(Whorf, 1956). (While often over-simplified to "language limits perception," the Whorf hypothesis states more generally that over-learned categories or concepts influence perception, which is clearly relevant to CP.)

Taken together, the above examples from speech, auditory non-speech, and non-auditory processes provide evidence that supports categorical perception as (a) a general phenomenon that is (b) an experience-driven mediator of (c) top-down perceptual processing.

1.2 The cortex and processing streams

Before specifically addressing categorical perception from a neuroscientific standpoint *(Section 1.3)*, I shall first examine the neuroscience of perception more generally, including its bottom-up and top-down components. The cortical processing of perceptual category information can then be placed within its proper context.

1.2.1 The modular cortex

The idea that the brain is modular —that different brain areas perform different functions— can infamously be traced back to Franz Joseph Gall and the phrenologists of the 1800s (Gall & Spurzheim, 1809). Phrenology, which ascribes mental traits based on the shape of the skull, has no predictive power and has long been relegated to a pseudoscience (Simpson, 2005). However, phrenology was rooted in accurate concepts: that the brain is the "organ of the mind" and that separate mental faculties exist in separate brain areas that are in communication with one another. The first true scientific link between a specific brain structure and its function came from Broca (1861), who described a patient who lost a specific ability (to produce fluent speech), which was linked to damage in a particular brain region (the left inferior frontal lobe). The lesion study approach, linking specific psychological deficits to brain damage, has now been used effectively for well over a century. Separately, Brodmann (1909) showed that the human cortex could be defined, anatomically, into tens of unique regions (Brodmann found approximately 50, while modern brain maps sub-divide many of these regions). Building on lesion studies, a striking example of function correlating predictably with anatomy was provided by Penfield (Penfield & Boldrey, 1937), whose neurosurgery experiments showed that direct current stimulation of the pre- and post-central gyri could elicit motor and somatosensory effects, respectively, with a topographic representation of the body. In short, there is a foundational and

time-tested literature demonstrating that the human cortex is sub-divided into many discrete functional/anatomical regions. What exactly each region contributes and how these regions interact are of course ongoing questions in cognitive neuroscience.

A full understanding of the nature of information in the human brain —how it is organized and how it is processed and transformed— is one of the overarching goals of neuroscience. Arguably the dominant (and most salient) source of such information originates from our various sensory systems, which carry enormous quantities of data about both the external world and the body's internal state. However, despite the large amount of data streaming into the CNS (e.g. ~1 million fibers per optic nerve, tens of thousands per vestibulocochlear nerve), this information is subsequently *multiplied* upon entering the cortex via extensive axonal branching, suggesting the brain relies heavily upon *parallel* processing (Van Essen & Maunsell, 1983). A result is cortical primary sensory areas (A1, V1) that have (a) many more neurons than their peripheral analogs (the spiral ganglia, the retinas' ganglion cell layers) and which (b) already display certain kinds of specialization (e.g. sound frequency tuning (Merzenich & Brugge, 1973), visual orientation selectivity (Hubel & Wiesel, 1962)). The large space and energy requirements of these cells suggest that the nervous system has evolved to place a very high priority on the ability to perform differential parallel processing on multiple "early" copies of sensory information.

It is also important to quickly note that there is an enormous amount of processing performed on incoming sensory information *before* it has reached the cerebral cortex (e.g. in the brainstem, thalamus, etc., which also receive cortical feedback via efferent connections). However, these regions and processes are generally outside the scope of this thesis and, with regard to *categorical* speech perception, it has been demonstrated that evoked patterns of late cortical activity contain categorical information, whereas brainstem responses generally reflect acoustic properties of the speech waveform (Bidelman, Moreno, & Alain, 2013). That being said, it is worth pointing out early on that even "early" primary sensory regions of the cerebral cortex are working with highly modified representations of the external world.

1.2.2 The visual ventral processing stream

We now turn our attention to what happens to information after it reaches sensory cortex, framing the discussion in terms of *processing streams*. A processing stream is a multi-stop pathway in which information is repeatedly received, processed, transformed, and sent along to the next stage in a hierarchy — which is to say *conceptually serial* in nature. Streams are not unique to cortex, as shown by the landmark visual processing study by Hubel (1959) which demonstrated that, while photoreceptors respond optimally to light (without regard for its structure), thalamic neurons have a center-surround organization, and those of V1 prefer *bars* of light, indicating that structure-generating processes take place in the retina and thalamus. Hubel and Wiesel's hierarchical model postulated that such seemingly non-intuitive receptive fields could be constructed by summing the responses of earlier neurons in the stream. And successive processing stages need not live in separate brain regions as Hubel & Wiesel (1962) also showed, in their study of what came to be known as "simple" and "complex" cells, both contained within area V1. However, as was later discovered, higher-level visual processing requires the contribution of successive processing stages outside of V1 (as detailed in a review by Zeki (1978)).

As the initial concept of a ventral stream emerged in the visual system, I will start the discussion there before moving to the auditory system. As described by Zeki (1978), the striate cortex (V1) was originally referred to as "visuo-sensory" cortex, whereas prestriate areas (i.e. Brodmann Areas 18 and 19, etc.) were known as the "visuo-psychic" band. To Zeki, this "terminology implied that 'sensation' occurred at the level of the striate cortex and 'perception' at that of the pre-striate cortex." Such suggestions were proven false by later research, an example being blindsight, in which V1 damage leads to loss of conscious perception, but direct LGN-extrastriate connections mediate certain behavioral responses to visual inputs (Schmid et al., 2010). Multiple primate studies (Baker, Blumstein, & Goodglass, 1981; Zeki, 1971; 1976)

(review by Van Essen & Maunsell (1983)) subsequently showed that so-called "pre-striate cortex" was actually several different functional regions, including what became to be known as areas V2, V3, V3A, V4, and V5. The latter two of which, areas V4 and V5 (V5 also known as area "MT"), were found to be differentially implicated in the processing of color and motion, respectively. This finding was important, as it showed an obvious "division of labor" between different portions of the extra-striate processing pathways. These separate areas receive and operate on different sorts of information transmitted from earlier nodes in the stream; for example, area V4 does not strongly represent the peripheral visual fields whereas area V5 does, in line with sensory apparatus physiology and evolutionary needs (i.e. that color-sensitive cones are primary located more centrally in the retina, and that it is crucial to perceive motion in the visual periphery, respectively).

Observations of these extra-striate dissociations are the historical beginnings of what has become known as the *ventral and dorsal streams* of information processing. As noted above, processing in the brain is hierarchical (serial) but also parallel. The "two streams hypothesis" (Mishkin, Ungerleider, & Macko, 1983), which will be discussed in this section, is a major example of both serial and parallel processes happening on a massive scale. While a serial processing model's "this then that" structure is conceptually simple (at least in its most basic form), parallel processing's "divide and conquer" approach is anything but. For example, one major problem that needs to be overcome is that of *binding* (Baars, 1997): how the brain links together disparate information about the "same thing." One approach that can be taken to help analyze such parallelism is to view the cortex in terms of its goals. One such goal is *recognition*: What is the identity of a sensed object? (Is it familiar? Threatening?) These processes appear to be tied to ventrally directed processing streams, which travel into the inferior (visual) and lateral (auditory) temporal cortices, as well as IPL/posterior insula (somatosensation). I will begin by discussing this ventral stream of processing (sometimes called the "what" stream) before taking

up a second, and more interactive, major processing goal: the manipulation and transformation of the sensory information (Goodale & Milner, 1992) (served by the dorsal stream, *section 1.2.5)*.

The bulk of research into the visual ventral stream has focused on shape and pattern perception and most of this research, in both humans and non-human primates, has centered on the inferior temporal (IT) cortex. Studies from the early 1980s had shown that, for example, individual neurons in the monkey IT cortex were selectively responsive to faces and hands (C. G. Gross, Rocha-Miranda, & Bender, 1972; Perrett, Rolls, & Caan, 1982). However, perhaps the best example of a true ventral *stream* of processing was provided by Tanaka, Saito, Fukada, & Moriya, (1991), who showed dramatic differences in the responses of neurons in the posterior vs. anterior portions of IT cortex (i.e. more upstream vs. downstream regions). Whereas most neurons in posterior IT cortex were maximally activated by slits or spots of light, the majority of anterior IT cells were classified as "elaborate," which the authors defined as "maximally activated by some pattern feature more complex than oriented contours, colored blocks, or simple textures" and did not respond to slits or spots. Stimuli that effectively drove elaborate cells included star shapes, triangles with vertical light/dark stripes, and circles with several slim bars emerging from them (the latter figure somewhat resembling a hand). Zeki (1978) and others (C. G. Gross & De Schonen, 1992), meanwhile, have noted that as one moves downstream (i.e. ventrally), single cells have larger and larger receptive fields, indicating that they are no longer processing a small slice of visual space and instead are driven by particular visual features anywhere in a large region of space.

The single-cell findings highlighted above are complemented by human brain lesion data relating to the syndrome of visual agnosia, an impairment of recognition not due to basic visual deficits (Farah, 2004). Visual agnosia comes in two primary varieties, apperceptive and associative, according to whether the deficit is in recognition of "pure" visual form (apperceptive) or of linking together that form with its meaning (associative). (Associative visual agnosia includes its most famous subtype, prosopagnosia, which is a deficit in the recognition of

faces.) In accordance with this discussion of the hierarchically-organized ventral stream, associative agnosias, thought to be the more complex variety, are often found with IT cortical lesions more ventral and anterior to those of apperceptive agnosias, which are generally found in or near extrastriate Brodmann areas 18 and 19 (Heilman & Edward Valenstein, 2011). Neuroimaging findings by Nancy Kanwisher and colleagues have highlighted distinct IT brain areas specifically responsive to the presentation of faces (Kanwisher, McDermott, & Chun, 1997) or places (Epstein & Kanwisher, 1998). One major target of the visual (as well as auditory and somatosensory) ventral stream is the hippocampus (and its related medial temporal lobe structures). These structures may serve a dual role of (1) "binding" together perceptions from the various sensory modalities into a uniform whole and (2) fixing perceptions in long-term memory (Eichenbaum, Sauvage, Fortin, Komorowski, & Lipton, 2012). Separately, information is passed to the inferior frontal gyrus (IFG)/ ventrolateral prefrontal cortex (VLPFC) (O Scalaidhe, 1997; Takahashi, Ohki, & Kim, 2013) via the extreme capsule (Petrides & Pandya, 2009), where it is further processed and integrated (Courtney, 1998; S. C. Rao, 1997). This frontal region is also a target of the visual dorsal stream (Takahashi et al., 2013) and, as I will discuss later in this section, the IFG/VLPFC/ventral premotor region is also a convergence zone for the *auditory* dorsal and ventral streams.

1.2.3 The auditory ventral stream

Compared to the visual system, significantly less is known about the auditory ventral stream. Analogous to early visual processing regions (striate and certain pre-striate areas), the auditory system, at least in non-human primates, has a "core" region, itself comprising a few separate functional subregions: A1, R, and RT (Kikuchi, Horwitz, & Mishkin, 2010). These regions receive direct input from the auditory thalamus (the ventral division of the medial geniculate nucleus (MGN)) and have a highly developed cortical layer 4 (Rauschecker & Tian, 2000), so

can be considered the most analogous auditory region to the striate visual cortex. In monkeys, additional non-primary regions have been identified in what has become known as the "belt" and "parabelt" areas of auditory cortex (Rauschecker & Tian, 2000). Belt regions are highly interconnected with the core regions, while also receiving direct projections from non-ventral divisions of the MGN (Hackett, 2007). An overview of the monkey core/belt/parabelt auditory cortex is provided in *Figure 1.1*, from Bendor & Wang (2008). Projections from non-core auditory regions, in turn, have been found to target the monkey homologues of Broca's area (BA44 and BA45) in ventrolateral prefrontal cortex (VLPFC) via the extreme capsule (Petrides & Pandya, 2009). This frontal region is also a target of dorsal stream projections, discussed later in this section, via the superior longitudinal fasciculus (SLF) (Petrides & Pandya, 2009), thus forming a ventral/dorsal convergence zone. Separately, auditory belt/parabelt regions target medial temporal lobe (MTL) memory structures, such as the entorhinal cortex (Munoz-Lopez, Mohedano-Moriano, & Insausti, 2010).

Generally, belt (and surrounding parabelt) regions contain neurons that best respond to stimuli that are more complex than those which cause best responses within the core region. Whereas simple tones of a single frequency are highly effective for driving core neuronal responses (Merzenich & Brugge, 1973) (with core sub-regions all tonotopically organized along best frequency axes), belt/parabelt neurons require the use of stimuli that are more complex in either their spectral or temporal characteristics (Rauschecker, Tian, & Hauser, 1995). Rauschecker & Tian (2000) make convincing visual → auditory analogies for: dots of light → pure tones; bars of light → band-passed noise (BPN, in which noise from only a select frequency range is presented); and moving light stimuli → frequency-modulated (FM) sweeps.) Both BPN (Rauschecker et al., 1995) and FM sweeps (Rauschecker, 1998) are effective in stimulating lateral belt neurons.

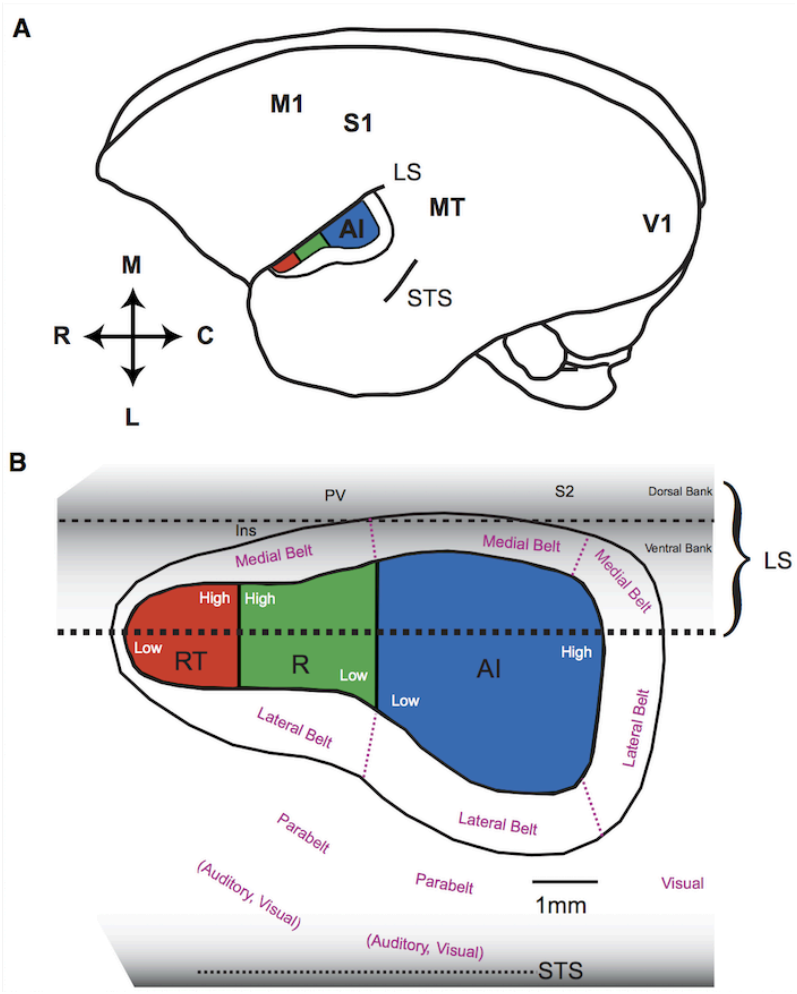

Figure 1.1 Auditory cortex in the monkey

From Bendor & Wang (2008), titled "Model of the organization of auditory cortex in marmosets." This figure clearly depicts the three "core" sub-regions of the primate auditory cortex, in addition to various surrounding belt and parabelt areas. Abbreviations: LS, lateral sulcus; S2, secondary somatosensory area; PV, parietal ventral area; Ins, insula; AI, primary auditory cortex; R, rostral field; RT, rostral temporal field; STS, superior temporal sulcus; M, medial; R, rostral; C, caudal; L, lateral; V1, primary visual cortex; M1, primary motor cortex; S1, primary somatosensory cortex; MT, middle temporal area.

Complex natural sounds such as monkey vocalizations were also found to be very effective in driving single neurons in the lateral belt (Rauschecker, 1998), and the somewhat non-selective responses of these cells to several different calls implies that the lateral belt areas are situated "midstream." The same study demonstrated intriguing non-linear spectral or temporal combinations of sound needed to drive certain neurons, including a neuron whose response required the simultaneous presence of both high- and low-pass-filtered versions of a vocalization, and a separate neuron that reliably fired only after stimulation with *both* syllables of a bisyllabic call. These results provide evidence that these ventrally positioned neurons must be integrating information from more upstream processes. Rauschecker and colleagues proposed a stream that originates in the core and passes information to the middle lateral belt area (ML). The ML in turn sends outputs to the anterolateral (AL) and caudolateral (CL) belt areas, which are most selective for monkey calls. A separate line of research suggests that, in addition to these (ventro-)lateral processes, there is a more *rostral* component of the ventral stream that is also highly-involved in stimulus identification (Kikuchi et al., 2010). Bendor & Wang (2008) suggest that these two divisions may divide up analysis into spectral vs. temporal processing, respectively.

As mentioned above, higher-level auditory areas project to the VLPFC. This region has been shown to exhibit sensitivity to learned categories, particular when the categorical distinctions are task-relevant. Neuronal activity in the monkey VLPFC has been shown to correlate with behavioral choices (Russ, Orr, & Cohen, 2008) more-so than to auditory perceptual features (J. H. Lee, 2009). Relatedly, Gifford, MacLean, Hauser, & Cohen (2005) and Cohen, Hauser, & Russ (2006) found VLPFC neurons that discriminated between food calls associated with distinct functional classes, but not between calls conveying the same functional information (despite the fact that such calls were perceptually discriminable). Tsunada, Lee, & Cohen (2011) showed that the response properties of VLPFC neurons were modulated by behavioral choices, but those of superior temporal neurons were not. Fritz, David, Radtke-Schuller, Yin, & Shamma (2010)

provided evidence for the establishment of dynamic functional connections between frontal and auditory areas during listening tasks, supporting two-way directionality (i.e. feedback connections) in the ventral stream. Thus, it seems that the VLPFC component of the ventral stream, rather than subserving perceptual extraction of relevant features, is involved in perceptually relevant goal-directed actions. As this region is also a dorsal stream target (and thus receives information with a wide variety of properties), it is well positioned to form "unified" representations of perceptual information, perhaps a prerequisite for such goal-directed behaviors (Griffiths, Warren, Scott, Nelken, & King, 2004).

The story is more clouded in humans. Recent research has demonstrated tonotopically-organized human analogs of primate core areas A1 and R on or around Heschl's Gyrus (HG, see review by Baumann, Petkov, & Griffiths (2013)), as well as possibly region RT (Moerel, De Martino, & Formisano, 2012). Functional specifics of the human belt and parabelt region analogs have been much harder to discern, with most researchers instead referencing superior temporal areas based on anatomy: either cytoarchitectonic (i.e. by Brodmann area (BA)) or macroscopic (e.g. the planum temporale (PT) and planum polare (PP), on the superior surface of the superior temporal gyrus (STG) posterior and anterior to HG, respectively). While there is a large amount of controversy on the topic (Baumann et al., 2013), areas of the PT and PP which are adjacent to HG/BA41, including BA42 and BA52, can be considered auditory "belt" regions (Baumann et al., 2013). *Figure 1.2*, from Baumann et al. (2013) and adapted from Hackett (2007), provides a nice overview of various definitions of core/belt/parabelt in the human brain. As with non-human primates, human auditory association areas project to the posterior frontal cortex, via the arcuate and the extreme capsule fasciculi (A. S. Dick & Tremblay, 2012; Makris & Pandya, 2008), with the latter best-positioned to conduct true "ventral" projections from anterolateral temporal to ventrolateral frontal areas. An overview of the human auditory ventral (and dorsal) processing streams is provided in *Figure 1.3* via two popular and related models from Hickok & Poeppel (2007) and Rauschecker & Scott (2009).

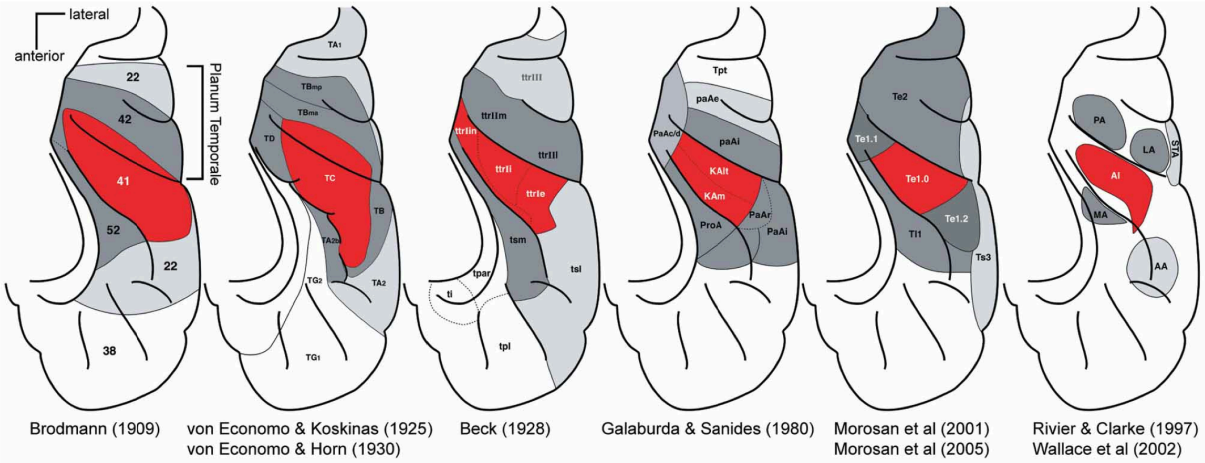

Figure 1.2 Human auditory cortical regions

From Baumann et al. (2013), titled "Parcelations of the human superior temporal cortex by different investigators." This figure was adapted by Baumann et al. from Hackett (2007). There are various ways to define the sub-regions of auditory cortex, going back to (Brodmann, 1909), including via cytoarchitectonics, gross anatomical features, or functional mapping. While these different schematics have clear differences, there is a global correspondence between core regions along much of Heschl's Gyrus (HG), belt areas adjacent to the core, and parabelt regions located further posterior, anterior, and lateral.

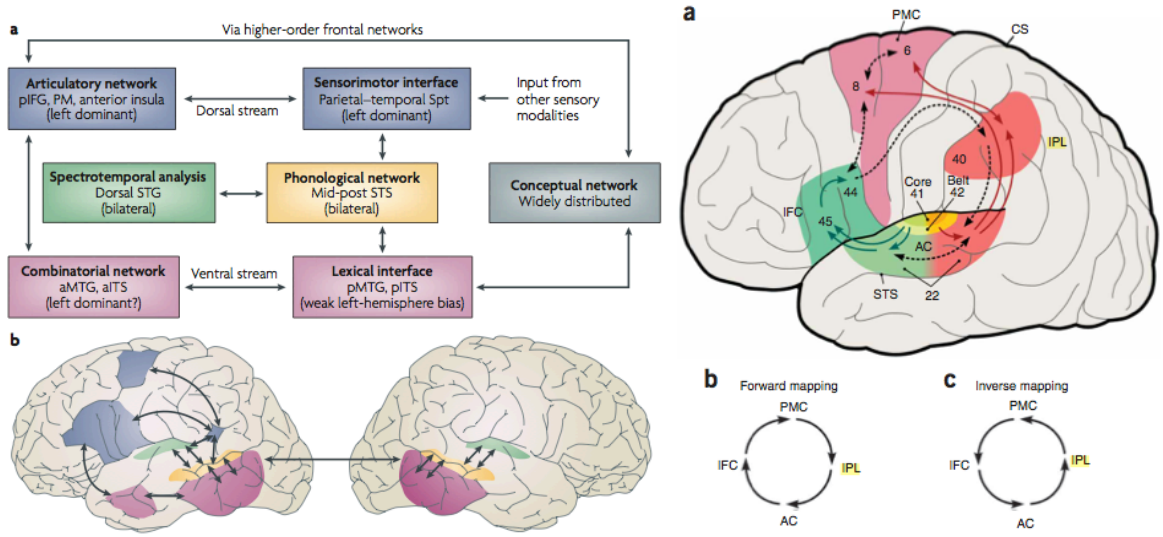

Figure 1.3 Auditory ventral and dorsal streams

Left, from Hickok & Poeppel (2007), titled "The dual-stream model of the functional anatomy of language." Right, from Rauschecker & Scott (2009), titled "Dual auditory processing scheme of the human brain and the role of internal models in sensory systems." These two influential models of dual-stream processing for sound have many regions in common, including the superior temporal sulcus (STS) and posterior prefrontal and premotor regions. Hickok et al.'s model is more bilaterally distributed, while both models highlight frontal sites as critical convergence zones for the dorsal and ventral streams.

Complex sounds that need to be successfully identified by humans (and which would theoretically require advanced processing within the auditory ventral stream) include those from speech, music, and the natural environment. These topics, particularly speech and music (and their relative processing within the two cerebral hemispheres) will be discussed at length in the next section (1.3). Here, I would like to briefly address some functions in which the auditory ventral stream has been implicated. These functions, critically, all have to do with identification of sound objects and properties.

Extending the results from non-human primates to humans via neuroimaging, pure tones have been found to robustly activate primary auditory cortex/HG (Da Costa et al., 2011). BPNs activate a wider network than pure tones, notably the lateral belt (Rauschecker, 1998), and speech sounds activate an even wider network than do BPNs, extending into lateral parabelt regions (i.e. the STS) (Binder et al., 2000). In a study that examined pitch and melody (R. D. Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002), compared to pitch-free sounds, pitched sounds more strongly activated only the lateral portions of HG. Stimuli where the pitched varied (i.e. to produce melodies) have been found to activate the STG, planum polare (PP), and anterior planum temporale (PT) (R. D. Patterson et al., 2002; Warren & Griffiths, 2003). The study by Warren & Griffiths showed a clear double dissociation between these ventral stream results and a site in the posterior PT that was linked to spatial locations of stimuli. This latter finding is in accord with the dorsal stream, discussed later in this section.

Turning to even more complex sound stimuli, Belin, Zatorre, Lafaille, Ahad, & Pike (2000) observed voice selective regions bilaterally in the superior bank of the anterior STS (with a right hemispheric bias), while Formisano, De Martino, Bonte, & Goebel (2008) found that the majority of informative voxels allowing for discrimination of speaker identity were located in the right STS. These voice-sensitive regions may be analogous to the single-cell monkey vocalization data (in fact, voice-sensitive areas in the monkey auditory ventral stream have been confirmed via fMRI (Petkov et al., 2008)) and, separately, to visual face-processing regions

found in the fusiform gyrus. Lewis, Brefczynski, Phinney, Janik, & DeYoe (2005) found a region of the middle STG that preferentially processes animate over inanimate complex environmental sounds. Sub-lexical (i.e. sound, but not meaning) studies of speech, such as those using phonemes, have shown large zones of activity bilaterally in the STG and STS, with somewhat debated biases toward the left hemisphere and the STS (see review by Hickok & Poeppel (2007)). This same review supports the view that even more ventrally-located temporal regions underlie the lexical/semantic interface. Okada et al. (2010) showed that core auditory regions were most sensitive to acoustic features of stimuli, whereas STS regions were sensitive to the *intelligibility* of speech (which was originally demonstrated by Scott, Blank, Rosen, & Wise (2000)) and not particularly sensitive to acoustic variation.

VLPFC regions such as Broca's Area have also been implicated in the processing of speech perception (Y. S. Lee, Turkeltaub, Granger, & Raizada, 2012; Zatorre, Evans, Meyer, & Gjedde, 1992; Zatorre, Meyer, Gjedde, & Evans, 1996). This region, as stated above, is targeted by both the ventral and dorsal streams (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009). The non-human primate literature, introduced above, has implicated the VLPFC in goal-directed behavior that relies upon abstracted perceptual information, rather than in the "pure" *perceptual* processes. This concept, generally, fits with the literature on human agnosias (higher-level perceptual deficits), which are almost universally linked to posterior (i.e. non-frontal) lesion sites (see review by Bauer (2006)). To summarize this section, there is compelling evidence, both in humans and non-human primates, for an auditory ventral stream of processing which originates in A1, proceeds through non-core areas in the STG, STS and middle-lateral temporal lobes, and eventually targets MTL memory structures and the VLPFC, the latter implicated in perceptually-mediated goals and behaviors. However, there is still considerable debate over the nature of hierarchical ventral stream processing, as well as how information processed in the ventral and

dorsal streams becomes integrated: issues that will be addressed in the upcoming two sections (1.2.4 and 1.2.5, respectively).

1.2.4 Debate over representation

Before moving on to the dorsal stream, I would like to briefly discuss an ongoing scientific debate concerning the roles and limits of cortical processing hierarchies in perception. There are multiple aspects to this debate. One argument concerns *how abstract* the representations of *single neurons* become. A second and related debate concerns the duties of "early" cortical regions (such as A1 and V1) in the representation of complex/abstract perceptual categories.

The first debate is the less controversial of the two and builds off the assumption that, after receiving information from earlier nodes, "higher" cortical areas (i.e. those that are more downstream) are responsible for the recognition of (and differentiation between) various percepts. Inherent to the hierarchical processing model is the idea that, pyramid-like, single units higher-up in the processing chain represent a greater amount of information in a more abstract manner, as compared to lower-level units. Using Hubel and Wiesel's visual model as an example, it would take several LGN cells with spatially aligned receptive fields to represent a bar of light. However, if these cells all converge upon one V1 cell, this *single* cortical cell may now represent the presence of this bar of light. Likewise, two perpendicularly oriented "bar" neurons may converge on a new cell, which codes for the presence of an X-like shape. This is "sparse coding," in which increasingly complex objects are coded for by smaller and smaller numbers of neurons. The present debate concerns exactly *how sparse* things get quite far down the ventral stream. One extreme position is embodied by what Jerome Lettvin famously defined as the "grandmother cell" (C. G. Gross, 2002): "a neuron that would respond only to a specific, complex, and meaningful stimulus, that is, to a single percept or even a single concept." Neurons with responses very much like "grandmother" have been found in human IT cortex (Quiroga, Reddy, Kreiman, Koch, & Fried, 2005), although Quiroga's group, themselves, argue against the

extreme "grandmother" interpretation of these data (Quiroga, Kreiman, Koch, & Fried, 2008). According to a more generally-accepted model (originally put forward by Hebb via his "cell assembly" (Hebb, 1949)), perception relies upon the pooled responses of fairly large numbers of neurons. This view, to which I subscribe, can still be thought of as quite hierarchical and fairly sparse, but does not require a processing stream to converge upon and activate one single, critical neuron.

The second debate, meanwhile, concerns the very existence (or, at least, the importance) of hierarchical processing. The basic hierarchical model, described above, allows for myriad feedforward *and* feedback connections between modules, but generally supports the idea that complexity of processing increases as information is passed further downstream (such assumptions underlie many ventral stream models, including that described by Hickok & Poeppel (2007) for speech processing). The hierarchical processing model, to a large degree, forms the basis for explanations that link specific perceptual deficits with relatively focal brain lesions (e.g. prosopagnosia with the fusiform gyrus), as well as for the wealth of fMRI literature linking well-controlled and complex stimuli presentation to BOLD activation in non-core sensory areas. Counter to the classic hierarchical model is what I will refer to as the "flat" model. This view postulates that large multi-region areas of cortex, which notably include "early" sensory regions, participate in complex perceptual processing. Here, instead of being restricted to downstream regions, ensembles of active neurons spanning *early (e.g. HG) and late (e.g. STS)* sensory areas are required to represent object identity. Perhaps the most compelling auditory study arguing for the flat model was performed by Kilian-Hütten, Valente, Vroomen, & Formisano (2011), who showed that physically-identical speech sounds that were perceived differently (following short-term priming procedures) could be decoded using voxels primarily from the left STG, and not the STS. Notably, however, only a minority of these voxels fell with HG, itself, with the remainder being found posterior and lateral to HG and within the superior temporal plane (i.e. in belt and, perhaps, parabelt regions). Separately, due to its inherently poor

temporal resolution, fMRI data supporting a more distributed model cannot easily differentiate between true "flat" processing vs. "echoes" in early sensory cortex provided by hierarchical models' feedback loops. A potential middle ground between the hierarchical and flat positions might be provided by reverse hierarchy theory (RHT), introduced in the previous section (1.1). RHT proposes that different neural processes underlie immediate vs. "scrutin[ized]" perception (Hochstein & Ahissar, 2002; Nahum, Nelken, & Ahissar, 2008), potentially bridging the gap between auditory and categorical processing. The theory that different neural substrates serve immediate vs. non-immediate perception will come up again in the next section, which concerns the dorsal stream.

1.2.5 The dorsal stream

Compared to the ventral stream, the dorsal stream is both less well defined and more controversial. This second stream was originally put forward as the "where" stream for visual object localization (Mishkin & Ungerleider, 1982). Mishkin and colleagues reported a double dissociation in monkeys (Mishkin et al., 1983), in which ablation of anterior IT regions resulted in an inability to discriminate visual objects based on shape (but not location), whereas posterior parietal lesions led to an inability to discriminate based on object location (but not shape). Several years later, a radically altered model of the dorsal stream was put forward by Goodale & Milner (1992), who argued that, instead of merely subserving identification of objection location, the visual dorsal stream was instead about *actions* and their requisite sensorimotor transformations. Considerable evidence has accumulated in support of this latter view (see reviews by McIntosh & Schenk (2009; Milner & Goodale (2008)), predominantly in the visual domain and focused on movements such as grasping and reaching (Cavina-Pratesi et al., 2010). The dorsal stream also encompasses the contentious "mirror system" (Rizzolatti & Craighero, 2004), where *observation* of action elicits responses in premotor and posterior parietal (Goodale,

2005) regions. Milner and Goodale's revised model explicitly names the ventral stream as the mediator of "conscious experience" as well as that of "'unconscious' or 'preconscious' perception of objects and events, which refers to mental representations that potentially *could* reach conscious awareness." Meanwhile, they claim that responsibility of the dorsal stream is *implementation* of action. This distinction gets at the heart of ongoing philosophical debates over what constitutes perception (as well as consciousness, more globally), highlighted in a recent psycho-philosophical review (Clark, 2009). Before moving on to the auditory system, I note that a dorsal/ventral stream dissociation has also been found in the somatosensory system, with the ventral component involving the parietal operculum/insula, with the dorsal component, as in the visual system, involving the posterior parietal cortex (see review by Dijkerman & de Haan (2007)).

An early proposal that there may be a "what"/"where" dissociation between streams in the auditory system was made by Rauschecker & Tian (2000) who, in a study of monkey auditory belt areas, showed that greater location selectivity was found in neurons of the caudolateral (CL) area, as opposed to those of the anterolateral (AL) and mid-lateral (ML) regions. This finding was in line with other studies which had demonstrated that caudally-located neurons were sensitive to space (Kaas & Hackett, 2000; Morel, Garraghty, & Kaas, 1993). Neurons of area AL, conversely, were shown to be more selective for monkey vocalization stimuli. Based on the neural selectivity as well as anatomical connectivity data, Rauschecker & Tian (2000) proposed a model where core regions pass information on to area ML, which in turn either sends further-processed information to CL ("where" stream) or both AL and CL ("what" stream). Early imaging research in humans implicated the right parietal lobe in auditory spatial (Weeks et al., 1999; Zatorre, Bouffard, Ahad, & Belin, 2002b) and motion (Griffiths, Green, Rees, & Rees, 2000) analysis, and an auditory what/where dissociation was observed in human ventral vs. dorsal cortical regions (Alain, Arnott, Hevenor, Graham, & Grady, 2001). Interestingly, the

review by Rauschecker and Tian concludes by discussing an apparent paradox between monkey and human data: whereas the monkey "what"/"where" dissociation seemed to follow a relatively clean ventral/dorsal mapping, in humans, the posterior (i.e. dorsal) regions of the superior temporal plane, the PT, is at the center of Wernicke's area (Wernicke, 1874), a speech perception region that seems to fit in much better with the "what" stream.

An auditory adaptation of Goodale's revised visual model (Warren, Wise, & Warren, 2005) provided a solution to the paradox: if the auditory dorsal stream processed audio-motor transformations/interactions (as opposed to merely analyzing sounds for spatial location), this would at least partially explain the position and involvement of Wernicke's area, as speech involves dynamic interactions between auditory and motor processes. Indeed, as far as the auditory system goes, the majority of the discussion surrounding the auditory "two streams" hypothesis has been reframed as a debate primarily concerning *speech* processing (see reviews by Hickok & Poeppel (2007); Rauschecker & Scott (2009) and *Figure 1.3*). These reviews propose fairly elaborate temporal-parietal-frontal models, complete with feedforward/feedback connections, as underlying various components of speech processing. Hickok's group puts forward a specific area, parietal-temporal "Spt" of the left hemisphere, as a "jumping off" point for the speech dorsal stream, which they argue is left hemisphere dominant (Hickok & Poeppel, 2007; Hickok, Okada, & Serences, 2008). Notably, Spt, which is thought to serve as an audio-motor interface, is located at the *posterior* end of the PT where it meets the parietal operculum (Hickok, Buchsbaum, Humphries, & Muftuler, 2003). In the human lesion literature, parietal-based circuits underlying audio-motor transformations have been linked to speech repetition deficits observed in conduction aphasia (Baldo, Klostermann, & Dronkers, 2008; Sidiropoulos, Ackermann, Wannke, & Hertrich, 2010).

The involvement of a dorsal perceptual speech processing stream hearkens back to the so-called "motor theory" of perception (Liberman et al., 1967), which postulated that the ability to generate sounds limits what can be perceived (thus inducing categorical perception of speech

sounds). As previously discussed, the strong form of the motor theory of perception has been shelved, as studies have shown that non-speakers such as infants (Eimas et al., 1971) and chinchillas (Kuhl, 1978) perceive speech sounds categorically. However, various studies have implicated the left ventrolateral prefrontal cortex (VLPFC) in speech perception, including phonemic experiments using PET (Zatorre et al., 1992) and fMRI (Y. S. Lee et al., 2012). Additionally, the involvement of this area was shown to be enhanced in high-noise (i.e. difficult listening) environments (Du, Buchsbaum, Grady, & Alain, 2014), suggesting that the dorsal stream and its motor circuitry may "lend a hand" when called upon to disambiguate certain perceptual situations (particularly when phonemic segmentation is required (Burton, Small, & Blumstein, 2000; Zatorre et al., 1996)), supporting a weak version of the motor theory of perception. These processes are reminiscent of (and perhaps related to) the "perception with scrutiny" accounted for by reverse hierarchical theory (Hochstein & Ahissar, 2002). (As an aside: the VLPFC, introduced above in the ventral stream section (1.2.3), has variably been labeled as being a dorsal (Hickok & Poeppel, 2007) or ventral (Rauschecker & Scott, 2009) stream structure by two-stream speech perception models. As previously stated, this region is a target of *both* streams, so its classification as either "dorsal" or "ventral" is somewhat a matter of semantics. More dorsally-positioned frontal regions, such as the dorsal premotor cortex, an unambiguous "dorsal stream" area, have also been implicated in speech perception (Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007).)

A dorsal stream that subserves audio-motor interactions is not limited to the domain of speech. In fact, Hickok's group showed that area Spt was involved in audio-motor interactions for melodies as well as for speech sounds (Hickok et al., 2003). Griffiths & Warren (2002) consider the PT, generally, to be a "computational hub," that can organize sounds according to their potential motor relevance and pass them down the dorsal stream, accordingly. Looking downstream, parietal and dorsal frontal regions may be performing processes related to

attentional orienting, including interfacing with other sensory modalities (Arnott & Alain, 2011). In the parietal lobe, the intraparietal sulcus (IPS), implicated in the reach/grasping literature mentioned above as well as with visual imagery tasks such as mental rotation (see review by Zacks (2008)), has recently been linked to auditory manipulation/transformation tasks, such as melody transposition (Foster & Zatorre, 2010) and temporal reversal (Foster, Halpern, & Zatorre, 2013). The mirror system, first detailed in the visual domain, has also been linked to auditory perceptual processing (Gazzola & Keysers, 2009; Kohler, 2002). Considering the motoric salience of sounds, Lahav, Saltzman, & Schlaug (2007) showed that perception of piano melodies in non-musicians did not activate premotor areas, but came to do so after short-term piano training, lending evidence to the "do pathway" dorsal stream model (Warren et al., 2005). To summarize, while there is much to still be discovered about the dorsal stream, it is clear that it is (1) involved in sensory-motor transformations in addition to more "pure" spatial analyses, (2) not restricted to the visual-motor system, and (3) not restricted to the speech domain in the auditory system.

1.3 Speech, music, and the two cerebral hemispheres

Before introducing the present investigation (section 1.5) and relevant methods (section 1.4), I would like first to introduce a general topic of the differential functions between the two cerebral hemispheres, with a focus on auditory processes. This topic will encompass the neural bases of speech plus some general knowledge concerning the functional properties of left vs. right hemispheres. I will also introduce the topic of music neuroscience as well as contemporary theories of why the left and right hemispheres may be differentially involved in speech and music processing. As discussed in the prior section (1.2), cortical processing of sound begins with spectrotemporal analysis in the primary auditory cortex and subsequently recruits differential hierarchically arranged structures as needed (with the major given example being of

dorsal vs. ventral stream structures). As I will now discuss, differential auditory processing in the left vs. right hemispheres is a second major dichotomy. Speech perception (including categorical perception) appears to be primarily a left hemisphere phenomenon, which in turn has implications on where *non-speech* auditory categorical perception may be taking place: a major question of this dissertation.

1.3.1 Language and cortex

It has been suggested for a long period of time, at least as far back as Broca and Wernicke (Broca, 1861; Wernicke, 1874), that the neural bases of language are lateralized to the left hemisphere. Considerable evidence in support of this view was added in the mid-to-late 20th century by use of the Wada test, in which one hemisphere is temporarily anesethetized (Wada & Rasmussen, 1960), and direct cortical stimulation studies of surgical patients (G. A. Ojemann, 1991; G. Ojemann, Ojemann, Lettich, & Berger, 1989). Rasmussen and Milner (1975; 1977) showed that essentially all right-handed (and most left-handed and ambidextrous) patients had speech representation weighted toward their left hemispheres. Penfield and Rasmussen (1949) had earlier presented evidence for "pure" speech areas: regions that were not "motor" in that their electrical stimulation did not induce vocalizations, yet halted speech production, itself. Such areas were found in left hemisphere regions roughly corresponding to Broca's Area in the inferior frontal gyrus (IFG), Wernicke's Area in the STG, and a third region in the inferior parietal lobule (which some have come to call "Geschwind's Territory" (Catani, Jones, & Ffytche, 2004)). Other lines of research in neurological patients demonstrated that left hemisphere dominance in speech is more pronounced for language production compared to comprehension. Such a production/comprehension dichotomy was vividly demonstrated by compelling "split brain" studies by Gazzaniga (1967; 1998) which showed that patients who had undergone callosotomies (severing the corpus callosum and effectively disconnecting the right

and left cerebral hemispheres) could read and comprehend simple written words/phrases presented only to their right hemispheres, but could not use the right hemisphere to speak those words. That being said, even considering only comprehension, the overwhelming literature points toward the left hemisphere as dominant for speech perception. While determining the lateralization of *auditory* language perception in these patients is less clear cut (as, unlike in the visual system, information from left and right fields is not so clearly segregated by sub-cortical circuitry), dichotic listening tasks (Kimura, 1967; B. Milner, Taylor, & Sperry, 1968; Springer & Gazzaniga, 1975; Zatorre, 1989) confirm a bias toward left hemisphere processes.

In the last couple of decades, neuroimaging has replaced lesion studies as the primary tool with which to examine the cortical basis of language processing. These techniques provide a unique window into the healthy, functioning brain. Many neuroimaging experiments using various methods (fMRI, PET, MEG, etc.) have confirmed the left hemisphere's dominant role in speech, while also demonstrating that both hemispheres contribute to the overall speech network (see review by Hickok & Poeppel (2007)). That being said, certain elements of speech perception may rely more heavily on right hemisphere circuitry, such as perception of voices (Belin et al., 2000; Formisano et al., 2008) and certain elements of prosody (Kotz, Meyer, & Paulmann, 2006; M. Meyer, Alter, Friederici, Lohmann, & Cramon, 2002; Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006).

The speech lateralization research is part of a broader literature concerning the general functions for which one or the other hemisphere plays a dominant role. Whereas the left hemisphere has long been thought to be the "language side," the right hemisphere has traditionally been considered dominant for spatial processing. Perhaps the most arresting evidence for this is the phenomenon of hemispatial neglect, in which patients with parietal lobe lesions fail to attend to approximately half of the visuo-spatial scene (see review by Vallar (1998)). This syndrome is relatively common in patients with damage to the right hemisphere, but quite rare following damage to the left (Heilman, Watson, & Valenstein, 1993), with the

imbalance thought to be due to the hemispheres' relative strengths: whereas the right hemisphere can effectively compensate for damage to the left (thus, the uncommon observation of right field neglect), the reverse is not true. The *left* parietal cortex, meanwhile has been implicated in disorders such as ideomotor and ideational apraxia (R. G. Gross & Grossman, 2008; Wheaton & Hallett, 2007), which generally require interpretation of linguistic instructions (e.g. to make a certain gesture). General left/right linguistic/spatial dissociations have also been highlighted in the working memory literature. Impairments in language-based phonological working memory have been linked to the left hemisphere structures, including posterior temporal/inferior parietal and inferior frontal regions (see review by Smith & Jonides (1997)), whereas analogous structures on the right have been implicated in visuo-spatial working memory (Smith, Jonides, Marshuetz, & Koeppe, 1998).

Of course, not all observation of left/right brain differences have to do with purely linguistic and/or spatial processing. Perception of faces, while bilateral, is thought to rely more on the right hemisphere (Meng, Cherian, Singal, & Sinha, 2012), although this may be in part due to the spatial nature of the task. Conversely, mathematical abilities, while also bilateral, may be more of a left hemisphere phenomenon (Dehaene, Piazza, Pinel, & Cohen, 2003), potentially due to their inherent links to language. Looking to the clinical literature, a hyperactive right hemisphere (and hypoactive left hemisphere) has been linked to depression (Hecht, 2010), part of a wider literature that links various emotional states more strongly to the left or right hemispheres (Craig, 2005). (As an aside: the idea that the two sides of the brain are broadly different from one another has spread to popular culture; a search of the book department at Amazon.com for "brain left right" reveals more than 20,000 results, with titles such as "*Left Brain, Right Stuff: How Leaders Make Winning Decisions*" (Rosenzweig, 2014), "*Raising a Left-Brain Child in a Right-Brain World*" (Beals, 2009), and "*At Left Brain Turn Right: An Uncommon Path to Shutting Up Your Inner Critic, Giving Fear the Finger & Having an Amazing Life!"* (Meindl, 2012). The

premise of such books is that there is a clear division between left (linguistic, rational, analytical) and right (artistic, creative, spontaneous) hemispheres, which govern overall personality traits. While there is no doubt that the two cortical hemispheres do show many specializations, there is little actual science behind such sweeping claims of over-arching left- vs. right-brain dominance (Nielsen, Zielinski, Ferguson, Lainhart, & Anderson, 2013).)

Moving on from the topic of left vs. right, I would like to briefly outline some of the nodes in the cortical speech/language network and their functional roles. Many of these regions will also come up again in my review of non-speech auditory processing, primarily via music. Historically, the frontal areas have been linked to speech production and temporal areas to perception (Wernicke, 1874) (although see the above discussion of dorsal stream evidence in perception, as well as the review by (Hickok & Poeppel (2007)), with the parietal areas linked to audio-motor transformations. In the previous section (1.2), I discussed the auditory dual stream model (Hickok & Poeppel, 2007; Rauschecker & Tian, 2000; Warren et al., 2005), which touches upon many of the major cortical speech perceptual regions. As formulated, the model is a primarily hierarchical one, with early cortical areas performing spectro-temporal analysis, before recruitment of the divergent processing streams. Areas of cortex implicated to some degree in speech processing include bilateral STG; bilateral ventral stream structures such as the STS, middle/inferior temporal regions, and the (ambiguously "ventral") inferior frontal gyrus (IFG); and left-lateralized dorsal stream structures such as the inferior parietal lobule (angular and supramarginal gyri), premotor cortex, and supplementary motor area (SMA). Perceptual speech areas can, very broadly, be divided into those dealing with pre-lexical processing (i.e. acoustics, phonemes) or those that engage lexical/semantic elements. Whereas tapping into lexical/semantic information seems to recruit the middle/inferior temporal regions (Bates et al., 2003), there is some degree of consensus that pre-lexical phonological units are first accessed in the STS and that these processes are bilateral (though weighted toward the left hemisphere) (Hickok & Poeppel, 2007). Such studies have generally been conducted by contrasting activity

related to speech sounds with non-intelligible but physically-matched controls such as sinewave analogs (Vouloumanos, Kiehl, Werker, & Liddle, 2001) and spectrally-rotated speech-like sounds (Liebenthal, Binder, Spitzer, Possing, & Medler, 2005; Scott et al., 2000). More recent research has confirmed the role of the left STS in intelligible speech (Okada et al., 2010), while also demonstrating the involvement of the homologous region on the right.

Discussion of phonological processing brings us full-circle to the topic of categorical perception, as speech phonemes are perceived categorically (as discussed in section 1.1). Liebenthal et al.'s imaging study from above (Liebenthal et al., 2005), in fact, was explicitly examining categorical perception of speech phoneme pairs, taken from a continuum between /ba/ and /da/. A spectrally-inverted control sound continuum produced samples that, while sounding speech-like, were not perceived categorically, and the contrast of phoneme > non-phoneme discrimination highlighted the left STS. A related phoneme study by Joanisse, Zevin, & McCandliss (2007) looked for adaptation effects following presentation of category-spanning oddball sounds (as contrasted with within-category oddballs) and also highlighted the left STS/MTG region (with smaller but significant activity observed in the left IPL, a putative dorsal stream area). In an intracranial EEG recording study, Chang et al. (2010) observed categorical processing in the lateral/posterior STG (although limitations of the technique did not allow for the electrodes to penetrate the STS or Sylvian Fissure).

Other studies relating to speech CP (Blumstein, Myers, & Rissman, 2005; Y. S. Lee et al., 2012; Myers, Blumstein, Walsh, & Eliassen, 2009; Raizada & Poldrack, 2007) have yielded more diverse findings, most notably the involvement of inferior frontal regions, which are convergence zones of the dorsal and ventral streams (Hickok & Poeppel, 2007; Petrides & Pandya, 2009; Rauschecker & Scott, 2009). Raizada's study highlighted left parietal and right posterior prefrontal regions as "amplification zones": areas that were more discriminative for between- compared to within-category phoneme pairs. Myers' study used a phonetic habituation

paradigm and observed regions of greater activity for between- vs. within-category adaptation in the left and right IFG and the left STG and STS. Blumstein et al. highlighted the left IFG as showing a graded response for phonemes that approach the voice onset time categorical boundary. Lee et al. used multivariate pattern analysis (MVPA) to determine regions that contained information that discriminated between the speech phonemes /ba/ and /da/, finding such regions in the STS, IPL, and Broca's Area (IFG), all on the left side. As previously discussed, Du et al. (2014) implicated both the IFG and superior temporal areas in speech CP, with the frontal region playing a larger role in noisier listening environments.

To summarize this section, it appears that CP of speech sounds calls upon both ventral and dorsal processing streams and that left hemispheric circuits seem to be dominant. Specifically, two regions, one each from the ventral (STS) and dorsal (IFG) processing streams, appear to be of particular relevance to categorical processing, with another dorsal stream region (the IPL) also implicated. As previously discussed in sections 1.2.3 and 1.2.5, the ventral and dorsal streams and their relevant structures are thought to play quite different roles in the analysis of sensory information, with ventral regions tied to object identification/recognition and dorsal regions involved in audio-motor transformations (Hickok & Poeppel, 2007). Here, the STS results may be mediating the conscious perceptual recognition of the phonemes, with the IPL/IFG sites putatively tapping into motoric processes related to an articulatory code. The preferential involvement of the IFG in ambiguous (Blumstein et al., 2005) and noisy (Du et al., 2014) scenarios suggests that such motoric processes may be "called upon" when needed, in order to supplement temporal lobe perceptual processes.

1.3.2 Music and the right hemisphere

Transitioning now to a discussion of the neural correlates of music perception, one of the oft-debated topics in left vs. right hemispheric function, introduced above, is their relative roles in speech vs. music processing. It is worth stating up front that this is not an either/or distinction, as

both domains are clearly processed by extensive networks in *both* hemispheres. As discussed in the prior section (*1.3.1*), there is considerable literature pointing toward a left-dominance for speech processing, although the right hemisphere's role is significant (Federmeier, Wlotko, & Meyer, 2008). Music is an intriguing topic of research as it, alongside speech, is unique to humans, found in all known human societies (D. E. Brown, 1991), and is the species' most sophisticated use of sound. While a less frequent topic of research than speech, the two domains nonetheless share many features, with each being complex information-conveying structures built from relatively small units: phonemes in speech, tones in music (Patel, 2007). There are further speech → music parallels at higher levels of organization, such as words → melodies and sentences → songs (Zatorre, Belin, & Penhune, 2002a), along with parallels in timbre (speaker → instrument) and time/amplitude dynamics (prosody → rhythm). There are, of course, major differences between the two systems, the foremost being speech's unique ability to convey highly precise information in the form of propositional statements (Massé, Harnad, Picard, & St-Louis, 2013), while aspects of music, such as tonality and metrical organization, are not found in spoken speech. This spectrum of similarities to and differences with speech makes music an appealing topic of neuroscientific investigation, as many findings thought to be specific to speech may (or may not!) in fact be much more general phenomena.

Like speech, music perception involves an extensive and bilateral network of cortical structures, including the superior/middle temporal cortices; motor and premotor regions; the inferior frontal gyri; and limbic and sub-cortical structures (see review by Peretz & Zatorre (2005)). As with speech, studies of patients with brain lesions provided some early evidence for a lateralization of musical processing, in this case towards the right hemisphere (B. Milner, 1962). Zatorre (1985) showed a link between right temporal lobe lesions and the ability to perform a discrimination task involving a single-note change within simple melodies, as compared to control groups with left hemisphere temporal or frontal lesions. Similar findings

implicating the right temporal lobe in tonal processing were found by Samson & Zatorre (1988). Peretz (1990) found that patients with right temporal lobe damage had specific problems in processing of musical contour. Meanwhile, behavioral (Ibbotson & Morton, 1981) and lesion (Fries & Swihart, 1990) studies of rhythm perception have shown a left/right dissociation between the ability to tap a rhythm (left hemisphere-based) vs. extract a beat (right hemisphere-based), indicating that the right hemisphere does not dominate every aspect of music processing. Indeed, evidence points to a link between musical aptitude and inter-hemispheric communication, as shown by Schlaug, Jäncke, Huang, & Staiger (1995)'s demonstration of a correlation between corpus callosum size and musicianship. However, in contrast to speech, music perception appears to rely more heavily upon right hemisphere circuitry, particularly in the superior temporal lobes (Peretz & Zatorre, 2005). This right lateralization is especially robust when considering the musical property of pitch.

As this thesis primarily involves the study of pitch (in the form of musical tones), I would like to specifically address sound frequency processing (upon which the perception of pitch relies), as well as sounds that build upon pitch perception, such as chords and melodies.

The processing of sound frequency is, of course, not limited to music, as the perception of speech and environmental sounds rely upon complex interactions between frequency- and time-based elements. As previously discussed, core auditory regions along Heschl's Gyrus respond to single tones and the various sub-regions of A1 have been each found to contain tonotopically-graded maps (Baumann et al., 2013; Moerel et al., 2012; Schönwiesner, Dechent, Voit, Petkov, & Krumbholz, 2014). The mental construct of *pitch*, meanwhile, requires the extraction and comparison of energies at various frequencies, which must be integrated into a unified pitch percept. The quest to find the "pitch center" has been somewhat contentious, but recent converging evidence seems to point to a region located bilaterally around the lateral HG and anterior-lateral PT (R. D. Patterson et al., 2002; Penagos, Melcher, & Oxenham, 2004). Dissociable from pitch, but still built upon frequency information, is timbre, a multidimensional

quality of sound having to do with its "tone," "voice," or "color." Timbre, the perception of which recruits core and non-core bilateral auditory regions (Menon et al., 2002), is, for example, what allows us to tell apart two musical instruments (or, alternatively, two speakers), even when they are playing the identical musical note.

Considering stimuli that require integration of multiple pitches begins to move us away from discussion of more domain-general properties of the auditory system towards musical processing, specifically. Musical sounds, for which perception requires the extraction and comparison of multiple pitches, include melodies/arpeggios, where tones are presented sequentially, as well as stimuli where multiple tones are presented simultaneously, such as harmonic intervals and chords. Processing of sequentially presented tones has been most strongly linked to the right auditory cortex. Johnsrude, Penhune, & Zatorre (2000) demonstrated that patients with right superior temporal lesions that specifically encroached upon HG were impaired in discriminating which of two tones was higher-pitched (although they could successfully make same/different discriminations on the same stimuli). Patterson et al. (2002) showed that, compared to steady-pitched sounds, melodies activated the right STG and PP. Hyde, Peretz, & Zatorre (2008) showed that the right PT was sensitive to changes in distance between two pitches. More recently, using MVPA, Lee, Janata, Frost, Hanke, & Granger (2011) found the right STS to contain information related to the direction (up vs. down) of a melodic contour. Simultaneously-presented chords, though less researched, have also been tied to right temporal processing (Koelsch, Gunter, Schröger, & Friederici, 2003). To summarize, there appears to be converging evidence that, in contrast to left-lateralized speech processes, musical aspects of pitch processing preferentially recruits circuitry in the *right* temporal lobe.

While the speech and music literature demonstrate, broadly, a dissociation between and left and right hemispheric processes, there remains an outstanding question of —*why*— this is the case. More specifically, there is debate over whether the two temporal lobes are actually tuned for more basic sound features that are differentially expressed in these two auditory domains.

There is, in fact, a large body of research suggesting that the left and right temporal lobes are differentially sensitive to temporally- vs. spectrally-complex sounds, respectively. The theory that the left hemisphere's dominance for speech processes was an evolutionary development for processing rapidly changing sensory and motor information was original proposed by Tallal, Miller, & Fitch (1993). Zatorre et al. (2002a) and Zatorre & Gandour (2008) extended this theory to encompass a temporal/spectral dissociation between left and right hemisphere processes, respectively. Zatorre & Belin (2001) used PET to directly test whether the hemispheres showed differential responses to sounds that expressed either temporal or spectral variation and, while demonstrating that both hemispheres play a role in both processes, observed the sort of left/right dissociation described above. Poeppel (2003) offered a related and complementary theory, based on the idea that the left and right temporal cortices operated upon different "time integration windows." Poeppel's Asymmetric Sampling in Time (AST) hypothesis suggests that the left temporal cortex is specialized to integrate information over short time scales (20-40ms, ideal for extraction of speech formant transitions) whereas the right side works over longer periods (150-250ms windows, on the scale of syllables and intonation contours). Recent research, such as that into cortical oscillations by Giraud & Poeppel (2012), provides compelling evidence that the left and right auditory cortices differ in the temporal windows (i.e. the "granularity") in which they "package" incoming sensory information.

I agree with this timescale asymmetry hypothesis in principle, although it may overemphasize such bilateral specialization as having evolved *for speech*. The right hemisphere temporal window, in particular, could subserve a wide variety of non-speech processes, whether environmental, musical, or otherwise. While the putative left hemisphere integration window seems to nicely fit those unique characteristics of speech (namely its fast temporal dynamics, particularly with regards to consonant perception), it remains an open question whether this specialization evolved —for— speech, or whether speech "hijacked" a specialized processor that was already in place. The fact that non-speaking animals such as chinchillas can make use of

those rapidly-changing features of the speech signal (Kuhl, 1978) suggests that such neural circuitry may have evolved prior to speech, rather than in service to it. A recent human neuroimaging study used an implicit category training paradigm with temporally-complex artificial non-speech sounds, and showed recruitment of such speech-sensitive regions in the left STS (Leech, Holt, Devlin, & Dick, 2009). This lends support to the idea that non-speech sounds containing speech-like temporal properties are processed in the left hemisphere. Meanwhile, multiple recent MRI experiments have implicated the right auditory cortex in processes requiring a fine-grained analysis of *spectral* energy (Boemio, Fromm, Braun, & Poeppel, 2005; Hyde et al., 2008; Schönwiesner, Rübsamen, & Cramon, 2005). While there is ongoing scientific debate about what exactly the left and right auditory cortices are doing differently from one another, the left/right temporal/spectral dissociation discussed above offers a fairly thorough explanation of why CP of speech sounds has been observed primarily in the *left* temporal lobe. Moreover, this theory suggests that categorical perception of certain sounds, namely those which are primarily defined by their spectral energy rather than their temporal dynamics, may be rooted in separate neural circuitry.

1.4 Neuroimaging methods and multivariate pattern analysis

Before discussing the specifics of the present investigation (section 1.5), I would like to first introduce a methodology, multivariate pattern analysis (MVPA), which was utilized in the second and third experiments (chapters 3 and 4) of this thesis, and compare MVPA with the more mainstream general linear model (GLM) approach, used in the first experiment (chapter 2). I will argue that the two methods offer complementary windows into higher cognitive function. As a description of MVPA and rationales for its use require a somewhat extensive background, I have chosen to introduce fMRI methods separately, prior to introducing the thesis questions, proper.

1.4.1 fMRI's advantages and limitations

Cognitive neuroscience (auditory or otherwise) relies to large degree on fMRI to provide brain-based biological data to complement and augment behavioral/psychological data. While biological data is available in other forms, including via psychophysiological (e.g. heart rate, skin conductance) and other brain-based measures (EEG, MEG, PET, lesion studies), fMRI has great advantages, the foremost of which is that it provides a non-invasive and fine-grained three-dimensional window into the healthy functioning brain.

Traditional fMRI analyses are based on the univariate contrast approach (Friston, Jezzard, & Turner, 1994; Worsley et al., 2002), which use the GLM to generate individual parameter estimates for every sampled voxel of the brain. These values are, in turn, used to compare across conditions using a "subtraction" or more generally a contrast (i.e. which voxels show significantly more signal in the experimental condition of interest vs. the control condition?). Thus, each voxel is treated as its own "island" for analysis and the direction of statistical inference is a "forward" one (i.e. the experimental design / task / stimuli are used to predict BOLD activity). As three-dimensional spatial smoothing is generally applied to the data (for a variety of reasons, discussed below), the GLM is primarily sensitive to the observation of *regions* of activity, rather than to isolated voxels. Thus GLM's strength lies in its ability to highlight regions of the brain that globally activate in certain experimental conditions, but not for baseline/control conditions that are lacking in some quality that hypothetically requires recruitment of certain neural circuitry. This method has been incredibly valuable to the field, as the fMRI-GLM pairing has formed the basis for thousands of cognitive neuroscience studies over the past two decades.

There are, however, certain issues and limitations with the standard GLM approach. The first major issue has to due with noise and (a lack of) sensitivity. The BOLD response, while well-validated, lives within a noisy signal (Logothetis, 2008; Logothetis, Pauls, Augath, Trinath, &

Oeltermann, 2001; Zarahn, Aguirre, & D'Esposito, 1997). While consistently robust for basic questions of perception (e.g. measuring the brain response to sounds that differ dramatically in their physical characteristics), BOLD differences between stimuli that differ in a more subtly cognitive manner may sometimes get lost in the sea of noise. In fact, there is research that suggests that fMRI data suffers from an epidemic of Type 2 errors that are widespread throughout the imaging literature (Lieberman & Cunningham, 2009). Spatial smoothing, generally employed to increase BOLD's signal-to-noise ratio, may actually *hinder* observation of significant activity if (a) there are relatively few highly active voxels (whose signals become averaged with that of inactive voxels) or (b) voxels that are more activated in experimental vs. control conditions are blurred together with neighboring voxels that are more highly activated in control than experimental conditions.

The second major limitation of the GLM approach concerns the independent examination of each voxel, mentioned above. (Note: use of the word "independent" here is a heuristic: voxels are not analyzed in a *truly* independent manner due to (a) spatial smoothing generally employed as a preprocessing step and (b) random-field theory-based statistics, which take into account the activity of neighboring voxels in order to reign in the extremely conservative statistical thresholds produced from Bonferroni corrections (Worsley et al., 2002). Procedures such as motion correction and resampling also contribute to non-independence of the data. However, in practice and principle, the GLM is a univariate approach, which analyzes a single location at a time.) This strategy reflects theoretical models of how the brain actually operates to a partial extent. One the one hand, there is considerable evidence behind the idea that the cortex is made up of modules (*see section 1.2.1*), hence the drive to link together univariate activity from a certain brain location to a particular behavior/percept. However, it is also well-established that (1) these modules are interconnected and interactive and (2) that individual brain regions process information in a locally-distributed manner (the "cell assembly" of Hebb (1949)). The GLM is generally blind to (1) (although certain insights can be made via the use of functional or effective

connectivity analyses (Friston, 1994)) as well as (2), assuming the distribution of processes takes place over a region significantly larger than a single voxel.

1.4.2 Multivariate pattern analysis - origins

Multivariate pattern analysis (MVPA) attempts to circumvent the GLM's limitations via the simultaneous analysis of many spatially distributed voxels paired with a "reverse" direction of inference. Over the past several years, the neuroimaging field has begun to heavily integrate MVPA, which is also referred to as a "brain reading" or "decoding" style of approach (K. A. Norman, Polyn, Detre, & Haxby, 2006). Decoding-based approaches to analysis, a subset of the larger "machine learning" field, have been used for many years in various fields both related and unrelated to neuroimaging. A well-known example is the use of classifiers in handwriting recognition software for identification of individual letters (Ahmad, Khalia, Viard-Gaudin, & Poisson, 2004). In the audio/music field, machine learning has been used for such varied purposes as identification of particular instruments in recorded music (McKay & Fujinaga, 2005) and conversion of electric guitar waveforms into midi note data (Yoo & Fujinaga, 1999). In the neuropsychology domain prior to its development for fMRI, classifier-based approaches had been successfully applied to simultaneously-acquired multi-neuron electrophysiological data in animal studies (Tsao, Freiwald, Tootell, & Livingstone, 2006) as well as to EEG data in humans (Peters, Pfurtscheller, & Flyvbjerg, 1998), including for use in brain-computer interfaces (Birbaumer & Cohen, 2007; Farwell & Donchin, 1988; Lotte, Congedo, Lécuyer, Lamarche, & Arnaldi, 2007).

In its simplest conceptual form, a learning algorithm is a piece of software that is "trained" on a set of data and then "tested" for efficacy on novel data (*see Figure 1.4*). Each sample of data (which I will be referring to as an "example") is made up of multiple individual components

(called “features"). Using the handwriting recognition example from above, each instance of a written letter is one example made up features (the individual pixels of the digitized image). During a training phase, a classifier learns which qualities of which features are most associated with a certain condition (e.g. the number "1" has certain qualities such as its orientation primarily along a single dimension, which clearly distinguish it from the number "0", which is characterized by an orientation along two dimensions and a closed loop, creating two isolated areas of background). After training, the learned model is tested on *novel* data that it has not previously had access to (e.g. “0”s and “1”s written by a different author) and assessed for the accuracy of these guesses. A well-trained classifier should be able to accurately predict the correct category membership of such novel examples at significantly above-chance levels.

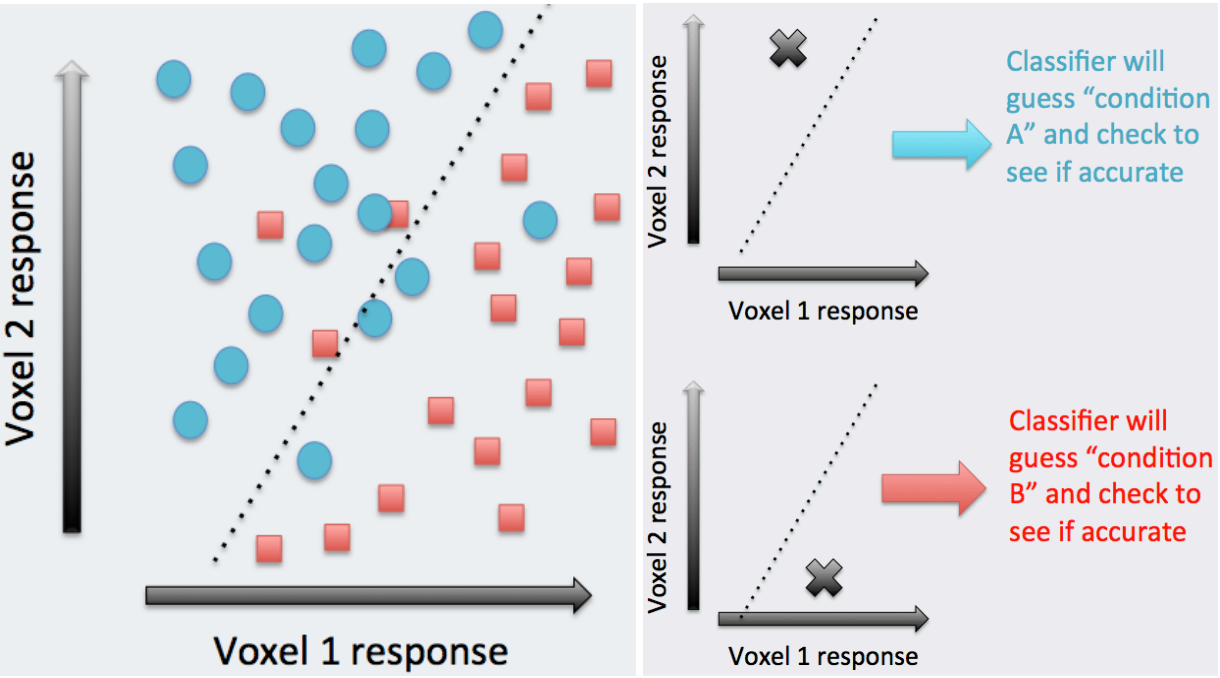

Figure 1.4 MVPA paradigm

Multivariate pattern analysis (MVPA) training (left) and testing (right) procedures. Voxel values from BOLD volumes were used to train software classifiers (as depicted in the left panel), in this case a support vector machine (SVM), using Python's PyMVPA toolbox (Hanke et al., 2009). While many voxels are simultaneously considered (hence the "M" in MVPA), for simplicity, the graphs show only an example set of responses from two voxels to Condition A (blue) and Condition B (red), which are plotted along two axes. (A and B could represent two distinct sounds, tasks, etc.) The SVM algorithm decides where to "draw" a decision boundary: a line in our example, a plane if three voxels were considered, and a "hyperplane" in multi dimensional space. For testing (right panel), the boundary calculated during training was then used to assess accuracy of the classifier on an independent dataset. Multi-voxel activity is checked against the decision boundary and a guess is made, which is compared to the *actual* condition label. If the classifier makes the correct decision at a rate significantly greater than chance, it follows that those voxels included in the analysis carry information content sufficient to disambiguate two conditions from one another.

1.4.3 MVPA with fMRI

There are many parallels between the above-mentioned handwriting recognition example and typical BOLD data: BOLD images are recorded as a series of discrete brain volumes, with each volume comprised of individual volumetric pixels (voxels). Likewise, in MVPA terms, each brain volume can be treated as an example and each voxel as a single feature. The first major application of machine learning methodology to fMRI data was done in a study of the visual system (Haxby et al., 2001). In this study, classifiers were trained on BOLD data from the ventral IT cortex to decode which of a set of stimuli subjects were looking at, where the stimuli differed from one another in category membership (faces, cats, man-made objects, etc.). The trained classifiers were able to determine the category membership of novel BOLD volumes at a rate significantly greater than chance. The authors determined that the neural representations of these categories were both "widely distributed and overlapping" (an interpretation very much in accordance with Hebb's cell assembly theory), which set the stage for much of the fMRI MVPA research that followed.

Another illustrative example of MVPA's high level of sensitivity, in a situation where univariate methods failed to detect significance differences between experimental conditions, can be seen in a visual study performed by Kamitani & Tong (2005). Here, the authors sought to observe BOLD differences following perception of eight different visual gratings, each with a different orientation. Orientation columns in visual cortex are believed to exist at a sub-voxel scale, with any one voxel containing neural populations that will respond robustly to stimuli at multiple orientations. Thus, it follows that a subtraction analysis (BOLD response to orientation A > BOLD response to orientation B) may not be sensitive to these between condition differences, as the voxels of interest will be similarly-responsive for both percepts. Indeed, the authors ran such univariate analyses and did not find significant activity differences. MVPA, however, showed classifiers could be trained to perform at accuracy levels that were significantly above chance level, indicating that the BOLD data in these visual regions did contain

information that correlated with differences in perception. The classifiers were thought to be working via *small-yet-consistent* within-voxel responses (e.g. voxel 1 generally responds a small amount more for condition A than for condition B) and between-voxel differences (e.g. voxel 2, located adjacent to voxel 1, generally responds slightly more for condition B than for condition A). Note that these differences were too small to be reflected in the GLM analysis and, furthermore, were likely obliterated during the smoothing process. While Kamitani & Tong (2005) showed that MVPA could reflect differences in percepts, Haynes & Rees (2005) took this a step further, using MVPA to show that classifiers could successfully differentiate between two physically distinct visual stimuli that the subjects, themselves, could not. In other words, differences in activity patterns in primary visual cortex reflected a difference between visual stimuli that subjects did not have conscious access to (as evidenced by chance-level behavioral discrimination performance).

Intriguing MVPA findings are, of course, not limited to fMRI studies of the visual system. A striking example in the auditory system was provided by Formisano et al. in a landmark 2008 paper. The authors, using a combination of 3 utterances made by 3 speakers, were able to demonstrate differential voxel patterns used to successfully decode the identity of the sound vs. the identity of the speaker, with the latter showing a solid right hemispheric bias. And MVPA need not be limited to studies of perception. A 2008 study (Soon, Brass, Heinze, & Haynes, 2008) employed a design in which subjects where asked to variably click the left or right buttons on an MR-compatible mouse. Subjects were completely free to determine which button to click and when to make their choices. MVPA showed that voxel patterns in fronto-polar cortex indicated which choice a subject would make *up to 10 seconds prior* to subjects' conscious awareness of their impending decisions (a result which was corrected for the time lag of the hemodynamic response). The MVPA results of the above-mentioned studies were not visible to GLMs, which were tested via separate analyses.

1.4.4 Univariate and multivariate methods compared

The landmark studies mentioned above illustrate the power of MVPA techniques and highlight some new theoretical concepts they can test. Most univariate studies attempt to differentiate between conditions that either have or lack some essential property (or vary the degree that quality is present via parametric approaches), e.g. testing for brain regions that are speech sensitive, by comparing BOLD responses to speech phonemes ("have") with responses following warped phonemes that are not perceived as speech ("lack") (Liebenthal et al., 2005). In contrast, Kamitani & Tong (2005) were able to analyze BOLD differences between percepts of different grating orientations, without having to decide *a priori* which orientation was the exemplar to which all others should be compared.

Thus, MVPA, allows for *direct* comparisons between what I will refer to as "sibling" conditions: a set of two (or more) conditions in which there is no true "control" or "null" condition that lacks some fundamental property possessed by the others. (To use a visual example, the same image colored blue, red and yellow may be thought of as three sibling conditions, whereas a grayscale version of the same image, due to its lack of a fundamental property (in this case, color), would be a "control" rather than a sibling.) This point is an important one when one considers the relative complexity of various experimental designs. As an illustrative example, consider two previously-mentioned studies by Liebenthal et al. (2005) and Lee et al. (2012). Both used fMRI to look at the neural bases of categorical perception of speech phonemes. Liebenthal et al.'s study employed (1) a quite complex acoustically-matched set of warped control stimuli upon which to perform contrast analyses, and (2) an active ABX discrimination task, in which subjects were required to hold multiple sounds in working memory and make an explicit choice (as well as press a button to indicate that choice). In contrast, the study by Lee et al. used neither a control stimuli set nor an active task: subjects passively listened to phonemes while performing only a simple orthogonal task designed to maintain alertness.

Without MVPA, both (1) and (2) were required to observe significant between-condition differences in the BOLD signal. One could argue that, if one is primarily concerned with examining automatic perceptual processes, the latter study's use of MVPA allowed for a more ecologically valid experiment.

That being said, there is clearly a role for both GLM- and MVPA-style research. Both approaches, despite their large methodological differences, may strengthen theories by showing converging evidence, as the left STS findings from both Liebenthal et al. (2005) and Lee et al. (2012) demonstrate. However, the search for activation (GLM) vs. information (MVPA) may also implicate separate portions of a larger network (e.g. Liebenthal's study found phonemic activation in the cingulate, Lee's in Broca's Area). This idea was directly tested by Jimura & Poldrack (2011), who found only moderate spatial correlations between the two methods. Empirical studies that have employed MVPAs and GLMs have shown that some regions that contain decodable information *do* activate from a resting baseline, whereas others do not (Soon, He, Bode, & Haynes, 2013). Such dichotomies are of theoretical interest if, for example, you observe multiple activated regions and believe that certain regions are involved in encoding of features of interest, whereas others regions are involved in some other task demand. As stated at the beginning of this section, I believe that MVPA and GLM analyses offer *complementary windows* into the workings of the brain.

1.4.5 MVPA choices

There are numerous issues and choices inherent in running an MVPA study which I will not discuss here in depth (but see the excellent practical review by Pereira, Mitchell, & Botvinick (2009), and other MVPA reviews (Etzel, Valchev, & Keysers, 2011; Haynes & Rees, 2006; Mur, Bandettini, & Kriegeskorte, 2008)). These reviews discuss issues such as temporal averaging, the choice of classification algorithm, and "over-fitting" of data (i.e. a lack of generalizability from

training to testing data). One important issue that I would like to briefly discuss is that of "feature elimination." Machine learning analyses work best in scenarios that contain many examples and relatively few features. fMRI yields the opposite scenario: many thousands of voxels paired with relatively few volumes, the latter limited by the length of the TR, the number of experimental conditions, and the duration of time a subject can ethically be scanned for. Thus, successful decoding generally requires a reduction of features prior to performing the MVPA, proper. There are at least 4 general methods for performing such a dimensionality reduction: (1) approaches such as independent component analysis (ICA); (2) pre-MVPA pruning of voxels, such as via t-thresholding against a baseline condition, running an ANOVA, or performing recursive feature elimination (RFE, in which *another* MVPA is used to discard voxels that do not contribute to successful decoding (De Martino et al., 2008)); (3) ROI analyses, in which voxels from only a specific pre-defined anatomical or functional regions are considered (Etzel, Gazzola, & Keysers, 2009); or (4) a searchlight approach (Kriegeskorte, Goebel, & Bandettini, 2006), in which the entire MVPA is performed on small sphere of voxels, which is then repeated iteratively across the entire cortex. All of the above approaches have certain pros and cons, which generally include a tradeoff between sensitivity (accuracy above chance) and specificity (the ability to tie such accuracy to a particular region or regions). I have generally chosen to employ the searchlight approach, as it allows for a high degree of spatial specificity and does not require *a priori* decisions about which region of the brain in which to look. While it also suffers from certain limitations (inability to use patterns from two distant regions in the same analysis; the problem of multiple comparisons and overly conservative statistical thresholding), these limitations are also present in GLM analyses, whereas MVPA opens up certain testable hypotheses that cannot be answered via a univariate approach. In certain situations, I have also used an ROI approach, as information may extend through regions larger than a searchlight sphere and/or there is a high level of spatial variability among participants. Due to the theoretical and practical differences between analyzing for information and activation (discussed above), I

have generally chosen to stay away from methodological choices that use activation thresholding as a precursor to MVPA.

1.5 The present investigation

The overarching goal of the present research was to clarify the functional bases of perceptual auditory processing in the human brain. As previously described, there is considerable language research indicating that the left hemisphere plays a primary role in speech perception and, more specifically, that the left STS, a ventral stream region, plays a major role in the perception of speech categories, whereas the left dorsal areas such as IPS and VLPFC perform audio-motor functions that are speech-relevant.

Considerably less research has been performed into *non*-speech auditory perception, both relating to perception of sound categories and sound-action representation. Thus, such left-lateralized neural findings for speech could be a consequence of speech-specific mechanisms. However, considering the temporal ventral stream, alternative hypotheses have been put forward, such as the possibility that the left may be a more generalized processor of auditory CP (Liebenthal et al., 2005)). Likewise for the dorsal stream, the observed left dominance (Hickok & Poeppel, 2007) may or may not be highly dependent on speech processes. Unless one contrasts speech with some other domain having similar properties it is difficult to decide on general vs. specific models.

Specifically, the goal of the present thesis was to examine the cortical network involved in the perception of musical sounds. A portion of this thesis (Studies 1 and 2) employed *categorically*-perceived sounds, in which behavioral indices showing unambiguous hallmarks of CP were used to tie the stimuli to a hierarchical organization of cortical sound processing. Such hierarchical networks could lie in either hemisphere, although prior speech/music research led us to hypothesize right hemispheric dominance. As such perceptual networks could lie in either the ventral or dorsal streams, we also performed an experiment that explicitly involved both

perceptual *and* motoric elements (Study 3), in an attempt to dissociate the kinds of processing occurring in temporal, parietal, and frontal circuitry. Ensuring the interpretability of the fMRI data required the use of well-selected auditory stimuli and control sounds, well-validated behavioral tasks, and pre-screened and highly characterized study participants. These elements, paired with cutting-edge fMRI analytical methods formed the cornerstone of the research.

A unique feature of this research, and common to all three experiments, was the extensive pre-screening of study participants, who were all highly trained musicians. The participants of studies 1 and 2 played a wide variety of instruments, whereas study 3 enrolled only pianists, and all were currently practicing or performing musicians, many of whom were enrolled in music degree programs at McGill University. Such selection was important as research has indicated that musicians show a much more robust behavioral CP effect than non-musicians (Locke & Kellar, 1973; Zatorre & Halpern, 1979) and that dorsal stream recruitment for music perception requires musical training (Lahav et al., 2007). Whereas such a sampling implies that any particular findings cannot be *directly* generalized to the broader population (namely, non-musicians), this was not the aim of the research. Instead, we sought to demonstrate the involvement of neural circuitry that relies upon long-term memory representations of non-speech sounds, selecting a population with the training necessary to have formed those representations.

Common to the entire line of research was the use of the musical interval as a basic building block for auditory stimulation. As discussed previously (*section 1.1.3*), it is the ratio of frequencies between two (or more) notes that serve as the foundation for Western musical harmony and melody. Whereas most musicians (as well as some non-musicians) can effortlessly name a two-tone musical interval, only those extremely rare individuals with absolute (aka "perfect") pitch have the ability to label a single tone played in isolation (Bachem, 1937); such individuals were not included in any of the present studies. Thus, it is the interval which forms the basic building block for the present research into music perception. Intervals, however, may come in many guises, both time-based (i.e. sounded simultaneously or sequentially) and

structural (in isolation, or as part of a larger musical structure such as a chord, phrase, etc.). This flexibility was exploited via differential use of musical intervals across the three experiments, including 3-note harmonic/simultaneous chords (experiment 1), 2-note melodic/sequential intervals (experiment 2), and 3-note melodic chords (experiment 3), each chosen to suit the particular needs of the experiment in question.

In experiments 1 and 2, participants were asked to perform a series of behavioral CP-related judgements, either online during scanning (experiment 1) or offline (experiments 1 and 2). Experiment 3, alternatively, was a piano keyboard-based motor task, so participants instead were behaviorally screened based on their ability to perform the piano task.

The CP tasks consisted of (1) identification and (2) discrimination paradigms. Identification tasks require a categorical selection following presentation of a stimulus, e.g. "Does the current interval sound more like interval A or interval B?" As described in *section 1.1*, categorically perceived sounds show identification plateaus surrounding category centers paired with sharp boundaries between two categories (i.e. sigmoidal rather than linear identification functions). However, due to short-term anchoring effects (Acker, Pastore, & Hall, 1995), non-categorically-perceived sounds may also show deviations from the linear function. Hence, discrimination, rather than identification, tasks are considered the "gold standard" for establishing true categorical perception. Discrimination tasks may take many forms, such as selection of same/different following presentation of two sounds or, in an "ABX" task, selection of whether sound "X" matches either sound "A" or sound "B." As previously described, for categorically perceived sounds, accuracy should be relatively good when two discriminated sounds are members of two different perceptual categories and relatively poor when both sounds belong to the same category. Discrimination functions were used as a behavioral screen for each of the first two experiments. Additionally, ABX discrimination was used "online" as the active behavioral task for a portion of experiment 1.

In contrast, the piano-based task was not designed to test CP, since that had been examined already in the first two experiments, but instead served as a means to dissociate the various

components of dorsal from ventral stream processes. Auditory ventral stream processes, likely (but not definitively) categorical, were hypothesized to be linked to the various combinations of tone percepts produced by the keyboard. Dorsal stream processes, meanwhile, were thought to be more related to the fingering combinations used to play those tones, rather than the tones themselves. Thus, the piano task was primarily designed to probe the mechanisms behind musicians' general abilities to perceive and produce structured musical sounds.

Our predictions at the outset of the first experiment were fairly specific: we expected to observe a link between behavioral categorical perception of musical intervals and BOLD activity in the right superior temporal sulcus. As discussed in *sections 1.2* and *1.3* the STS is generally regarded as part of the ventral stream for conscious identification, with the right hemisphere being preferentially implicated in detailed spectral analysis, requisite in pitch/interval perception. Moreover, the STS in the left hemisphere has been specifically implicated in speech-based categorical perceptual processing. While we primarily expected to see right STS activity, particularly for experiments 1 and 2 (which were direct tests of *categorical* perception) we did not rule out the observation of *bilateral* ventral stream processing, in the light of the bilateral circuitry of the ventral speech perception network (Hickok & Poeppel, 2007). Additionally, for various reasons outlined in section 1.2, we suspected the possible involvement of a dorsal processing stream, involved in the transformation of auditory information into an abstract and/or motor code.

Chapter 2 - A role for the right superior temporal sulcus in categorical perception of musical chords

Klein ME, Zatorre RJ. 2011. A role for the right superior temporal sulcus in categorical perception of musical chords. *Neuropsychologia*. 49:878–887.

2.1 Preface

This chapter describes an experiment conducted to examine the neural correlates of categorically perceived musical chords. Categorical perception (CP) has been behaviorally demonstrated in both music and speech (Section 1.1), and the neural correlates of speech CP have been linked to the left temporal lobe's ventral processing stream (Sections 1.2 and 1.3.1). As, compared to speech, many musical processes have been demonstrated to show a right hemispheric bias (Section 1.3.2), we tested the hypothesis that CP of musical chords would preferentially engage the *right* auditory ventral stream, particularly the superior temporal sulcus (STS). This manuscript was published in a 2011 issue of *Neuropsychologia* (Klein ME, Zatorre RJ. 2011. A role for the right superior temporal sulcus in categorical perception of musical chords. *Neuropsychologia*. 49:878–887).

2.2 Abstract

Categorical perception (CP) is a mechanism whereby non-identical stimuli that have the same underlying meaning become invariantly represented in the brain. Through behavioral identification and discrimination tasks, CP has been demonstrated to occur broadly across the auditory modality, including in perception of speech (e.g. phonemes) and music (e.g. chords) stimuli. Several functional imaging studies have linked CP of speech with activity in multiple regions of the left superior temporal sulcus (STS). As language processing is generally left-hemisphere dominant and, conversely, fine-grained spectral processing shows a right hemispheric bias, we hypothesized that CP of musical stimuli would be associated with right STS activity. Here, we used functional magnetic resonance imaging (fMRI) to test healthy, musically-trained volunteers as they (a) underwent a musical chord adaptation/habituation paradigm and (b) performed an active discrimination task on within- and between-category chord pairs, as well as an acoustically-matched, more continuously-perceived orthogonal sound set. As

predicted, greater right STS activity was linked to categorical processing in both experimental paradigms. The results suggest that the left and right STS are functionally specialized and that the right STS may take on a key role in CP of spectrally complex sounds.

2.3 Introduction

Categorical perception (CP) is a phenomenon that occurs when signals that vary over a continuous physical scale are perceived as belonging to a small number of discrete groups. CP can be considered the converse of the default process of continuous perception, in which signals are perceived along a smooth continuum and are not lumped into categories. Two hallmarks of CP are (a) distinct categories with obvious boundaries that can be observed during labeling tasks and (b) a peak in discriminability between stimuli near a boundary, with complementary troughs far from boundaries (Liberman et al., 1957).

Formation and use of categories is thought to serve multiple related perceptual purposes. CP allows the perceptual system to quickly abstract complicated information – in the realm of speech, spectrally complex and rapidly changing acoustic signals – into "bins" for further downstream use. Put another way, the brain labels a speech sound as belonging to a certain phonemic category (e.g. /da/ or /ta/) and then can build words from these phonemes, as opposed to having to store and manipulate the much more complex auditory representation relayed from the brainstem.

Relatedly, this process provides a theoretically simple solution to the problem of acoustic variation between speech utterances. In the context of a particular phoneme, individual speech utterances vary considerably between speakers and, to a lesser extent, from act to act performed by the same speaker. Because no two voicing of a phoneme can be identical, though, linguistically, it makes sense to treat them as such, CP provides the means for a pre-conscious decision in favor of one of among a relatively small number of categories. CP was initially thought to be specific to speech processing (Mattingly et al., 1971). Liberman et al. (1957)

detailed the presence of non-linear features in subjects' identification and discrimination abilities, which show, respectively, how reliably a specific signal will be labeled as having membership in a certain category and the degree to which two neighboring signals along a certain portion of a continuous physical spectrum are differentiable. The theory that CP is a product of learning/familiarity was given traction by studies, beginning with Goto (1971), which showed that subjects perceived phonemes from their first language significantly more categorically than non-native speech contrasts (a well known example being the meaningful distinction between /l/ and /r/ in English, but not in Japanese).

Up until this point, the bulk of experiments looking at CP used stimuli that were exclusively linguistic and drew conclusions about the phenomenon that were specific to the speech domain. However, studies in the 1970s and '80s broadened the literature from the speech domain to the psychology of music, by looking at perception of musical intervals and chords with regard to category membership (with obvious examples being minor vs. major distinctions). Musically, the frequency ratio between a base note and its third defines the two-note interval (or chord if there are three or more notes) as being "minor" or "major." Burns & Ward (1978) showed categorical perception of intervals, as seen in identification and discrimination plots. Subjects showed troughs in discrimination ability in locations that correlated with interval category centers. While Burns and Ward's study focused on melodic (i.e. sequential) note presentation, Zatorre & Halpern (1979) showed that the same phenomenon occurred in harmonic (i.e. simultaneous) intervals. Additionally, the authors showed that CP of musical intervals was much more prevalent in trained musicians than in subjects who did not have significant musical training. Zatorre (1983) also addressed the putative existence of (and relationship between) "auditory" and "categorical" memory processing stages by selectively interfering with only the former. The experimental manipulation seemed to spare a "binary variable" that constituted the categorical memory.

In the past few years, numerous functional imaging studies have examined the neural correlates of CP in subjects performing linguistic tasks, with results generally implicating the left superior temporal sulcus (STS). The left and right STS each are large regions, spanning posteriorly-to-anteriorly from y-values of less than –40 (MNI space) to near the temporal pole, respectively, and encompassing large portions of Brodmann areas 21 (inferior STS/middle temporal gyrus (MTG)) and BA22 (superior STS/superior temporal gyrus (STG)) as well as smaller regions of BA38 and BA39 (temporal pole and angular gyrus, respectively). Here, we refer to STS regions most proximal to Heschl's gyrus as middle STS (mSTS) (y-values of approximately –25 to –5) and label the anterior STS (aSTS) and posterior STS (pSTS) accordingly. Liebenthal et al. (2005) compared blood–oxygen-level dependent (BOLD) responses in subjects who were discriminating phonemes in addition to a warped, non- phonemic continuum of comparably complex sounds that did not sound like English-language phonemes and could not be associated with pre-learned categories. Contrasting BOLD activity in the two conditions highlighted two peaks in the anterior/middle and posterior STS. An adaptation (i.e. short-interval habituation) paradigm Joanisse et al. (2007) looked at BOLD activity contrasting conditions where oddball stimuli either did or did not cross a categorical boundary. The authors found greater BOLD activity for the between-category condition in the left STS, positioned between the peaks found by Liebenthal et al. The general correspondence of results between these two studies was notable, as the former utilized an active discrimination task and the latter a non-overt paradigm based upon a hypothesis of dishabituation/neural rebound, a design more common to ERP/MEG studies (Zevin & McCandliss, 2005).

Another recent study (Leech et al., 2009) showed that the left STS is likely involved more generally in CP and not merely limited to speech categorization. Subjects were trained on a video game, wherein certain fast-transforming complex sounds were indicative of an imminent game-play action. Study participants did not report these "acoustically-complex, artificial, and non-linguistic" stimuli as sounding speech-like. Because presentation of the sounds preceded (and

predicted) specific upcoming events and required behavioral responses, acquisition of these new non-linguistic categories would be helpful with game performance. Participants who best learned these novel categories showed the greatest pre- to post-training change in BOLD response in the left pSTS, as observed during passive listening to these same stimuli. Thus, the authors concluded that CP correlated with left STS activity reflects auditory expertise in domains not limited to just language, and is susceptible to learning.

The common thread between these imaging studies is the observation of significant BOLD activity in the left STS. The authors generally support the theory that the left STS is strategically positioned in the midst of the auditory "ventral stream" (Rauschecker & Tian, 2000), between more primary areas involved in the analysis of physical features of speech/other complex sounds and higher- order auditory cortex located in the left MTG and parts of the STS located more anteriorly. Liebenthal et al. suggest that phonemic recoding may be the earliest speech signal analysis that is lateralized to the left and that the STS is the actual "point of transition" – where sound starts to become speech. The implication here is that the category maps, themselves, reside within the left STS and that the observed BOLD signal, at least in part, reflects activity of the neurons that comprise the maps.

While the above imaging experiments of speech perception, as well as the study by Leech et al., make a very convincing case for a major role of the left STS in CP, they paint an incomplete picture of the phenomenon. The commonality between those studies is that they look for a BOLD response following categorization of rapidly transforming, temporally complex sounds. These findings cannot necessarily be taken as having highlighted the neural basis of all auditory categorical perception. Namely, they say little concerning acoustic stimuli lacking dynamic spectral variation, of which musical intervals are a prime example (and one that has already been shown to be perceived categorically). The idea of quickly- vs. slowly-varying auditory signals relates to theories of hemispheric specialization, in particular that the left hemisphere is tuned for perception of fast-changing signals (and thus is well-suited for speech) while the right hemisphere is tuned for higher spectral resolution. This theory –that left and right hemispheres,

respectively, subserve these two parallel and complementary functions– was put forward by Zatorre et al. (2002a) as well as Poeppel (2003), whose argument was framed around putative "time integration windows" that are preferred by each of the two respective cortices. In this vein, numerous studies have shown that the right hemisphere is preferentially active for stimuli containing small variations in spectral energy (Boemio et al., 2005; Hyde et al., 2008; Schönwiesner et al., 2005; Zatorre & Belin, 2001). Thus, an imaging study that seeks to highlight brain areas involved in categorization of musical chords may implicate neural networks in the right temporal lobe responsible for a more inclusive concept of categorical perception. One can also make an alternate hypothesis that musical categories, such as minor and major, are mediated linguistically and thus rely heavily on the left STS for their percepts as categories. However, as any such linguistic labeling is predicated upon fine-tuned spectral analysis/extraction, it follows that some sort of pre-categorical → categorical transformation must occur prior to associations with lexical elements, and that such a transformation is more likely to be primarily carried out by the right temporal lobe.

Here, we used fMRI to test the prediction that greater activity in or near the right STS of highly-trained musician subjects would be observed following presentation of stimuli comprised of chords from a larger number of musical categories. Such a finding would (a) suggest that there is something intrinsic to this brain region, bilaterally, that allows for transformations from nonspecific raw signal into pre-defined, cortically-based category and (b) lend credibility to theories that the relative strengths of the right and left temporal lobes are grounded in a differential sensitivity to slowly- and quickly-evolving sounds, respectively. While the specifics of any such findings (i.e. right STS activity associated with musical categories in musically-trained subjects) might not generalize to the population at large directly, observation of the predicted result would speak to a differential readiness/ability of the right vs. left STS to take on such a role in CP of spectrally complex sounds. In addition to looking at differences between minor/major 2-category vs. single-category conditions, we created a set of acoustically matched

orthogonal sounds to serve as an additional experimental control. These orthogonal stimuli use absolute pitch cues and lack association with any learned musical categories. We predicted that, compared to the experimental triads, these orthogonal triads would be perceived in a less categorical manner, as measured by identification and discrimination scores. Finally, seeking converging evidence of functional localization, we employed two discrete experimental protocols: (1) an adaptation/oddball paradigm in which subjects were not asked to make explicit judgments related to category membership and (2) an ABX discrimination paradigm where overt, keyed responses were required.

2.4 Methods

2.4.1 Participants

We enrolled 35 participants in a behavioral pre-test. All subjects were right-handed, age 18–50, and did not claim to possess absolute pitch abilities. All were musicians with 4+ years of formal training on an instrument and claimed to be currently performing or practicing. All subjects gave informed consent to participate in this study, in accordance with procedures approved by the Research Ethics Committees of the McConnell Brain Imaging Centre and the Montreal Neurological Institute. Because we were interested in maximizing the likelihood of measuring the neural substrates of CP, following our pre-test, 19 of the 35 participants were excluded from further participation due to lack of sufficiently clear CP-like discrimination functions (see Section 2.5, for specifics of inclusion criteria). Additionally, two subjects who met these criteria chose not to participate in the imaging study and four more were eventually excluded due to failure to comply with instructions during scanner sessions. Thus, the imaging data are from a final cohort of 10 participants.

2.4.2 Stimuli

The behavioral pre-test involved two parallel sound sets, each containing 11 discrete triads (*see Figure 2.1*). We generated an experimental and an orthogonal set, which shared one common triad. All of the triads were composed of three simultaneous 500 ms sine-wave tones (i.e. harmonic triads) that were generated using Audacity software and were derived from equally-tempered semitones (in which an octave lies 1200 cents above a starting frequency and each 100 cents signifies a 1/2 tone shift). Sound intensity was adjusted to each subject's comfort level and every triad was presented using a 50 ms linear ramp-up/down. The experimental sound set consisted of triads that ranged from true minor (middle note 300 cents above base note) to true major (middle note 400 cents above base note), in 10-cent increments (i.e. 300, 310, . . ., 390, 400). For all triads in the experimental set, the high note (musically, the 5th) was positioned 700 cents above the low/base note. Note that, for all triads in this set, the low and high notes were fixed at the same frequencies (G-natural at 392 Hz and the D-natural at 587.3 Hz) and only the middle note varied, from B-flat (300 cents above G-natural/466.2 Hz) to B-natural (400 cents above G-natural/496.8 Hz).

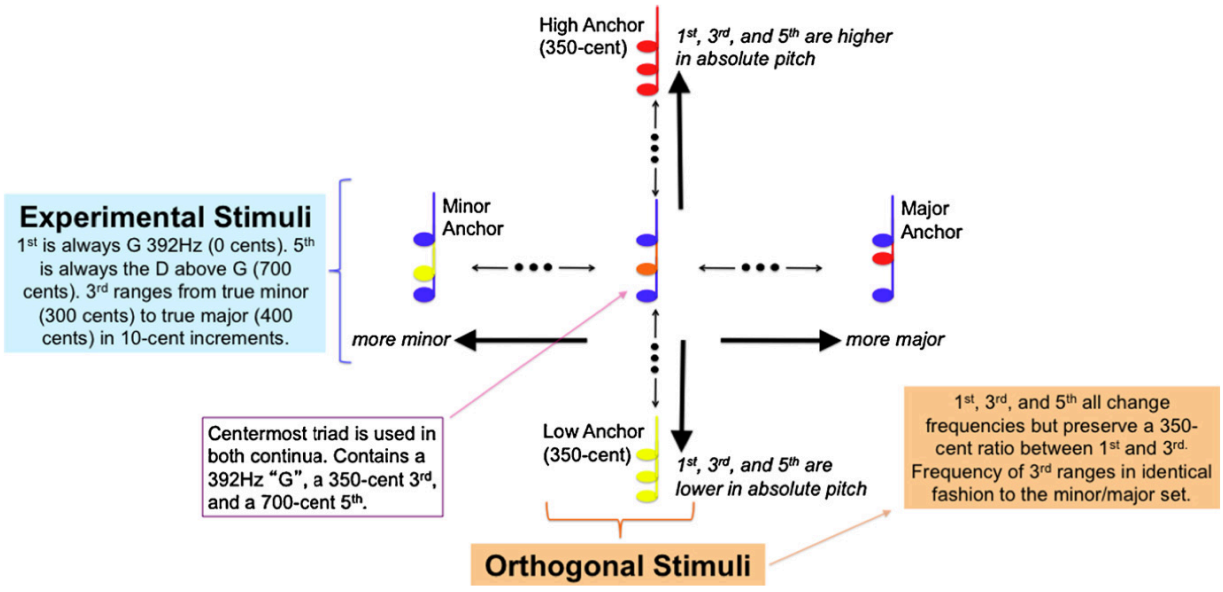

Figure 2.1: Two triad sets

Experimental stimuli are represented horizontally. Moving from left to right, the triads become progressively more major (from 300 cents to 400 cents, in 10 cent increments). This was done by varying the frequency of the middle note, while the frequencies of the bottom and top notes remain constant. The mid-most triad (350 cents) is shared with the 2nd stimuli set. Orthogonal stimuli are represented vertically. Moving from bottom to top, triads become progressively higher in frequency; this is true for all three notes of the triads (as opposed to only the middle note, as in the experimental set). Because the frequency ratio for the three notes of each triad is held constant, these orthogonal triads do not differ from one-another in the minor/major dimension.

The orthogonal stimuli set was constructed in parallel to the experimental set. Our intent was to create a series of triads that did not span the categorical boundary between minor/major, while remaining as acoustically related to the experimental stimuli as possible. As it is the ratio between the musical 1st and 3rd that determines the minor or major quality of the triad, we kept this ratio fixed at 350 cents (i.e. 1:~1.22) for all triads in the orthogonal set. The 350-cent triad was chosen as it represents the midpoint on the minor/major continuum and does not clearly belong to either the former or latter category, as shown by identification ratings (see Section 2.4.3). As with the experimental set, the middle notes of these 11 triads ranged from B-flat (466.2 Hz) to B-natural (496.8 Hz). However, in order to keep consistent a 350-cent interval between low and middle tones, it was necessary to vary the frequency of the low tone from triad to triad. This is in direct contrast to the experimental set, where the frequency of the low tone was always fixed at 392 Hz. As the middle tone varied from 466.2 Hz to 496.8 Hz, the low tone varied from 380.8 Hz (between G-flat and G-natural) to 405.8 Hz (between G-natural and G-sharp). Likewise, the high tone (5th), which was always positioned 700 cents above the base tone, varied in the orthogonal sound set, from 570.6 Hz to 608 Hz. While all three tones vary in frequency from triad to triad within this sound set, the frequency ratio between the three tones is held constant. As a result, these orthogonal triads, unlike the experimental triads, do not differ from one another along the minor/major dimension, but instead differ on the basis of their absolute frequency. In order to keep a consistent naming scheme, individual triads from both sound sets will be referred to on a scale from 0 to 100 cents, which represents the distance above the low anchor triad from either set. However, it is important to note that this distance refers either to pitch-variation of the middle note (experimental triads) or of all three notes (orthogonal triads), depending on the sound set.

For the pre-test, sounds presentation and data collection were conducted using Max/MSP software (Cycling ‘74 Inc., http://www.cycling74.com) and Sennheiser HD280 Pro headphones. In-scanner tasks were administered with Presentation software (Neurobehavioral Systems,

http://www.neurobs.com) and MR-Confon Peltor Optimex magnetic resonance-compatible headphones.

2.4.3 Pre-test tasks

Subjects performed identification and discrimination tasks of both sound sets as part of a behavioral pre-test, conducted inside a sound booth. Prior to performing the identification task, subjects listened to repeating and alternating presentations of the two endpoint-triads. These endpoint (a.k.a. "anchor") triads were the true minor and major triads for experimental set identification, or the two analogous triads if the subjects were performing the task on the orthogonal set. The order of presentation was counter-balanced so that half of the subjects first heard the experimental triads and half the orthogonal triads. During the fMRI portion of the experiment, subjects performed a similar discrimination task, and also underwent an adaptation/oddball paradigm.

After familiarization with the anchor triads, subjects were presented with trials that contained a single triad that could come randomly from anywhere in the set. They were then asked to rate that triad on a scale of 1–6: (1) subject is sure triad is closer to low anchor, (2) subject thinks the triad is closer to the low anchor, but is not positive, (3) subject is fairly unsure, but if pressed to guess, would place the triad closer to the low anchor, (with (4), (5), and (6) the complementary choices for the high anchor). Subjects had unlimited time to make their selections and, following each choice, were given a 2-s silent period prior to presentation of the next triad. Each of the 11 possible triads from a given set was presented 12 times in a pseudo-random order.

Following the identification task, subjects performed an ABX discrimination task on the same triad set. For each trial in this task, subjects heard three triads, each separated from the next by 500 ms of silence. In this task, "A" could be any one of the 11 possible triads; "B" would be a triad, 2 steps away from "A" (either up or down) on the continuum; and "X" would be a repetition of either "A" or "B." An example from the experimental set would be presentation of a

30-cent triad ("A"), followed by a 10-cent triad ("B") and another 10-cent triad ("X"). After each presentation, subjects were asked to click "1" if they believed X matched A or to click "2" if they believed X matched B. In the above example, a response of "2" is correct. Following each response, there was a 2-s silent period prior to the next trial. Subjects were not provided with correct/incorrect feedback. There were an even number of X = A and X = B trials as well as an even number of trials where A>B or B>A, in terms of frequency/position in the stimuli set. Each of the 9 possible complementary triad pairs from a given set was presented 12 times in a pseudo-random order. Each subject performed two identification tasks and two discrimination tasks for each triad set.

In order to qualify for the fMRI portion of the study, a subject had to show a (a) discrimination performance peak for the minor/major triad set that was 25%+ better than the average of their within-category endpoints and (b) 50/70 cent discrimination rate that was not significantly lower than their peak performance (whether that peak was found at 40/60, 60/80, etc.). The second criterion was included because, as the large majority of subjects' performance peaks were found at 50/70, this pair was selected to become the between-category condition used in-scanner. 16 of the initial 35 subjects met both of the above criteria. Of these 16, two subjects declined to participate in the fMRI section. Data from four further subjects who were scanned were excluded from the imaging analyses due to subjects' failure to comply with in-scanner instructions (i.e. required behavioral responses that were absent or inconsistent). Thus, our imaging data come from a final cohort of 10.

2.4.4 MRI procedures

Each participant underwent an anatomical scan and two functional imaging runs. Each run consisted of eight blocks of triads: four each for the adaptation (ADPT) and discrimination (DISC) protocols (see below for details of each protocol). For each protocol, two blocks contained triads from only the experimental sound set (EXP) and two contained triads from only

the orthogonal sound set (ORT). Run "A" was ordered DISCexp>ADPTexp> DISCort>ADPTort>ADPTexp>DISCexp>ADPTort>DISCort. Run "B" was ordered ADPTort > DISCort > ADPTexp > DISCexp > DISCort > ADPTort > DISCexp > ADPTexp. We used a counterbalanced design so that half the subjects underwent run "A" then "B" and half "B" then "A."

Blocks were separated from one another by two silent trials where no sounds were played, followed by a "cue" trial, where subjects were told which protocol to follow in the upcoming block. Each run contained a total of 166 10-s trials: 76 from the adaptation experiment (19 per block × 4 blocks); 64 from the discrimination experiment (16 per block × 4 blocks); 18 silent; and 8 cue. Trials using the middle-frequency triad pair of each stimuli set (50/70) were presented twice as often as those from either the low- or high-frequency pairs (0/20 and 80/100, respectively). Triad pairs from the pre-test, other than 0/20, 50/70, and 80/100, were not used for the imaging experiment as we sought to contrast the most boundary-spanning (50/70) and least boundary-spanning (0/20 and 80/100) conditions.

Adaptation paradigm: A single ADPT block contained 19 trials and used triads from only one of the two sound sets. Each trial (*see Figure 2.2*) was one of two types. Repeating type (REP) was presented as A–A–A–A–A, where the same triad was presented 5X, with 500 ms silent gaps between sounds. Changing ("oddball") type (CHG) was presented as A–A–A–A–B, where one triad was presented four times followed by a second triad that was presented once. As with REP, there were 500ms silent gaps between sounds. In any given trial, A and B were complementary triads from a pair (ex: if A = 70, B = 50). REP and CHG trials were presented with equal frequency and in a random order. Of each block's 19 trials, 4 contained triads from the 0/20 pair, 4 from the 80/100 pair, and 8 from the 50/70 pair.

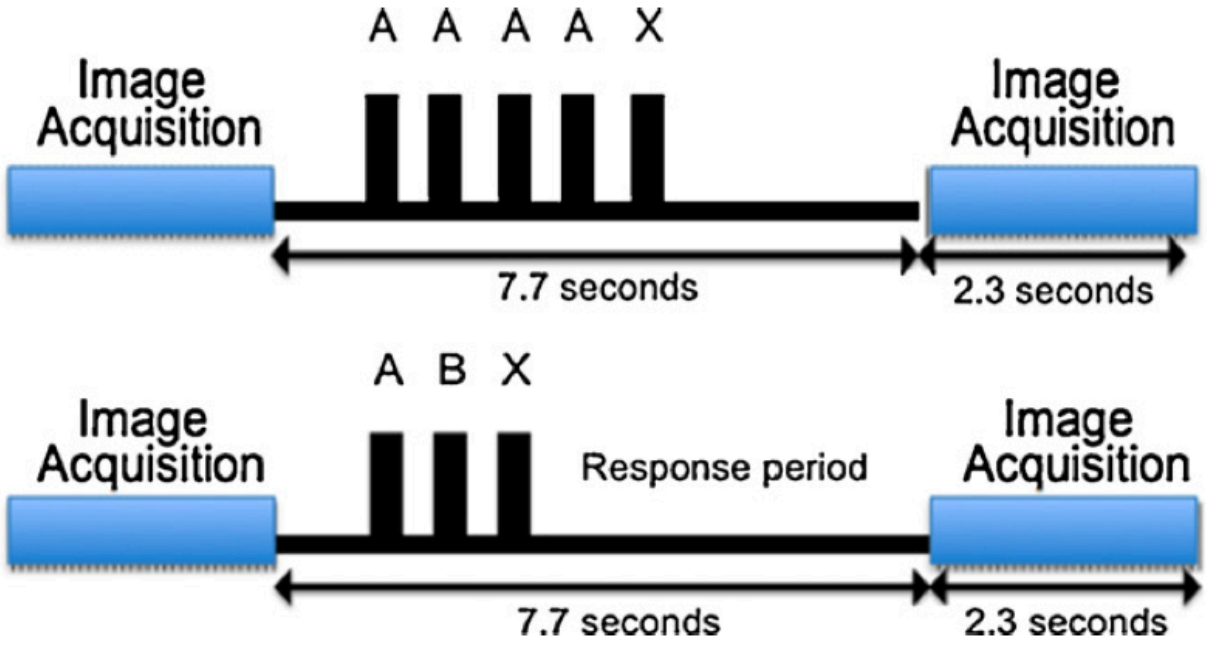

Figure 2.2: Single trials from adaptation and discrimination experiments

Single trials from adaptation (top) and discrimination (bottom) experiments. Each 10s trial was comprised of 2.3s for image acquisition following 7.7s for sound presentation and behavioral responses. Longer durations of stimuli during adaptation trials were offset by the lack of a need for a response period. Trials occurred in blocks containing only those of same type (e.g. discrimination of experimental triads, adaptation using orthogonal triads, etc.).

The remaining 3 trials per block were employed for a separate purpose. The adaptation paradigm, itself, required no overt responses from subjects. However, in order to ensure that they remained alert and were attentive to the sounds, we had subjects undergo each ADPT block under the guise of an overt "loudness" task. Subjects were requested to make a key-press if a trial's final triad was heard as being quieter than the preceding 4. Thus, in addition to the 16 trials mentioned above (in which all 5 triads were of equal intensity), 3 trials contained final triads that were of 1/4 the amplitude of the first 4. While behavioral responses were checked for compliance with the loudness task, we did not analyze fMRI data collected from these trials. For this paradigm, subjects were not specifically instructed to listen for whether the final triad was of different pitch quality than the first 4.

Discrimination paradigm: The in-scanner ABX discrimination task (*see Figure 2.2*) was similar to that described for the pre-test. As mentioned above, one difference was that subjects heard and discriminated only the 0/20, 50/70, and 80/100 pairs from each set. A second difference was that, where the pre-test allowed for a response period of unlimited duration, the fMRI task required a response before the onset of BOLD volume acquisition. This period of relative quiet ranged between 3.8 and 4.8 s in duration and, following presentation of triad X, subjects were asked to respond as "quickly as possible" by pressing one of two buttons on an MRI-compatible controller, depending on whether they heard X as matching A (choice 1) or X as matching B (choice 2). Of each DISC block's 16 trials, 4 contained triads from the 0/20 pair, 4 from the 80/100 pair, and 8 from the 50/70 pair.

Image collection and analysis: Images were acquired on a 1.5 T Siemens Sonata scanner. A high-resolution (voxel = 1 mm3) T1-weighted scan was obtained for anatomical localization. During two functional runs, one whole-head frame of 36 contiguous T2*-weighted images was acquired in ascending, interleaved fashion (TR=10s, 64X64 matrix, voxel size=8mm3 (2mm×2mm×2mm)). We used a sparse-sampling procedure (Belin, Zatorre, Hoge, Evans, &

Pike, 1999): tasks were performed between the 2.3s acquisitions to prevent scanner noise from interfering with the auditory stimuli. Sound samples were presented near the beginning of the 7.7s non-acquisition window. Relative timings between scan acquisitions and tasks were systematically varied or "jittered" by up to ±500 ms to maximize the likelihood of obtaining the peak of the hemodynamic response for each task.

All BOLD images were realigned with the third frame of the first run to correct for motion artifacts. To increase the signal-to-noise ratio, images were smoothed with a 6-mm full-width at half-maximum (FWHM) isotropic Gaussian kernel. Image analyses were conducted utilizing the general linear model via fMRISTAT as outlined by Worsley et al. (2002). Motion-correction parameters were used as covariates in fMRISTAT to further account for motion artifacts in the imaging results. In-house software was used to non-linearly transform each subject's images into standardized space using the MNI/ICBM 152 template, prior to conducting the group analyses (Collins, Neelin, Peters, & Evans, 1994; Mazziotta et al., 2001). Peaks from the full-brain analysis were considered significant if above a threshold of $t > 4.57$, which was corrected for multiple comparisons ($p = 0.05$). The program stat summary assessed the threshold for significance by selecting the minimum among the values given by a Bonferroni correction, random field theory, and the discrete local maximum (Worsley, 2005). We report peaks of neural activity if their voxel or cluster p-values are <0.05. For a portion of our fMRI analysis, we pre-defined a region spanning the right STS. Within this predicted area we report any peaks that were significant above an uncorrected threshold of $p=0.001$. We performed the location-based analysis because our primary prediction, based upon multiple streams of prior research, focused on this specific right temporal region. As the speech/language literature has highlighted activity peaks over multiple areas in the left STS, we delineated the entire right STS, spanning from the most posterior to most anterior regions of the sulcus. The STS was manually segmented based upon anatomical landmarks: (a) from posterior to anterior for as long as the sulcus was clearly visible (near angular gyrus ($Y = –46$) to near temporal pole ($Y = 6$)); (b) dorsal/ventral from the

most central/superficial point of the STG to that of the MTG; and (c) encompassing the entire sulcus (superficial to white matter).

2.5 Results

2.5.1 Behavioral results

A two-way ANOVA performed on identification ratings from all 35 subjects during the pre-test (*see Figure 2.3*) showed a significant interaction effect between sound set and stimulus frequency ($F = 8.848$, $p < 0.001$). Tukey's honestly significant difference (HSD) post hoc tests showed a significant difference in mean rating of the experimental vs. orthogonal triads at frequencies of 40, 50, 60, and 70 cents ($p < 0.05$ for all), but not at the left-most (0, 10, 20, and 30 cents) and right-most (80, 90, and 100 cents) ends of the functions.

For discrimination performance from all 35 subjects during the pre-test (*see Figure 2.4*), a two-way ANOVA showed a significant interaction effect between sound set and stimulus frequency ($F = 5.154$, $p<0.001$). Tukey's honestly significant difference (HSD) post hoc tests showed a significant difference in discrimination performance of the experimental vs. orthogonal triads at frequency pairs of 0/20, 10/30, 20/40, 30/50, 70/90, and 80/100 cents ($p < 0.05$ for all), but not at the center of the functions (40/60, 50/70, and 60/80 cents). To confirm that the experimental triads were being perceived in a categorical-like manner, further Tukey's HSD post hoc tests showed that peak discrimination performance of this sound set (50/70 comparison, 84% accuracy) was significantly better than at the 0/20 (56% accuracy, $p < 0.05$) and 80/100 (56% accuracy, $p < 0.05$) endpoints. Performance at the two endpoints did not differ significantly from one another. The orthogonal triads were discriminated with a peak accuracy of 85% (50/70) and endpoint accuracies of 69% (0/20 and 80/100).

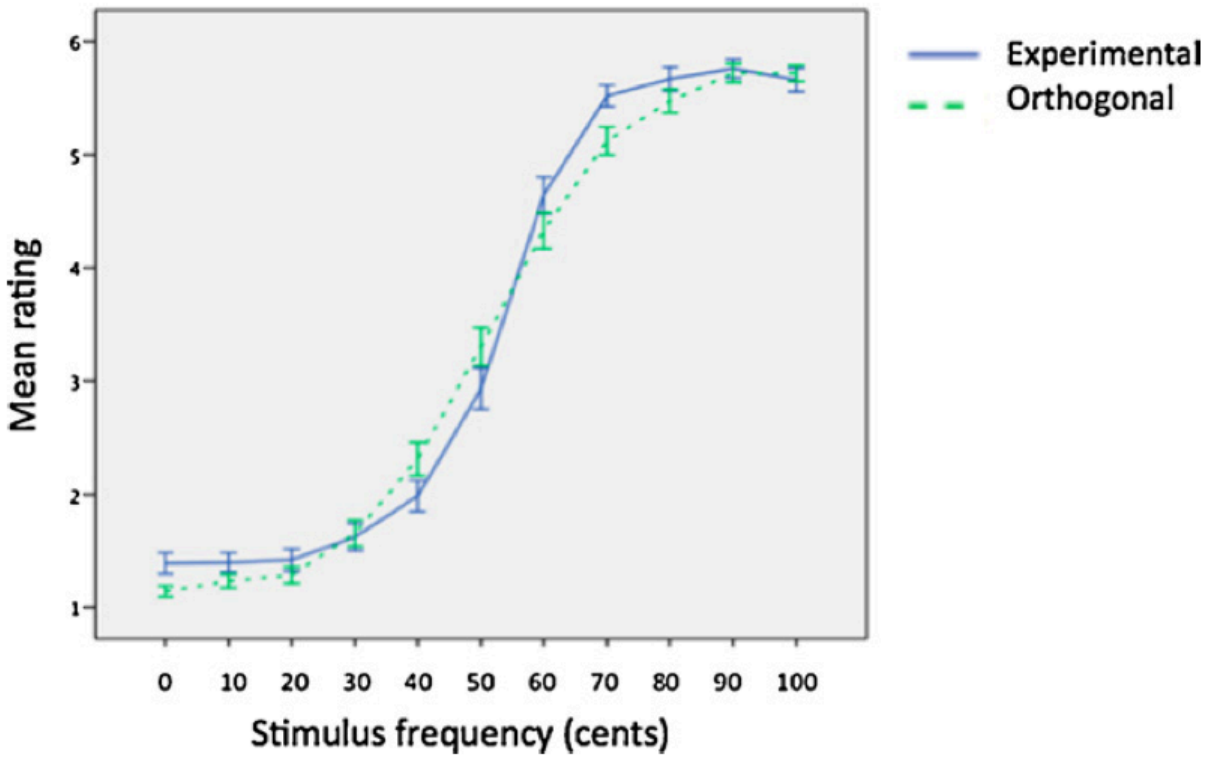

Figure 2.3: Identification performance

Mean ratings are from a scale of 1 to 6, as presented triads are perceived as resembling the low to high anchor triads of each sound set, respectively. X-axis values represent cents above a minor triad as determined by the middle note (experimental) or cents of each of the three notes above the lowest-frequency triad (orthogonal). Error bars show 95% confidence intervals.

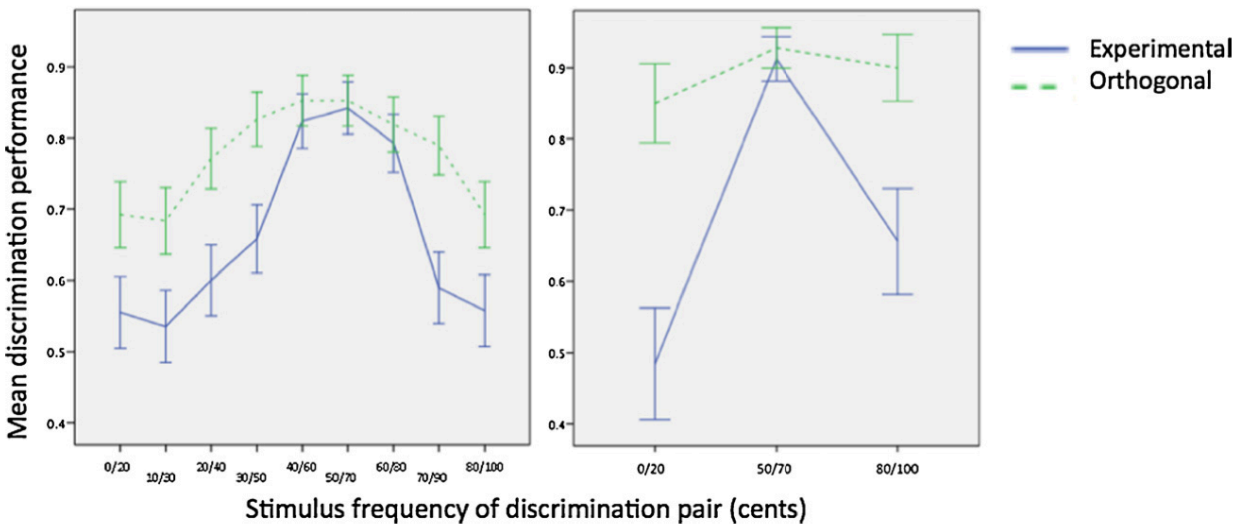

Figure 2.4: Discrimination performance

Discrimination performance from pre-test (left) and scanner session (right). Discrimination scores are out of 1 (100% accuracy). X-axis shows position in frequency space of triads within a given sound set. X-axis values represent cents above a minor triad as determined by the middle note (experimental) or cents of each of the three notes above the lowest-frequency triad (orthogonal). Error bars show 95% confidence intervals.

In-scanner discrimination data (*see Figure 2.4*) are from the final cohort of 10 subjects. A two-way ANOVA showed a significant interaction effect between sound set and stimulus frequency ($F = 29.385$, $p < 0.001$). Tukey's honestly significant difference (HSD) post hoc tests showed a significant difference in discrimination performance of the experimental vs. orthogonal triads at frequency pairs of 0/20 as well as 80/100 cents ($p < 0.05$ for both), but not at 50/70 cents. Once again, to confirm that the experimental triads were being perceived in a categorical-like manner, Tukey's HSD post hoc tests showed that peak discrimination performance of this sound set (50/70 comparison, 91% accuracy) was significantly higher than at the 0/20 (48% accuracy, $p < 0.05$) and 80/100 (66% accuracy, $p < 0.05$) endpoints. Unlike in the pre-test, in-scanner performance at the 0/20 endpoint was significantly below performance at the 80/100 endpoint ($p < 0.05$). These 10 subjects did not show a similar performance pattern during the pre-test, discriminating experimental triads at 91% (50/70), 59% (0/20), and 51% (80/100). The orthogonal triads were discriminated in-scanner with a peak accuracy of 93% (50/70) and endpoint accuracies of 85% and 90% (0/20 and 80/100, respectively).

2.5.2 fMRI results

We analyzed a total of six contrasts: three for each experimental paradigm. The contrasts were chosen to employ as much parallelism as possible between the two paradigms. However, it is important to note that certain elements do not exactly translate across the experiments. The adaptation paradigm primarily looked at oddball-related habituation effects across the two sound sets. The discrimination paradigm was more closely tied to an active behavior. Relatedly, because the observed in-scanner discrimination of major-category (but not minor-category) triads was better than what was expected based upon pre-test behavioral results, we chose to focus on the minor- and between-category discrimination pairs for this second paradigm. This was done with the intent of maximizing the chances of observing BOLD activity related to CP, which was the primary goal of the study. Separately, in order to complement the right STS sub-analysis

described in the Section 2, a similar region-specific analysis was conducted in the left STS, although we did not predict activity in the latter area. No significant peaks were observed using the same threshold criteria ($p < 0.001$ uncorrected).

Adaptation paradigm: The first contrast from our adaptation paradigm (Adapt1) compared BOLD activity from all oddball experimental trials with repeating experimental trials, after subtraction of the analogous orthogonal volumes: [[EXPCHG – EXPREP] – [ORTCHG – ORTREP]]. A significant peak was found in the right aSTS ($x = 60$, $y = 4$, $z = -8$; $t = 5.66$, *see Figure 2.5*).

The second contrast (Adapt2) looked at BOLD activity following EXPCHG stimuli that crossed the minor/major categorical boundary (i.e. the 50/70 pair, between-category: "BW") minus the analogous ORTCHG trials: [EXPCHG-50/70 – ORTCHG-50/70]. This contrast showed a peak that was significant at the whole-brain level in the left intraparietal sulcus (IPS)/ inferior parietal lobule ($x = -44$, $y = -56$, $z = 50$; $t = 4.60$). A sub-threshold peak in a similar right-hemispheric region was also observed ($x = 52$, $y = -46$, $z = 44$; $t = 3.34$).

The third adaptation paradigm contrast (Adapt3) looked at between- and within-category experimental oddball conditions: [EXPCHG-50/70 – EXPCHG-0/20, 80/100]. No significant peaks were observed. This contrast was primarily conducted for congruence with Disc3 (below), a main contrast from the discrimination experiment.

Discrimination paradigm: Disc1, a discrimination paradigm contrast that was employed to parallel Adapt1, did not show any significant peaks. This contrast compared activity following discrimination of all experimental triads with that of activity following discrimination of all orthogonal triads: [EXP – ORT].

Disc2, which was constructed to parallel Adapt2, did not yield any significant peaks. While Adapt2 compared 50/70 oddballs across the two sound sets, Disc2 simply compared BOLD

activity following discrimination of the experimental and orthogonal 50/70 pairs: [EXP50/70 – ORT50/70].

The primary discrimination contrast, Disc3, compared between-category and within-category (minor) conditions ([EXP50/70 –EXP0/20]) and showed a significant peak within the right middle/posterior STS (x = 44, y = –26, z = –4; t = 3.39, significant via right STS sub-analysis, *see Figure 2.5*). We also note the presence of a large, though sub-threshold, peak nearby in the right STG (x = 50, y = –26, z = 14; t = 4.31).

Region	*x*	*y*	*z*	*t*	Contrast
Right STS	60	4	−8	5.66	Adapt1
Cerebellum	−44	−48	−40	4.75	Adapt2
Left occipital	−20	−92	30	4.62	Adapt1
Left IPS/inferior parietal lobule	−44	−56	50	4.60	Adapt2
Right STS*	44	−26	−4	3.39	Disc3

Table 2.1: Peak BOLD effects

All peaks are significant at the whole-brain level ($p < 0.05$, corrected), except for the second right STS peak. * Observation of statistical significance via anatomically segmented right STS region-based analysis ($p < 0.001$ uncorrected).

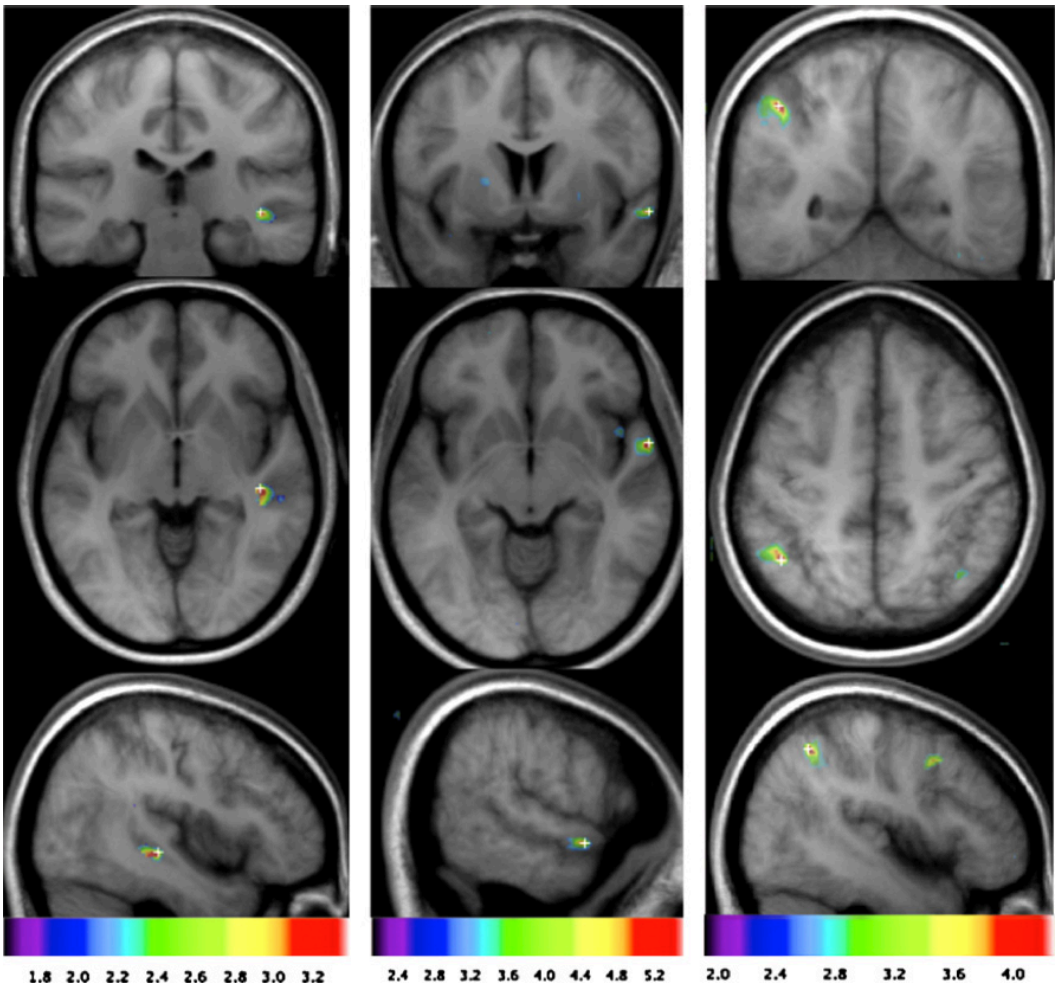

Figure 2.5: BOLD peaks

Contrast Disc3 (left) from our discrimination protocol (right STS sub-analysis) compares BOLD activity following discrimination of between-category experimental triads minus discrimination of within-category (minor) triads (EXP50/70 – EXP0/20). This comparison is meant to isolate activity arising following presentation of multiple categories (i.e. minor and major) vs. a single category. A peak (t = 3.39, right STS sub-analysis) was observed in the right middle/posterior STS (x = 44, y = –26, z = –4). Contrast Adapt1 (centre) from our adaptation protocol compared BOLD activity following presentation of all experimental oddball trials (EXP-CHG) with non-oddball trials (EXP-REP), after subtraction of similar volumes from the orthogonal sound set ((ORT-CHG) – (ORT-REP)). This comparison is meant to isolate a rebound from adaptation, but only when such a rebound taps into neural substrates that contain category information. A peak (t = 5.66) was observed in the right aSTS (x = 60, y = 4, z = –8). Contrast Adapt2 (right) compared BOLD activity following presentation of boundary spanning experimental oddball trials (EXP-CHG50/70) with the analogous orthogonal volumes (ORT-CHG50/70). This comparison is also meant to isolate a rebound from adaptation, but only when associated with a second and distinct musical category. A peak (t = 4.60) was observed in the left IPS/ inferior parietal lobule (x = –44, y = –56, z = 50). All anatomical underlays are from the nonlinearly registered average of the 10 subjects tested.

2.6 Discussion

2.6.1 Behavioral performance

Overall behavioral performance of subjects, observed both during the pre-test and in the scanner, yielded data that show all the signs of classic CP functions. This categorical effect was much stronger for the experimental than the orthogonal triads, suggesting that the latter successfully functioned as an appropriate control. Identification functions for the experimental and orthogonal triad sets showed a significant interaction effect, with subsequent post hoc tests indicating that the differences came primarily from the center of the plots. Mean identification ratings at 40 and 50 cents were significantly closer to the low anchor for the experimental vs. the orthogonal triads. The opposite was true at 60 and 70 cents, suggesting the experimental function showed more of the "quick transition" that is hallmark of a boundary region between categories. Subjects were required to respond in terms of a triad's "closeness" to one anchor vs. the other based on a rating scale. The orthogonal identification ratings, while less categorical than the experimental, did not take the form of a perfectly linear function as triads increased in frequency. We believe that this finding reflects anchoring effects, which likely are due either to a response bias (i.e. subjects' tendency not to respond as "unsure") and/or perceptual factors involving auditory memory or volatility of the mental representations of the anchor sounds (Acker et al., 1995). Regardless of any such effects, CP was demonstrably stronger in the experimental identification function, thus providing evidence that we were using a proper orthogonal control.

Discrimination data confirmed the findings from identification. Although we used n = 10 for our in-scanner task, data were reported from all 35 pre-test subjects in order to show that observed CP effects were general to our entire sample of musicians. In order to best distinguish the neural substrates of CP, the 10 best-performing subjects were scanned and analyzed, and in-scanner discrimination data from these subjects were also reported. Both data sets showed a peaked, CP-like function for the experimental sounds and less CP-like functions for the

orthogonal sounds: a result that echoed our identification findings. The experimental pre-test function showed within-category performances slightly above chance (56% accuracy), with the performance peak at the 50/70-cent comparison (84% accuracy). This peak accuracy was almost identical to that from the orthogonal stimulus function (85%), which also occurred at the 50/70-cent comparison. As with identification, the orthogonal plot does not appear as a purely continuous perceptual function, which in this case would be a flat line. Instead, it contains endpoint troughs, which likely are due to the same anchoring effects spoken about above. It is of note that the discrimination peaks of the two sound sets (91% and 93% for experimental and orthogonal, respectively) are almost identical, suggesting that any BOLD differences observed when contrasting these two conditions are likely not a performance effect of the behavioral task.

The in-scanner behavioral functions follow the same general pattern as those from the pre-test, with certain differences. First, the three orthogonal triad pairs were discriminated with more consistent (and higher) accuracy than during the pre-test, which is likely an effect of practice/exposure. This same flattening of the function was not observed for the experimental stimuli, which appear to have been perceived even more categorically during the scanner session. Both of these points speak to a likely dominance of category-based processing: in other words, task-based short-term practice effects could not compete with over-learned CP, which has been acquired throughout participants' entire lifetimes. While some degree of the performance increase from pre-test to scanner may be due to subjects being tested on only 6 triad pairs in the latter sessions (a subset of the 18 pre-test pairs), this alone cannot fully explain the differential changes observed between the experimental vs. orthogonal sound sets. A final difference was a performance imbalance between discrimination of triads taken from the minor and major ends of the continuum (48% and 66%, respectively), which had been discriminated at essentially identical rates by the n = 35 population at pre-test (56% for both). As stated in the results section, this was not due to an issue with the n = 10 subsample, which actually showed the reverse performance trend during the pre-test (59% for minor, 51% for major). Because this last finding was both unexpected and difficult to explain, we felt it appropriate to use only minor-category

fMRI trials for discrimination protocol contrasts, as our main intent was to measure the neural correlates of CP by comparing clear within- vs. between-category conditions. Despite these small differences between pre-test and scanner session data, we feel, as with identification, that the discrimination results as a whole confirm that CP effects for the experimental triads were demonstrably stronger, providing additional evidence that the orthogonal triads functioned as a proper control for use in imaging contrasts.

2.6.2 Right temporal activity

In the present study, our goal was to test whether regions in the right STS are preferentially active for stimuli containing more musical category information. As predicted, the right STS showed such BOLD responses, which were present across both of our experimental paradigms.

The first adaptation paradigm contrast (Adapt1) elicited a large BOLD peak in the aSTS and Disc3 showed a significant peak in the middle/posterior right STS (*see Figure 2.5*; latter peak assessed via the location-based analysis). Taken together, the peaks elicited across both experimental paradigms suggest that observed activity in this right temporal region is a real effect. The large anterior peak is located in a position that is roughly symmetrical to the more anterior of two left STS peaks from Liebenthal et al. (2005) (x = –60, y = –8, z = –3). Liebenthal et al. compared BOLD activity following discrimination judgments of phonemes against a warped, acoustically matched set of non-speech-like sounds. Like the contrast used by Liebenthal, et al., Adapt1 compared both within- and between-category experimental stimuli against stimuli from an orthogonal control condition. Likewise, the more posterior right STS peak shows general correspondence with those of Liebenthal et al. (x = –56, y = –31, z = 3) as well as Joanisse et al. (2007)(x = –66, y = –26, z = 7 and x = –64, y = –25, z = –7) (n.b. Peak locations listed for Liebenthal et al. and Joanisse et al. are in Talairach coordinates, though discrepancy from MNI coordinates are minor).

Liebenthal et al. have proposed that phonemic recoding may be the earliest kind of speech processing that is truly lateralized to the left temporal lobe. Liebenthal et al.'s and Joanisse et al.'s phonemic CP results provide evidence that the middle/anterior left superior temporal region is where this recoding takes place, a conclusion that has been supported by other imaging studies of phonemic perception (Hutchison, Blumstein, & Myers, 2008; Obleser, Zimmermann, Van Meter, & Rauschecker, 2007b). The left pSTS training effect observed by Leech et al. (2009) (x = –54, y = –37, z = –1), which dealt exclusively with temporally-complex non-speech sounds, indicates that this left hemispheric specialization may be more general in nature. Looking to the more anterior STS, studies have implicated both the left and bilateral aSTS in higher-order speech processes that contribute to phrase- or sentence-level comprehension (e.g. phonetic, semantic, syntactic) (Davis & Johnsrude, 2003; Humphries, Willard, Buchsbaum, & Hickok, 2001; Narain et al., 2003). These ultra-phonemic processes, which lie farther down the putative "ventral stream," are also likely making use of certain types of speech category information (e.g. noun vs. verb). Our results, which contrast (a) category-containing stimuli against stimuli perceived significantly less categorically, as well as (b) between-category stimuli against within-category stimuli, show analogous right hemispheric activity to the left temporal peaks of the speech literature. As our control stimuli were selected to be well-matched for spectral complexity, we believe that the observed right STS BOLD signals are truly reflective of pitch-based categorical processing, which extends prior findings that show a more general right auditory cortex bias for fine-grained spectral processing (Hyde et al., 2008; Zatorre & Belin, 2001).

The ventral and dorsal streams make up the individual components of the "two-stream hypothesis" that was originally put forward by Mishkin & Ungerleider (1982). The theory was initially formulated with respect to the visual system and argued for a ventral "what" pathway that handles identification of objects, as well as a dorsal "where" pathway that deals with objects' locations in space. As part of the hypothesis, the ventral and dorsal streams are thought

to be primarily mediated by the temporal and parietal lobes, respectively, with more abstract representations of objects existing further from primary sensory areas. This theory has been extended to the auditory domain (Rauschecker & Tian, 2000), with more recent two-pathway models involving abstraction beyond simple what vs. where components to encompass sensory-motor aspects of processing (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009). Sensitivity to features of auditory objects has been linked to antero-ventral areas of right temporal cortex (i.e. ventral stream) (Zatorre, Bouffard, & Belin, 2004) and, generally, the category-centric exploration of phoneme identification/discrimination and resultant left STS findings fall under the broad heading of "ventral stream."

The right STS activity observed in our discrimination paradigm may to some degree represent higher neural processing demands following exposure to a greater number of categories, as it was observed following discrimination of boundary-spanning triad pairs (2 categories), after contrasting with within-category minor pairs (1 category). However, employing an active discrimination task raises the possibility that the observed STS activity may reflect task-related use of any categorical information, as opposed to "pure" category percepts, themselves. This issue was addressed via our adaptation paradigm, where subjects were not instructed to judge sounds for category/pitch quality. The Adapt1 contrast, which yielded the large right aSTS peak, grouped together 1- and 2-category experimental triad pairs, which were then compared with all orthogonal pairs. We note that the two paradigms each have different degrees of memory load and attentional requirements. In the discrimination task, subjects paid more explicit attention to the experimentally-relevant features of the triads, though they were not instructed to listen specifically for the "quality" of sounds (merely to compare/choose among them). While the orthogonal AAAAX task (related to loudness) was easier and required different and likely fewer attentional processes, it ensured that subjects' focus was still on the auditory modality. Regarding memory load, performance of both tasks likely utilized working memory as well as echoic memory. If there were no musical categories, the ABX task could be performed via echoic memory, without any need to remember A (i.e. B either matches X or does not match

X). For within-category comparisons, the most successful strategy likely involves a shift in focus toward sensory memory as soon as B is heard (with the opposite being true of between-category comparisons). While the discrimination task is the more demanding of the two, both tasks, in a sense, really only require one triad to be "kept in mind" prior to presentation of X, with such tracking likely involving a blend of memory-types.

It is of note that the Adapt1 adaptation contrast compared oddball and repeating trials, after subtraction of the orthogonal from experimental volumes. Based on the behavioral data, the orthogonal triad pairs were even more discriminable than the experimental pairs, so it is improbable that participants simply could not perceive the orthogonal oddball ("change") trials as sounding different from repeating trials. It may be the case that observed anterior activity follows equally from single- and multi-category stimuli, but is less related to non-categorizable stimuli. This hypothesis could explain the lack of such an anterior peak in the Disc3 discrimination contrast, which did not use a control from the less-categorically-perceived sound set. It is of note that Liebenthal et al.'s results, which include both middle/posterior and middle/anterior STS peaks, were also from a contrast of both 1- and 2-category experimental stimuli against category-free orthogonal stimuli. A second possibility is that the aSTS may be involved in combining category information relayed from the middle/posterior STS with pre-categorical auditory information, thus making it most sensitive to changes that are specific to already-binned objects.

We believe that the sum of these results provide evidence for a role of the right STS in perception of spectrally-complex auditory categories. As mentioned in Section 1, while these specific results do not generalize beyond subjects with musical training who show strong behavioral CP traits, they do suggest a predisposition of the right STS to take on a larger role than the left. We feel that, most likely, the functional results presented here arise via a combination of a specialization of right temporal lobe, present in a large proportion of the general population, and a specific sort of training/learning that capitalizes on this hemispheric bias. Questions remain, including the degree to which temporal regions respond to single vs.

multiple categories, as well as the degree to which category representations are distinct or overlap with one another. Taken as a whole, the body of literature strongly suggests that bilateral ventral streams, and more specifically the left and right STS, underlie auditory categorical perception. However, observation of auditory category-related BOLD activity seems to be a subtle phenomenon, with some studies yielding significant peaks only via a large number of participants and a subset of contrasts (e.g. Liebenthal et al. scanned 25 subjects and observed significant STS activity for a phonemic vs. non-phonemic contrast, but not for a between-category vs. within-category contrast). Additionally, many auditory CP studies employ temporal lobe-ROI analyses in addition to looking at whole-brain activity (Hutchison et al., 2008; Joanisse et al., 2007). Likewise, while we observed one very clear BOLD peak in the right aSTS, the more posterior right STS peak was detected using a relatively liberal threshold for significance. However, our STS peaks show general right/left location correspondence to those from the speech literature. It may be the case that traditional "A minus B" univariate analyses of BOLD signal will often lack the sensitivity needed to differentiate between certain closely-related auditory categories, whether they are specific to music (e.g. minor vs. major), speech (e.g. /ta/ vs. /da/), voice (male vs. female), etc. Recently, there has been a movement toward using multivariate information-based approaches to the localization of brain function. By looking at multiple neighboring voxels simultaneously, a "searchlight" of the brain may determine whether regionally-specific activity patterns can successfully predict and classify future events (Kriegeskorte et al., 2006). It follows that categorical maps, while distributed beyond individual voxels, may still be localizable to anatomically distinct regions (Staeren, Renvall, De Martino, Goebel, & Formisano, 2009). The study by Staeren et al. showed that activity in bilateral STS regions could be used as an effective predictor of both auditory object category (e.g. cat vs. guitar sounds) and fundamental frequency, with a significant degree of regional overlap between these two independent variables. These classifier-based results provide further evidence for a pivotal role of the STS in perception of category, while also suggesting that observation of

distributed patterns of activity, though still regionally local, may be critical to the identification of more detailed and precise category maps.

2.6.3 Intraparietal sulcus

Bilateral activity in the IPS was observed in the second adaptation paradigm contrast, Adapt2, which compared oddball stimuli that crossed the minor/major boundary and the analogous oddballs from the orthogonal set. This was not a result that we had predicted: neither Liebenthal et al.'s nor Joanisse et al.'s phoneme studies had reported significant BOLD activity in either IPS. This region deserves additional examination with regard to what role it may be playing in CP of musical stimuli. The IPS is part of what has classically been considered the "dorsal stream" (Culham & Kanwisher, 2001). Some recent studies have suggested that the IPS may play a large role in dealing with the frequency relationships between stimuli. Rinne et al. (2007) observed IPS recruitment to large pitch shifts in sound discrimination tasks. Zarate & Zatorre (2008) and Zarate, Wood, & Zatorre (2010) showed that the IPS may play a major role in auditory feedback monitoring for vocal regulation following pitch-shifts and, additionally, may interact with the right pSTS to extract the directionality of such a pitch-shift. Another recent study (Foster & Zatorre, 2010) showed that performance of a task that involved transposition of melodies correlated with BOLD activity in the right IPS. This latter finding points to a role of the IPS in the cognition of relative pitch. Since interval categories are based upon frequency ratio relationships (and not the absolute frequency distance between two notes), it would follow that CP for chords may preferentially recruit neural networks that make use of interval "quality." In other words, the IPS may be recruited when comparing stimuli that differ in interval type (minor vs. major), but may not be utilized to as great an extent when such a quality is missing (e.g. in our orthogonal triads that differ in terms of absolute pitch space, but not in terms of "minor-" or "major-ness"). The above contrast from the adaptation protocol, which compares major/minor

and orthogonal triad pairs of approximately equal discriminability (based on behavioral data), provides evidence for such recruitment. It is of note that the discrimination paradigm contrast, Disc3, which does not compare relative vs. absolute pitch conditions, lacks significant BOLD activity in either IPS. Thus it may be the case that the IPS is preferentially recruited to help manipulate musical category information, but is relatively less sensitive to which particular category or categories are present at any given time. Musical categories, including chords (minor, major, etc.) and intervals (3rd, 4th, 5th, etc.), differ along a spectrum that has a dimension of perceptual "size" (e.g. a 5th is perceived as being a larger interval than a 3rd). On the contrary, phonemes are not intuitively thought of in terms of size, or any other linear dimension (i.e. /ta/ cannot be thought of as larger than /da/) and hence lack inherent underlying ordering. The absence of analogy, in this particular dimension, between musical and phonemic categories may explain the lack of observed IPS activity in prior studies of speech categorization.

2.6.4 Conclusion

The present data provide evidence for the involvement of the right STS in CP of spectrally-complex auditory stimuli. The results support models of hemispheric specialization for differential spectral resolution, as well as the role of a ventral stream as the basis of CP of numerous stimulus types.

2.7 Acknowledgements

We thank Marc Bouffard for technical assistance, and Dr. Jeffrey Binder for ideas about the control condition, as well as the staff of the McConnell Brain Imaging Centre for help in acquiring the data. This work was supported by funding from the Canadian Institutes of Health Research.

Chapter 3 - Representations of invariant musical categories are decodable by pattern analysis in superior temporal and intraparietal sulci

Klein ME, Zatorre RJ. 2014. Representations of invariant musical categories are decodable by pattern analysis of locally distributed BOLD responses in superior temporal and intraparietal sulci. *Cereb Cortex*. doi:10.1093/cercor/bhu003

3.1 Preface

This chapter describes a study conducted to follow-up the results of Study 1 by linking categorical perception (CP) more directly to automatic perceptual responses in the brain. Additionally, while the first experiment demonstrated that music CP *activated* regions in both ventral and dorsal streams of cortical processing, Study 2 was aimed at testing for the presence and location of category specific *information* in the cortex. To achieve these aims, we employed multivariate pattern analyses (MVPA), which utilize fine-grained differences in the spatial patterns of fMRI BOLD responses to "decode" between perceptual states. As in the first experiment, we continued to enroll expert musicians who, via psychophysical measurements, were shown to demonstrate robust CP for musical intervals. We hypothesized that the ventral and dorsal regions highlighted in Study 1 would contain information allowing for the decoding of musical categories (e.g. minor from major thirds), but not for sounds that varied according to non-categorical parameters (e.g. in absolute pitch space). This manuscript was published in a 2014 issue of *Cerebral Cortex* (Klein ME, Zatorre RJ. 2014. Representations of invariant musical categories are decodable by pattern analysis of locally distributed BOLD responses in superior temporal and intraparietal sulci. Cereb Cortex. doi:10.1093/cercor/bhu003).

3.2 Abstract

In categorical perception (CP), continuous physical signals are mapped to discrete perceptual bins: mental categories not found in the physical world. CP has been demonstrated across multiple sensory modalities and, in audition, for certain over-learned speech and musical sounds. The neural basis of auditory CP, however, remains ambiguous, including its robustness in nonspeech processes and the relative roles of left/right hemispheres; primary/non-primary cortices; and ventral/dorsal perceptual processing streams. Here, highly trained musicians listened to 2-tone musical intervals, which they perceive categorically while undergoing

functional magnetic resonance imaging. Multivariate pattern analyses were performed after grouping sounds by interval quality (determined by frequency ratio between tones) or pitch height (perceived noncategorically, frequency ratios remain constant). Distributed activity patterns in spheres of voxels were used to determine sound sample identities. For intervals, significant decoding accuracy was observed in the right superior temporal and left intraparietal sulci, with smaller peaks observed homologously in contralateral hemispheres. For pitch height, no significant decoding accuracy was observed, consistent with the non-CP of this dimension. These results suggest that similar mechanisms are operative for nonspeech categories as for speech; espouse roles for 2 segregated processing streams; and support hierarchical processing models for CP.

3.3 Introduction

An overarching feature of perception is the awareness of stimuli as "whole" objects, rather than complex amalgams of ambiguous physical signals. A specific aspect of this phenomenon occurs for certain classes of stimuli that are subject to "categorical perception" (CP), whereby continuous physical signals are mapped onto discrete mental categories, mediated by long-term memory. CP was first behaviorally demonstrated in speech perception (Liberman et al., 1957) and later in nonspeech and non-auditory domains, including perception of musical intervals (Burns & Ward, 1978; Zatorre & Halpern, 1979) and color (Bornstein & Korda, 1984), implicating it as a more general phenomenon. The neural substrates of CP remain unclear, but increasing evidence indicates that it may be mediated by 2 dissociable streams of information processing: (1) A more perceptual ventral system focused on object identification/recognition and (2) a dorsal system related to motor production, with requisite linkages to the premotor/motor system (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009).

Over the past decade, functional neuroimaging studies of CP have implicated subregions of the left superior temporal sulcus (STS) (Joanisse et al., 2007; Leech et al., 2009; Liebenthal et al., 2005), thought to be part of a ventral stream, as well as portions of the posterior superior temporal gyri (STG) and left parietal and frontal lobes, thought to be nodes in a motor-related dorsal stream (Hutchison et al., 2008; Myers et al., 2009; Raizada & Poldrack, 2007). Most of these studies, however, have employed speech (or speech-like) stimuli, leading to what may be an overgeneralization of the predominantly left hemispheric results. A study examining blood oxygen level-dependent (BOLD) responses to categorically perceived musical intervals implicated the right STS and left intraparietal sulcus (IPS) (Klein & Zatorre, 2011), indicating that these cortical streams may also be recruited for nonspeech categorical processing. The wide variety of intra- and extra-STS peaks is likely due in part to design choices (specific in-scanner experimental tasks, control conditions, and contrasts), leading to differences in networks observed for any one task/contrast (a situation complicated by the range of sensitivity available via univariate and multivariate analysis methodologies). This literature, and the resultant interpretation of imaging results, is further complicated by the strictness with which true CP is behaviorally defined; many studies report data for identification, but not discrimination tasks, while the latter is the only way to ascertain that the processing of category information in some way dominates perception (Repp, 1984). Thus, while evidence has begun to mount implicating the STS in categorical processing, the totality of the neural circuitry underlying both speech and nonspeech auditory category perception remains an open question.

To examine the neural basis of nonspeech auditory CP while minimizing potential confounds due to the nature of tasks and control stimuli, we utilized multivariate pattern analyses (MVPAs), which consider data from spatially distributed patterns of brain activity to differentiate between experimental conditions (Haynes & Rees, 2006; Mur et al., 2008; Pereira et al., 2009). MVPA's enhanced sensitivity over univariate General Linear Model (GLM) analyses allows for (a) comparison between "sibling" conditions of interest from the same underlying continuum, as

opposed to use of "null" conditions lacking some essential quality (e.g. direct comparison of 2 speech phonemes without the need for acoustically matched controls that are not perceived as speech sounds) and (b) utilization of fairly passive scanning protocols, free of major behavioral task requirements. Using a local pattern analysis "searchlight" approach (Kriegeskorte et al., 2006), we sought to distinguish between brain regions carrying decodable information about the categorical quality of musical intervals from any regions underlying non-categorical processing of pitch height. Compared with speech stimuli, musical intervals are nonlinguistic, acoustically simple, and allow for experimental and orthogonal differentiability based on the same feature (tone frequency). Thus, the use of musical intervals allows for the possibility to dissociate bottom-up, absolute pitch-based effects (present in both stimuli dimensions in roughly equal quantity) from top-down, categorical memory-based effects (present in the interval quality—but not the absolute frequency— dimension).

Unlike prior imaging studies of CP, we employed a combination of (a) behavioral identification and discrimination tasks to be certain that true CP was demonstrated; (b) 3 categories per continuum, in order to be certain that observations were not due to anchoring/range effects (Simon, 1978); and (c) an orthogonal control dimension, which circumvent confounds due to differences in the physical features of stimuli. Because analyses decoding only single exemplars of musical intervals would not allow us to dissociate which component of the results were due to categorical differences as opposed to acoustic differences between the sounds, the classifiers were trained and tested on multiple exemplars of each interval varying in absolute pitch (i.e. roved in the orthogonal dimension), and these MVPA results were compared with those from the orthogonal analysis based on the pitch height dimension, which was not predicted to be categorically perceived. Classification of categorical qualities was hypothesized to occur in the superior temporal and intraparietal sulci, with successful pitch height decoding predicted in the STG.

3.4 Materials and methods

3.4.1 Study participants

We recruited 37 trained musician participants (22 females, minimum 5 years formal training and currently practicing or performing); the majority of whom came from McGill University's undergraduate and graduate music student populations and none of whom possessed absolute pitch abilities. Of this cohort, we selected 10 participants (4 females, average 13 years of musical training, 8 instrumentalists, and 2 singers) who showed the greatest degree of CP, as determined by discrimination task performance (see "pre-scanning behavioral tasks" below). All participants gave their informed consent. Ethical approval was granted by the Montreal Neurological Institute Ethics Review Board.

3.4.2 Pretest sound stimuli

Each experimental stimulus was composed of a 2-tone melodic (i.e. sequential) interval. Each 750-ms complex tone was synthesized in Audacity and Max/MSP software out of 5 harmonics with amplitudes inversely proportional to the harmonic number. A volume envelope was applied (initial 50 ms ramp from 0% to 100% and final 50 ms ramp from 100% to 0%) in order to avoid onset and offset percussive clicks, and sound intensity was adjusted to each subject's comfort level. The two 750-ms tones in a given interval were separated by a 500-ms silent gap, resulting in 2000-ms long intervals (only 1500 ms of which contained sound). The second tone always the higher-pitched of the two.

A musical interval in common Western musical practice is defined by the frequency ratio (measured in terms of a logarithmic frequency variable termed "cents") between its constituent tones, rather than by the absolute frequencies of the tones. This feature allows us to construct intervals that are invariant in the category they belong to, but are made from tones with different

frequencies. The stimulus set we constructed thus varied along 2 orthogonal dimensions. In the first dimension ("interval quality"), the frequency ratio between the higher- and lower-pitched tones varied, with ratios derived from equally tempered semitones (in which each 100 cents corresponds to a semitone, and the 3 intervals we used, minor third, major third, and perfect fourth, correspond to 300, 400, and 500 cents, respectively). These values ranged from 287.5 to 512.5 cents, with stimuli generated at 12.5-cent increments (*see Figure 3.1*). This range spanned and included minor thirds, major thirds, and perfect fourths, all of which are common and important intervals in Western music.

In the second ("pitch height") dimension, which is orthogonal to the first, the frequency values of the intervals were roved in absolute pitch space (e.g. a 400-cent major third can be generated with base notes of C-natural, C-sharp, mistuned notes between C-natural and C-sharp, or any other frequency). Thus, without affecting the quality of the intervals along a minor third <-> major third <-> perfect fourth dimension, intervals were generated with base notes that varied from 259.7 Hz (slightly below middle C) to 295.8 Hz (slightly above the D 2 semitones above middle C). The second note of each interval was then independently calculated according to whichever frequency ratio (from the interval quality dimension) we wished to implement. Thus, interval quality could be manipulated without affecting the absolute pitch of intervals, and vice versa.

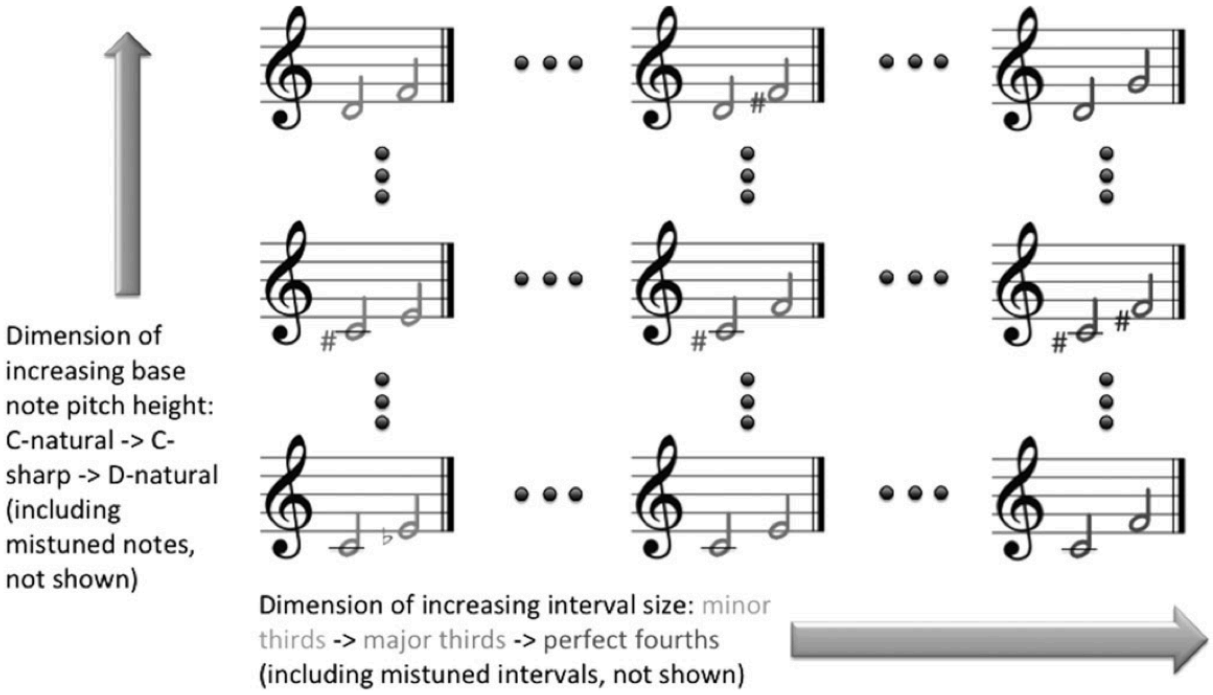

Figure 3.1: sound stimuli

Schematic of auditory stimuli used in the behavioral and imaging experiments. The imaging study used only 9 pictured stimuli, while the behavioral pretest included those 9 in addition to many additional sounds that were "mistuned" between standard frequencies and standard frequency ratios (indicated by ellipses). Movement along the x-axis indicates a change in interval size (i.e. frequency ratio between 2 notes), but no change in the pitch of base notes. Movement along the y-axis indicates a change in the pitch of both notes, but no change in the frequency ratio of an interval's 2 notes.

In general, the chosen approach to examining CP was to (a) create a set of sounds that were shown to be perceived categorically, (b) create an orthogonal extension of this first set that was acoustically well matched but not categorically perceived, (c) take the "hallmark" exemplars from each spectrum and present them within an functional magnetic resonance imaging (fMRI) paradigm, and (d) examine differences in how well machine learning algorithms were able to decode within the experimental versus the orthogonal sets. CP, specifically, was screened for in the behavioral experiment ((a) and (b)) by making use of the continuous feature space in both the interval size and pitch height dimensions. Afterwards, a subset of 9 of these sounds was used during the fMRI experiment: the 3 "true" (non-mistuned) intervals (300-cent minor third; 400-cent major third; and 500-cent perfect fourth), each of which were synthesized with base notes of exactly C-natural, C-sharp, and D-natural (3 × 3 design, *see Figure 3.1*). We chose to use an approach comparing and contrasting primary and orthogonal stimulus dimensions (interval quality vs. absolute pitch), as both could be manipulated via the same simple feature: frequency of constituent tones. Links could then be made between behavioral divergence and differing patterns of fMRI results. Three-category classification was chosen over more common 2-category experimental designs (which are often required in speech experiments due to the multidimensional nature of phoneme space) in order to: (1) generalize imaging results beyond a single pair of musical categories and (2) demonstrate behavioral CP that is clearly differentiable from anchoring/endpoint effects, mediated by short-term memory (see Hary & Massaro (1982) and Schouten (2003) for common criticisms of 2-category perceptual tasks).

3.4.3 Pre-scanning behavioral tasks

In our behavioral pretest, study participants were asked to perform a series of 4 tasks (2 identification tasks and 2 discrimination tasks). For each of these tasks, participants performed a practice run (2–5 min) to ensure that they were comfortable with the response interface and understood the instructions. For each type of task (e.g. identification, which was performed

twice), participants heard the identical set of stimuli both times, but they were asked to attend to different qualities of the sounds (e.g. "listen for interval quality" or "listen for pitch of base note"). The experiment was counter-balanced, so that half of the participants performed tasks (1) and (2) prior to (3) and (4), with the other half first performing (3) and (4).

1. Identification of interval quality.

Prior to performing the task, participants were asked to listen to a series of exemplars of each of the 3 true interval qualities. Ten examples of minor thirds were presented, all of which had 300-cent frequency ratios but varied randomly in pitch height, while the phrase "minor thirds" was displayed on the screen. This was immediately followed by 10 examples of major thirds and perfect fourths, respectively. For the task proper, participants were asked to simply assign each interval with a label by pressing a keyboard key: "j" for minor third; "k" for major third; and "l" for perfect fourth. Participants were asked to select whichever label an interval was closest to. Responses were not under time constraints, but participants were asked to make their selections as quickly as they could comfortably do so. After a response was logged, the next trial would begin after a delay of 2000 ms. For the practice run only, responses were followed by a visual displaying the participants' choice (e.g. "you selected major third") and the actual physical property of the interval (e.g. "the interval was closest to a major third"). No feedback was provided during post-practice runs. Nineteen intervals were presented in a pseudorandom order, with each interval type presented 4 times for a total of 78 trials. The pitch height for each interval was generated pseudorandomly.

2. Discrimination of interval quality.

Participants were presented with pairs of intervals and asked to judge which of the 2 intervals was "wider" (i.e. whether the first- or second-presented interval had more separation between low and high notes). This instruction therefore does not constrain the listeners' judgment with respect to the categories that they may be familiar with. Participants were instructed to press "j"

or "k" if they believed that the first- or second-presented interval met this criterion, respectively. The ratio between the 2 intervals of a trial always differed by 25 cents. Trials were balanced so that "j" and "k" were the correct responses an equal number of times, and so that the interval with the higher-pitched base note appeared first or second an equal number of times. As in (1), the intervals were presented in a pseudorandom order. The orthogonal dimension of pitch height for each interval was generated pseudorandomly, with an additional stipulation that the base notes of the 2 intervals in any one trial must differ by at least 37.5 cents in order to safeguard against the possibility of participants basing their judgments solely on the pitch of the intervals' top notes (in a situation where both intervals used identical or near-identical base notes). As in the identification task, participants first performed a practice run, where they were given visual feedback after each trial (e.g. "incorrect: you selected the first interval and the second interval was wider"). No feedback was provided during the 5 post-practice runs (each run containing 17 trials, one for each discrimination pair, presented in a pseudorandom order).

3. Identification of pitch height.

The stimuli used in this task were identical to those from (1). Participants were asked to attend not to quality of the intervals (minor, major, and perfect), but instead to the pitch of the base notes. (The 2-tone intervals were still used, but participants were instructed that they could ignore the top tone of each interval.) Prior to performing the task, participants were asked to listen to a series of exemplars of each of 3 base notes: C-natural, C-sharp, and D-natural. Ten examples of intervals with base notes of C-natural were presented, all of which had variable top notes, while the phrase "C-naturals" was displayed on the screen. This was then immediately followed by 10 examples of C-sharps and D-naturals, respectively. For the task proper, participants were asked to simply assign each presented base note with a label by pressing a keyboard key: "j" for C-natural; "k" for C-sharp; and "l" for D-natural. Participants were asked to select whichever label the presented sound was closest to. Feedback was given for a practice

run (e.g. "you selected C-sharp, the presented sound was closest to D-natural"), but not the post-practice runs. All other methods were identical to those used in (1).

4. Discrimination of pitch height.

Participants were presented with pairs of intervals and asked to judge which of the 2 intervals had a higher-pitched base note. As in (3), subjects were told that they could complete the task successfully without considering the top notes of the intervals, which were chosen pseudorandomly. As in all prior tasks, participants first performed a practice run, where they were given visual feedback after each trial (e.g. "correct: you selected the first interval and the first interval had the higher-pitched base note"). All other methods were identical to those used in (1–3).

Participants were chosen for the MRI experiment based on the degree of difference between peak and trough discrimination accuracy in task (2). Specifically, participants were screened to have an "M"-shaped interval quality discrimination function, with performance troughs near category centers (e.g. near 400 cents/"major third") and performance peaks far from these centers (e.g. near 450 cents/midway between "major third" and "perfect fourth"). This function shape, with discrimination accuracy peaks near hypothesized category boundaries, is characteristic of CP in speech and other domains (Burns & Ward, 1978; Liberman et al., 1957). Performance peaks are thought to occur when the 2 stimuli in a discrimination task pair span such a boundary, with long-term memory systems assigning "all or nothing" labels to the sounds, which perceptually diverge.

3.4.4 fMRI tasks and data acquisition

MRI volumes were acquired on a 3-T Siemens Magnetom Trio scanner. A high-resolution (voxel = 1 mm^3) T1-weighted anatomical scan was obtained for each participant. For each

functional trial, one whole-head frame of 39 contiguous T2*-weighted images was acquired in an ascending, interleaved fashion (time repetition = 9.5s, time echo = 30 ms, 64 × 64 matrix, voxel size = 3.5 mm isotropic), yielding a total of up to 351 BOLD volumes per subject (9 runs × 39 volumes/run). fMRI scanning was performed via a sparse temporal sampling protocol (Belin, Zatorre, Hoge, Evans, & Pike, 1999), where each trial consisted of 2000 ms of data acquisition that followed 7500 ms of relative quiet. In 90% of trials, a single melodic interval was presented 3 times for a total of 6 tones during this quiet time period, with each 750 ms tone followed by 500 ms of silence, and 250 ms of silence bookending the initial and final tones. Unlike the behavioral pretest, which utilized pitches that were mistuned between standard notes and ratios that were mistuned between semitones, the MRI protocol employed only intervals that started on 3 standard base notes ("middle" C natural, C sharp, and D natural) and used 3 standard interval ratios (300-cent minor thirds, 400-cent major thirds, and 500-cent perfect fourths). This 3 × 3 design yielded a set of 9 unique sound samples as stimuli. Subjects were not asked to explicitly or implicitly identify intervals according to the interval quality or base note. Instead, they performed an orthogonal task in which they were asked to listen attentively and to press a response button upon hearing a trial that contained only 5 tones instead of 6 (10% of trials). Such oddball/catch trials were used as a check on attention/alertness and these imaging data were discarded. This experimental protocol was chosen above an overt identification or discrimination task in order to look at processes that occur relatively automatically.

Each functional run consisted of 39 trials (and thus generated 39 BOLD volumes). After an initial silent trial, 4 pairs of silent baseline trials (9 silent trials in total) were interspersed between 3 sets of 10 experimental trials (one trial for each of the 9 unique sound samples, and one catch trial). These 10 trials were presented in a pseudorandom order, with the main constraint being that any one interval could not follow a trial using the same interval type or base note (e.g. a major third starting on D natural could not follow a major third starting on C sharp or C natural, and could not follow a minor third starting on D natural or a perfect fourth starting on D natural). This constraint was imposed to avoid potentially confounding adaptation effects.

Nine 39-trial runs were conducted, each of which contained sounds in a unique order of presentation. Each participant underwent each of the 9 runs, with half the participants performing the runs in the opposite order from the other half. Of the 10 participants enrolled in the MRI study, 6 completed the protocol exactly as planned. For 3 of the 10 participants, one run had to be discarded due to inattention (failure to press response button for at least 2 of the run's 3 catch trials). For 1 of those 3 participants, an additional run had to be discarded due to failure to comply with the instructions. The fourth participant's data had to be discarded due to an equipment malfunction.

3.4.5 GLM analyses

A set of GLM analyses were performed in order to (1) determine cortical regions that were activated by sound (i.e. sound > silence contrast) and (2) to perform between-condition subtractions (e.g. major > minor) to compare with MVPA results. Standard GLM-based analyses were performed using FSL4's fMRI expert analysis tool (FEAT) (http://www.fmrib.ox.ac.uk/fsl/feat5/index.html). Preprocessing steps consisted of motion correction using MCFLIRT; non-brain removal using brain extraction tool (BET); and spatial smoothing using a Gaussian kernel of full-width at half-maximum (FWHM) 7.0 mm. For each analysis (interval quality or pitch height), a design matrix was generated with one predictor for each category of stimulus (e.g. in the column for "minor," an "1" was assigned for all volumes following the presentation of minor intervals and a "0" for all other volumes). As part of FEAT, native space images were registered to the Montreal Neurological Institute (MNI) space using FNIRT. Following the first-level analysis, individual subjects' runs were combined using a second-level, fixed-effects analysis. Third-level between-subjects analyses were performed using FSL's FLAME mixed-effects model. Specific one-tailed contrasts were performed twice for each of 3 condition pairs in both the interval quality (e.g. minor > major) and the pitch height (e.g. C-natural > C-sharp) analyses. Z-(Gaussianised T/F) statistic images were thresholded using

Gaussian Random Field theory-based maximum height thresholding with a (corrected) significance threshold of P = 0.05 (Worsley et al., 2002). (Note that these analyses were performed once using the entire cortical space, and a second time on a restricted region of interest (*see Section 3.4.6*) in order to provide the fairest possible comparison with MVPA results.)

3.4.6 MVPA procedures

Prior to the main analyses, motion correction was performed by realigning all BOLD images with the first frame of the first run following the T1-weighted scan (generally the fifth or sixth functional run) using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). An MVPA was performed on single-subject data in native space, prior to nonlinear registration using the MNI/ICBM152 template (performed with FSL4's FNIRT tool: http://www.fmrib.ox.ac.uk/fsl/fnirt/index.html), and a standard top-level between-subjects analysis, performed with SPM8.

The MVPAs were performed using the Python programming language's PyMVPA toolbox (Hanke et al., 2009) and LibSVM's linear support vector machine (SVM) implementation (http://www.csie.ntu. edu.tw/~cjlin/libsvm/). Each participant's runs were concatenated to form a single long 4D time series (up to 351 3D volumes). Note that no spatial smoothing/blurring was performed on the functional data prior to MVPA. A text file was generated assigning each volume a run (1–9) and a condition (minor, major, or perfect for the interval quality analysis; C-natural, C-sharp, and D-natural for the pitch analysis). Within each run, we performed (a) linear detrending to remove signal changes due to slow drift and (b) z-scoring to place voxel values within a normal range (Pereira et al., 2009). As SVMs are pairwise classifiers, we ran individual analyses on pairs of 2 conditions (e.g. minor vs. major; C-sharp vs. D-natural). The final preprocessing step was to perform temporal averaging (Mourao-Miranda, Reynaud, McGlone,

Calvert, & Brammer, 2006) on the BOLD data; we used 3 -> 1 averaging, combining three images (e.g. all perfect fourths from the first 1/3 of a functional run) into a single image. SVMs for interval quality comparisons were both trained and tested using intervals from all three pitch height classes; the reverse is true for pitch-height analyses. Classification was performed using leave-one-out cross-validation, where a classifier was trained on data from 8 of the functional runs and tested on data from the 9th, and the procedure was then repeated 8 times testing on a novel run each time. SVM classification was performed using a searchlight procedure (Kriegeskorte et al., 2006), whereby the decoding algorithm considers only voxels from a small sphere of space (radius = 3 voxels, up to 123 voxels in a sphere). (While accuracy has been shown to generally increase along with the size of searchlight spheres (Oosterhof, Wiestler, Downing, & Diedrichsen, 2011), we chose a radius of 3 voxels as a compromise between classifier performance and spatial specificity.) An accuracy score (percentage above chance (50%) that the classifier was able to successfully identify category) was calculated using an average of the 9 cross-validation folds, and this value was assigned to the center voxel of the sphere. This procedure was repeated using every brain voxel as a searchlight center (~35,000–45,000 spheres), yielding local accuracy maps for the entire brain. As the primary interest was in observing abstracted category representation (and not that of specific sound pairs), at this stage accuracy maps for each subject were averaged across the 2 pairwise classifications (i.e. minor third/major third maps were averaged with major third/perfect fourth maps). Minor third/perfect fourth classification was not performed, as these 2 stimuli sets differed more so from one another in physical and category distances (2 semitones) than the other 2 pairs (1 semitone) and would have added an additional confound to the analysis. A parallel averaging step was performed for the pitch height analysis: accuracy maps for C-natural/C-sharp discrimination were combined with those for C-sharp/D-natural discrimination. We note that certain MVPA studies that compare all possible decoding pairs (Formisano et al., 2008; Kilian-Hütten et al., 2011) often examine identity of auditory objects that have no inherent "ordinal" quality (i.e. whereas perfect

fourths are larger than major thirds, which are larger than minor thirds, voice identities differ from one another in myriad ways that are difficult to rank), and thus do not need to consider this particular confound.

Prior to performing group-level analyses, participants' brain masks were generated with FSL4. The averaged accuracy values, which served as effect sizes for the group-level analysis, were then linearly transformed into a subject's native anatomical space before being non-linearly transformed into standard space using FSL4's linear and non-linear registration tools (FLIRT and FNIRT). While there is an inherent smoothness to the searchlight MVPA procedure, at this stage we explicitly smoothed each subject's accuracy maps (a 7-mm FWHM isotropic Gaussian kernel) in order to best account for inter-subject brain variability and to perform and interpret group-level statistics. The registered and smoothed accuracy maps were then input into SPM8, which output group-level t-statistics for each voxel.

The threshold for statistical significance was set voxel-wise at $t > 7.98$ (corrected for multiple comparisons, family-wise error (FWE) < 0.05, $n = 9$). While data were collected and are presented for the entire cortex, significance testing was performed on a restricted volume in line with the a priori hypothesis of involvement of the right STS/left IPS, based on results from an earlier study (Klein & Zatorre, 2011). This mask was created off a standard MNI152 anatomical image by delimiting the full extent of the gray matter in these regions. This approach was used due to a lack of consensus in the MVPA/searchlight literature about a methodology for setting accurate group significance thresholds that are not extremely conservative (a full-brain, purely between-subjects ($n = 9$) analysis using random field theory thresholding yields a t-statistic cutoff above $t = 16$). Stated another way, there is no set method outlined for determining a "smoothed variance ratio" (Worsley et al., 2002) for these data, as is often implemented in standard GLM analyses. Because of this ambiguity and for completeness we have also reported all peaks comprised of voxels significant at $P < 0.001$ (uncorrected), with at least 10 contiguous voxels meeting this criterion. All of these peaks would generally be considered statistically significant in a standard GLM analysis (t-statistics > ~5) and, while some do not meet the very

conservative threshold used here, we believe that the nearly symmetric positioning and large spatial extents of the parietal and temporal peaks (*see Section 3.5.4*) lend weight to the argument that a substantial portion of these $t < 7.98$ results are not merely false positives.

Separately, as a check on searchlight statistical procedures, the voxels within the previously described ROI mask (left IPS and right STS, anatomically defined based on the results found in Klein & Zatorre (2011)) were also used within an "Monte Carlo" permutation test. For each subject, the identity labels for training examples were randomly scrambled and tested 1000× in order to generate subject-by-subject null distributions. These analyses yielded a single ROI decoding value for each subject (determined without label scrambling), which could be (a) compared with the subjects' null distributions to generate subject-wise P-values and (b) input into a group-level t-test. Three-category MVPAs (m3/M3/P4, 33.3% chance accuracy) were performed in order to generate and report a single P-value per subject. While the experiment was not specifically designed to test for single-subject significance (and complex feature elimination procedures were not employed), these permutation tests were performed to provide converging evidence for categorical decoding using markedly different procedures than with the primary searchlight analyses.

3.5 Results

3.5.1 Identification

Figure 3.2 shows identification functions for 3 representative subjects for both interval quality (*Figure 3.2b*) and pitch height (*Figure 3.2c*). Graphs are shown for individuals, in addition to the $n = 10$ group data (*Figure 3.2a*), as averaging necessarily obscures the sharp boundaries of the functions due to individual variability in the location of boundaries and category centers. Three obvious labeling plateaus are evident in the plot for interval quality, but not for pitch height. To quantify the degree to which participants' identification task responses

were "categorical," we first generated a "triple-plateau" function, which served as a model "perfectly" categorical response. Identification responses were recoded as 1, 2, and 3 for minor third, major third, and perfect fourth, respectively (likewise for the pitch height task). The model function was created by labeling intervals from 287.5 to 337.5 cents as "1," 350 cents as "1.5," 362.5 to 437.5 cents as "2," 450 cents as "2.5," and 462.5 to 512.5 cents as "3." The 1.5 and 2.5 values were chosen as these sound stimuli were physically exactly half way between the exemplar sound tokens. For each participant, we calculated difference scores between that participant's response function and the ideal function (one each for interval quality and pitch identification). Participants performed the interval quality identification in a significantly more categorical manner than the pitch task, as judged by proximity to the model function ($df = 36$, paired-sample 1-tailed t-test, $P = 0.00018$). While we ultimately selected our 10 MRI participants based on their discrimination task responses, this group also performed the interval quality task in a significantly more categorical manner than the screened-out cohort of 27 participants ($df = 35$, unpaired-sample 1-tailed t-test, $P = 0.0035$).

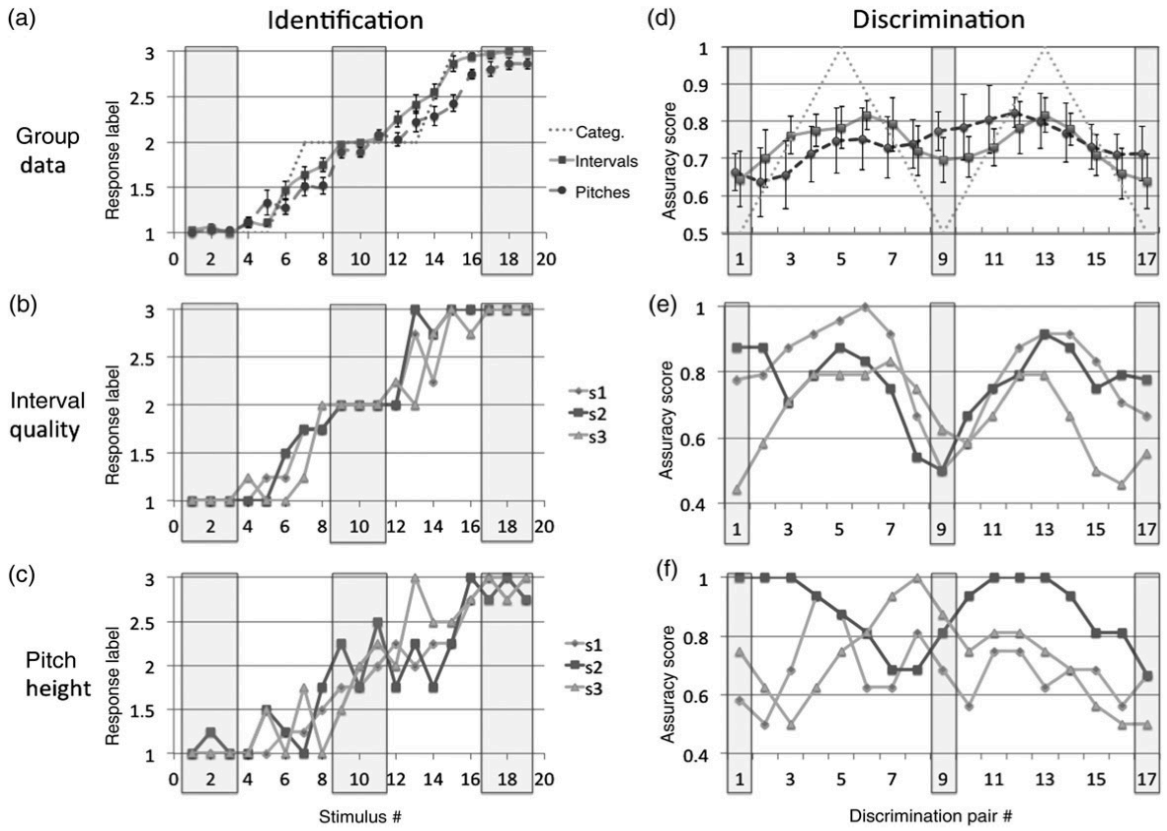

Figure 3.2: Behavioral results

Behavioral results for identification (a–c, left column) and discrimination (d–f, right column). The top row depicts n=10 group data for the subjects enrolled in the imaging experiment; the middle row depicts interval quality identification and discrimination for 3 representative subjects (s1, s2, and s3); and the bottom row depicts pitch height identification and discrimination for those same 3 subjects. For the "identification" charts, the x-axis represent the stimulus number, corresponding to musical interval size (frequency ratio) and the y-axis, the participants' averaged responses (where 1, 2, and 3 = identification as low, mid, and high tokens, respectively). Theoretical category centers for the stimuli were at x-axis positions 2, 10, and 18, which represent canonical intervals or pitch heights. As in (b), which depicts results solely from the interval quality analysis, the x-axis positions 2, 10, and 18 correspond to the interval qualities of minor third (300 cents), major third (400 cents), and perfect fourth (500 cents), respectively. In (c), which depicts results from the pitch height analysis, those same 3 x-axis positions correspond to base note pitch heights of C-natural, C-sharp, and D-natural, respectively. (a) contains results from both analyses. Approximate theoretical category centers, defined for simplicity as the standard/token intervals or pitches ±1 mistuning from the standard, are indicated in gray boxes (e.g. stimulus 17, 18, and 19, corresponding to intervals of 487.5, 500, and 512.5 cents, respectively). For the group data, theoretically perfect categorical functions (gray dotted lines) are shown alongside perceptual functions for interval quality (blue solid lines) and pitch height (red dashed lines), with error bars showing SEMs. For the "discrimination" charts, the x-axes also represent stimuli number, although these stimuli are now "pairs" rather than single intervals (e.g. stimulus 9 for the interval quality analysis depicts trials for discrimination of 387.5 vs. 412.5 cents). The y-axes are averaged accuracies of subjects' responses, where 1 indicates 100% correct and 0.5 indicates chance-level performance. For both the identification and discrimination tasks, group pitch height data deviated from the ideal categorical functions to a significantly greater degree than the interval quality data. For the individuals, clear identification labeling plateaus were observed for interval quality (b) but not for pitch height (c) and, likewise, clear M-shaped discrimination functions were observed for interval quality (e) but not for pitch height (f).

3.5.2 Discrimination

For the interval quality discrimination task, we screened for subjects showing an M-shaped function with peaks at or near theoretical categorical boundaries (e.g. 337.5 vs. 362.5 cent discrimination) and troughs at or near the categorical centers (e.g. 387.5 vs. 412.5 cent discrimination) (*see Figure 3.2d*, dotted gray line). This task proved very difficult for participants due to the variability in pitch space (i.e. while the 2 intervals in a given trial were always 25 cents apart, the base notes of those intervals could be separated by as much as 2 semitones, with an average spacing of about 1 semitone). However, a subset of our sample did show this M-shaped function [and to a significantly greater degree than for pitch height discrimination (df = 9, paired-sample 1-tailed t-test, P = 0.0058)] (*see Figure 3.2e and 3.2f* for discrimination functions of 3 individuals, presented for identical reasons as stated above). These same subjects discriminated all interval pairs near category boundaries ("between-categories") with significantly greater success than those near category centers ("within-categories") (80% correct vs. 67% correct, paired-sample 1-tailed t-test, df = 9, P = 0.0016). The same effect was not present for pitch height discrimination (78% correct vs. 72% correct, paired-sample 1-tailed t-test, df = 9, P = 0.25), with lower accuracies occurring only near the ends of the function (i.e. an "inverted U," not an M-shaped function), indicative of short-term memory/attentional-based anchoring effects. These 10 subjects were then enrolled in the functional imaging experiment.

We did not expect the majority of musicians to show a clear categorical discrimination function, due to prior research (Zatorre & Halpern, 1979), indicating a very high degree of task difficulty when using intervals with roving pitches. Our sample of 37 was screened not with an intention to generalize results to a larger population, but instead to select for those individuals demonstrating clearest evidence for CP of musical categories. While it is highly likely that the use of a non-roving lower pitch would have greatly enhanced the observable CP qualities of task performance in more subjects, it would not have allowed for simple abstraction beyond specific frequencies and note pairs.

3.5.3 GLM analysis

A contrast of all sound > silence (excluding volumes following presentation of the rare/oddball 5-tone stimuli) revealed 3 large significant clusters: (1) the right STG/STS (3242 voxels); (2) the left STG (2290 voxels); and (3) the left/right supplementary motor area (974 voxels, cluster spans the inter-hemispheric fissure, but contains more voxels in the left hemisphere). No statistically significant group activation peaks were observed anywhere in the brain for any pairwise contrast in either the interval quality or pitch height analyses, which was predicted due to the high degree of physical similarity between all 9 sound samples used in the imaging experiment.

3.5.4 Searchlight analysis

Group-level searchlight results showed significant accuracy peaks in the right STS ($t = 9.34$; $x, y, z = 48, -14, -14$) and left IPS ($t = 9.93$; $x, y, z = -30, -50, 46$; *see Figure 3.3*). No other brain regions contained voxels that surpassed $t = 7.98$. The number of contiguous voxels that passed a $P < 0.001$ uncorrected threshold were similar in these 2 regions: 66 voxels (left IPS) and 53 voxels (right STS). We also note, both because of spatial extent and approximately symmetrical locations to the 2 significant peaks, a region in the right IPS ($tmax = 5.24$; $x, y, z = 36, -54, 46$; 57 contiguous voxels at $P < 0.001$ uncorrected) and the left STS ($tmax = 5.09$; $x, y, z = -50, -14, -16$; 15 contiguous voxels at $P < 0.001$ uncorrected). No other cortical regions contained 10 or more contiguous voxels surpassing the $P < 0.001$ uncorrected threshold. No significant group-level accuracies were observed anywhere in the brain for the pitch height discrimination pairings.

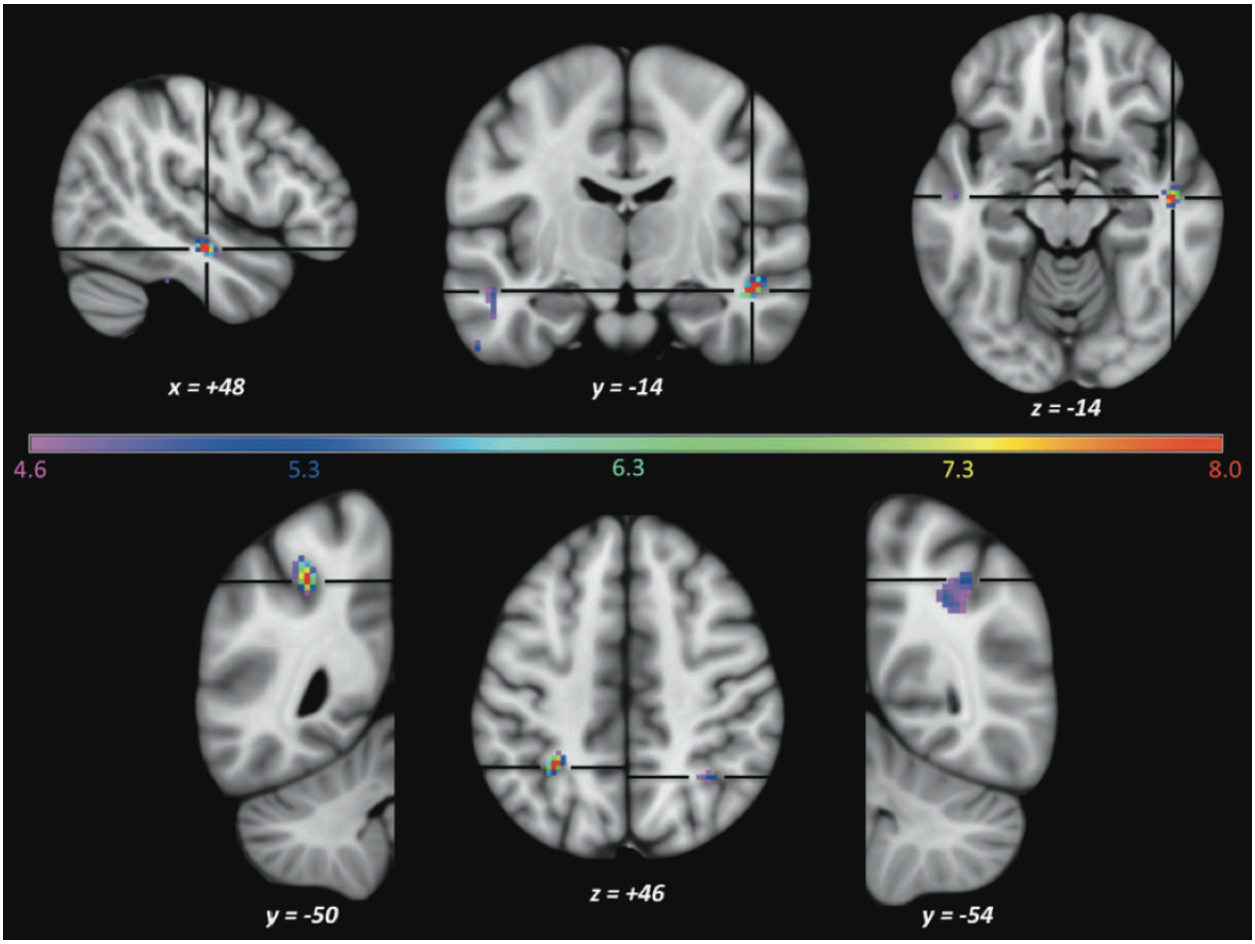

Figure 3.3: Searchlight imaging results

Group-level (n = 9) statistical peaks for the searchlight decoding analysis for interval quality, overlaid on an MNI152 0.5-mm T1 anatomical image. Colored voxels indicate t-statistics ranging from t = 4.6 (violet, P = 0.001 uncorrected) to t = 8.0 (red, P = 0.05 corrected for multiple comparisons). The top panel shows results from the right (and left) STS. The bottom panel shows results from the left and right IPS. All voxels depicted in deep red (situated in the left IPS and right STS) are statistically significant (t > 7.98).

We next examined raw classification accuracies (i.e. effect sizes) in the peak voxels of these 4 regions. Looking at 9-subject averages (chance-level accuracy = 50%), we observed accuracies of 55.8% in the right STS (individuals ranged between 53.1% and 59.1%), 56.1% in the left STS (range 49.0 – 59.7%), 57.1% in the right IPS (range 49.9–63.0%), and 55.2% in the left IPS (range 53.2 – 58.0%; *see Figure 3.4*). These individual values are presented for description only: Statistical significance testing was assessed solely via group analyses (performed naively over the entire cortical space).

We note that the larger t-values in the right STS/left IPS appear to be driven by smaller variability (rather than larger effect sizes) compared with the analogous peaks in the opposite hemispheres. The overall average decoding accuracy of all cortical searchlight spheres in all 9 subjects for the minor/ major and major/perfect discriminations was near chance at 50.5%, which suggests a combination of a chance distribution (centered at 50% correct, underlying the vast majority of spheres) and the smaller number of information-containing spheres (accuracy > 50% correct). This indicates no consistent brain-wide over-fitting in the decoding analyses, which would have led to artificially high "null" decoding averages.

Inter-subject variability gave rise to small spatial dissociations between maximum average accuracy peaks (i.e. group beta values) and statistical peaks (i.e. group t-values). In the same local neighborhoods as the statistical peaks, we observed local peaks in average classification of 57.6% in the right STS (x, y, z = +52, –2, –14); 56.9% in the left STS (–50, –10, –16), 58.5% in the right IPS (+32, –54, +42); and 57.7% in the left IPS (–34, –48, +50). For the left STS, right IPS, and left IPS, the spatial distances between beta and t-statistic peaks are very small (4, <6, and <7 mm, respectively). While this distance is somewhat larger for the right STS (12–13 mm), the maximum group average peak is still clearly positioned in the sulcus (in a more anterior position).

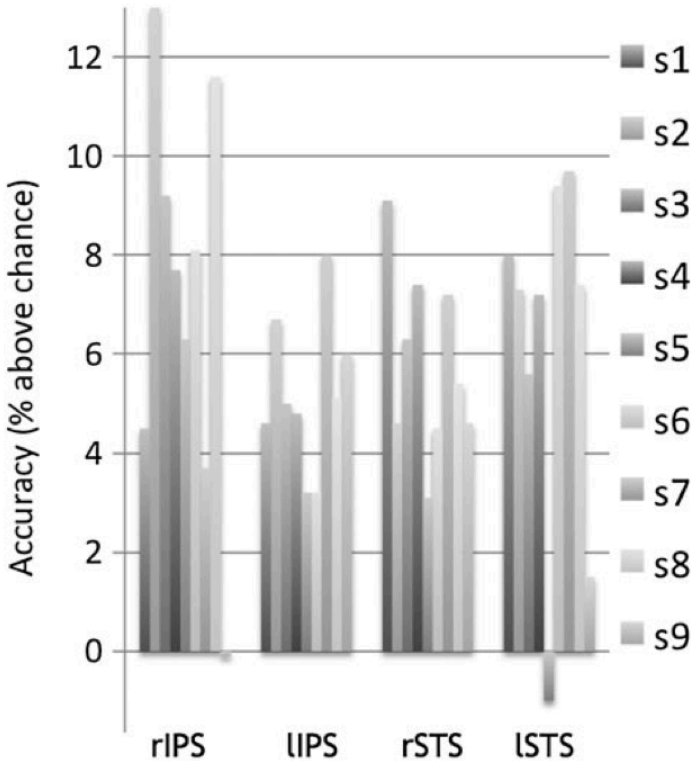

Figure 3.4

Single-sphere decoding accuracies (% above chance) for each of the 9 fMRI study participants at locations determined by group statistical peaks. (Individuals' maximally decodable spheres have higher accuracies, but variable locations.) The locations of the sphere centers in the MNI space are x, y, z = 36, –54, 46 (right IPS); x, y, z = –30, –50, 46 (left IPS); x, y, z = 48, –14, –14 (right STS); and x, y, z = –50, –14, –16 (left STS).

3.5.5 Permutation test

We observed decoding accuracies above chance for the 9 individuals at +2.9% (P = 0.139), +3.8% (P = 0.099), +4.2% (P = 0.075), +3.4% (P = 0.135), +6.6% (P = 0.020), +3.8% (P = 0.097), +2.4% (P = 0.166), –4.2% (P = 0.682), and +4.1% (P = 0.085). Even though the experimental protocol was not designed to test significance at the single-subject level (and additional feature elimination was not performed within the ROI), 8 of 9 subjects showed positive trends (P < 0.17). Furthermore, just as with the searchlight analyses, permutation testing suggested small yet consistent effects in individuals, which reached statistical significance when considered as a group. Inputting the 9 decoding accuracies into a 1-tailed single-sample t-test, we observed a significant group-level effect at P = 0.008 (degrees of freedom = 8).

3.6 Discussion

CP has repeatedly been demonstrated to be a robust behavioral phenomenon using both speech and certain non-speech auditory stimuli. However, a combination of the limitations of available experimental protocols and analytic methods, as well as a general focus on speech-specific process, has left ambiguous the identification of its full neural correlates. The sample of trained musicians presented here demonstrates behavioral CP functions for musical intervals, while MVPA of their functional brain data implicates local information-containing regions in the superior temporal and intraparietal sulci in the representation of abstract musical interval categories. The right STS and left IPS were also highlighted in an earlier study (Klein & Zatorre, 2011), despite the use of dramatically different experimental designs (active discrimination vs. a more passive orthogonal task) and analysis strategies (magnitude-based contrast analysis vs. multivariate classification algorithms). These regions thus demonstrate locally distributed response patterns linked to specific musical categories and theoretically comprise important

regions in a cortical network for sound categorization. These results argue that such a network is recruited automatically for some types of non-speech auditory processing.

The STS may serve a critical early role in a ventral stream of information processing, with particular links having been made between left STS and phoneme perception (Joanisse et al., 2007; Liebenthal et al., 2005). The present results suggest bilateral STS processing for musical intervals, with a right hemispheric bias, thus generalizing the role of the STS beyond the speech modality. The right STS may subserve an early "post-auditory" stage of processing (Pisoni, 1975; Zatorre, 1983), where continuous acoustic signals are converted to invariant "all-or-nothing" codes. These invariant, over-learned categorical memories may be mediated by Hebbian neural population codes (Hebb, 1949) distributed over many of voxels in a region. Triggering of these population codes may result in robust invariant BOLD responses, visible above noise to classification algorithms. While these analyses do not allow a full review of the spatial extent of these putative population codes, a sufficient portion (as defined by decoding success) of the circuits appears to exist at scales similar to the size of the searchlight spheres (~123 3.5 mm3 isotropic voxels, about 5 ml). The left STS, less implicated here, could be performing a parallel stream of categorical processing, tuned for different features of the signal. Alternatively, left STS response patterns could be representative of (a) inter-hemispheric communication or (b) access to the verbal lexicon, as these musical categories cannot be completely dissociated from their names (e.g. "minor"). We do not believe that the STS is the exclusive mediator of musical CP, but instead plays a dominant role in the ventral stream component of categorical processing.

The use of multiple intervals with variable pitch height ensured that these putative category maps represented abstract features beyond specific sound samples, instead reflecting learned relative pitch relationships between musical notes. Classifiers were trained and tested blind to the specific pitch classes (pitch height) of the musical intervals and thus were only able to utilize information related to category membership, rather than absolute pitch information. In fact, as

specific tones were not repeated within interval categories, but were reused across categories, the SVMs had to learn to largely "disregard" absolute frequency-driven features. While categorical distinctions are not requisite for successful MVPA, null imaging results from the pitch height analysis suggest that, here, categorical quality is the distinguishing feature detectable by the classifier. The absence of MVPA results in the pitch height analysis could be due to the use of sound stimuli that (a) were highly physically similar to one another and (b) exhibited considerable overlap in the frequencies of their note pairs, and suggests that top-down, memory-based processing may be the critical component in eliciting a robust stable BOLD response pattern in the interval quality analysis. While subjects did show some ability to "identify" sounds based on pitch height, they did not demonstrate the clear labeling plateaus consistently found for interval identification (*see Figure 3.2*). This finding, as well as the lack of an M-shaped function for pitch height discrimination, is highly indicative of short-term anchoring effects, as opposed to access to an over-learned long term memory store. Likewise, the lack of significant orthogonal MVPA results suggests that fMRI classification success may rely heavily on the degree of perceptual differentiability, which may, in turn, originate from either bottom-up or top-down processes. (Although, at least in the visual domain, BOLD data have been used to decode certain physical stimuli even in the absence of conscious perceptual differences (Haynes & Rees, 2005; Kamitani & Tong, 2005).) A recent MVPA study (Lee et al., 2011) of non-musician subjects using melodic musical stimuli did not yield significant decoding for minor vs. major sounds, suggesting that these categorical processes are highly experience-driven, in accordance with previous behavioral studies demonstrating little or no categorical musical perception in non-musicians (Burns & Ward, 1978; Zatorre & Halpern, 1979).

A recent speech study (Kilian-Hütten et al., 2011), meanwhile, used a categorical "midpoint" approach to demonstrate CP via auditory recalibration in the absence of acoustic differences between stimuli. We considered a related approach for this study: demonstration of MVPA differences following the presentation of stimuli that varied by a single continuous physical

parameter, yet were perceived as members of 2 discrete categories. This alternative approach is powerful in that it minimizes acoustically driven confounds, but does not easily generalize beyond the examined category pair. Thus, in order to drive generalizability, we chose to demonstrate categorical versus non-categorical processing via an orthogonal, absolute-pitch-roving dimension. Unlike other auditory "objects," where absolute frequency may be largely unrelated to the dimension of interest (e.g. sound identity (Staeren et al., 2009)), simple pitch values, critically, define the category identity of musical intervals and thus can be manipulated to form the basis for both the experimental and orthogonal stimuli dimensions. Thus, to make the MVPA results as generalizable as possible, we chose to test the categorical component of the analysis across 3 categories, and to dissociate the acoustically driven, non-categorical component by way of a second tone frequency-based dimension.

The decoding results, which provide evidence that such categorical information is present in the STS (but do not show any such evidence for the STG), stand in contrast with recent studies, suggesting that early auditory areas mediate complex, object-based processing (Kilian-Hütten et al., 2011; Ley et al., 2012; Staeren et al., 2009). These recent studies are all excellent demonstrations of MVPA's ability to reveal that auditory cortex is involved in classification of sounds, but say less about true categorical—"perception"—as classically defined, as none reported results from behavioral discrimination tasks. The STS results presented here support a hierarchical auditory ventral stream processing model (Hickok & Poeppel, 2007), which is not necessarily contradictory to architectures that may also contain myriad feedback/forward connections and parallel processing stages. The null results in and around Heschl's Gyrus (HG) may be due in part to the use of standard BOLD voxel size (3.5 mm isotrophic, as opposed to <2 mm voxels used in certain studies) or a high degree of variance in the shape/location of tonotopic maps in individuals. (Voxel size here was chosen as a compromise between relatively small size and full-brain coverage.) We therefore do not dismiss the idea that early auditory areas play a nontrivial role in categorical sound processing, as we report only null evidence in this study.

However, some of the differences between our relatively focal STS results and those of other auditory MVPA studies mentioned (relatively distributed over large portions of HG/ STG/STS) could be due to the strict categorical nature of the utilized sound stimuli. While stimuli such as cats/guitars (Staeren et al., 2009) or syllables spoken by different voices (Formisano et al., 2008) are clearly differentiable and "identifiable" with near-perfect accuracy, they have not been shown to display all the hallmarks of CP as originally defined (Liberman et al., 1957), where discrimination is limited by identification (or, at least "partially" limited, according to revisions of the theory (Pisoni, 1971; Zatorre, 1983)). It is therefore plausible that these multidimensional "cognitive categories" rely heavily on supra-perceptual processes, which, in turn, use more widely distributed neural networks ("categorical cognition" as opposed to "categorical perception"). The behavioral data presented here (clear 3-category identification functions with aligned M-shaped discrimination functions) reflect the fixed, specific nature of over-learned musical categories and not a more general configuration of features. These 2 processing models —distributed versus hierarchical— may not be mutually exclusive, with the former putatively more applicable in situations where categories are less well-established or more like natural semantic categories (as demonstrated by Staeren et al. (2009)), and the latter for more purely "perceptual" categorization.

The IPS results suggest the involvement of a dorsal stream of information processing. Unlike those in the STS, the IPS peaks fall well outside brain regions highlighted in the all sounds > silence contrast, suggesting that decoding in these parietal regions is performed on "supra-auditory" information (and, separately, argues against ubiquitous use of activation masks as a first step in feature elimination methodologies, in accordance with findings from Jimura & Poldrack (2011)). The dorsal stream, originally postulated as the spatial processing system (Mishkin & Ungerleider, 1982), is now more often considered to underlie the transformation and combination of information between sensory modalities (e.g. Culham & Kanwisher (2001)) and into motor and execution codes (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009). The IPS

specifically has also been implicated in high-level sound transformations that require relative pitch processing (Foster & Zatorre, 2010). The IPS peaks may thus be reflecting information that is still sensory, but no longer strictly auditory and on route to interfacing with the motor system. The motor theory of perception (Galantucci, Fowler, & Turvey, 2006) is particularly relevant here, as our subjects all had extensive instrumental musical training. It follows that these individuals have formed strong associations between categories of musical sounds and the sets of movements required to make such sounds (Zatorre, Chen, & Penhune, 2007). A recent fMRI repetition suppression paradigm of expert pianists (Brown et al., 2013) demonstrated the involvement of the IPS in auditory-motor transformations for correct positioning of fingers, which, in combination with the presented results, implicates the IPS as a crucial "audio"-motor interface (in addition to its more well-established role in visuo-motor processing).

The location of the parietal peaks, particularly the left IPS, invites comparison with the more ventral area "spt." Spt is believed to form part of the auditory dorsal stream (Hickok & Poeppel, 2007), is considered a "sensorimotor interface," and has been implicated in both speech production and perception (Hickok et al., 2008). Dorsally streaming music- versus speech-related information is likely destined for shared yet distinct frontal regions, with these results suggesting that spatially distinct processes emerge early. Furthermore, with the exception of few instruments (notably voice), music production relies heavily on the hands: this is notable as the parietal peaks observed here lie in the IPS, which is believed to underpin transformations between vision and limb and hand/grip movements (Cavina-Pratesi et al., 2010). However, in opposition to speech perception/production (with its strong one-to-one correspondence between sound/movement), a musical interval can be played using a variety of gestures requiring myriad sets of fingers/notes/instruments. It follows that frontal lobe perceptual decoding, such as the phonemic decoding reported by (Lee et al., 2012), may require motor specificity beyond that provided for by abstract musical categories.

In summary, the STS and IPS results presented here, along with earlier fMRI data for musical interval categorization (Klein & Zatorre, 2011) and multiple speech studies, indicate the likely presence of 2 streams of auditory information processing for CP. The right STS, a critical component of the putative ventral stream, may underlie successful identification and recognition of simple musical categories, with the presented bilateral (but right lateralized) pattern of results complementing the speech phoneme CP literature (Wolmetz, Poeppel, & Rapp, 2011). In contrast, the dorsal IPS nodes may reflect a transformation stage between unimodal auditory and motoric information. These current analyses do not indicate the degree to which these streams remain separate entities or interact (and, if so, how). Finally, these results demonstrate the power of MVPA to enable mapping of highly automatic cognitive/perceptual processes, even in the absence of demanding behavioral tasks, which generally require larger working memory loads and complex control conditions, both of which may confound imaging results.

3.7 Acknowledgements

Funding: This work was supported by the Canadian Institutes of Health Research (CIHR) (grant nos MOP14995 and MOP11541) and by the Canada Fund for Innovation (grant no. 12246), and by infrastructure support from the Fonds de Recherche du Québec Nature et Technologies via the Centre for Research in Brain, Language and Music.

Notes: We thank the staff of the McConnell Brain Imaging Centre for help in acquiring imaging data, as well as various members of the PyMVPA mailing list for assistance with analysis scripts. Conflict of Interest: None declared.

Chapter 4 - fMRI pattern analysis of played vs. perceived piano sequences in dorsal vs. ventral cortical streams

Klein ME, Hollinger AD, Zatorre RJ. (in preparation). fMRI pattern analysis of played vs. perceived piano sequences in dorsal vs. ventral cortical streams.

4.1 Preface

In the first two studies (Chapters 2 and 3), I demonstrated a role for both the superior temporal and intraparietal sulci in categorical perception of musical intervals and chords. In the present study, the goal was to examine the differential contributions of these ventral and dorsal regions, as well as their potential interactions with top-down processes originating in frontal cortex. This was achieved via an active paradigm, in which participants not only listened to musical sequences, but also produced them via an MRI compatible piano keyboard.

4.2 Abstract

The plurality of ways in which incoming sensory information is processed by the cerebral cortex is a major topic of ongoing investigation in perceptual neuroscience. A major discovery, first observed in the visual system and later extended to audition, is the division of such processing into at least two major processing streams: ventral and dorsal. Whereas the ventral stream's role seems to be identification of perceptual objects, the dorsal stream may perform as a generalized transformer of information within and across modalities, including sensory-*motor* transformations. Here, we sought to dissociate the roles of the two streams in perceptual processing with fMRI, employing musical stimuli and a piano-based performance task. Trained pianists played two different tone sequences using two different fingering patterns during scanning. Piano performance was either paired with realistic auditory feedback or no feedback (silence), and participants passively perceived the piano sounds in separate experimental blocks. Multivariate pattern analysis (MVPA) was performed on the subsequent BOLD data, decoded for information related to finger motion (independently of sounds heard) or sounds (independently of motion). Decoding peaks were observed in several regions, notably in the right superior temporal sulcus (STS), left intraparietal sulcus (IPS), and right premotor and inferior frontal cortex. Whereas the STS and IPS patterns were more related to perception (i.e. auditory,

but not motor, decoding), the right frontal results seemed to require audio-motor integration: this area contained decodable information for both sound and motion, but *only* for the combined (audio-motor) portion of the study. Taken together, these results highlight various functional nodes in the dorsal and ventral streams of the auditory perceptual network, including contributions of right ventrolateral frontal cortex to perceptually linked actions.

4.3 Introduction

The strategies by which the cerebral cortex organizes and processes sensory information has long been a topic of scientific interest. One major advance was made by Mishkin et al. (1983) and Mishkin & Ungerleider (1982), who demonstrated a stark double dissociation in lesion studies of monkeys: those with temporal/ventral lesions were impaired in visual object recognition, whereas those with parietal/dorsal lesions were primarily affected in their spatial perception. The dorsal component of the "two streams" hypothesis, as it came to be known, later underwent an influential revision by Goodale & Milner (1992), who argued that the dorsal stream should be thought of as the "do" pathway, allowing for the manipulation and transformation of sensory information. Thus, the two streams have gradually become defined according to their putative operational qualities: perception for recognition (ventral) and perception for action (dorsal), rather than according to stimulus features. In the visual system, the ventral stream leads primarily to the infero-temporal (IT) cortex and eventual targets in inferior frontal cortex, whereas the dorsal stream involves the parietal, premotor and dorsolateral frontal lobes (Ungerleider & Haxby, 1994). The broader visual dorsal stream concept inherently invokes aspects of the motor system (including the somewhat controversial "mirror" system (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996)) and encapsulates various processes

including mental rotation (Harris & Miniussi, 2003; Tagaris et al., 1997; Zacks, 2008) and reaching/grasping (Cavina-Pratesi et al., 2010; Culham & Kanwisher, 2001).

There is a large body of research in humans and non-human primates indicating that, as in the visual system, there is a ventrally-oriented stream (Bendor & Wang, 2008; Kaas & Hackett, 2000; Rauschecker et al., 1995; Rauschecker & Tian, 2000; Romanski et al., 1999) underlying the perception of auditory objects (Bizley & Cohen, 2013). These ventral stream regions include temporal lateral belt and parabelt areas (Kaas & Hackett, 2000), as well as certain ventral frontal regions (Romanski et al., 1999). Processes that have been linked to the ventral stream include perception of pitch (Hyde et al., 2008; Johnsrude et al., 2000; Schneider et al., 2005), voice (Belin et al., 2000), complex auditory objects (Zatorre et al., 2004) and musical stimuli (Abrams et al., 2011; Klein & Zatorre, 2011; 2014; Lee et al., 2011), with certain right vs. left lateralization observed. The largest amount of research into the human auditory ventral stream, however, examines the perception of speech, which has highlighted various regions of the superior temporal gyri (STG) and sulci (STS) (as well as inferior temporal/frontal cortex), typically lateralized to the left (see review by Hickok & Poeppel (2007)). This speech research focus is to be expected, considering language's predominant role in auditory compared to visual processing.

Interestingly, certain speech and music perception studies have also highlighted dorsal cortical regions. A dorsal stream for cortical auditory processing was first proposed in line with the visual "where" model (Rauschecker & Tian, 2000) and subsequently revised to what Warren et al. (2005) termed a "do" pathway for the transformation and manipulation of auditory sensory information. As with the ventral stream, the vast majority of research into the dorsal auditory stream has been conducted via speech, which has yielded certain unexpected controversies. One major debate revolved around the putative "motor theory" of perception (Liberman et al., 1967), which was originally postulated on behavioral, not neurobiological, grounds and states that the

ability to *make* speech sounds (i.e. involvement of motoric processing) is a limiting factor in *perception*. While the strong form of this theory has generally been discarded in light subsequent research (Galantucci et al., 2006; Kuhl, 1978), considerable work has implicated posterior (primarily premotor) frontal cortex —both dorsal and ventral— in speech perception (Du et al., 2014; Hickok & Poeppel, 2007; Rauschecker & Scott, 2009; Zatorre et al., 1992). Separately, certain music perceptual tasks have been linked to the parietal lobe's intraparietal sulcus (IPS) (Foster et al., 2013; Foster & Zatorre, 2010; Klein & Zatorre, 2011; 2014). And, as with visual processing, a frontal-parietal action-representational system has been implicated in auditory processing (Kohler, 2002; Lahav et al., 2007). This system was previously highlighted using a repetition suppression paradigm with piano performance, such that IPS, dorsal, and ventral frontal regions all showed modulation as a function of repeated execution of melodies (Brown et al., 2013). In line with Warren's "do pathway" interpretation of the dorsal stream, both speech and music are sounds that humans *make* as well as perceive. Thus, it should not be surprising that perception of both classes of sounds may automatically engage particular nodes of this stream in accordance with their particular learned associations with audio-motor integrative requirements. However, the specifics of these processes are still poorly understood.

Speech is characterized by a fairly one-to-one correspondence between perception and action; there is a specific way to produce a specific sound and, conversely, a perceived sound implies a particular movement. Musical sounds however, need not have such restraints. A particular piano key, for example, can be played with any finger, while combinations of notes may be played with various combinations of fingerings. Likewise, a particular fingering combination need not elicit one specific sets of notes. (Still, for trained musicians, there is a clear connection between processing in auditory and motor cortices: compared to non-musicians, passive listening activates the motor system, while feedback-free playing activates temporal-lobe circuitry (Bangert et al., 2006).) Thus, it seems that music perception, like speech, relies upon dorsal pathways (Zatorre et al., 2007) in a way that is perhaps more flexible and abstract than the

recoding that takes place for speech. This distinction, between flexible and relatively direct mapping, parallels more general models of the premotor cortex, which draw a similar distinction between its dorsal (PMd) and ventral (PMv) sub-domains. The flexible processes in the PMd are thought to subserve abstract sensori-motor integration (Hoshi & Tanji, 2007; Zatorre et al., 2007) and conditional motor associations, which, via learning, link initially arbitrary sensory cues with specific motor commands (Petrides, 1985; Wise, di Pellegrino, & Boussaoud, 1996). The PMv, in contrast, may be more involved in processes that require predominantly *direct* links between sensation and action (Hoshi & Tanji, 2007; Keysers et al., 2003; Kohler, 2002), of which speech is a prime example (Hickok & Poeppel, 2007). It follows that those aspects of the dorsal stream that underlie musical processing may rely more heavily on dorsal than ventral premotor networks.

The flexibility provided by music perception/production, combined with the relative ease by which motor-sourced and auditory-sourced processing may be dissociated, makes music an ideal candidate with which to probe the roles of the dorsal/ventral streams in perception. As mentioned above, past perceptual studies from our lab of simple harmonic (Klein & Zatorre, 2011) and melodic (Klein & Zatorre, 2014) musical sequences have revealed both ventral *and* dorsal foci of neural activition/information. In particular, the 2014 study revealed locally-distributed patterns of information (Kriegeskorte et al., 2006) in both the STS (ventral) and IPS (dorsal) following a passive listening paradigm, with this information linked to abstract meaningful musical categories (e.g. minor thirds, major thirds), but not to absolute pitch height (e.g. c-natural, c-sharp). Here, in order to dissociate processes subserved by the two streams, we employed an active musical performance task, utilizing an MR-compatible piano keyboard (Hollinger, Steele, Penhune, Zatorre, & Wanderley, 2007). The study required trained pianists to perform two different fingering patterns, each of which could be used to generate two separate sound sequences (the two dimensions thus being orthogonal to one another). The experiment was divided into three conditions: a “combined” audio-motor condition in which active performance

generated sound feedback; a "motor only" condition, in which performance did not result in auditory feedback; and an "audio only" condition, in which participants passively listened to tone-sequences similar to what they were asked to perform. Through use of the orthogonal 2x2 design, paired with these three divisions, we hoped to determine the various contributions of ventral and dorsal processes to perception. Specifically, we hypothesized that the ventral stream, particularly the right STS, would be sensitive to sound identity (irrespective of motor involvement), whereas information in the dorsal stream (premotor and/or posterior parietal) would be influenced by motor commands in both the combined and motor conditions. As the two sound conditions (as well as the two motor sequences) were not hypothesized to show region-by-region differences in *activation*, multivariate pattern analyses (MVPA) were employed to decode between conditions. Unlike GLM analyses, which require such region-wide differences, MVPA can utilize the fine-grained patterns in a signal to dissociate between two (or more) conditions that appear highly similar when viewed only at a coarser scale.

4.4 Methods

4.4.1 Participants

We recruited 11 highly-trained, right-handed pianists (6 females, minimum 5 years formal training and currently practicing or performing); the majority of whom came from McGill University's undergraduate and graduate music student populations and none of whom possessed absolute pitch abilities. All participants gave their informed consent. Ethical approval was granted by the Montreal Neurological Institute Research Ethics Board.

4.4.2 Piano keyboard, hardware, and software

Study participants were trained to use a 2-octave MR-compatible optical piano keyboard, designed and built at McGill University's Centre for Interdisciplinary Research in Music Media and Technology (Hollinger & Wanderley, 2013; Hollinger et al., 2007) (*Figure 4.1*). The keyboard position was individually fitted to each participant using a specially-designed plexiglass keyboard mount, which allowed for easy access to the weighted piano keys and a good range of hand/arm motion, despite lying supine inside the scanner bore. Hardware circuitry, including an LED light-source and phototransistor, were located in a control room adjacent to the MR scanner. This circuitry connected to the keyboard through a set of fiber optic cables, which were passed through a small port in the wall connecting the control and scanner rooms. Additional hardware and electronic details of the MR keyboard are detailed by Hollinger & Wanderley (2013) and Hollinger et al. (2007). Note-onset messages were communicated via USB to the experimental computer: a Macbook running OSX version 10.7 and Python version 2.7. The experiment was run using version 1.7 of Python's PsychoPy toolbox (Peirce, 2007) paired with in-house Python scripts for interaction with the MR keyboard. During certain time windows, key presses triggered one-second-long piano samples of the corresponding note, which were created using Live version 8 (Ableton AG, Berlin, Germany). The sounds samples were output using a Duet 2 sound card (Apogee Electronics Corp., Santa Monica, USA) and presented binaurally via MRI-compatible headphones (S14 Insert Earphones, Sensimetrics).

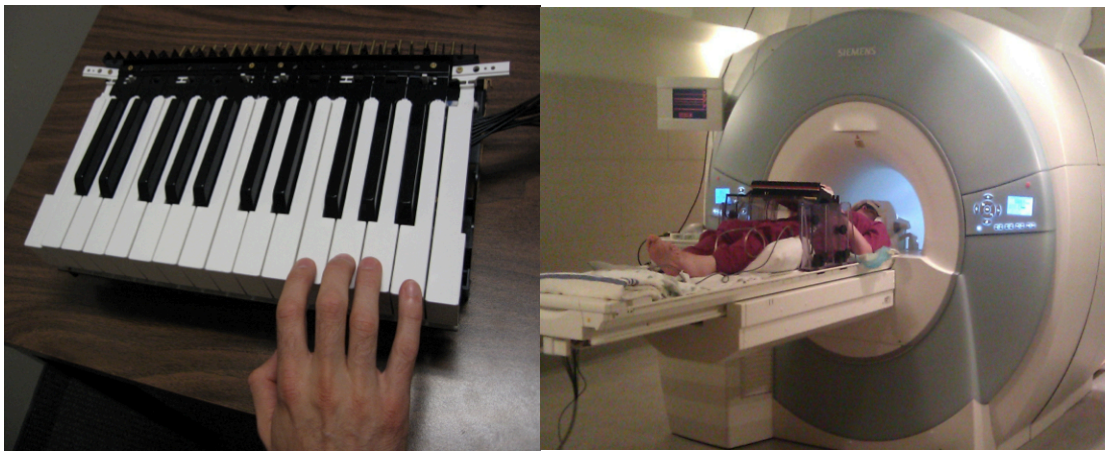

Figure 4.1 – MRI compatible piano keyboard

Left: MR-compatible piano keyboard used to create realistic audio-motor experiences for musically-trained participants. Key press information was relayed via fiber optics to circuitry in a control room, where it was then converted to note-onset messages and routed to the experimental computer. Right: Positioning of the keyboard with participant on scanner bed, ready to be moved into bore.

4.4.3 fMRI details

Study participants underwent anatomical and functional imaging in a 3 Tesla Siemens Magnetom Trio with a 32-channel head coil. Five T2* functional runs (3mm isotropic) were interspersed with an anatomical T1 acquisition (1mm isotropic, TR/TE = 2300ms/ 2.98ms). Each functional run consisted of 59 trials, each of 11.76 seconds duration. Of the 59 trials, 32 required piano playing, 8 were passive listening to piano melodies, and 10 were used to cue the beginning (or end) of blocks, with the remaining 9 used to establish baseline/resting levels. Individual trials were broken up into four 2.94 second bins, with each bin corresponding to a single BOLD volume acquired using an interleaved silent steady state (ISSS) approach (Schwarzbauer, Davis, Rodd, & Johnsrude, 2006) (*Figure 4.2*). ISSS allows for the interspersed use of two kinds of sequences: (1) EPI pulse sequences that are in line with standard BOLD sequences used in continuous or sparse temporal sampling (Belin et al., 1999; Hall et al., 1999) paradigms, and (2) "silent" pulse sequences, which do not yield usable BOLD data, but have the advantage of maintaining a steady state of magnetization in the signal. The end product of this approach is the ability to acquire multiple T2* volumes back-to-back, without the differential T2* brightness issues which can negatively affect standard "sparse clustered" approaches. As we performed multivariate pattern analyses (MVPA, *see Section 4.4.6*) on the data, acquiring multiple volumes/trial was advantageous as MVPA's robustness is highly sensitive to the size and signal-to-noise of the dataset (Pereira et al., 2009). Specifically, twice as much data could be collected without doubling the length of the experiment, while at the same time maintaining periods of relative quiet in which to present auditory stimuli.

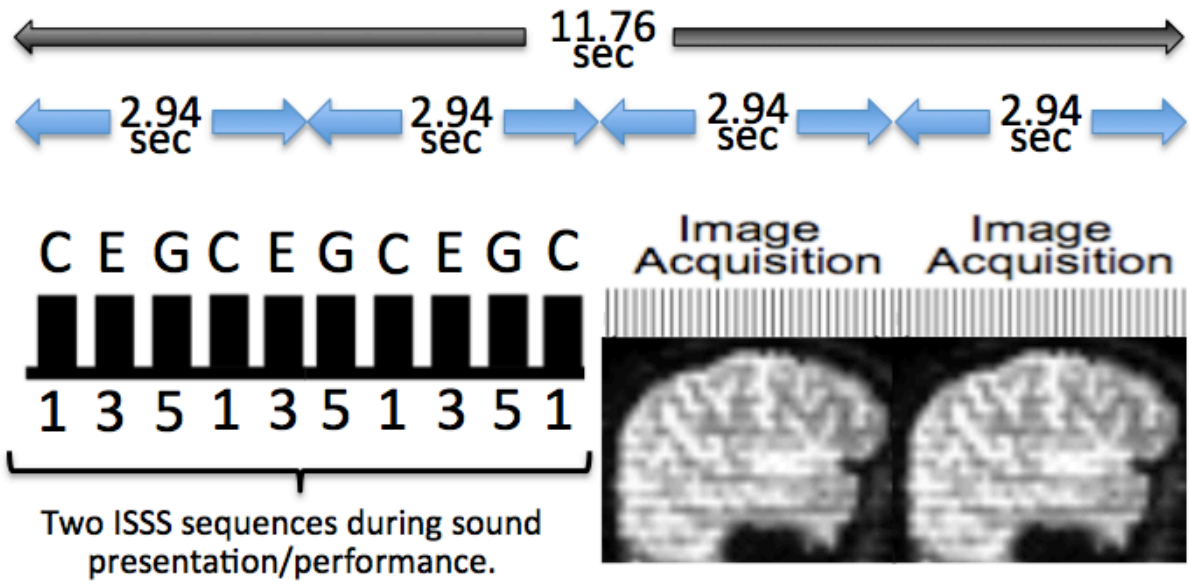

Figure 4.2 – fMRI protocol

fMRI trials each lasted 11.76 seconds, with half that time spent collecting BOLD data (loud background period, 2 volumes, 2.94 s per volume) and half for piano performance and/or stimulus presentation (quiet background period, 2 volumes, during which ten notes would be played: illustrated here is one sequence, keys C, E, and G played with fingers 1, 2, and 3). During the piano performance blocks, participants were trained to commence playing immediately following the *offset* of the loud scanner noise from the preceding trial. The quiet background was created using "silent steady state" sequences (Schwarzbauer et al., 2006), in order to maintain steady state magnetization in the signal. Each trial thus produced two volumes that were used for analyses, as well as two steady-state volumes that were discarded.

4.4.4 Experimental design

As the primary goal of the experiment was to examine differences between dorsal and ventral stream components of perception, we employed a 2x2 design, in which two distinct fingering patterns could be employed to make either of two piano melodies. In particular, participants used either the 1st, 3rd and 5th or the 1st, 2nd, and 4th fingers of their right hands to produce 10-note sequences comprised of C-, E-, and G-naturals (major chord starting on middle C) or B-, D-, and F-naturals (diminished chord) (*Figure 4.3, top*). These two melodic (i.e. arpeggiated presentation of notes) chords were chosen for their strong difference in perceptual salience, proximity on the piano keyboard, and identical distancing between keys. The fingering patterns were chosen as both are very common three key sequences performed with a single hand. Separately, the experiment was parceled into three further divisions: trials where participants played the piano and received auditory feedback ("combined" condition), trials where participants played the piano but *did not receive* auditory feedback ("motor only" condition), and trials that did not utilize the piano and instead consisted purely of passive listening to matched sequences of notes ("auditory only" condition).

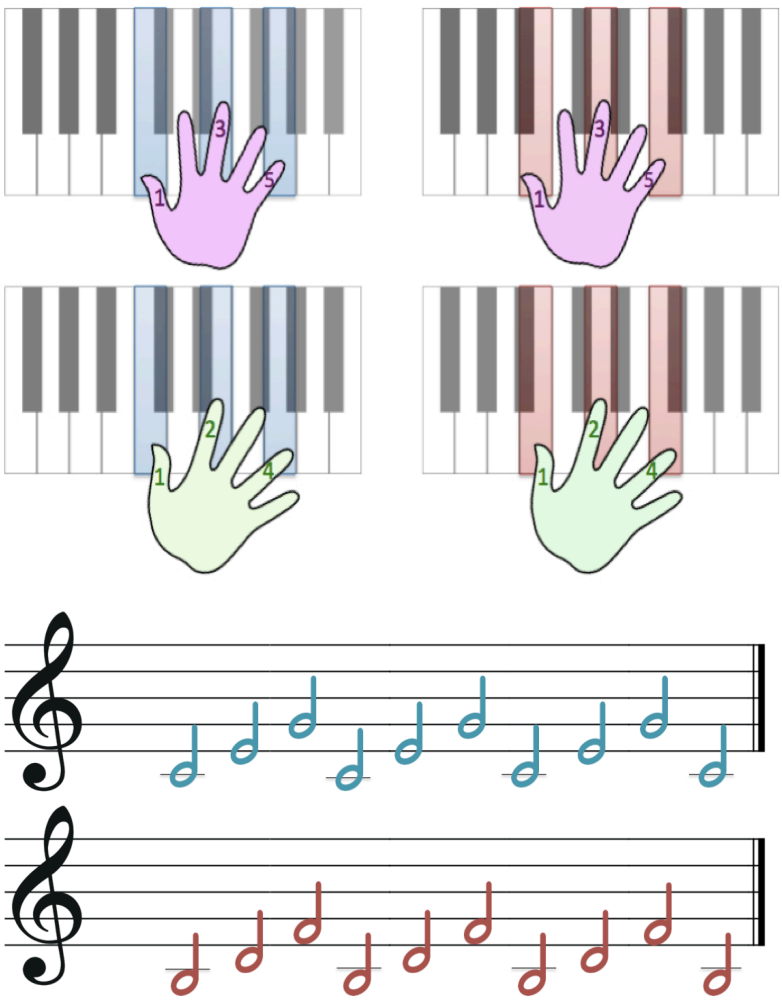

Figure 4.3 – Piano task and sound stimuli

Participants were instructed, block by block, to play a specific three-key (10-note) combination using a specific set of three fingers. Thus, there was a 2x2 design: two key combinations (C/E/G/C/E/G/C/E/G/C (blue keys in figure) or B/D/F/B/D/F/B/D/F/B (red keys in figure)) via two fingering patterns (1/3/5/1/3/5/1/3/5/1 (purple hands in figure) or 1/2/4/1/2/4/1/2/4/1 (green hands in figure)). Via the MRI keyboard and Python software environment, each key press either triggered a piano tone sample matched to the piano key (audio-motor condition) or did not produce any auditory feedback (motor-alone condition). These same piano sound samples, arranged in matched patterns of 10 notes, were also used in the audio-alone condition, in which participants passively listened but did not play the keyboard.

A blocked design was used, as the active motor task required subjects to repeatedly adjust their hand positions to perform precise fingering patterns. Because such adjustments took several seconds (hindered by participants' supine position and closed eyes), it was not possible within acceptable time frames to run the experiment in a true event-related fashion, which would require a re-positioning cue/movement trial prior to each task trial. Participants were instead cued at the beginning of a block to position a specific combination of fingers on a specific set of keys (e.g. "With your first, third, and fifth fingers, play 'c,' 'e,' and 'g'"). The initiation of playing in each individual trial was cued by the *offset* of scanner noise from the prior trial: as soon as this noise stopped, participants had 5.88 seconds of relative quiet in which to perform the 10-note sequences, which consisted of ascending melodic triads (beginning on B or C) repeated 3X plus a final note identical to the first (*Figure 4.3, bottom*). (All participants underwent a training session several days before their MRI sessions, in which they were familiarized with the task and keyboard, and in which pre-recorded scanning noise was presented over speakers. By the end of the training, all participants were able to consistently pace their playing with approximately two notes played per second. A minimal number of fMRI trials had to be disregarded due to too fast/slow of a tempo.) Each block consisted of a cue trial followed by eight trials requiring performance, plus a final cue trial that instructed participants to stop playing. Of the eight trials, half contained real-time auditory feedback ("combined" condition) and half were silent ("motor only" condition), which were ordered in a pseudo-random fashion. Each functional run contained four such blocks: one for each unique combination of fingering pattern and piano keys. The audio-only trials, mentioned above, were presented in separate blocks. For these blocks, participants were instructed to rest their hands motionless in their laps. There were two such blocks per functional run and the sound conditions were presented in a pseudo-random order. The resulting dataset could thus be divided according to sounds heard *or* by fingering patterns, and each third of the data (combined, motor only, or auditory only conditions) could be examined separately.

4.4.5 GLM analyses

Standard GLM analyses were performed in order to examine overall activation levels for the global auditory, motor, and combined conditions, as well as to compare between the subordinate conditions (e.g. c-e-g vs. b-d-f). Trials where participants had played too few, too many, or wrong notes were discarded from the analysis, however these events were rare (average of 3.1 (~2%) trials per subject, standard error of 1.0). fMRI data processing was carried out using FEAT (FMRI Expert Analysis Tool) Version 6.00, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). For 1st level analyses: registration to high-resolution structural and/or standard space images was carried out using FLIRT (Jenkinson, 2001; 2002). Registration from high resolution structural to standard space was then further refined using FNIRT nonlinear registration (Andersson, 2007a; 2007b). The following pre-statistics processing was applied; motion correction using MCFLIRT (Jenkinson, 2002); non-brain removal using BET (Smith, 2002); spatial smoothing using a Gaussian kernel of FWHM 6.0mm; grand-mean intensity normalisation of the entire 4D dataset by a single multiplicative factor; highpass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma = 50.0s). Time-series statistical analysis was carried out using FILM with local autocorrelation correction (Woolrich, 2001). Z (Gaussianised T/F) statistic images were thresholded using clusters determined by $Z>3.09$ ($p = 0.001$) and a (corrected) cluster significance threshold of $P = 0.05$ (Worsley, 2001). For 2nd level (within-subject) analyses, individual runs were combined using a fixed-effects analysis. 3rd level (between-subjects) analyses were run using FSL's FLAME mixed-effects model.

4.4.6 Multivariate pattern analysis (MVPA)

Pre-processing: Pre-processing, prior to MVPA, was performed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/) due to its more aggressive motion correction parameters (compared to FSL). In particular, FSL's MCFLIRT tool, while practical for within-

run motion correction as performed in standard GLM analyses, does not optimally account for larger head movements found *between* functional runs. (FSL deals with this problem by registering each run separately to a standard brain.) MVPA benefits from operating in a participant's native functional space and, as cross-validation requires a high degree of alignment between runs, it was necessary to directly correct all runs for motion, concatenated into a single long 4D image (up to 295 3D volumes). No spatial smoothing was performed on the functional data prior to MVPA. Temporal averaging (Mourao-Miranda et al., 2006) was performed by combining the two BOLD images acquired in each trial into a single volume. Within each run, we performed (a) linear detrending to remove signal changes due to slow drift and (b) z-scoring to place voxel values within a normal range (Pereira et al., 2009).

MVPA proper: MVPA was performed using the Python programming language's PyMVPA toolbox (Hanke et al., 2009) and LibSVM's linear support vector machine (SVM) implementation (http://www.csie.ntu. edu.tw/~cjlin/libsvm/). Separate classification analyses were performed on the auditory-only data (c-e-g vs. b-d-f); motor-only data (1-3-5 vs. 1-2-4); and combined condition data (c-e-g vs. b-d-f ignoring motor information, as well as 1-3-5 vs. 1-2-4 ignoring sound information). Classification was performed using leave-one-out cross-validation, where a classifier was trained on data from four of the functional runs and tested on data from the fifth, and the procedure was then repeated four times testing on a novel run each time and the results then averaged. Two different styles of MVPA were performed. First, we conducted a set of region of interest (ROI) analyses across 10 regions of the brain defined *a priori*: the left Heschl's Gyrus (HG); superior temporal sulcus (STS); intraparietal sulcus (IPS); dorsal premotor cortex and supplementary motor area (PMd); ventral premotor cortex and the pars opercularis / Brodmann Area 44 (PMv/44); and the analogous regions in the right hemisphere (*Figure 4.4*). These areas were chosen due to their known role in auditory (HG) or motor (PMd, PMv/44) processing, or past evidence of their role in musical perception (Klein & Zatorre, 2011; 2014). ROI masks were created in standard space (the MNI/ICBM152 template) via anatomical "painting" (STS and IPS) or the use of standard atlases (HG, PMd, PMv/44)

provided with FSL or MRIcron (http://www.mccauslandcenter.sc.edu/mricro/mricron). These masks were then warped from standard into native functional space using parameters determined by FSL's FNIRT tool. MVPA yielded a single score per subject per region, which was compared to chance-level accuracy (50% for all analyses) and input into a single-sample t-test. As detailed above, separate MVPAs were performed for each division of the experiment: fingering patterns in the combined and motor conditions, sound stimuli in the combined and auditory conditions. Following the ROI analyses, we performed a searchlight set of analyses (Kriegeskorte et al., 2006), whereby the decoding algorithm considers only voxels from a small sphere of space (radius = 3 voxels, up to 123 voxels in a sphere), which is then performed iteratively across many such spheres. Whereas the ROI analyses were performed on ROIs defined *a priori*, the searchlight analyses are more exploratory in nature and serve as an unbiased method with which to explore information covering the entire cortex. "Accuracy maps" are created by assigning the decoding accuracy above chance for each sphere to the center voxel of that sphere.

At the moment, there is no "gold-standard" practice for how to report between-subjects data from a searchlight analysis. As the images are not explicitly smoothed as in a GLM analysis, gaussian random field theory-based corrections yielding a "smoothed variance ratio" (Worsley et al., 2002) are not applicable, with pure Bonferroni-based corrections far too conservative (for n=11, whole-brain t-thresholds can be in the mid-teens). In order to bridge the gap between false positives and negatives, we used a threshold/binarize/tally approach, somewhat similar to that outlined (via a non-searchlight methodology) by Staeren et al. (2009). First, accuracy maps, which were in each participant's own 3mm (isomorphic) native functional space, were warped to a 3mm MN152 template using nearest neighbor interpolation. Second, these single-subject maps were binarized at 5% accuracy above chance ($p = 0.2$, uncorrected). Third, values at each voxel in the standardized/binarized images were summed, yielding an integer score between 0 (< 55% accuracy in all subjects) and 11 (> 55% accuracy in all subjects). These tallied images were then thresholded at a value of 8 of 11 participants ($p = 0.0004$). However, as we are only reporting clusters of greater than 33 contiguous voxels (the size of a sphere with 2-voxel radius), the actual probability of a false positive is considerably lower than this value.

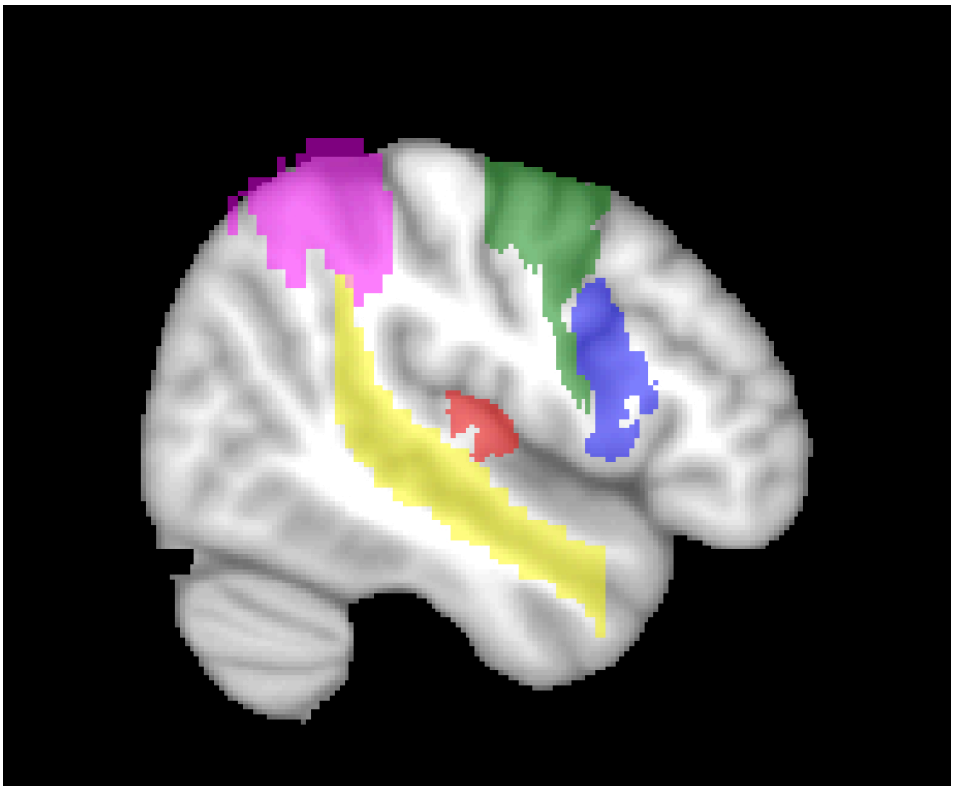

Figure 4.4 – Region-of-interest masks

Region-of-interest (ROI) masks were created *a priori* in 5 regions bilaterally, resulting in 10 separate masks. As the masks were nearly symmetrical, only a single hemisphere is shown in sagittal section. The ROIs were for Heschl's Gyrus (HG, red); superior temporal sulcus (STS, yellow); intraparietal sulcus (IPS, magenta); dorsal premotor cortex and supplementary motor area (PMd, green); ventral premotor cortex and the pars opercularis / Brodmann Area 44 (PMv/44, blue). These masks were created in standard (MNI) space and then warped into each participant's native functional space using parameters determined by FSL's FNIRT tool.

4.5 Results

4.5.1 GLM results

We performed univariate GLM analyses with FSL, in order to assess and compare global activation patterns for the auditory-alone, motor-alone, and combined conditions vs. the resting (baseline) condition, which are reported below. Univariate analyses were also performed between subordinate conditions (e.g. 1-3-5 vs. 1-2-4 fingering patterns). No significant univariate effects were observed when contrasting between the two sounds or between the two fingering patterns in any of the three global conditions (auditory-only, motor-only, or combined). These null findings were expected due to the high degree of similarity between the sound stimuli (and between the motor sequences).

For the **auditory-only analysis** (*Figure 4.5, top row*), as was expected, significant clusters of bilateral activity were observed in the superior temporal gyrus (STG, including Heschyl's Gyrus (HG)), with the right hemisphere cluster extending into the superior temporal sulcus (STS). Left hemisphere activity was also observed in somatosensori-motor regions, including the left hemisphere's inferior frontal sulcus (IFS) and gyrus (IFG), the post-central gyrus (extending into the post-central and central sulci), and premotor cortex (including the supplementary motor area (SMA)). A left hemisphere peak was also observed in the intraparietal sulcus (IPS). In the right hemisphere, peaks were observed in the pre-central gyrus, lateral occipital lobe, and the hippocampus. Additionally, bilateral activity was observed in the cerebellum and the orbito-frontal cortex. For the **motor analysis** (*Figure 4.5, second row*), a very large left hemisphere cluster spanning the right hand area of the pre- and post-central gyri was observed, consistent with the piano task. Activity that is likely part of the greater interconnected somatosensori-motor system was also observed bilaterally in the cerebellum, SMA, and IFG; in the inferior parietal lobule (IPL) and the parietal operculum of the left hemisphere; and in the right hemisphere's post-central sulcus/gyrus. Additionally, bilateral activity was observed in the occipital lobes'

lingual gyri. The **combined analysis** showed peaks largely in the same regions as listed above (*Figure 4.5, third row*). In order to examine potential audio-motor integration areas, we performed a subtraction analysis to reveal voxels activated by the combined condition, but not activated by *either* the auditory *or* the motor conditions. (In other words, we binarized the thresholded activation maps for all three images and then subtracted both the auditory and motor images from the combined image.) A large cluster of left hemisphere voxels elicited via this method was found in the posterior lateral sulcus (in or close to area "spt"), with an analogous region observed contralaterally in the right hemisphere's supra-marginal gyrus (SMG) (*Figure 4.5, bottom row*). Other areas highlighted were found in the right IFG and cerebellum; the left pre-central gyrus, IPL, and anterior IPS; and bilaterally in the anterior STG.

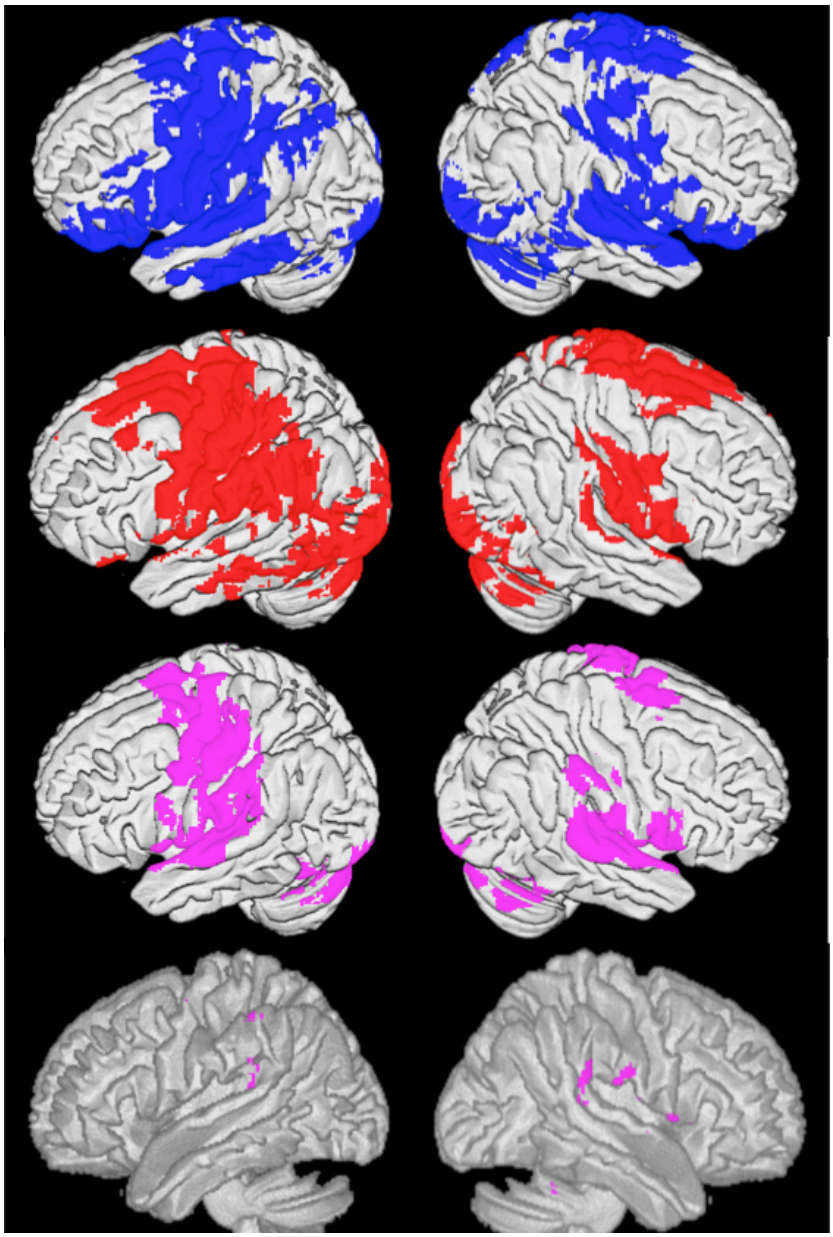

Figure 4.5 – GLM results

Univariate GLM results, projected onto white matter surface renderings of the left and right hemispheres, depict significant activation from baseline for auditory-alone (top, blue); motor- alone (second row, red); and audio-motor (third row, magenta) conditions. The bottom row shows regions that activated in the audio-motor condition, but did not activate in *either* the auditory-alone *or* motor-alone conditions, thus being candidates for audio-motor interactive processing. These various activation maps implicate regions previously tied to auditory and/or motor processing: most notably the superior temporal cortices; the inferior parietal lobule and intraparietal sulcus; and posterior prefrontal and motor areas. The majority of voxels significantly activated in only the combined condition were located in inferior parietal cortex, bilaterally.

4.5.2 MVPA region of interest analysis

Similar to the GLM, we performed separate MVPA analyses for the auditory-only, motor-only, and combined conditions. However, here, instead of reporting activation patterns vs. baseline (rest), we are comparing the results *between* conditions of interest. For the auditory-only condition, the MVPA results are for regions that decode the sounds CEG and BDF from one another. For the (silent) motor condition, the comparison is between the 1-3-5 and 1-2-4 fingering patterns. For the combined condition, the decoding was split into two orthogonal dimensions: (1) decoding for sound (CEG vs. BDF) irrespective of motor patterns, and (2) decoding for motor patterns (1-3-5 vs. 1-2-4) irrespective of sound stimuli. As discussed in the methods section, we performed MVPA on 5 separate regions of interest for each of the two hemispheres: Heschl's Gyrus (HG), ventral premotor/posterior IFG (PMv/44), the intraparietal sulcus (IPS), dorsal premotor (PMd), and the STS. Accuracies were compared to chance level (50%) and input into a t-test (n = 11, 1-tailed) in order to assess significance (reported for both $p \leq 0.05$ and $p \leq 0.01$). As there was a large degree of variability between individuals, smaller effect sizes (average accuracy scores) coupled with lower variance sometimes led to larger t-values than higher effect sizes that were coupled with high variance.

For the **auditory-only** condition, the only ROI that decoded sounds from one another significantly above chance level was in the in right STS (54.1% accuracy, $p \leq 0.01$). No ROIs were able to decode between the **motor-only** conditions (*Figure 4.6, top chart*). ROIs were considerably more decodable when examining the **combined** condition (*Figure 4.6, bottom chart*). In the right hemisphere, fingering patterns were able to be decoded in the PMv/44 region (56.8% accuracy, $p \leq 0.01$), with a trend that just missed statistical significance found in PMd (54.9% accuracy, $p = 0.07$). Meanwhile, sound stimuli were decodable in several ROIs. In the left hemisphere, we observed information-containing patterns in the IPS (56.3% accuracy, $p \leq 0.01$) and PMd (55.7% accuracy, $p \leq 0.05$). In the right hemisphere, all 5 regions showed significant decoding for sound: HG (56.9% accuracy, $p \leq 0.01$), PMv/44 (58.1% accuracy, $p \leq 0.01$), IPS (55.2% accuracy, $p \leq 0.05$), PMd (54.7% accuracy, $p \leq 0.01$), and STS (57.9% accuracy, $p \leq 0.01$).

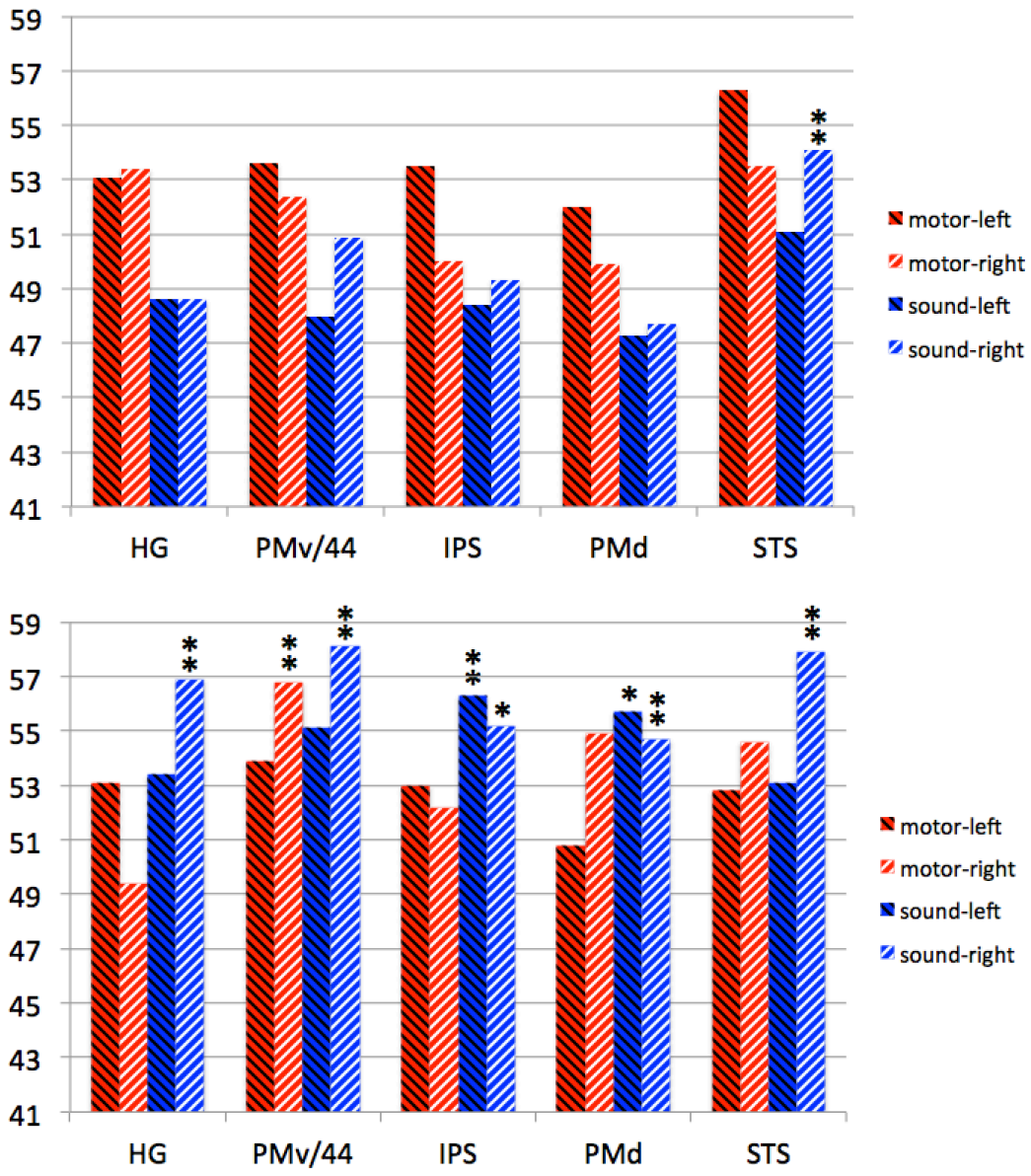

Figure 4.6 – Region-of-interest decoding results

Decoding accuracies for the various ROIs in the motor-alone and audio-alone conditions (top) and for the motor-combined and audio-combined conditions (bottom). Motor analyses are depicted in red, auditory analyses in blue. Left hemisphere regions are shown by stripes directed up to the left, with stripes directed up to the right for the right hemisphere. Chance-level decoding was 50% for all analyses. Statistical significance was assessed for each ROI against the chance-level baseline, with one asterisk (*) signifying a p-value of ≤ 0.05 and two asterisks depicting p ≤ 0.01. Across the board, decoding was enhanced in the combined conditions (bottom chart) vs. the isolated audio or motor conditions (top chart).

4.5.3 MVPA searchlight analysis

As discussed in the Methods section, we performed an exploratory searchlight analysis in order to complement the *a priori* ROI results. *Table 4.1* contains all clusters of greater than 33 voxels (the size a sphere with 2-voxel radius), where all voxels in the cluster surpassed 55% accuracy in 8 of the 11 subjects. Additional smaller clusters (< 33 voxels) were also observed and are not reported here. As in the ROI analysis, we present data for sound decoding in the auditory-only condition; motor decoding in the motor-only condition; and both sound and motor decoding in the combined (audio-motor) condition. Generally, there was a large degree of correspondence between the ROI and searchlight results. For example, just as sound decoding in the combined condition showed the most significance in the ROI analysis (i.e. 7 of 10 regions showed above-chance decoding), the same analysis via the searchlight methodology showed the greatest number of supra-threshold clusters. Additionally, many of these clusters fell within the previously defined ROIs.

For the **auditory-only** condition we observed a single supra-threshold cluster (37 voxels) in the right STS (*Figure 4.7, right*). For the **motor-only** condition we also observed a single supra-threshold cluster (163 voxels), this one located around the left lateral middle/inferior temporo-occipital border, spanning BA19 and BA37 (*Figure 4.7, left*).

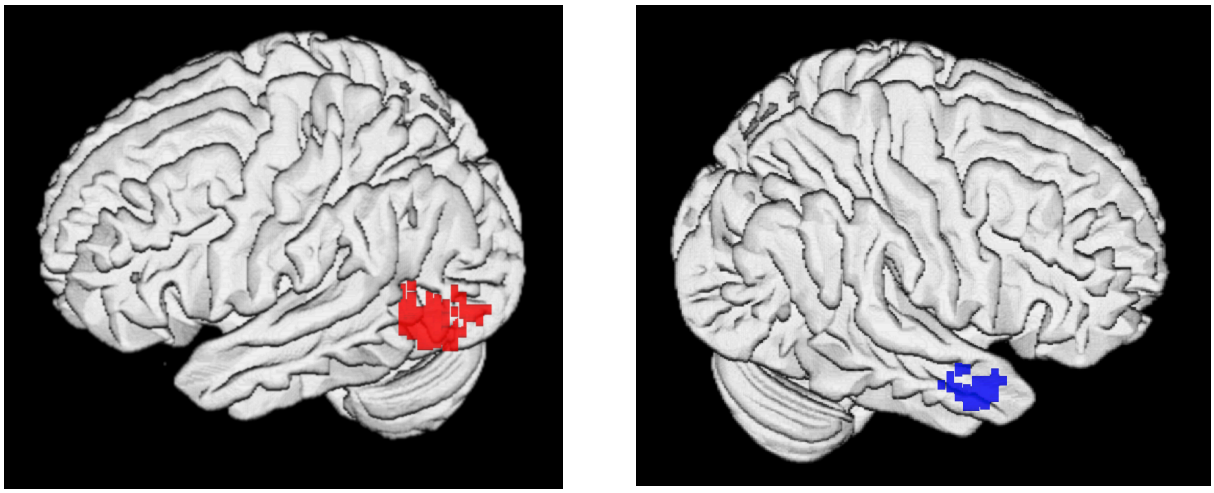

Figure 4.7 – MVPA searchlight results for motor-alone and sound-alone

Voxel clusters that showed a high degree of decoding across multiple subjects for the motor- alone (red, left) and audio-alone (blue, right) conditions. One region was highlighted for each analysis: the left lateral middle/inferior temporal-occipital border for motor decoding and, for sound decoding, the right superior temporal sulcus.

Turning to the **combined** (audio-motor) condition, we observed 3 supra-threshold clusters for the motor decoding analysis (*Figure 4.8*). The largest cluster (147 voxels) spanned multiple regions, across the right frontal operculum, anterior insula and anterior STG. A second large cluster (134 voxels) was located proximally to the largest cluster from the motor-only analysis (left lateral temporo-occipital region). A third large cluster (97 voxels) spanned the right ventral premotor/ posterior IFG region (likely related to the significant decoding in this region in the ROI analysis). As stated above, the greatest number of supra-threshold clusters was observed in the combined condition auditory analysis (*Figure 4.9*). Similar to the motor-combined analysis, the largest cluster from the auditory-combined analysis (339 voxels) spanned the right frontal operculum, anterior insula, and anterior STG, while also extending into the STS. Other large clusters were located in the right hippocampal formation (72 voxels); left MTG/ITS (46 voxels); left lateral temporo-occipital region (40 voxels); left intraparietal sulcus (IPS, 38 voxels); and right parieto-occipital sulcus (34 voxels).

Cluster Index	Voxels	MAX X	MAX Y	MAX Z	COG X	COG Y	COG Z	Region
Audio alone								
1	37	62	2	-22	57	2	-18	right STS
Audio combined								
1	339	50	-4	-19	49	4	0	right frontal operculum, anterior insula, anterior STG, STS
2	72	40	-19	-19	32	-22	-16	right hippocampal formation
3	46	-54	-10	-22	-54	-11	-22	left MTG/ITS
4	40	-38	-64	-7	-40	-64	0	left lateral temporo-occipital region
5	38	-29	-74	47	-30	-72	49	left IPS
6	34	4	-64	20	11	-67	27	right parieto-occipital sulcus
Motor alone								
1	163	-35	-80	-13	-42	-73	-5	left middle/inferior temporal/occipital region (BA19/37)
Motor combined								
1	147	34	23	-4	40	19	-11	right frontal operculum, anterior insula and STG
2	134	-54	-55	-13	-58	-53	-8	left middle/inferior temporal/occipital region (BA19/37)
3	97	56	5	26	53	5	30	right ventral premotor/ posterior IFG region

Table 4.1 – Searchlight MVPA clusters

A list of all searchlight clusters larger than 33 voxels (the size of a sphere with 2-voxel radius), where all voxels in the cluster surpassed 55% accuracy in at least 8 of the 11 subjects. Separate results are presented for audio and motor decoding in the alone and combined conditions. MAX X, Y, and Z are the coordinates within the cluster that had the highest inter- subject overlap. COG X, Y, Z is the center-of-gravity of the cluster

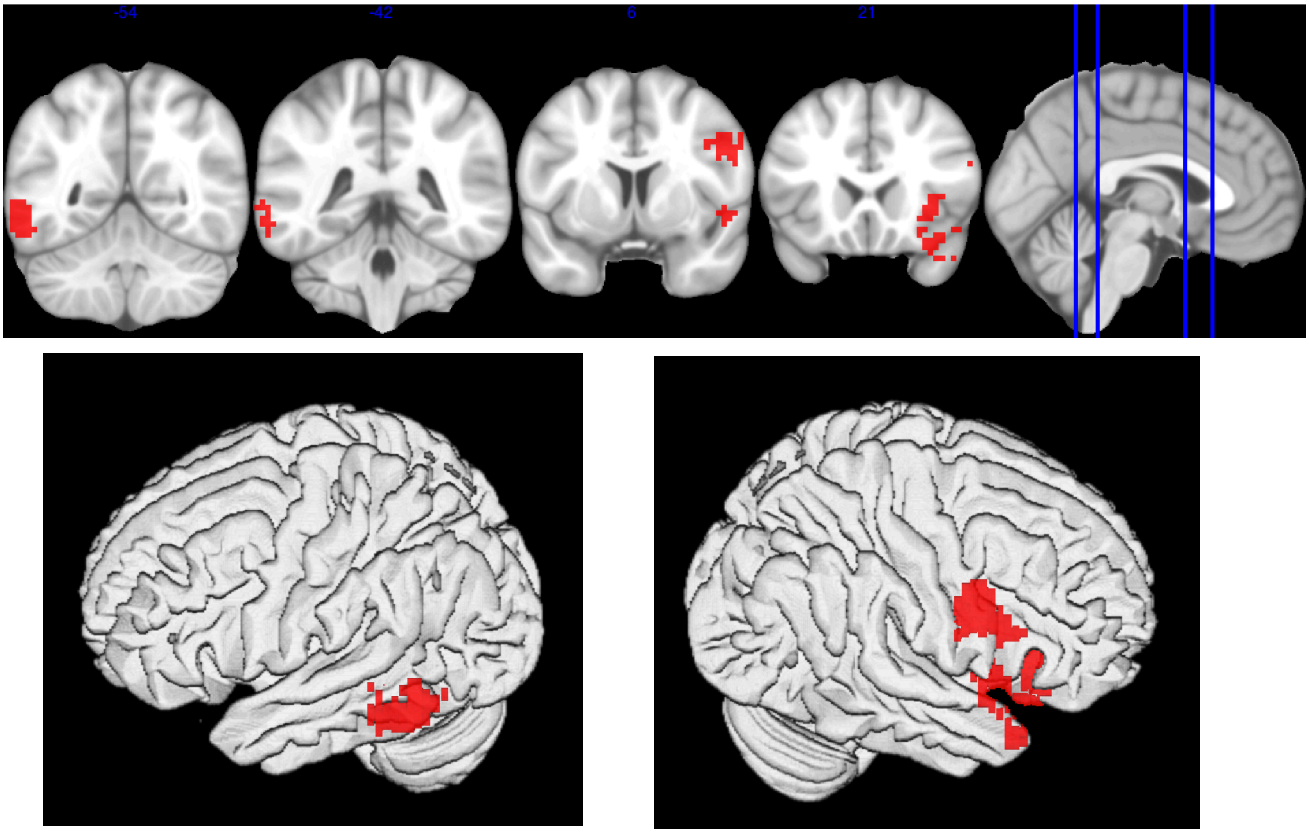

Figure 4.8 – MVPA searchlight results for motor-combined decoding

Voxel clusters that showed a high degree of motor decoding across multiple subjects in the combined condition in a surface rendering (bottom panels) and in coronal sections (top panel: MNI Y-values of -54, -42, 6, and 21 from left to right). Three clusters were observed: a multi-region cluster spanning the right frontal operculum, anterior insula, and anterior superior temporal gyrus; in the left lateral occipital-temporal border (similar to that observed in the motor-only analysis, Figure 4.7); and in the right ventral premotor (PMv)/ posterior inferior frontal gyrus (IFG). The latter cluster was highly overlapping with the PMv/44 region-of-interest, which also showed significant decoding for this condition.

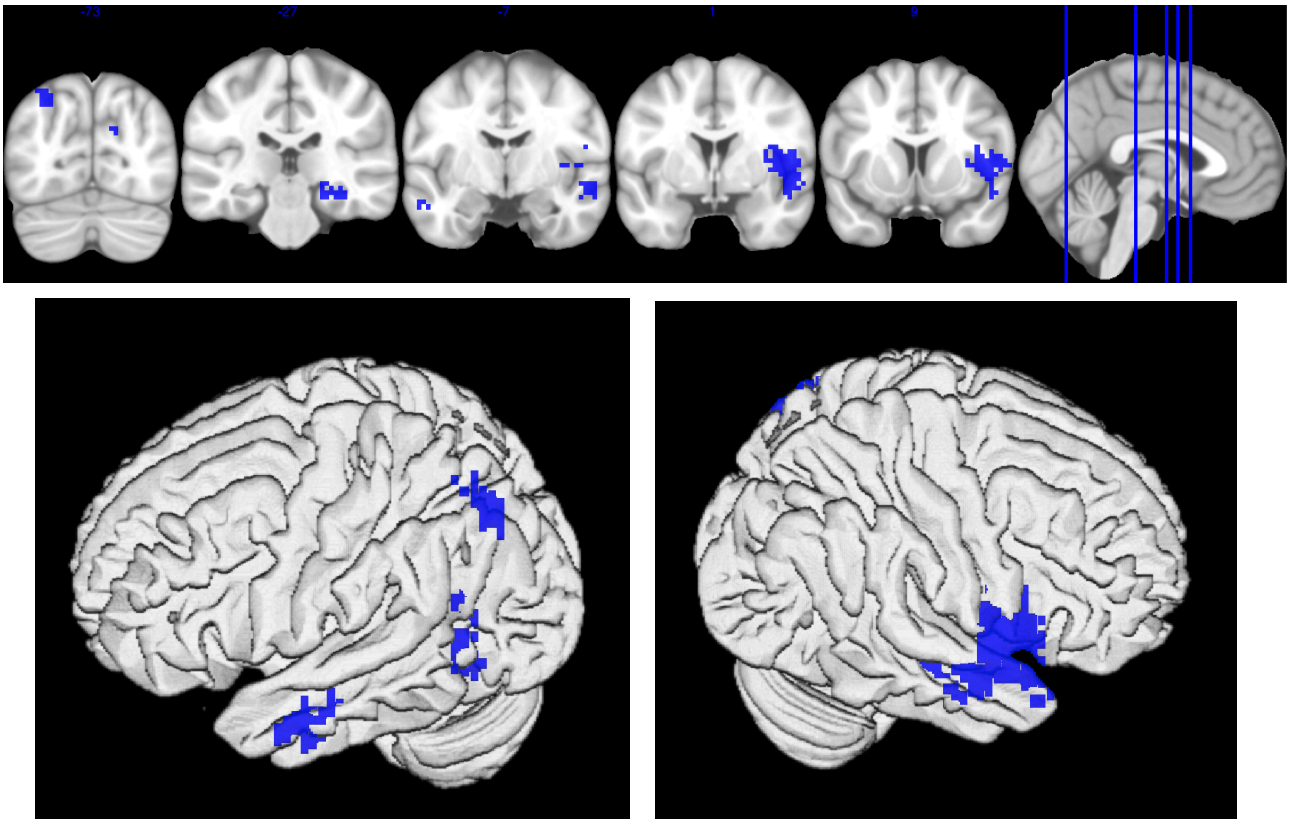

Figure 4.9 – MVPA searchlight results for sound-combined decoding

Voxel clusters that showed a high degree of audio decoding across multiple subjects in the combined condition in a surface rendering (bottom panels) and in coronal sections (top panel: MNI Y-values of -73, -27, -7, 1, and 9 from left to right). Six clusters were observed, with the largest spanning the right frontal operculum, anterior insula, anterior STG, and STS. The other clusters were located in the right hippocampal region; left anterior middle temporal gyrus / inferior frontal sulcus; left lateral temporal-occipital region; left intraparietal sulcus; and right parieto-occipital sulcus.

4.6 Discussion

4.6.1 GLM activation-based results

The GLM analyses were run in order to be able to compare global/regional activity levels between conditions (e.g. auditory > baseline vs. motor > baseline); to look for regions that activated *only* in the combined condition (thus being candidate areas for audio-motor integration); and to compare activated regions with those information-containing regions implicated via MVPA.

The results, broadly, showed expected patterns of activity for the motor-only, audio-only, and combined conditions of the experiment, predominantly in somatosensori-motor and auditory regions of the cortex. And, considering the motor-relevance of the sound stimuli (and vice versa), the large degree of overlap that was observed between the three conditions is not surprising. As stated in the introduction, there is an extensive literature linking perception of various types of auditory stimuli (including music) to parieto-frontal regions that comprise a dorsal stream of processing (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009; Warren et al., 2005; Zatorre et al., 2007). Conversely, sound-implying motion without concurrent audio feedback (such as playing a silent piano keyboard) has been previously found to activate auditory regions in the temporal lobes (Bangert et al., 2006; Baumann et al., 2007).

Most interesting are the regions highlighted in the combined activation analysis, which were not present in either the auditory- or motor-only analyses (*Figure 4.5, bottom row*). The largest clusters of voxels highlighted via this method were found bilaterally near the temporo-parietal junction, in the left posterior terminus of the Sylvian fissure and the right SMG. This region, on the left, overlaps with area spt, which is considered to be a major hub for inter-relating auditory and motor information (Hickok et al., 2003; 2008). Interestingly, the integrative processes subserved by spt are considered to be relatively specific to the vocalization system (Hickok et al.,

2008; Pa & Hickok, 2008) (both in humans and non-human primates), so its observation here suggests a more general role in audio-motor integration. Observation of a proximal, though more superficial, region in the right hemisphere suggests that, for musical stimuli, audio-motor integration may be occurring bilaterally and in parallel. (We note that anatomical regions near the temporal-parietal junction, including the planum temporale (PT), are famously asymmetric (Rubens, Mahowald, & Hutton, 1976) and that functional regions analogous to left PT on the right may be found in more superior positions (Binder, Frost, Hammeke, Rao, & Cox, 1996) because of the upswing of the Sylvian fissure on the right compared to the left (Westbury, Zatorre, & Evans, 1999).) This same analysis also highlighted portions of the right inferior frontal gyrus (IFG), which is positioned at the junction of the dorsal and ventral auditory streams (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009). It is likely this convergence of processing streams that led to this area's integrative audio-motor role and, as discussed later in this section, the right IFG was the predominant cortical region that showed both audio and motor decoding, but *only* for the combined (audio-motor) condition.

Certain results from the GLM analyses were somewhat surprising. For the auditory-only analysis, perhaps the most unexpected was the magnitude of response in the left post-central gyrus (S1), in/around the area representing the right hand. For this condition, participants had been instructed to listen attentively, but to hold their hands in "loose fists" in their lap while being careful to not move their fingers along with the piano sounds. The presence of S1 activity coupled with relatively little activity in M1 suggests that, in the absence of permitted motor activity, participants may have instead been relying upon tactile imagery. Conversely, the motor-only results showed less evidence of *auditory* cortex activity, with small peaks observed in the left MTG and right posterior PT. This was somewhat surprising, as prior experiments employing musical performance without auditory feedback (Bangert et al., 2006; Baumann et al., 2007) did show clearer peaks in auditory regions, albeit in non-primary auditory association cortex. The lack of superior temporal activity here could be an artifact of the experimental design, as motor-

only and motor-combined trials were interspersed pseudorandomly within the same blocks. It is therefore plausible that motor-only following motor-combined trials could be evoking a suppression effect in auditory cortex. (During the motor-only trials, participants had not been instructed to imagine the "missing" sounds.) We did observe, however, motor-only BOLD activation in visual areas such as the lingual gyrus; thus, the eyes-closed no-feedback environment may have led to the recruitment of visual regions for purposes of visuo-spatial imagery.

4.6.2 The STS

As with our previous experiments of musical interval perception (Klein & Zatorre, 2011; 2014), significant involvement of the right STS was observed: both in the ROI and searchlight analyses. While the present experiment was not designed as an explicit test of *categorical* perception, as in the prior studies, it seems likely, in light of the past results, that the STS decoding results are due to abstract/categorical differences between the chords (major vs. diminished), rather than reflecting acoustic differences between the tones (which instead would be expected in and around HG). There was little evidence of STS involvement in the motor task, with non-significant decoding results in the ROI analysis and only a few (and very anterior) voxels appearing in the searchlight analysis near the right STG/STS temporal pole. Thus, we believe these results provide evidence for a ventral stream of auditory processing for sound *identity* (Hickok & Poeppel, 2007; Rauschecker & Tian, 2000). Notably, however, STS auditory decoding in both the ROI and searchlight analyses was markedly enhanced in the combined condition, as compared to the sound-alone condition. (Such was also the case in other pre-defined ROIs such as the IPS, as well as certain other regions only examined via the exploratory searchlight analysis.) This enhancement was somewhat surprising as, for sound decoding in the combined condition, the neural correlates of *both* movements should be present in *both* sides of the classification, potentially making the two sets of BOLD images more similar to one another

(which, in turn, should decrease decoding accuracy). The fact that the opposite was observed suggests that (1) motor information did not substantively "pollute" ventrally-located information-containing patterns in the auditory signal, and (2) something about the active motor task enhanced the global sound decoding, possibly related to attentional load and/or working memory processes. While acknowledging the psychological/philosophical importance of "binding" together various aspects of sensory information, processed in parallel, into a gestalt whole (Clark, 2009; Golledge, Hilgetag, & Tovée, 1996), the two points above suggests that certain aspects of information processed in dorsal and ventral streams may, in fact, be amplified by dorsal/ventral interactions.

4.6.3 The IPS

Similar to previous fMRI experiments of interval perception (Klein & Zatorre, 2011; 2014), we found evidence of left IPS involvement in both the ROI and searchlight analyses, with less (but still present) evidence implicating the right IPS. As with the STS results, IPS decoding was found for sound, but not motor, conditions and was much more evident in the combined than the auditory-alone condition. The overall pattern of these results partially refutes and partially supports our hypothesis that the IPS/dorsal stream would contain information related primarily to motor (not auditory) conditions. First, there was no evidence of IPS motor pattern decoding in the motor-alone or combined conditions. This suggest that, despite its position in the dorsal stream of processing, the IPS is truly acting here as a *sensory* area, with less relevance to pure motoric processing (i.e. top-down motor commands). On the other hand, significant IPS auditory decoding results were only observed in the *combined* condition, suggesting an audio-motor integrative role for this region, in line with repetition suppression findings from a previous fMRI study of pianists by (Brown et al., 2013). The absence of IPS audio-only results could be due to a lack of need for the experimental sounds to be transformed into a "normalized" model, as sound categories in this experiment did not vary in pitch height, as they did in the previous study. This

interpretation is in line with the findings of Foster & Zatorre (2010) that the IPS is recruited only for transposed melodic patterns, and not for patterns with invariant pitch height. The IPS may perform multiple dorsal stream roles: sensori-*motor* transformations (such as for visual reaching/grasping (Cavina-Pratesi et al., 2010)) as well as multi-/supra-modal *sensory* abstractions and transformations (such as visual mental rotation (Harris & Miniussi, 2003; Tagaris et al., 1997), and musical tone-pattern manipulations (Foster et al., 2013)). Such posterior parietal processes may form part of an active working memory process (Berryhill & Olson, 2008; Bledowski, Rahm, & Rowe, 2009; Champod & Petrides, 2007; Smith et al., 1998) or be more automatic in nature.

4.6.4 Activation and information

Before discussing the remaining MVPA results, it is worth briefly discussing a dichotomy between areas highlighted in activation (i.e. GLM) vs. information (MVPA) analyses. As in previous studies from our group (Klein & Zatorre, 2014) and others (Soon et al., 2013), there may be a dissociation between areas highlighted in activation vs. information maps; here, the IPS (as well as medial temporal structures, discussed later) is such a region. This dichotomy was investigated directly by Jimura & Poldrack (2011), who came to largely the same conclusion: there is only a moderate degree of correlation between patterns of region-wide activation and fine-grained information. We note that certain non-searchlight MVPA procedures (examining brain areas significantly larger than those used in the present investigation) often require some sort of feature reduction (i.e. reduction in the number of analyzed voxels) in order to produce accurate decoding (De Martino et al., 2008); this is sometimes accomplished via t-thresholding vs. baseline. While some information-containing regions certainly show activation from baseline, this does not always appear to be the case, possibly due to high levels of "resting" activation in certain areas of cortex. The reverse case of what is described above may also occur: we observed

activation from baseline in areas that did *not* show MVPA effects, including area spt. Such activated —but not decodable— regions may be performing certain global operations that do not differ in a stimulus-by-stimulus manner (in spt's case, such a role may involve the general binding together of a produced sound's auditory/motor aspects into a unified framework). And, of course, some regions may be both activated *and* information containing (we observed such a correspondence in the ventral premotor cortex/IFG, discussed in the next section). Such regions are both broadly activated by the experimental paradigm and differentially utilized between conditions, the latter only being visible at the level of fine-grained spatial patterns. MVPA and univariate approaches thus appear to offer complementary views of the neural bases of mental processing. These two views sometimes highlight overlapping brain areas, but may also reveal separate nodes of larger-scale neural networks.

4.6.5 Frontal lobe results: premotor cortex and IFG

For the combined-condition MVPA —both auditory and motor analyses— premotor and inferior frontal areas contained decodable information. For simplicity, we will refer to dorsal premotor (PMd) as those regions spanning BA6 located superior to the inferior frontal sulcus, although a small portion of the mask extends into the inferior part of the pre-central sulcus. PMv/44, meanwhile, covers the remainder of ventral/anterior premotor cortex, along with the posterior portion of the IFG. Neighboring regions, namely the frontal operculum and the anterior insula, were not examined in the ROI analysis but were highlighted via the searchlight.

First, this effect appears to be a predominantly *right hemisphere* phenomenon, as there was minimal evidence of left frontal decoding (no searchlight clusters, only significant ($p < 0.05$) in the left PMd ROI for the auditory-combined analysis). In the right hemisphere, however, this region is highlighted in both the ROI analysis ($p \leq 0.01$ for the motor and sound analyses in PMv/44 and for the sound analysis in PMd) and the searchlight analysis (the largest clusters from both the auditory-combined and motor-combined decoding spanned this region). Second, the

presence of decodable information in this region seems to depend on audio-motor integration, as there is no evidence of frontal decoding (ROI or searchlight) in either the audio-alone or motor-alone conditions.

The ventral result is consistent with the GLM data indicating a portion of the right IFG as one of the few areas activated by the audio-motor condition, but neither by the sound-alone nor motor-alone conditions. PMv/44 may contain circuitry that is simultaneously coding for sound and motor properties, yet does not contain significant information for passively perceived sounds or movements unlinked to auditory feedback. This preference for combinatorial representations jibes with two stream speech perception models (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009) that point to the (left) ventral posterior frontal cortex as being a ventral/dorsal convergence zone (with the former and latter models labeling this region as belonging to the dorsal vs. the ventral stream, respectively). In monkeys, Petrides & Pandya (2009) have demonstrated distinct ventral and dorsal anatomical connections to the homologues of area 44. Speech research comparing perception and production has shown neighboring yet distinct ventral premotor regions which are preferentially activated by one or the other (Wilson, Saygin, Sereno, & Iacoboni, 2004), suggesting a convergent hierarchy. A TMS study showed enhanced motor corticospinal excitability following perception of hand-related action sounds, further suggesting integrative sensori-motor processing (Aziz Zadeh, Iacoboni, Zaidel, Wilson, & Mazziotta, 2004), with a similar amplification effect demonstrated for music (D'Ausilio, Altenmuller, Olivetti Belardinelli, & Lotze, 2006). Thus, while the IPS results seem to indicate the presence of primarily auditory (i.e. sensory) information, PMv/44 in contrast may be serving as a combinatorial audio-motor region exerting top-down influence over the dorsal stream. We believe that the present right-lateralized results provide similar evidence for *non*-speech stream integration in the right hemisphere, a theory supported by other music research which selectively implicated this ventral frontal region for action and action-coupled perception (Chen, Penhune, & Zatorre, 2008).

Relatedly, we note that mirror neurons were first observed in the analogous ventral region of the monkey brain (di Pellegrino et al., 1992) and their function, though contentious, is certainly related to sensory-motor interaction. Such a system has also been strongly linked to the auditory system in primates (Keysers et al., 2003; Kohler, 2002) and humans (Gazzola, Aziz Zadeh, & Keysers, 2006; Lahav et al., 2007). Others have suggested that an inferiorly-located network, including the vPM/IFG region, performs a "mirror-matching" function (Haslinger et al., 2005) that translates multi-sensory perception to action. Neurons with "mirror-like" responses have been found in monkeys in the premotor cortex (Gallese et al., 1996), parietal lobe (Fogassi et al., 2005), and STS (Jellema & Perrett, 2003), with similar functional imaging results found in humans (Rizzolatti & Craighero, 2004), raising the question of how these regions functionally differ from one another. Compared to the dorsal stream areas, the STS appears significantly less involved in the *action* side of things (Rizzolatti & Craighero, 2004), which is in accordance with the present decoding results. And compared to parietal sites (including the IPS), the PMv/44 sits in a "privileged" position at the junction of high levels of both dorsal and ventral streams (Rauschecker & Scott, 2009). Such access to highly abstract information of both varieties may underlie this region's robust "mirror" properties.

Right PMd, unlike PMv/44, was only highlighted (a) in sound (but not motor) decoding (although still exclusively in the combined condition) and (b) observed via the ROI (but not searchlight) approach. A potential explanation for (b) could be related to the size of the ROIs: ~250-350 voxels in PMv/44 and ~1200-1500 voxels in PMd (variable for each participant's particular brain anatomy). It is likely that the information contained in PMd is spread throughout the larger region, leaving less visible to individual searchlight spheres. Conversely, the size of PMv/44 is close to that of the searchlight (123 voxels), which explains the correspondence of the searchlight and ROI results in this region. PMd is thought to be involved in processing arbitrary stimulus-response associations, primarily linking sensory stimuli with movements (Hoshi &

Tanji, 2007; Petrides, 1985; Petrides, Alivisatos, Evans, & Meyer, 1993; Wise et al., 1996), which is clearly relevant to piano performance. Dorsal/rostral BA6 and adjacent BA8 together comprise the "posterior DLPFC," functionally distinct from the "mid-DLPFC" comprised of the more anteriorly-positioned areas 9, 46 and 9/46, which are thought to perform a larger role in tasks that require active working memory and self-monitoring (Petrides & Pandya, 1999). The rostral portion of dorsal premotor cortex is thought to subserve functions more similar to prefrontal than other motor cortex (Muhammad, Wallis, & Miller, 2006). The PMd/8 region has been implicated in the action representation of musical sequences (Lahav et al., 2007) and specifically linked to conditional associative memory for musical chords (Bermudez & Zatorre, 2005).

The auditory combined-condition decoding results observed in the present experiment may represent this sort of flexible audio → motor mapping, which varied from condition to condition (i.e. the same sounds were produced by multiple fingering patterns). Like the PMv, the PMd sits at an interesting crossroads between various pathways: in this case, at a high level of the sensory-dorsal stream (Rauschecker & Scott, 2009) and at the bottom of a hierarchical rostro-caudal axis in the frontal lobe that subserves goal-directed behavior (Badre & D'Esposito, 2009). Unlike the PMv, the PMd is not a major node in *ventral* stream processing (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009). And via its top-down effects on the PMv (Hoshi & Tanji, 2007), the PMd, while being positioned at a level too abstract for *direct* participation in mirror-like processing, may subserve the sort of flexibility required for non one-to-one sensori-motor mappings. The present results, which show stronger evidence of information in ventrolateral frontal regions, could be a product of the experimental paradigm: each particular motor-auditory association was determined at the very beginning of each fMRI block and then carried through for several trials. Thus, the PMd may have initially "set a program" in the more ventrally located

region, the results of which persisted throughout the block and were more visible to BOLD scanning.

4.6.6 Other information-containing areas

While there were a few non-hypothesized regions that showed searchlight decoding peaks, the largest and most pervasive among the various conditions were found in (1) the right hippocampus / parahippocampal gyrus (*Figure 4.9*), and (2) the left ventral occipital-temporal cortex (vOT) (*Figures 4.7, left, and 4.8*), which spans Brodmann Areas 37 and 19 on the ventrolateral aspect of the cerebrum.

Medial temporal lobe (MTL) structures that are critical for episodic memory are a target of the various sensory ventral streams (Amaral, Insausti, & Cowan, 1983; Munoz-Lopez et al., 2010; Squire & Zola-Morgan, 1991), so it is not surprising to have observed involvement of the right hippocampus and surrounding cortex in auditory perceptual decoding. (This region was not implicated in any of the motor decoding analyses, indicating its specificity to ventral stream processes.) A recent study (Kumar et al., 2014) using similar analytical techniques (MVPA with ROIs) demonstrated that newly-learned acoustic patterns could be decoded from information present in the hippocampus (as well as from the planum temporale, with an effect in the STS just missing statistical significance). A contemporary view of the episodic memory-hippocampal system suggests that the hippocampus is a general encoder of the *sequence* of events (Eichenbaum et al., 2012): the hippocampus is thought to represent prior knowledge of that sequence and acts to retrieve (and possibly modify) its contents. As our auditory stimuli differed primarily in terms of the sequences of notes (C then E then G, or B then D then F), the hippocampal-based information could represent such a memory trace being re-activated for each trial.

The vOT decoding peaks —found in the motor-alone, motor-combined, and audio-combined conditions— were also not expected. The lack of audio-alone results here, combined with the much larger results in motor-alone/combined vs. audio-combined, suggest that this region's function has more to do with motor sequencing than sound patterns. This region has broadly been implicated in a range of functions, including those related to language (Abrahams et al., 2003), memory (Slotnick & Schacter, 2004), and vision (Beer, Blakemore, Previc, & Liotti, 2002). Several studies have tied this region to aspects of visuospatial analysis, including visual discrimination of hand/finger gestures (Hermsdörfer et al., 2001), analysis of motion (Deutschländer et al., 2002; Dupont, Orban, De Bruyn, Verbruggen, & Mortelmans, 1994), and retrieval of structural knowledge about objects (Kellenbach, Hovius, & Patterson, 2005). (BA19 includes visual area MT/V5, which has been highly-implicated in processing of visual motion (Born & Bradley, 2005).) Our active motor task, conducted with closed eyes (and partially without auditory feedback), may be recruiting aspects of visual working memory (Baddeley, 2000) as participants perform the 10-note piano sequence, in the process tapping into these visual association structures at the OT border.

4.6.7 Summary

In summary, we employed a combination of univariate and multivariate analyses, which together highlighted a network of brain regions implicated in musical auditory representations, motor programs, as well as the interaction of the two. For the perceptual (auditory) decoding, information-containing regions included the right STS (with some degree of decoding also observed contra-laterally) and the left IPS: two areas previously-implicated in the categorical processing of musical intervals (Klein & Zatorre, 2011; 2014). Separately, right inferior frontal regions —the ventral premotor and ventrolateral prefrontal cortex— were highly implicated in

both auditory and motor processing. The audio-motor interactive condition seemed to greatly enhance decodable information content across the brain: in temporal, parietal, and frontal areas, likely related to the high degree of two-way connectivity between the various nodes of the two streams.

4.7 Acknowledgements

We thank the staff of the McConnell Brain Imaging Centre for help in designing scanner sequences and acquiring imaging data. This work was supported by funding to RJZ from the Canadian Institutes of Health Research, and from the Canada Fund for Innovation.

Chapter 5 - General discussion

5.1 Summary of motivations and findings

This thesis, via a set of three experiments, investigated the neural correlates for perception of musically meaningful sounds. This unifying theme was tested via a variety of online and offline tasks, music stimuli, MRI parameters and analytical methodologies. The global motivation was to gain insight and understanding into the ways in which the cerebral cortex parcels and processes complex non-speech auditory information. Quite a lot of recent (and not-so-recent) research has been conducted into cortical perception processing of both visual and auditory speech stimuli. Thus, issues of generalizability have emerged, as what's true in the visual modality may have variable relevance to non-visual perception, and what's true for speech may or may not be relevant to the auditory system more globally. Using non-speech, yet cognitively salient sound stimuli allowed us to compare perceptual processing *intra*-modally against speech, while also examining poly-modal sensory and motor responses in the parietal and frontal lobes.

To this end, we exclusively enrolled expert musicians with multiple years of musical training, who also kept up a current practice. This sample was selected because we were primarily interested in brains that have been shaped over time to perform complex operations on information that is highly relevant to the listener. While essentially all humans are experts in perceiving and producing speech in their native language, the story is different for music (despite music's universal presence in all known human societies (Brown, 1991; Patel, 2007)). In contemporary times, only a fraction of the general population has significant experience playing and performing with a musical instrument. And, although listening to music is quite common in the general population, only trained musicians have shaped their listening patterns via sustained practice in a way that enables them to quickly and effortlessly recognize the quality of a musical chord, violations in key, identity of a specific instrument, etc. Thus, examining the behavior and

neurobiology of musicians allows one to probe questions related to what the nervous system is *capable* of, rather than universal *specifics* of its functioning.

Studies 1 and 2 were highly focused on *categorical perception* (CP) of musical intervals and chords. Similar to the perception of stop consonants in speech (Liberman et al., 1957), perception of intervals and chords has been found to exhibit the hallmarks of behavioral CP: (1) identification of sounds over a wide physical continuum as members of a single category (with sharp boundaries between such categories), and (2) discrimination functions suggesting perceptual "compression" of same-category, but physically distinct sound pairs, with concurrent perceptual divergence of sounds that do not belong to the same category, even if, acoustically, the are fairly similar to one another (Liberman et al., 1957; 1967). These intervals/ chords are highly relevant to both music perception and performance: they comprise the basis for structured syntax, melody and harmony. Thus, the focus of the first two experiments was on (a) behavioral demonstration of CP for musical intervals, and (b) verification of the perceptual neural pathways by which these sound categories are extracted and transformed. Study 3 built off these concepts and was motivated by the theory that there are two segregated cortical perceptual streams that serve divergent purposes: (1) a ventral stream for perceptual feature extraction and (2) a dorsal stream for sensory-motor interaction. Whereas studies 1 and 2 used primarily perceptual paradigms, study 3 mixed perception with motor action, requiring participants to generate musical sequences on an MR-compatible piano keyboard. Thus, we hoped to probe how these multiple neural perceptual codes would be differentially affected by top-down executive commands originating in the frontal lobe.

Study 1 was actually two experiments in one: (1) an active task, in which participants explicitly discriminated between two chords, and (2) an adaptation protocol, in which one chord was presented several times, followed by final chord that was either identical to or distinct from the first set. Both (1) and (2) used the same sound set, with the primary difference being the effortful vs. automatic nature of the protocol. The sound set was comprised of three-tone chords that ranged from true minor to true major via several mistuned exemplars created at regular

steps. The mid-most exemplar (halfway between minor and major) was used to create an orthogonal sound set, in which all three tones of the chord were roved in absolute pitch space (i.e. "pitch height"), manipulating the perception of pitch, but *not* of category. These stimuli thus allowed for various comparisons dependent on the relationship of a chord pair: two physically non-identical sounds from the same perceptual category at the same pitch level; two physically non-identical sounds from two perceptual categories (i.e. minor or major); or two physically non-identical sounds that differed along the orthogonal (i.e. non-categorical) pitch dimension. As our primary interest was in categorical perceptual processing, we contrasted brain activity for (1) multi-category > single-category trials, and (2) categorical > orthogonal trials. The discrimination protocol revealed a significant peak of brain activation in the right middle/posterior superior temporal sulcus (STS). The adaptation protocol also revealed a peak in the right STS, located more anteriorly, plus a second peak in the left intraparietal sulcus (IPS). Based on prior auditory imaging results, we interpreted these sites as belonging to the ventral and dorsal processing streams, respectively.

Study 1 attempted to identify regions of cortex that *activated* for categorical comparisons: trials that evoked between-category comparisons (whether such comparisons were explicit or automatic) vs. comparisons that could not make use of categorical information. The STS and IPS results, present in certain contrasts but not others, paired with fMRI results from the speech literature, suggested to us that the CP-related BOLD signal is fairly subtle. Put differently, *all* of our chord stimuli seemed to be activating these perceptual circuits, with the multi-category conditions doing so only slightly more so. Thus, Study 2 was designed to look not for *activation* contrasts, but instead for regions where multi-voxel *patterns* in the BOLD response could be used to tell apart the perceptual conditions that generated them (i.e. "decoding") (Haynes & Rees, 2006). Such multivariate pattern analyses (MVPA) do not require region-wide activation between two conditions, thus allowing us to directly compare the BOLD response to two intervals, something that was not feasible using the univariate methods from Study 1. A secondary benefit of using MVPA was that it allowed for simplification of the in-scanner

experimental protocol, as participants were no longer asked to explicitly discriminate between stimuli (a process that involves not only perception, but also working memory and decision making). Study 2 used two-tone sequential intervals (minor thirds, major thirds, perfect fourths) and, as in Study 1, utilized an absolute pitch-based orthogonal dimension (i.e. any of the three interval types could start on c-natural, c-sharp, or d-natural). Thus, we could evaluate abstract category membership: a quality that did not depend on the presence of a specific set of tones with invariant pitches, but rather on the relation between the tones. Decoding could be performed to dissociate any pair of interval categories, regardless of their absolute pitch. The two primary regions of the brain found to decode between interval categories converged with the activation results of Study 1: the right STS and the left IPS. The analogous regions of the left STS and right IPS were also highlighted, although to a lesser degree. No other brain regions were found to decode for interval quality, and no significant decoding was observed anywhere in the brain for the orthogonal absolute pitch dimension.

Studies 1 and 2 showed highly convergent spatial results, despite the use of dramatically different experimental and analytical protocols. These STS and IPS nodes are in line with established theories of two streams of perceptual processing: a ventral stream thought to underlie conscious perception and feature extraction, and a dorsal stream thought to subserve mental transformations (both sensory and sensory-motor). However, while these studies allowed us to infer two spatially segregated processing streams, they did not allow us to dissociate the streams based on their function (i.e. separate hypothetical contributions of streams to sound feature extraction vs. transformation/recoding into a supra-auditory space). Thus, Study 3 built upon the perceptual foundation of the first two experiments, while also explicitly involving the motor system. Piano-based motor tasks, designed to be orthogonal to the sounds they produced, were postulated to primarily affect dorsal stream processing in the parietal and/or frontal cortex. MVPA was once again employed, this time to dissociate motor commands as well as auditory percepts. Some of the same areas were highlighted for auditory decoding (the bilateral STS, the left IPS), alongside a newly highlighted region in the right ventral premotor (PMv)/ ventrolateral

prefrontal cortex (VLPFC) that contained information relating to both sound and motor content. Decoding in these frontal regions seemed to require audio-motor interactivity, as significant results were not found in separate analyses in which sounds were passively perceived (i.e. no motor task) or the keyboard was played silently without auditory feedback.

The overall pattern of results suggest different roles for the STS, IPS and PMv/VLPFC regions in music perception. The consistency of the STS results —across active and passive paradigms, for simultaneously- and sequentially-presented stimuli, and for stimuli that did or did not rove in absolute pitch— leads to the conclusion that it is a critical node in the auditory ventral stream that subserves feature extraction and stimulus identification. The IPS, the engagement of which was also observed in all three studies, appears to be recruited in situations that either (1) require normalization to a standard (and potentially supra-auditory) space (e.g. relative > absolute pitch processing) or (2) putatively tap into a motoric code, potentially to disambiguate perceptual circumstances. Related to point (2), the frontal regions highlighted in Study 3 contain information related to both auditory and motor properties (but only for interactive conditions that pair motor commands with auditory results). Thus, we believe the right PMv/VLPFC subserves auditory-motor integrative processes, propelled by this region's positioning at the intersection of the ventral stream, dorsal stream, and frontal circuits, the latter involved in planning and directed behavior. The analysis of each of these three areas will be expanded in the upcoming section.

5.2 The role of the superior temporal sulcus in the ventral stream

As stated above, the STS (particularly on the right) was a constant finding across the three studies. Auditory imaging studies in humans (e.g. DeWitt & Rauschecker (2012); Liebenthal et al. (2005)), as well as findings from non-human primates (see review by Rauschecker & Scott (2009)), lead us to believe that the ventral stream, including belt/parabelt structures, is involved

in perceptual feature extraction, perhaps even containing the "maps" that distinguish perceptual categories from one another. Here, I will address three separate issues: the role of this ventral STS region from more primary regions on the superior temporal gyrus (STG); functional differences between STS subregions (i.e. anterior vs. posterior); and differential roles for the right and the left STS.

The potential contributions of various nodes of the auditory ventral stream, as well as the ongoing debate over hierarchical processing (or the lack thereof) were introduced in Chapter 1 (Sections 1.2.3 and 1.2.4, respectively). The three studies presented here provide compelling evidence for the STS's role in complex perceptual processing of music stimuli, but no such evidence for the STG (either "core" areas on Heschl's gyrus (HG) or surround belt regions). As discussed in Sections 1.2.3 and 1.3.2, core auditory regions have been linked to broad spectro-temporal analysis of sound, whereas certain portions of the surround belt cortex (lateral HG, anterolateral planum temporale (PT)) have been implicated specifically in pitch (Patterson et al., 2002; Penagos et al., 2004), the extraction of which presumably relies upon the more general frequency analyses conducted in HG. This can then be seen as two nodes in a stream: the "downstream" pitch regions processing information that has been handed off by the "upstream" primary region. Musical intervals/chords in turn require integration of multiple pitches (two for intervals, three or more for chords). Thus, it seems highly plausible that the STS is downstream from the pitch areas, which also is an anatomical fit (i.e. amongst those regions, the STS is furthest from HG and receives input from belt areas (Rauschecker & Scott, 2009)). As discussed in the middle thesis chapters, the lack of results with primary auditory cortex is potentially due to the high degree of physical similarity between the sound conditions (e.g. overlapping or identical tones in Studies 2 and 3, simple sounds generated from sine waves and harmonics in Studies 1 and 2). While I do not discount some potential contributions of "early" auditory areas to complex musical percepts, I believe that the results presented here make a strong case for the STS as the *critical* ventral stream node underlying such processes. Supporting this case, the STS (on the

left) has been the primary structure highlighted for phonemic processing (Joanisse et al., 2007; Liebenthal et al., 2005), which, behaviorally, shows remarkable similarity to musical intervals in various measures of categorical perception (Burns & Ward, 1978; Liberman et al., 1957). These results also jibe with single-cell recordings from monkeys, showing neurons that clearly seemed to be integrating spectral and/or temporal information received from upstream processes (Rauschecker, 1998). As will be discussed later in this section, a spectral vs. temporal distinction may be at the core of right vs. left temporal dissociations.

Related to the STG/STS discussion is the question of differential functional contributions of various sub-regions of the STS that lie upon its anterior/posterior axis (see review by Hein & Knight (2008)). In the speech literature, the roles of the middle/posterior STS (mpSTS) vs. anterior STS (aSTS) has been fairly contentious. (Separately, very posterior regions of the STS are thought to serve more multimodal audiovisual processes, see for example Man, Kaplan, Damasio, & Meyer (2012).) On the one hand, categorical phoneme perception research like that of Joanisse et al. (2007) and Liebenthal et al. (2005) supports phonemic mapping in the left mpSTS. This more posterior focus for the phonological network is also argued for by Hickok & Poeppel (2007). On the other hand, research into speech *intelligibility* has most strongly highlighted left anterior sites (S. Evans et al., 2014; Scott et al., 2000), although certain right and posterior left temporal regions have also been observed (Evans et al., 2014; Okada et al., 2010; Spitsyna, Warren, Scott, Turkheimer, & Wise, 2006). Intelligible speech, of course relies upon phonemic extraction, but, compared to unintelligible controls, also represents properties related to lexical access, syntax, prosody, etc. Thus, it is not surprising that these studies have highlighted a network that is more extensive (yet mpSTS inclusive) than that observed for CP. Separately, in both hemispheres but with a right predominance, voice perception has been linked to both anterior and posterior STS sites (Belin et al., 2000; Formisano et al., 2008), with different

functional aspects (e.g. familiar > non-familiar speakers) linked to various sub-regions along the anterior-posterior axis (Kriegstein & Giraud, 2004).

In the three studies that comprise this thesis, right temporal locations in the anterior *and* middle/posterior STS were highlighted (alongside less robust, but still present, evidence of left STS involvement). Study 1 highlighted the mpSTS for active discrimination and the aSTS for the adaptation/repetition protocol. Study 2 highlighted the mpSTS, whereas Study 3 highlighted aSTS regions. It seems to be the case that conditions with a greater number of tones (potentially with more structural complexity and/or presented over longer periods of time) may be most correlated with more anterior activation/information foci. This would explain the dissociation in Study 1 (3 chords for discrimination, 5 for adaptation), as well as that of Study 2 (6 tones / 2 unique) vs. Study 3 (10 tones / 3 unique). While somewhat speculative, this idea does inherently fit with network/stream models, where regions that lie further from primary auditory cortex are likely to be further "downstream" (Hickok & Poeppel, 2007), and thus subserve more complex processes. Similarly in speech, regions which are believed to have lexical/semantic/syntactic functioning are located more ventral/anterior to the mpSTS (Hickok & Poeppel, 2007). As there is support for hierarchical processing stages proceeding from HG to PT to STS (Kumar, Stephan, Warren, Friston, & Griffiths, 2005), the right mpSTS thus appears to be the best candidate for "first extraction" of music category information, as it is both the most "upstream" STS site and implicated in the simplest of our experimental conditions.

The dissociation of function between the left vs. right superior temporal lobes is a topic that I covered at length in the introduction (Section 1.3). It is clear that both music and speech activate extensive bilateral networks and that disparities in right vs. left processing may have less to do with "music vs. speech" per say, as opposed to more low-level specializations (Giraud & Poeppel, 2012; Poeppel, 2003; Zatorre & Belin, 2001). However, when considering *categorical* sound processing, Studies 1 and 2 presented here, alongside speech CP research (Joanisse et al., 2007; Liebenthal et al., 2005) do suggest a clear right vs. left dissociation in the STS. I believe

that the STS, both in the right and left hemispheres, is a critical region where bottom-up sensory information meets "over-learned" top-down representations stored in long-term memory. A study by Leech et al. (2009), in fact, demonstrated that creating such categorical memories via a learning paradigm was correlated with increased activity in the pSTS. Once extracted, such mental units (phonemes, intervals, etc.) can be used to create more complex structures, with the latter processes potentially invoking circuitry in both hemispheres. The observation of weaker but fairly symmetrically located responses in the left STS (e.g. Study 3), suggests that the right temporal lobe is not acting in isolation and is likely communicating with analogous bilateral structures via commissural fibers.

Summarizing this section, I present evidence that the right STS is the crucial structure for perception of musical categories. The right pSTS is likely receiving information from auditory belt regions in the STG, producing specific and robust responses that are unique to particular musical categories, and passing along this processed information to downstream regions (likely including more anterior regions of the STS). While the left pSTS receives projections from similar belt areas, it does not appear to contain robust maps for musical categories, instead displaying qualities relating to speech categories and other sounds that are spectro-temporally related to speech.

5.3 The intraparietal sulcus and the auditory dorsal stream

Like the STS, all three studies presented in this thesis have highlighted the left IPS, generally in the more superior/anterior portions of the sulcus. As stated in the introduction, regions of the left parietal lobe, including the IPS, have been highly-implicated in a dorsal stream of processing, both for vision (e.g. reaching/grasping (Cavina-Pratesi et al., 2010; Goodale, 2005) and for auditory speech (Rauschecker & Scott, 2009), with speech primarily linked to the inferior parietal lobule (IPL). The posterior parietal region (excluding the superior parietal lobule),

generally, is considered to be not only a dorsal stream structure, but also a poly-modal sensory region (Grefkes, Weiss, Zilles, & Fink, 2002), particularly for visual and tactile information (Bodegård, Geyer, Grefkes, Zilles, & Roland, 2001; Buelte et al., 2008; Grefkes & Fink, 2005). (The posterior end of the STS, mentioned above as an audio-visual integrative region (Man et al., 2012), actually extends to posterior portion of the IPL, where it is surrounded by the Angular Gyrus (AG).) Meanwhile, there also exists a dorsal vs. ventral connection gradient between the parietal and frontal lobes: rostral inferior parietal regions connect to the VLPFC, whereas more caudal/superior parietal cortex connects to the posterior DLPFC (Friederici, 2011; Petrides, 2005).

As discussed in detail in Chapter 4, the dorsal/ventral dissociation, at least in posterior-lateral *frontal* cortex, is thought to underlie distinct yet related processes for more "flexible" vs. more "direct" mapping. Whereas dorsal circuitry (including the posterior DLPFC and dorsal premotor cortex) is thought to subserve more abstract sensory-motor processes (Hoshi & Tanji, 2007; Zatorre et al., 2007), ventral regions (the VLPFC, ventral premotor cortex) are thought to drive more direct links between perception and action (Hoshi & Tanji, 2007; Keysers et al., 2003; Kohler, 2002), with the latter being more relevant to speech (Hickok & Poeppel, 2007). Ventral *parietal* regions of the left hemisphere also have been directly linked to speech dorsal stream processing, most notably area spt (technically a border structure spanning the junction of the most caudal portion of the PT with the IPL's supramarginal gyrus (SMG)) (Hickok et al., 2003; 2008), which has anatomical links with the left mpSTS (Hickok & Poeppel, 2007). It follows that musical sensory information, which requires more flexible audio-motor mapping than does speech, would rely upon more dorsally-located parietal sites as it courses through a dorsal perceptual processing stream. Considered differently, music information may require a recoding into supra-auditory/ multisensory space prior to being passed along to frontal planning/movement circuitry, with the IPS putatively the main substrate for such hetero-modal

(but still firmly *sensory*) processing. Separately, there is known to be a functional gradient between the superior parietal lobule (SPL), more involved in spatial processing, and the IPL, to which is ascribed primary non-spatial functions (Husain & Nachev, 2007). The IPS, comprising the border between the SPL and IPL, thus is well positioned to serve a quasi-spatial function, which is perhaps very befitting of musical perceptual processing. As discussed in Chapter 2, musical intervals differ in a dimension of perceptual "size" (5ths being larger than 4ths, etc.) (Rusconi, Kwan, Giordano, & Umilta, 2006). There is no obvious analogous dimension for phonemic categories in speech, despite a similar need to convert such perceptual categories into motor coordinates.

The three studies presented here all show strong evidence for left IPS involvement in music perception, with more variable evidence for right IPS involvement (primarily in Studies 2 and 3). Such a pattern of results is somewhat at odds with the common perception that music processing is generally right-hemisphere dominant, but largely fits with speech perceptual models that posit that the auditory dorsal stream is in fact much more left dominant than the auditory ventral stream (Hickok & Poeppel, 2007). Few studies of music perception have implicated the parietal dorsal stream, with the best examples being those of Foster et al. (Foster et al., 2013; Foster & Zatorre, 2010). In those studies, tasks requiring melody transposition and reversal, respectively, each highlighted bilateral IPS regions overlapping with those of the Studies 1-3. Unlike the STS and (as will be discussed) the inferior frontal regions, which both show strongly right-lateralized results, the bilateral IPS may be working in a relatively unified manner via its interconnection by way of the posterior corpus callosum. While the right parietal lobe is known to be dominant in general spatial processing (M. Kim et al., 1999; Vallar, 1998), the left is known to be preferentially involved in the manipulation of symbolic information (e.g. numerical processing (Dehaene et al., 2003)). Music perception may involve both: highly abstract "spatial" processing paired with a certain degree of arbitrary this-to-that mapping (e.g. the multiple legal mappings

between specific tones and the gestures required to produce those tones, which show an even greater degree of variability when considering different musical instruments).

Above, I have described various theoretical contributions of the IPS, with specific ideas for how they may relate to music processing. However, before moving on to the frontal lobe results, I would like to briefly clarify the putative role(s) of this region. First, I am a supporter of the parietal dorsal stream as the "do" pathway, as put forward by Goodale & Milner (1992) (visual) and Warren et al. (2005) (auditory). That said, I also firmly believe that the parietal lobe contains primarily *sensory/perceptual* structures, leaving the —actual— "doing" for prefrontal and motor cortex.

So what, then, is a "do" pathway that doesn't definitively "do"? To begin with, Study 1 (adaptation experiment) and Study 2 have shown that the IPS may be recruited in quite passive conditions, with no task-relevant working memory components. Thus, it seems that the IPS, like the ventral stream, can be engaged via automatic processes and is not exclusively a working memory area (as it has classically been thought of (Klingberg, 2006; Naghavi & Nyberg, 2005; Zimmer, 2008)). However, studies such as that by Champod & Petrides (2007) have clearly demonstrated involvement of the posterior parietal cortex (PPC, which includes the IPS) in working memory, implicating this region in the manipulation of information more so than monitoring such information (with the latter linked to the DLPFC). Thus the IPS may compute a multidimensional "space" into which sensory information can be transformed, and within which such information may be arranged and rearranged as needed. Such processes may be more automatic (as demonstrated in Studies 1 and 2) or more volitional, as demonstrated by the studies by Foster et al. Thus, the IPS regions from the first two studies may underlie processes directed at placing sensory information into an abstract code that serves as a common frame of reference. Such a "space" can be used to make implicit judgements about ordinal qualities of intervals that rely upon their relative size (i.e. a 5th is "larger" than a 3rd, even if the absolute pitches that comprise the 5th are lower than those of the 3rd).

Relatedly, such an abstract "space" with putative supra-auditory coding is an ideal candidate for interfacing with motoric (and more abstract planning) circuitry in the frontal lobes. Such an interface, perhaps serving functions more abstract in nature than the audio-motor interfacing believed to occur in area spt for speech processing (Hickok et al., 2003), may then act as a "translator" of sorts between unimodal sensory and motor cortices. Such a property may be called upon as needed in order to disambiguate perceptual information via motoric expertise, a function which has recently been tied to the auditory dorsal stream (Du et al., 2014; Obleser, Wise, Alex Dresner, & Scott, 2007a). While highly speculative, such interactions could be driving IPS sound decoding in Study 3's combined auditory-motor condition, which showed enhanced informational content compared to the passive auditory condition.

5.4 Two streams, convergence, and frontal cortex

The right ventral premotor cortex (PMv, ventral BA6, possibly extending to the border with dorsal BA6) and ventrolateral prefrontal cortex (VLPFC, BA44, possibly extending to area 45) were strongly implicated in Study 3, both for auditory and motor decoding. There was significant spatial overlap between the auditory and motor searchlight maps in the right frontal operculum (VLPFC cluster), plus observation of a more superior region for motor decoding near the posterior terminus of the inferior frontal sulcus (IFS), which typically serves as the border between what is defined as "dorsal" vs. "ventral" lateral frontal cortex. The PMv and VLPFC are sometimes grouped together under one or the other name, and the spatial limits of fMRI, as well as the analytical methods employed (which ascribe informational content over a spherical region to its centermost voxel) make it difficult to dissociate the processes occurring in one region vs. the other. Thus, I will discuss both in parallel, while noting that, as determined via non-human primate research, there are clear differences between these regions in terms of cytoarchitecture, function, and long-distance connectivity (Petrides, 2005).

The *left* VLPFC contains Broca's Area (Broca, 1861), perhaps the single most famous cortical region in the history of human neuroscience. Damage to Broca's area has been historically linked to speech production deficits (Broca's Aphasia), which starkly contrasts with the kind of speech comprehensions deficits observed following damage to the left posterior superior temporal lobe (Wernicke's Aphasia (Wernicke, 1874)). However, more recent lesion data (Basso, Casati, & Vignolo, 1977; Blumstein, Baker, & Goodglass, 1977) suggest a more complex role for Broca's area, as damage to this region also results in certain comprehension deficits. Neuroimaging studies of speech likewise confirmed a role for Broca's region in perceptual processes (Zatorre et al., 1992), including for complex syntactic processes (Friederici, 2011). Broca's area and its right hemisphere analog have also been repeatedly implicated in music processing, particularly with regard to syntax (Abrams et al., 2011; Koelsch, 2006; Levitin & Menon, 2003; Maess, Koelsch, Gunter, & Friederici, 2001).

In a separate field of research, the VLPFC has been repeatedly linked to controlled retrieval of knowledge, with the left VLPFC associated with semantic knowledge (Barredo, Öztekin, & Badre, 2013; Dobbins & Wagner, 2005) and the right VLPFC with spatial knowledge (Kostopoulos & Petrides, 2003). More specifically, volitional retrieval of knowledge has been linked to the anterior VLPFC (pars orbitalis), whereas the mid-VLPFC region (pars triangularis, including BA45) is then thought to be involved in choosing amongst competing options (i.e. resolving uncertainty) related to the retrieval (Badre, Poldrack, Paré-Blagoev, Insler, & Wagner, 2005). Such choices can then be passed posteriorly to regions more directly involved in motor functioning (BA44, BA6, primary motor cortex). Kostopoulos & Petrides (2003) conceptualize the function of the mid-VLPFC as "an executive control mechanism that directs attention to the relevant aspect of a stimulus in memory and silences the irrelevant aspect of that same stimulus in memory." The IFG results from Study 3, which overlap with the mid-VLPFC region, could

thus be representing processes linking together specific motor actions with specific auditory percepts: both expected and realized.

Looking "up," in addition to the rostro-caudal axis linking together sub-regions of the VLPFC, there are also dorsal-ventral connections linking the posterior VLPFC (BA44) with the posterior DLPFC (BA6 and BA8) (as well as linking the mid-VLPFC with mid-DLPFC, and linking ventral and dorsal premotor areas) (Petrides, 2005). Reilly (2010) terms these axes as differing along abstract/concrete (rostral/caudal) vs. "what"/"how" (ventral/dorsal) dimensions, with the latter distinction in line with the more general ventral/dorsal stream model discussed previously. Interestingly, the more superior frontal site observed in Study 3 for motor decoding lies at the border of ventro- and dorsolateral frontal cortex, which is sometimes called the inferior frontal junction (IFJ) (Brass, Derrfuss, Forstmann, & Cramon, 2005). This region has been implicated in task- or set-switching (see review by Brass et al. (2005)), the "flexible" sorts of processes generally ascribed to decidedly *dorsal* frontal regions. Bermudez & Zatorre (2005) implicated this region (which partially overlaps with BA6/8: the posterior DLPFC) plus more dorsal and rostral areas in conditional stimulus-response associations (Petrides, 2005) for musical stimuli. Thus, it is fitting that, relative to our more ventral findings, this ventral-dorsal border region was found to contain information *only* for motor actions, in line with the "how" from Reilly's "what"/"how" axis.

As stated above, the VLPFC proper, in contrast to the motor-specific decoding found more superiorly in the IFJ, contained considerable information related to both auditory and motor content. The VLPFC region, via it's dorsal- and anterior-directed connections with other regions in the frontal lobe, sits at the "bottom" of a rostral-caudal axis that is highly hierarchical in nature (Badre & D'Esposito, 2009), while also being positioned at the "top" of both the dorsal and ventral perceptual streams of processing (Hickok & Poeppel, 2007). All of these connections are bidirectional, as the VLPFC has reciprocal connections with DLPFC areas (Petrides, 2005), ventral stream areas in the superior temporal lobe, and parietal dorsal stream regions

(Rauschecker & Scott, 2009). Thus, the VLPFC straddles a multi-pathway integration zone, allowing it to serve in cognitive top-down *and* bottom-up capacities. It mediates aspects of directed behavior (Kostopoulos & Petrides, 2003) and expectations (Fadiga, Craighero, & D'Ausilio, 2009), while also receiving highly-abstracted sensory information (object- and action-focused (Grefkes & Fink, 2005; Rauschecker & Scott, 2009)) via the perceptual streams. The VLPFC may thus serve to integrate the sorts of multimodal processes found in Study 3 and, via back-propagation, amplify or silence particular aspects of the signal present in the perceptual streams. In other words, I feel there is no contradiction in considering this region to be both a volitional/planning/action structure and a quasi-perceptual structure: all are cognitive processes!

5.5 Future directions

The pattern of results which have emerged from the three studies in this thesis contribute to an understanding of the ventral and dorsal streams of auditory perceptual processing, as well as the interactions of those streams with the planning/motor system. These findings have also raised questions for future research, which I will now address. fMRI research, while an incredibly valuable tool for cognitive neuroscience, has certain limitations. Two of those limitations are that (1) it is primarily a tool to assess correlations, not causal inference, and (2) its spatial sensitivity far outstrips its capacity to resolve events in time.

Throughout the three studies, I have presented various pieces of evidence linking neural foci with cognitive and perceptual processes. However, in order to assess whether a specific region is truly necessary as well as sufficient to perform a specific function, additional methods are needed. In the non-human literature, highly precise lesion methods are often utilized. Such methods, in fact, have played a central role in initially dissociating the ventral and dorsal perceptual streams (Mishkin & Ungerleider, 1982), and in the various contributions of sub-regions of lateral prefrontal cortex (e.g. (Petrides, 1985; Petrides et al., 1993)). While similar

experimental lesion studies are not ethical in studying humans, transcranial magnetic stimulation (TMS) provides a decent facsimile, allowing one to temporarily perturb the normal activity of neuronal populations in a small cortical region. TMS works by inducing a magnetic field at the scalp, which produces synchronous firing in a sub-population of the neurons located under the stimulating coil, thus disrupting the normal pattern of activity (see review by Jahanshahi & Rothwell (2000)). The effects of repetitive TMS (rTMS) may last for tens of minutes following stimulation (Jahanshahi & Rothwell, 2000), allowing one to probe its effects on cognitive processes. rTMS could be used in multiple ways, in order to assess the contributions of various dorsal and ventral stream sites to auditory categorical perception. Considering a behavioral paradigm, rTMS could be selectively applied to either ventral (STS) or dorsal (IPS) sites, and its affects on CP identification or discrimination functions could be assessed. It may be that these regions, at least in certain listening environments, could be serving largely redundant processes, in which case rTMS over a single region would not have a significant effect. However, if these interconnected regions are highly non-independent (i.e. if, for example, the IPS relies upon output from the STS in order to perform its CP-related operations), ventral rTMS stimulation should have a larger behavioral effect than stimulating the dorsal stream site. A considerable amount of information is known about stream convergence in the frontal lobe (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009), but less is known about stream "cross-talk" in posterior regions of the cortex, so a TMS experiment could help shed light on this issue. As the perceptual "binding" of sensory information into a unified whole is of ongoing scientific (Golledge et al., 1996) and philosophical (Clark, 2009) interest, it may be highly informative to develop our understanding of the "early" vs. "late" bases of stream integration.

Like TMS, MEG also provides a useful complement to fMRI. MEG is highly-resolved in time (millisecond precision), but only has a moderate degree of spatial sensitivity (Baillet, Mosher, & Leahy, 2001). However, fMRI may be used in conjunction with MEG to improve its spatial precision (Dale et al., 2000). While MEG lacks MRI's millimeter precision, it is certainly

capable of dissociating information sourced in the parietal from superior temporal cortices (and thus can likely dissociate dorsal from ventral stream processes). As theories of the two streams ascribe to them different roles (i.e. "what" vs. "how"), it has also been postulated that these streams act at different speeds (J. Norman, 2003): a relatively slow ventral stream vs. a faster dorsal stream. Similar reasoning may explain why, for example, proprioceptive information from muscles/joints is conducted via faster/wider axons than that of fine-grained touch from skin receptors: speed is costly and the "what" (identification) stream can afford to wait a bit. MEG may allow us to test this empirically. Can the IPS region decode CP-related sound identities more quickly than the STS? (And, if so, is that information more fleeting? MEG has been used with MVPA to train a classifier at one time point and test it at another (Cichy, Pantazis, & Oliva, 2014), thus checking for the stability of information in the cortex.) Separately, does information conducted via the dorsal stream arrive in frontal cortex more quickly than via its ventral counterpart? If so, how does the frontal cortex resolve this temporary information imbalance?

Thinking beyond methodology, it would also be interesting to explore differences in auditory-motor interactions amongst various musical instruments. In Study 3, we utilized piano players, largely because the piano has an inherent flexibility in mappings between movement and sound (i.e. no 1 to 1 correspondence), which is highly contrasting with speech, in which specific actions produce specific auditory results. However, such specificity *is* largely present in a certain instruments, for example the saxophone, for which in general only a unique positioning of the fingers can produce a given tone. Separately, for many instruments (including the sax), there is no simple linear spatial mapping from key to pitch, such as that provided by the left -> right axis of the piano keyboard. Thus, the neural correlates (particularly with regard to the dorsal stream) of saxophone perception/performance may be more similar to that of speech or of the piano.

Bibliography

Abrahams, S., Goldstein, L. H., Simmons, A., Brammer, M. J., Williams, S. C. R., Giampietro, V. P., et al. (2003). Functional magnetic resonance imaging of verbal fluency and confrontation naming using compressed image acquisition to permit overt responses. *Human Brain Mapping*, *20*, 29–40. doi:10.1002/hbm.10126

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

Abramson, A. S. (1977). Noncategorical perception of tone categories in Thai. *The Journal of the Acoustical Society of America*, *61*(S1), S66–S66. doi:10.1121/1.2015837

Acker, B. E., Pastore, R. E., & Hall, M. D. (1995). Within-category discrimination of musical chords: perceptual magnet or anchor? *Perception & Psychophysics*, *57*(6), 863–874.

Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2008). Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1515), 285–299. doi:10.1098/rstb.2008.0253

Ahmad, A. R., Khalia, M., Viard-Gaudin, C., & Poisson, E. (2004). Online handwriting recognition using support vector machine (pp. 311–314). Presented at the IEEE Region 10 Conference, IEEE.

Alain, C., Arnott, S. R., Hevenor, S., Graham, S., & Grady, C. L. (2001). "What" and 'where' in the human auditory system. *Proceedings of the National Academy of Sciences*, *98*(21), 12301–12306.

Amaral, D. G., Insausti, R., & Cowan, W. M. (1983). Evidence for a direct projection from the superior temporal gyrus to the entorhinal cortex in the monkey. *Brain Research*, *275*(2), 263–277.

Arnott, S. R., & Alain, C. (2011). The auditory dorsal pathway: Orienting vision. *Neuroscience and Biobehavioral Reviews*, *35*(10), 2162–2173. doi:10.1016/j.neubiorev.2011.04.005

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation*, *2*, 89–195.

Aziz Zadeh, L., Iacoboni, M., Zaidel, E., Wilson, S., & Mazziotta, J. (2004). Left hemisphere motor facilitation in response to manual action sounds. *European Journal of Neuroscience*, *19*(9), 2609–2612. doi:10.1111/j.1460-9568.2004.03348.x

Baars, B. J. (1997). *In the Theater of Consciousness*. Oxford University Press.

Bachem, A. (1937). Various Types of Absolute Pitch. *J Acoust Soc Am*, *9*(2), 146–151. doi:10.1121/1.1915919

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

Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience*, *10*(9), 659–669. doi:10.1038/nrn2667

Badre, D., Poldrack, R. A., Paré-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable Controlled Retrieval and Generalized Selection Mechanisms in Ventrolateral Prefrontal Cortex. *Neuron*, *47*(6), 907–918. doi:10.1016/j.neuron.2005.07.023

Baillet, S., Mosher, J. C., & Leahy, R. M. (2001). Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, *18*(6), 14–30. doi:10.1109/79.962275

Baker, E., Blumstein, S. E., & Goodglass, H. (1981). Interaction between phonological and semantic factors in auditory comprehension. *Neuropsychologia*, *19*, 1–15.

Baldo, J. V., Klostermann, E. C., & Dronkers, N. F. (2008). It's either a cook or a baker: Patients with conduction aphasia get the gist but lose the trace. *Brain and Language*, *105*(2), 134–140. doi:10.1016/j.bandl.2007.12.007

Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., et al. (2006). Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *Neuroimage*, *30*(3), 917–926. doi:10.1016/j.neuroimage.2005.10.044

Barredo, J., Öztekin, I., & Badre, D. (2013). Ventral Fronto-Temporal Pathway Supporting Cognitive Control of Episodic Memory Retrieval. *Cerebral Cortex*, bht291. doi:10.1093/cercor/bht291

Basso, A., Casati, G., & Vignolo, L. A. (1977). Phonemic identification defect in aphasia. *Cortex*, *13*(1), 85–95. doi:10.1016/S0010-9452(77)80057-9

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., & Dronkers, N. F. (2003). Voxel-based lesion–symptom mapping. *Nature Neuroscience*, *6*(5), 448–450. doi:10.1038/nn1050

Bauer, R. M. (2006). The Agnosias. In P. J. Snyder, P. D. Nussbaum, & D. L. Robins, *Clinical neuropsychology: A pocket handbook for assessment* (pp. 508–533). Washington, DC: American Psychological Association (APA).

Baumann, S., Koeneke, S., Schmidt, C. F., Meyer, M., Lutz, K., & Jancke, L. (2007). A network for audio–motor coordination in skilled pianists and non-musicians. *Brain Research*, *1161*, 65–78. doi:10.1016/j.brainres.2007.05.045

Baumann, S., Petkov, C. I., & Griffiths, T. D. (2013). A unified framework for the organization of the primate auditory cortex. *Frontiers in Systems Neuroscience*, *7*. doi:10.3389/fnsys.2013.00011

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, *57*, 217–239.

Beals, K. (2009). Raising a Left-Brain Child in a Right-Brain World - Katharine Beals - Google Books.

Beer, J., Blakemore, C., Previc, F. H., & Liotti, M. (2002). Areas of the human brain activated by ambient visual motion, indicating three kinds of self-movement. *Experimental Brain Research*, *143*(1), 78–88. doi:10.1007/s00221-001-0947-y

Belin, P., Zatorre, R. J., Hoge, R., Evans, A. C., & Pike, B. (1999). Event-related fMRI of the auditory cortex. *Neuroimage*, *10*(4), 417–429.

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

Bendor, D., & Wang, X. (2008). Neural Response Properties of Primary, Rostral, and Rostrotemporal Core Fields in the Auditory Cortex of Marmoset Monkeys. *Journal of Neurophysiology*, *100*(2), 888–906. doi:10.1152/jn.00884.2007

Bermudez, P., & Zatorre, R. J. (2005). Conditional associative memory for musical stimuli in nonmusicians: implications for absolute pitch. *The Journal of Neuroscience*. doi:10.1523/JNEUROSCI.1560-05.2005

Berryhill, M. E., & Olson, I. R. (2008). Is the posterior parietal lobe involved in working memory retrieval? *Neuropsychologia*, *46*(7), 1775–1786. doi:10.1016/j.neuropsychologia.2008.03.005

Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage*, *79*(C), 201–212. doi:10.1016/j.neuroimage.2013.04.093

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., & Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, *10*(5), 512–528.

Binder, J. R., Frost, J. A., Hammeke, T. A., Rao, S. M., & Cox, R. W. (1996). Function of the left planum temporale in auditory and linguistic processing. *Brain*, *119*(4), 1239–1247.

Birbaumer, N., & Cohen, L. G. (2007). Brain–computer interfaces: communication and restoration of movement in paralysis. *The Journal of Physiology*, *579*(3), 621–636.

Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception.

Nature Publishing Group, *14*(10), 693–707. doi:10.1038/nrn3565

Bledowski, C., Rahm, B., & Rowe, J. B. (2009). What "Works" in Working Memory? Separate Systems for Selection and Updating of Critical Information. *Journal of Neuroscience*, *29*(43), 13735–13741. doi:10.1523/JNEUROSCI.2547-09.2009

Blumstein, S. E., Baker, E., & Goodglass, H. (1977). Phonological factors in auditory comprehension in aphasia. *Neuropsychologia*, *15*(1), 19–30. doi:10.1016/0028-3932(77)90111-7

Blumstein, S. E., Myers, E. B., & Rissman, J. (2005). The perception of voice onset time: An fMRI investigation of phonetic category structure. *Journal of Cognitive Neuroscience*, *17*(9), 1353–1366.

Bodegård, A., Geyer, S., Grefkes, C., Zilles, K., & Roland, P. E. (2001). Hierarchical processing of tactile shape in the human brain. *Neuron*, *31*(2), 317–328.

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, *8*(3), 389–395. doi:10.1038/nn1409

Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT. *Annual Review of Neuroscience*, *28*, 157–189.

Bornstein, M. H., & Korda, N. O. (1984). Discrimination and matching within and between hues measured by reaction times: some implications for categorical perception and levels of information processing. *Psychological Research*, *46*(3), 207–222. doi:10.1007/BF00308884

Brass, M., Derrfuss, J., Forstmann, B., & Cramon, D. (2005). The role of the inferior frontal junction area in cognitive control. *Trends in Cognitive Sciences*. doi:10.1016/j.tics.2005.05.013

Broca, P. (1861). Perte de la Parole, Ramollissement Chronique et Destruction Partielle du Lobe Antérieur Gauche du Cerveau. *Bull Soc Anthropol*.

Brodmann, K. (1909). Brodmann: Contributions to the histologic localisation... - Google Scholar. *Journal Für Psychologie Und Neurologie*.

Brown, D. E. (1991). *Human universals*. New York: McGraw-Hill.

Brown, R. M., Chen, J. L., Hollinger, A., Penhune, V. B., Palmer, C., & Zatorre, R. J. (2013). Repetition suppression in auditory–motor regions to pitch and temporal structure in music. *Journal of Cognitive Neuroscience*, *25*(2), 313–328.

Buelte, D., Meister, I. G., Staedtgen, M., Dambeck, N., Sparing, R., Grefkes, C., & Boroojerdi, B. (2008). The role of the anterior intraparietal sulcus in crossmodal processing of object features in humans: An rTMS study. *Brain Research*, *1217*, 110–118.

doi:10.1016/j.brainres.2008.03.075

Burns, E. M., & Ward, W. D. (1978). Categorical perception—phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *The Journal of the Acoustical Society of America*, *63*, 456.

Burton, M., Small, S., & Blumstein, S. (2000). The role of segmentation in phonological processing: an fMRI investigation. *Cognitive Neuroscience, Journal of*, *12*(4), 679–690.

Catani, M., Jones, D. K., & ffytche, D. H. (2004). Perisylvian language networks of the human brain. *Annals of Neurology*, *57*(1), 8–16. doi:10.1002/ana.20319

Cavina-Pratesi, C., Monaco, S., Fattori, P., Galletti, C., McAdam, T. D., Quinlan, D. J., et al. (2010). Functional magnetic resonance imaging reveals the neural substrates of arm transport and grip formation in reach-to-grasp actions in humans. *Journal of Neuroscience*, *30*(31), 10306–10323. doi:10.1523/JNEUROSCI.2023-10.2010

Champod, A. S., & Petrides, M. (2007). Dissociable roles of the posterior parietal and the prefrontal cortex in manipulation and monitoring processes. *Proceedings of the National Academy of Sciences*, *104*(37), 14837–14842.

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Publishing Group*, *13*(11), 1428–1432. doi:10.1038/nn.2641

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Listening to Musical Rhythms Recruits Motor Regions of the Brain. *Cerebral Cortex*, *18*(12), 2844–2854. doi:10.1093/cercor/bhn042

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

Clark, A. (2009). Perception, action, and experience: Unraveling the golden braid. *Neuropsychologia*, *47*(6), 1460–1468. doi:10.1016/j.neuropsychologia.2008.10.020

Cohen, Y. E., Hauser, M. D., & Russ, B. E. (2006). Spontaneous processing of abstract categorical information in the ventrolateral prefrontal cortex. *Biology Letters*, *2*(2), 261–265. doi:10.1006/anbe.1999.1416

Courtney, S. M. (1998). An Area Specialized for Spatial Working Memory in Human Frontal Cortex. *Science*, *279*(5355), 1347–1351. doi:10.1126/science.279.5355.1347

Craig, A. D. B. (2005). Forebrain emotional asymmetry: a neuroanatomical basis? *Trends in Cognitive Sciences*, *9*(12), 566–571. doi:10.1016/j.tics.2005.10.005

Culham, J. C., & Kanwisher, N. G. (2001). Neuroimaging of cognitive functions in human parietal cortex. *Current Opinion in Neurobiology*, *11*(2), 157–163.

D H Hubel, T. N. W. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, *148*(3), 574.

D'Ausilio, A., Altenmuller, E., Olivetti Belardinelli, M., & Lotze, M. (2006). Cross-modal plasticity of the motor cortex while listening to a rehearsed musical piece. *European Journal of Neuroscience*, *24*(3), 955–958. doi:10.1111/j.1460-9568.2006.04960.x

Da Costa, S., van der Zwaag, W., Marques, J. P., Frackowiak, R. S. J., Clarke, S., & Saenz, M. (2011). Human Primary Auditory Cortex Follows the Shape of Heschl's Gyrus. *Journal of Neuroscience*, *31*(40), 14067–14075. doi:10.1523/JNEUROSCI.2000-11.2011

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., & Halgren, E. (2000). Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*, *26*(1), 55–67. doi:10.1016/S0896-6273(00)81138-1

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

De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R., & Formisano, E. (2008). Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *Neuroimage*, *43*(1), 44–58. doi:10.1016/j.neuroimage.2008.06.037

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, *20*(3-6), 487–506. doi:10.1080/02643290244000239

Deutschländer, A., Bense, S., Stephan, T., Schwaiger, M., Brandt, T., & Dieterich, M. (2002). Sensory system interactions during simultaneous vestibular and visual stimulation in PET. *Human Brain Mapping*, *16*(2), 92–103. doi:10.1002/hbm.10030

DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(8), E505–14. doi:10.1073/pnas.1113427109

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, *91*(1), 176–180. doi:10.1007/BF00230027

Dick, A. S., & Tremblay, P. (2012). Beyond the arcuate fasciculus: consensus and controversy in the connectional anatomy of language. *Brain*, *135*(12), 3529–3550. doi:10.1093/brain/aws222

Dijkerman, H. C., & de Haan, E. H. F. (2007). Cambridge Journals Online - Behavioral and Brain Sciences - Abstract - Somatosensory processes subserving perception and action.

Behavioral and Brain Sciences, *30*(02), 189. doi:10.1017/S0140525X07001392

Dobbins, I. G., & Wagner, A. D. (2005). Domain-general and domain-sensitive prefrontal mechanisms for recollecting events and detecting novelty. *Cerebral Cortex*, *15*(11), 1768–1778. doi:10.1093/cercor/bhi054

Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.1318738111

Dupont, P., Orban, G. A., De Bruyn, B., Verbruggen, A., & Mortelmans, L. (1994). Many areas in the human brain respond to visual motion. *Journal of Neurophysiology*, *72*(3), 1420–1424.

Eichenbaum, H., Sauvage, M., Fortin, N., Komorowski, R., & Lipton, P. (2012). Towards a functional organization of episodic memory in the medial temporal lobe. *Neuroscience and Biobehavioral Reviews*, *36*(7), 1597–1608.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303–306.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601. doi:10.1038/33402

Etzel, J. A., Gazzola, V., & Keysers, C. (2009). An introduction to anatomical ROI-based fMRI classification analysis. *Brain Research*, *1282*(C), 114–125. doi:10.1016/j.brainres.2009.05.090

Etzel, J. A., Valchev, N., & Keysers, C. (2011). The impact of certain methodological choices on multivariate analysis of fMRI data with support vector machines. *Neuroimage*, *54*(2), 1159–1167.

Evans, S., Kyong, J. S., Rosen, S., Golestani, N., Warren, J. E., McGettigan, C., et al. (2014). The Pathways for Intelligible Speech: Multivariate and Univariate Perspectives. *Cerebral Cortex*, *24*(9), 2350–2361. doi:10.1093/cercor/bht083

Fadiga, L., Craighero, L., & D'Ausilio, A. (2009). Broca's area in language, action, and music. *Annals of the New York Academy of Sciences*, *1169*(1), 448–458.

Farah, M. J. (2004). *Visual Agnosia*. Cambridge, Mass.: MIT Press.

Farwell, L. A., & Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, *70*(6), 510–523.

Federmeier, K. D., Wlotko, E. W., & Meyer, A. M. (2008). What's 'Right' in Language Comprehension: Event-Related Potentials Reveal Right Hemisphere Language Capabilities.

Language and Linguistics Compass, *2*(1), 1–17. doi:10.1111/j.1749-818X.2007.00042.x

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal Lobe: From Action Organization to Intention Understanding. *Science*, *308*(5722), 662–667. doi:10.2307/3841989?ref=no-x-route:567d74fafa9cdb3b57d8665716e2b073

Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). " Who" Is Saying“ What?” Brain-Based Decoding of Human Voice and Speech. *Science*, *322*(5903), 970.

Foster, N. E. V., Halpern, A. R., & Zatorre, R. J. (2013). Common parietal activation in musical mental transformations across pitch and time. *Neuroimage*, *75*(C), 27–35. doi:10.1016/j.neuroimage.2013.02.044

Foster, N. E., & Zatorre, R. J. (2010). A role for the intraparietal sulcus in transforming musical pitch information. *Cerebral Cortex*, *20*(6), 1350–1359. doi:10.1093/cercor/bhp199

Friederici, A. D. (2011). The Brain Basis of Language Processing: From Structure to Function. *Physiological Reviews*, *91*(4), 1357–1392. doi:10.1152/physrev.00006.2011

Fries, W., & Swihart, A. A. (1990). Disturbance of rhythm sense following right hemisphere damage. *Neuropsychologia*, *28*(12), 1317–1323. doi:10.1016/0028-3932(90)90047-R

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: a synthesis. *Human Brain Mapping*, *2*(1-2), 56–78.

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

Fritz, J. B., David, S. V., Radtke-Schuller, S., Yin, P., & Shamma, S. A. (2010). Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. *Nature Publishing Group*, *13*(8), 1011–1019. doi:10.1038/nn.2598

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, *5*(4), 171–189.

Fujisaki, H., & Kawashima, T. (1968). The influence of various factors on the identification and discrimination of synthetic speech sounds. *6th International Congress on Acoustics*, *2*, 95–98.

Fujisaki, H., & Kawashima, T. (1969). On the modes and mechanisms of speech perception. *Annual Report of the Engineering Research Institute*, *28*, 67–73.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(3), 361–377.

Gall, F. J., & Spurzheim, J. G. (1809). *Recherches sur le système nerveux en général, et sur celui du cerveau en particulier*.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*(2), 593–609.

Gazzaniga, M. S. (1967). The Split Brain in Man. *Scientific American*, *217*(2), 24–29.

Gazzaniga, M. S. (1998). The split brain revisited. *Scientific American*, *279*(1), 50–55.

Gazzola, V., & Keysers, C. (2009). The Observation and Execution of Actions Share Motor and Somatosensory Voxels in all Tested Subjects: Single-Subject Analyses of Unsmoothed fMRI Data. *Cerebral Cortex*, *19*(6), 1239–1255. doi:10.1093/cercor/bhn181

Gazzola, V., Aziz Zadeh, L., & Keysers, C. (2006). Empathy and the Somatotopic Auditory Mirror System in Humans. *Current Biology*, *16*(18), 1824–1829. doi:10.1016/j.cub.2006.07.072

Gibson, J. J. (1950). *The perception of the visual world*. Oxford, England: Houghton Mifflin.

Gibson, J. J. (1966). *The Senses Considered as Perpetual Systems*. Oxford, England: Houghton Mifflin.

Gifford, G. W., III, MacLean, K. A., Hauser, M. D., & Cohen, Y. E. (2005). The neurophysiology of functionally meaningful categories: macaque ventrolateral prefrontal cortex plays a critical role in spontaneous categorization of species-specific vocalizations. *Journal of Cognitive Neuroscience*, *17*(9), 1471–1482.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf Hypothesis Is Supported in the Right Visual Field but Not the Left. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 489–494. doi:10.2307/30048329?ref=search-gateway:c9f3a3a987021ebaf5656c8caeb6ccd9

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. doi:10.1038/nn.3063

Golledge, H. D., Hilgetag, C. C., & Tovée, M. J. (1996). Information processing: A solution to the binding problem? *Current Biology*, *6*(9), 1092–1095.

Goodale, M. A. (2005). Action Insight: The Role of the Dorsal Stream in the Perception of Grasping. *Neuron*, *47*(3), 328–329. doi:10.1016/j.neuron.2005.07.010

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

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “L” and ‘R’. *Neuropsychologia*, *9*, 317–323.

Grefkes, C., & Fink, G. R. (2005). REVIEW: The functional organization of the intraparietal sulcus in humans and monkeys. *Journal of Anatomy*, *207*(1), 3–17.

Grefkes, C., Weiss, P. H., Zilles, K., & Fink, G. R. (2002). Crossmodal processing of object features in human anterior intraparietal cortex: an fMRI study implies equivalencies between humans and monkeys. *Neuron*, *35*(1), 173–184.

Gregory, R. (1970). *The intelligent eye*. London, Englad: Weidenfeld & Nicolson.

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

Griffiths, T. D., Green, G. G. R., Rees, A., & Rees, G. (2000). Human brain areas involved in the analysis of auditory movement. *Human Brain Mapping*, *9*(2), 72–80.

Griffiths, T. D., Warren, J. D., Scott, S. K., Nelken, I., & King, A. J. (2004). Cortical processing of complex sound: a way forward? *Trends in Neurosciences*, *27*(4), 181–185. doi:10.1016/j.tins.2004.02.005

Gross, C. G. (2002). Genealogy of the "Grandmother Cell." *The Neuroscientist*, *8*(5), 512–518. doi:10.1177/107385802237175

Gross, C. G., & De Schonen, S. (1992). Representation of Visual Stimuli in Inferior Temporal Cortex [and Discussion]. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *335*(1273), 3–10.

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. *Journal of Neurophysiology*, *35*(1), 96–111.

Gross, R. G., & Grossman, M. (2008). Update on apraxia. *Current Neurology and Neuroscience Reports*, *8*(6), 490–496.

Hackett, T. A. (2007). Organization of the thalamocortical auditory pathways in primates. In R. F. Burkard, J. J. Eggermont, & M. Don, *Auditory evoked potentials: basic principles and clinical application* (pp. 428–440). New York, NY: Lippincott, Williams, and Wilkins.

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., et al. (1999). Sparse temporal sampling in auditory fMRI. *Human Brain Mapping*, *7*(3), 213–223.

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V., & Pollmann, S. (2009). PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics*, *7*(1), 37–53.

Harnad, S., Hanson, S. J., & Lubin, J. (1995). Learned Categorical Perception in Neural Nets: Implications for Symbol Grounding. In V. Honavar & L. Uhr, *Symbol Processors and Connectionist Network Models in Artificial Intelligence and Cognitive Modelling: Steps Toward Principled Integration*. New York: Academic Press.

Harris, I. M., & Miniussi, C. (2003). Parietal lobe contribution to mental rotation demonstrated

with rTMS. *Journal of Cognitive Neuroscience*, *15*(3), 315–323.

Hary, J. M., & Massaro, D. W. (1982). Categorical results do not imply categorical perception. *Attention, Perception, & Psychophysics*, *32*(5), 409–418.

Haslinger, B., Erhard, P., Altenmüller, E., Schroeder, U., Boecker, H., & Ceballos-Baumann, A. O. (2005). Transmodal sensorimotor networks during action observation in professional pianists. *Journal of Cognitive Neuroscience*, *17*(2), 282–293.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425–2430. doi:10.1126/science.1063736

Haynes, J.-D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience*, *8*(5), 686–691. doi:10.1038/nn1445

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, *7*(7), 523–534. doi:10.1038/nrn1931

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory, 1st Edition* (1st ed.). Wiley.

Hecht, D. (2010). Depression and the hyperactive right-hemisphere. *Neuroscience Research*, *68*(2), 77–87. doi:10.1016/j.neures.2010.06.013

Heilman, K. M., & Edward Valenstein, M. D. (2011). *Clinical Neuropsychology*. New York, USA: Oxford University Press.

Heilman, K. M., Watson, R. T., & Valenstein, E. (1993). Neglect and related disorders. In K. M. Heilman & E. Valenstein, *Clinical Neuropsychology 4th Edition* (pp. 243–293). New York, USA: Oxford University Press.

Hein, G., & Knight, R. (2008). Superior temporal sulcus—it's my area: or is it? *Cognitive Neuroscience*, *20*(12), 2125–2136.

Helmholtz, H. (1867). Concerning the perceptions in general. In G. Karsten, *Allgemeinen Encyclopädie der Physik*. Leipzig, Germany.

Hermsdörfer, J., Goldenberg, G., Wachsmuth, C., Conrad, B., Ceballos-Baumann, A. O., Bartenstein, P., et al. (2001). Cortical Correlates of Gesture Processing: Clues to the Cerebral Mechanisms Underlying Apraxia during the Imitation of Meaningless Gestures. *Neuroimage*, *14*(1), 149–161. doi:10.1006/nimg.2001.0796

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

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

Neuroscience, *15*(5), 673–682. doi:10.1093/brain/124.1.83

Hickok, G., Okada, K., & Serences, J. T. (2008). Area Spt in the Human Planum Temporale Supports Sensory-Motor Integration for Speech Processing. *Journal of Neurophysiology*, *101*(5), 2725–2732. doi:10.1152/jn.91099.2008

Hochstein, S., & Ahissar, M. (2002). View from the Top. *Neuron*, *36*(5), 791–804. doi:10.1016/S0896-6273(02)01091-7

Hollinger, A. D., & Wanderley, M. M. (2013). MRI-compatible optically-sensed cello (pp. 1–4). Presented at the 2013 IEEE Sensors, IEEE. doi:10.1109/ICSENS.2013.6688614

Hollinger, A., Steele, C., Penhune, V., Zatorre, R., & Wanderley, M. (2007). fMRI-compatible electronic controllers (p. 246). Presented at the the 7th international conference, New York, New York, USA: ACM Press. doi:10.1145/1279740.1279790

Hoshi, E., & Tanji, J. (2007). Distinctions between dorsal and ventral premotor areas: anatomical connectivity and functional properties. *Current Opinion in Neurobiology*, *17*(2), 234–242. doi:10.1016/j.conb.2007.02.003

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*, 106–154.

Humphries, C., Willard, K., Buchsbaum, B., & Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: an fMRI study. *Neuroreport*, *12*(8), 1749–1752.

Husain, M., & Nachev, P. (2007). Space and the parietal cortex. *Trends in Cognitive Sciences*, *11*(1), 30–36. doi:10.1016/j.tics.2006.10.011

Hutchison, E. R., Blumstein, S. E., & Myers, E. B. (2008). An event-related fMRI investigation of voice-onset time discrimination. *Neuroimage*, *40*(1), 342–352. doi:10.1016/j.neuroimage.2007.10.064

Hyde, K. L., Peretz, I., & Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia*, *46*(2), 632–639. doi:10.1016/j.neuropsychologia.2007.09.004

Ibbotson, N. R., & Morton, J. (1981). Rhythm and dominance. *Cognition*, *9*, 125–138.

Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences*, *106*(4), 1245–1248.

Jahanshahi, M., & Rothwell, J. (2000). Transcranial magnetic stimulation studies of cognition: an emerging field. *Experimental Brain Research*, *131*(1), 1–9. doi:10.1007/s002219900224

James, W. (1891). *The Principles of Psychology*. London, England: Macmillan & Co.

Jellema, T., & Perrett, D. I. (2003). Cells in monkey STS responsive to articulated body motions

and consequent static posture: a case of implied motion? *Neuropsychologia*, *41*(13), 1728–1737. doi:10.1016/S0028-3932(03)00175-1

Jimura, K., & Poldrack, R. A. (2011). Analyses of regional-average activation and multivoxel pattern information tell complementary stories. *Neuropsychologia*, *50*(4), 544–552. doi:10.1016/j.neuropsychologia.2011.11.007

Joanisse, M. F., Zevin, J. D., & McCandliss, B. D. (2007). Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using fMRI and a short-interval habituation trial paradigm. *Cerebral Cortex*, *17*(9), 2084–2093. doi:10.1093/cercor/bhl124

Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain*, *123*(1), 155–163.

Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences*, *97*(22), 11793–11799.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*(5), 679–685.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *The Journal of Neuroscience*, *17*(11), 4302–4311.

Kellenbach, M. L., Hovius, M., & Patterson, K. (2005). A pet study of visual and semantic knowledge about objects. *Cortex*, *41*(2), 121–132. doi:10.1016/S0010-9452(08)70887-6

Keysers, C., Kohler, E., Umilt, M. A., Nanetti, L., Fogassi, L., & Gallese, V. (2003). Audiovisual mirror neurons and action recognition. *Experimental Brain Research*, *153*(4), 628–636. doi:10.1007/s00221-003-1603-5

Kikuchi, Y., Horwitz, B., & Mishkin, M. (2010). Hierarchical Auditory Processing Directed Rostrally along the Monkey's Supratemporal Plane. *Journal of Neuroscience*, *30*(39), 13021–13030. doi:10.1523/JNEUROSCI.2267-10.2010

Kikutani, M., Roberson, D., & Hanley, J. R. (2010). Categorical Perception for Unfamiliar Faces: The Effect of Covert and Overt Face Learning. *Psychological Science*, *21*(6), 865–872. doi:10.1177/0956797610371964

Kilian-Hütten, N., Valente, G., Vroomen, J., & Formisano, E. (2011). Auditory Cortex Encodes the Perceptual Interpretation of Ambiguous Sound. *The Journal of Neuroscience*, *31*(5), 1715–1720.

Kim, M., Na, D. L., Kim, G. M., Adair, J. C., Lee, K. H., & Heilman, K. M. (1999). Ipsilesional neglect: behavioural and anatomical features. *Journal of Neurology, Neurosurgery & Psychiatry*, *67*(1), 35–38.

Kimura, D. (1967). Functional Asymmetry of the Brain in Dichotic Listening. *Cortex*, *3*(2), 163–178. doi:10.1016/S0010-9452(67)80010-8

Klein, M. E., & Zatorre, R. J. (2011). A role for the right superior temporal sulcus in categorical perception of musical chords. *Neuropsychologia*, *49*(5), 878–887. doi:10.1016/j.neuropsychologia.2011.01.008

Klein, M. E., & Zatorre, R. J. (2014). Representations of Invariant Musical Categories Are Decodable by Pattern Analysis of Locally Distributed BOLD Responses in Superior Temporal and Intraparietal Sulci. *Cerebral Cortex*. doi:10.1093/cercor/bhu003

Klingberg, T. (2006). Development of a superior frontal–intraparietal network for visuo-spatial working memory. *Neuropsychologia*, *44*(11), 2171–2177. doi:10.1016/j.neuropsychologia.2005.11.019

Koelsch, S. (2006). Significance of Broca's area and ventral premotor cortex for music-syntactic processing. *Cortex*, *42*(4), 518–520.

Koelsch, S., Gunter, T., Schröger, E., & Friederici, A. D. (2003). Processing Tonal Modulations: An ERP Study. *Journal of Cognitive Neuroscience*, *15*(8), 1149–1159.

Kohler, E. (2002). Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons. *Science*, *297*(5582), 846–848. doi:10.1126/science.1070311

Kostopoulos, P., & Petrides, M. (2003). The mid-ventrolateral prefrontal cortex: insights into its role in memory retrieval. *European Journal of Neuroscience*, *17*(7), 1489–1497. doi:10.1046/j.1460-9568.2003.02574.x

Kotz, S. A., Meyer, M., & Paulmann, S. (2006). Lateralization of emotional prosody in the brain: an overview and synopsis on the impact of study design. *Progress in Brain Research*, *156*, 285–294. doi:10.1016/S0079-6123(06)56015-7.3d

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, *103*(10), 3863.

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

Kuhl, P. K. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *The Journal of the Acoustical Society of America*, *63*(3), 905–917. doi:10.1121/1.381770

Kumar, S., Bonnici, H. M., Teki, S., Agus, T. R., Pressnitzer, D., Maguire, E. A., & Griffiths, T. D. (2014). Representations of specific acoustic patterns in the auditory cortex and hippocampus. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1791),

20141000–20141000. doi:10.1016/0006-8993(83)90987-3

Kumar, S., Stephan, K. E., Warren, J. D., Friston, K. J., & Griffiths, T. D. (2005). Hierarchical Processing of Auditory Objects in Humans. *PLoS Computational Biology*, *preprint*(2007), e100. doi:10.1371/journal.pcbi.0030100.eor

Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *Journal of Neuroscience*, *27*(2), 308–314. doi:10.1523/JNEUROSCI.4822-06.2007

Lee, J. H. (2009). Prefrontal activity predicts monkeys' decisions during an auditory category task. *Frontiers in Integrative Neuroscience*, *3*. doi:10.3389/neuro.07.016.2009

Lee, Y. S., Janata, P., Frost, C., Hanke, M., & Granger, R. (2011). Investigation of melodic contour processing in the brain using multivariate pattern-based fMRI. *Neuroimage*, *57*, 293–300.

Lee, Y. S., Turkeltaub, P., Granger, R., & Raizada, R. D. S. (2012). Categorical Speech Processing in Broca's Area: An fMRI Study Using Multivariate Pattern-Based Analysis. *Journal of Neuroscience*, *32*(11), 3942–3948

Leech, R., Holt, L. L., Devlin, J. T., & Dick, F. (2009). Expertise with Artificial Nonspeech Sounds Recruits Speech-Sensitive Cortical Regions. *Journal of Neuroscience*, *29*(16), 5234–5239. doi:10.1523/JNEUROSCI.5758-08.2009

Levitin, D. J., & Menon, V. (2003). Musical structure is processed in “language” areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. *Neuroimage*, *20*(4), 2142–2152. doi:10.1016/j.neuroimage.2003.08.016

Lewis, J. W., Brefczynski, J. A., Phinney, R. E., Janik, J. J., & DeYoe, E. A. (2005). Distinct Cortical Pathways for Processing Tool versus Animal Sounds. *The Journal of Neuroscience*, *25*(21), 5148–5158.

Ley, A., Vroomen, J., Hausfeld, L., Valente, G., De Weerd, P., & Formisano, E. (2012). Learning of new sound categories shapes neural response patterns in human auditory cortex. *Journal of Neuroscience*, *32*(38), 13273–13280. doi:10.1523/JNEUROSCI.0584-12.2012

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431. doi:10.1037/h0020279

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology: Human Perception and Performance*, *54*, 358–368.

Liebenthal, E., Binder, J., Spitzer, S., Possing, E., & Medler, D. (2005). Neural Substrates of Phonemic Perception. *Cerebral Cortex*, *15*(10), 1621–1631. doi:10.1093/cercor/bhi040

Lieberman, M. D., & Cunningham, W. A. (2009). Type I and Type II error concerns in fMRI research: re-balancing the scale. *Social Cognitive and Affective Neuroscience*, *4*(4), 423–428. doi:10.1093/scan/nsp052

Locke, S., & Kellar, L. (1973). Categorical perception in a non-linguistic mode. *Cortex*, *9*(4), 355–369.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, *453*(7197), 869–878. doi:10.1038/nature06976

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*(6843), 150–157.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, *4*.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nature Neuroscience*, *4*(5), 540–545.

Makris, N., & Pandya, D. N. (2008). The extreme capsule in humans and rethinking of the language circuitry. *Brain Structure and Function*, *213*(3), 343–358. doi:10.1007/s00429-008-0199-8

Man, K., Kaplan, J. T., Damasio, A., & Meyer, K. (2012). Sight and Sound Converge to Form Modality-Invariant Representations in Temporoparietal Cortex. *Journal of Neuroscience*, *32*(47), 16629–16636. doi:10.1523/JNEUROSCI.2342-12.2012

Massé, A. B., Harnad, S., Picard, O., & St-Louis, B. (2013). Symbol grounding and the origin of language. In C. Lefebvre, B. Comrie, & H. Cohen, *New perspectives on the origins of language*. John Benjamins Publishing Company.

Mattingly, I. G., Liberman, A. M., Syrdal, A. K., & Halwes, T. (1971). Discrimination in speech and nonspeech modes. *Cognitive Psychology*, *2*(2), 131–157. doi:10.1016/0010-0285(71)90006-5

McIntosh, R. D., & Schenk, T. (2009). Two visual streams for perception and action: Current trends. *Neuropsychologia*, *47*(6), 1391–1396. doi:10.1016/j.neuropsychologia.2009.02.009

McKay, C., & Fujinaga, I. (2005). Automatic music classification and the importance of instrument identification. Presented at the Proceedings of the Conference on Interdisciplinary Musicology.

Meindl, A. (2012). *At Left Brain Turn Right*. Los Angeles, USA: Meta Creative.

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The Essential

Role of Premotor Cortex in Speech Perception. *Current Biology*, *17*(19), 1692–1696. doi:10.1016/j.cub.2007.08.064

Meng, M., Cherian, T., Singal, G., & Sinha, P. (2012). Lateralization of face processing in the human brain. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1735), 2052–2061. doi:10.1093/cercor/bhq050

Menon, V., Levitin, D. J., Smith, B. K., Lembke, A., Krasnow, B. D., Glazer, D., et al. (2002). Neural correlates of timbre change in harmonic sounds. *Neuroimage*, *17*(4), 1742–1754. doi:10.1086/660123?ref=no-x-route:021e255edf2f4c9fe97c71d33ef6414a

Merzenich, M. M., & Brugge, J. F. (1973). Representation of the cochlear partition on the superior temporal plane of the macaque monkey. *Brain Research*, *50*(2), 275–296. doi:10.1016/0006-8993(73)90731-2

Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., & Cramon, Von, D. Y. (2002). FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Human Brain Mapping*, *17*(2), 73–88. doi:10.1002/hbm.10042

Miller, J. D. (1976). Discrimination and labeling of noise–buzz sequences with varying noise-lead times: An example of categorical perception. *The Journal of the Acoustical Society of America*, *60*(2), 410–417. doi:10.1121/1.381097

Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, *46*(3), 774–785. doi:10.1016/j.neuropsychologia.2007.10.005

Milner, B. (1962). Laterality effects in audition. In V. B. Mountcastle, *Interhemispheric Relations and Cerebral Dominance, Ed. by Vernon B. Mountcastle* (p. 294). Baltimore: Johns Hopkins Press.

Milner, B., Taylor, L., & Sperry, R. W. (1968). Lateralized suppression of dichotically presented digits after commissural section in man. *Science*, *161*(3837), 184–185.

Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural Brain Research*, *6*(1), 57–77.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, *6*, 414–417. doi:10.1016/0166-2236(83)90190-X

Moerel, M., De Martino, F., & Formisano, E. (2012). Processing of Natural Sounds in Human Auditory Cortex: Tonotopy, Spectral Tuning, and Relation to Voice Sensitivity. *Journal of Neuroscience*, *32*(41), 14205–14216. doi:10.1523/JNEUROSCI.1388-12.2012

Morel, A., Garraghty, P. E., & Kaas, J. H. (1993). Tonotopic organization, architectonic fields,

and connections of auditory cortex in macaque monkeys. *Journal of Comparative Neurology*, *335*(3), 437–459.

Mourao-Miranda, J., Reynaud, E., McGlone, F., Calvert, G., & Brammer, M. (2006). The impact of temporal compression and space selection on SVM analysis of single-subject and multi-subject fMRI data. *Neuroimage*, *33*(4), 1055–1065.

Muhammad, R., Wallis, J. D., & Miller, E. K. (2006). A comparison of abstract rules in the prefrontal cortex, premotor cortex, inferior temporal cortex, and striatum. *Journal of Cognitive Neuroscience*, *18*(6), 974–989.

Munoz-Lopez, M. M., Mohedano-Moriano, A., & Insausti, R. (2010). Anatomical Pathways for Auditory Memory in Primates. *Frontiers in Neuroanatomy*, *4*. doi:10.3389/fnana.2010.00129

Mur, M., Bandettini, P. A., & Kriegeskorte, N. (2008). Revealing representational content with pattern-information fMRI--an introductory guide. *Social Cognitive and Affective Neuroscience*, *4*(1), 101–109. doi:10.1093/scan/nsn044

Myers, E. B., Blumstein, S. E., Walsh, E., & Eliassen, J. (2009). Inferior Frontal Regions Underlie the Perception of Phonetic Category Invariance. *Psychological Science*, *20*(7), 895–903. doi:10.1111/j.1467-9280.2009.02380.x

Naghavi, H. R., & Nyberg, L. (2005). Common fronto-parietal activity in attention, memory, and consciousness: shared demands on integration? *Consciousness and Cognition*, *14*(2), 390–425.

Nahum, M., Nelken, I., & Ahissar, M. (2008). Low-level information and high-level perception: the case of speech in noise. *PLOS Biology*, *6*(5), e126. doi:10.1371/journal.pbio

Narain, C., Scott, S. K., Wise, R. J., Rosen, S., Leff, A., Iversen, S. D., & Matthews, P. M. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cerebral Cortex*, *13*(12), 1362–1368. doi:10.1093/cercor/bhg083

Nasir, S. M., & Ostry, D. J. (2009). Auditory plasticity and speech motor learning. *Proceedings of the National Academy of Sciences*, *106*(48), 20470–20475.

Nielsen, J. A., Zielinski, B. A., Ferguson, M. A., Lainhart, J. E., & Anderson, J. S. (2013). An Evaluation of the Left-Brain vs. Right-Brain Hypothesis with Resting State Functional Connectivity Magnetic Resonance Imaging. *PLoS ONE*, *8*(8), e71275. doi:10.1371/journal.pone.0071275.t003

Norman, J. (2003). Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches. *Behavioral and Brain Sciences*, *25*(01), 73–96. doi:10.1017/S0140525X0200002X

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

O Scalaidhe, S. P. (1997). Areal Segregation of Face-Processing Neurons in Prefrontal Cortex. *Science*, *278*(5340), 1135–1138. doi:10.1126/science.278.5340.1135

Obleser, J., Wise, R. J. S., Alex Dresner, M., & Scott, S. K. (2007a). Functional Integration across Brain Regions Improves Speech Perception under Adverse Listening Conditions. *Journal of Neuroscience*, *27*(9), 2283–2289. doi:10.1523/JNEUROSCI.4663-06.2007

Obleser, J., Zimmermann, J., Van Meter, J., & Rauschecker, J. P. (2007b). Multiple stages of auditory speech perception reflected in event-related FMRI. *Cerebral Cortex*, *17*(10), 2251–2257. doi:10.1093/cercor/bhl133

Ojemann, G. A. (1991). Cortical organization of language. *The Journal of Neuroscience*, *11*(8), 2281–2287.

Ojemann, G., Ojemann, J., Lettich, E., & Berger, M. (1989). Cortical language localization in left, dominant hemisphere: an electrical stimulation mapping investigation in 117 patients. *Journal of Neurosurgery*.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I., Saberi, K., et al. (2010). Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex*, *20*(10), 2486.

Oosterhof, N. N., Wiestler, T., Downing, P. E., & Diedrichsen, J. (2011). A comparison of volume-based and surface-based multi-voxel pattern analysis. *Neuroimage*, *56*(2), 593–600. doi:10.1016/j.neuroimage.2010.04.270

Pa, J., & Hickok, G. (2008). A parietal–temporal sensory–motor integration area for the human vocal tract: Evidence from an fMRI study of skilled musicians. *Neuropsychologia*, *46*(1), 362–368. doi:10.1016/j.neuropsychologia.2007.06.024

Patel, A. D. (2007). *Music, Language, and the Brain*. New York: Oxford University Press.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, *36*(4), 767–776.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1-2), 8–13. doi:10.1016/j.jneumeth.2006.11.017

Penagos, H., Melcher, J. R., & Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *The Journal of Neuroscience*, *24*(30), 6810–6815.

Penfield, W., & Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral

cortex of man as studied by electric stimulation. *Brain*, *60*(4), 389–443. doi:10.1093/brain/60.4.389

Penfield, W., & Rasmussen, T. (1949). Vocalization and arrest of speech. *Archives of Neurology & Psychiatry*, *61*(1), 21–27.

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage*, *45*(1), S199–S209.

Peretz, I. (1990). Processing of local and global musical information by unilateral brain-damaged patients. *Brain*, *113*, 1185–1205.

Peretz, I., & Zatorre, R. J. (2005). Brain Organization for Music Processing. *Annual Review of Psychology*, *56*(1), 89–114. doi:10.1146/annurev.psych.56.091103.070225

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex - Springer. *Experimental Brain Research*, *47*, 329–342.

Peters, B. O., Pfurtscheller, G., & Flyvbjerg, H. (1998). Mining multi-channel EEG for its information content: an ANN-based method for a brain–computer interface. *Neural Networks*, *11*(7), 1429–1433.

Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., & Logothetis, N. K. (2008). A voice region in the monkey brain. *Nature Neuroscience*, *11*(3), 367–374. doi:10.1038/nn2043

Petrides, M. (1985). Deficits in non-spatial conditional associative learning after periarcuate lesions in the monkey. *Behavioural Brain Research*, *16*(2), 95–101.

Petrides, M. (2005). The Rostral-Caudal Axis of Cognitive Control within the Lateral Frontal Cortex. In S. Dehaene, J. Duhamel, M. Hauser, & G. Rizzolatti, *From Monkey Brain to Human Brain* (pp. 293–314). Cambridge, USA: MIT Press.

Petrides, M., & Pandya, D. N. (1999). Dorsolateral prefrontal cortex: comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. *European Journal of Neuroscience*, *11*(3), 1011–1036.

Petrides, M., & Pandya, D. N. (2009). Distinct Parietal and Temporal Pathways to the Homologues of Broca's Area in the Monkey. *PLOS Biology*, *7*(8), e1000170. doi:10.1371/journal.pbio.1000170.s003

Petrides, M., Alivisatos, B., Evans, A. C., & Meyer, E. (1993). Dissociation of human mid-dorsolateral from posterior dorsolateral frontal cortex in memory processing. *Proceedings of the National Academy of Sciences*, *90*(3), 873–877.

Pisoni, D. B. (1971). *On the nature of categorical perception of speech sounds*. Doctoral thesis, Michigan Univ Ann Arbor.

Pisoni, D. B. (1975). Auditory short-term memory and vowel perception. *Memory & Cognition*, *3*(1), 7–18. doi:10.3758/BF03198202

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

Quiroga, R. Q., Kreiman, G., Koch, C., & Fried, I. (2008). Sparse but not "Grandmother-cell" coding in the medial temporal lobe. *Trends in Cognitive Sciences*, *12*(3), 87–91.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*. 435(23), 1102-1107

Raizada, R. D. S., & Poldrack, R. A. (2007). Selective Amplification of Stimulus Differences during Categorical Processing of Speech. *Neuron*, *56*(4), 726–740. doi:10.1016/j.neuron.2007.11.001

Rao, S. C. (1997). Integration of What and Where in the Primate Prefrontal Cortex. *Science*, *276*(5313), 821–824. doi:10.1126/science.276.5313.821

Rasmussen, T., & Milner, B. (1975). Clinical and Surgical Studies of the Cerebral Speech Areas in Man. In *Cerebral localization* (pp. 238–257). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-66204-1_19

Rasmussen, T., & Milner, B. (1977). The role of early left-brain injury in determining lateralization of cerebral speech functions. *Annals of the New York Academy of Sciences*, *299*(1), 355–369.

Rauschecker, J. P. (1998). Parallel Processing in the Auditory Cortex of Primates. *Audiology and Neuro-Otology*, *3*(2-3), 86–103. doi:10.1159/000013784

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Publishing Group*, *12*(6), 718–724. doi:10.1038/nn.2331

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of "what" and 'where' in auditory cortex. *Proceedings of the National Academy of Sciences*, *97*(22), 11800.

Rauschecker, J. P., Tian, B., & Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science*, *268*(5207), 111–114.

Raz, I. (1977). Categorical perception of nonspeech stimuli by musicians and nonmusicians. *J Acoust Soc Am*, *62*(S1), S60. doi:10.1121/1.2016288

Reilly, R. C. O. (2010). The What and How of prefrontalcortical organization. *Trends in Neurosciences*, *33*(8), 355–361. doi:10.1016/j.tins.2010.05.002

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. *Speech and Language: Advances in Basic Research and Practice*, *10*, 244–322.

Rinne, T., Kirjavainen, S., Salonen, O., Degerman, A., Kang, X., Woods, D. L., & Alho, K. (2007). Distributed cortical networks for focused auditory attention and distraction. *Neuroscience Letters*, *416*(3), 247–251. doi:10.1016/j.neulet.2007.01.077

Rizzolatti, G., & Craighero, L. (2004). The Mirror-Neuron System. *Annual Review of Neuroscience*, *27*(1), 169–192. doi:10.1146/annurev.neuro.27.070203.144230

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*(2), 131–141.

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neuroscience*, *2*(12), 1131–1136.

Rosenzweig, P. (2014). *Left Brain, Right Stuff*. London, England: Profile Books.

Rubens, A. B., Mahowald, M. W., & Hutton, J. T. (1976). Asymmetry of the lateral (sylvian) fissures in man. *Neurology*.

Rusconi, E., Kwan, B., Giordano, B. L., & Umilta, C. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition*.

Russ, B. E., Orr, L. E., & Cohen, Y. E. (2008). Prefrontal Neurons Predict Choices during an Auditory Same-Different Task. *Current Biology*, *18*(19), 1483–1488. doi:10.1016/j.cub.2008.08.054

Samson, S., & Zatorre, R. J. (1988). Melodic and harmonic discrimination following unilateral cerebral excision. *Brain and Cognition*, *7*, 348–360.

Schlaug, G., Jäncke, L., Huang, Y., & Staiger, J. F. (1995). Increased corpus callosum size in musicians. *Neuropsychologia*, *33*(8), 1047–1055.

Schmid, M. C., Mrowka, S. W., Turchi, J., Saunders, R. C., Wilke, M., Peters, A. J., et al. (2010). Blindsight depends on the lateral geniculate nucleus. *Nature*, *466*(7304), 373–377. doi:10.1038/nature09179

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., et al. (2005). Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nature Neuroscience*, *8*(9), 1241–1247. doi:10.1038/nn1530

Schouten, B. (2003). The end of categorical perception as we know it. *Speech Communication*, *41*(1), 71–80. doi:10.1016/S0167-6393(02)00094-8

Schönwiesner, M., Dechent, P., Voit, D., Petkov, C. I., & Krumbholz, K. (2014). Parcellation of Human and Monkey Core Auditory Cortex with fMRI Pattern Classification and Objective

Detection of Tonotopic Gradient Reversals. *Cerebral Cortex*. doi:10.1093/cercor/bhu124

Schönwiesner, M., Rübsamen, R., & Cramon, Von, D. Y. (2005). Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *European Journal of Neuroscience*, *22*(6), 1521–1528. doi:10.1111/j.1460-9568.2005.04315.x

Schwarzbauer, C., Davis, M. H., Rodd, J. M., & Johnsrude, I. (2006). Interleaved silent steady state (ISSS) imaging: a new sparse imaging method applied to auditory fMRI. *Neuroimage*, *29*(3), 774–782.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*(12), 2400–2406.

Sidiropoulos, K., Ackermann, H., Wannke, M., & Hertrich, I. (2010). Brain and Cognition. *Brain and Cognition*, *73*(3), 194–202. doi:10.1016/j.bandc.2010.05.003

Siegel, J. A., & Siegel, W. (1977). Categorical perception of tonal intervals: Musicians can't tellsharp from flat. *Perception & Psychophysics*, *21*(5), 399–407.

Simon, H. J. (1978). Selective anchoring and adaptation of phonetic and nonphonetic continua. *The Journal of the Acoustical Society of America*, *64*(5), 1338–1357. doi:10.1121/1.382101

Simpson, D. (2005). Phrenology and the neurosciences: contributions of FJ Gall and JG Spurzheim. *ANZ Journal of Surgery*, *75*(6), 475–482. doi:10.1111/j.1445-2197.2005.03426.x

Slotnick, S. D., & Schacter, D. L. (2004). A sensory signature that distinguishes true from false memories. *Nature Neuroscience*, *7*(6), 664–672. doi:10.1038/nn1252

Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology*, *33*(1), 5–42.

Smith, E. E., Jonides, J., Marshuetz, C., & Koeppe, R. A. (1998). Components of verbal working memory: evidence from neuroimaging. *Proceedings of the National Academy of Sciences*, *95*(3), 876–882.

Soon, C. S., Brass, M., Heinze, H.-J., & Haynes, J.-D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, *11*(5), 543–545. doi:10.1038/nn.2112

Soon, C. S., He, A. H., Bode, S., & Haynes, J.-D. (2013). Predicting free choices for abstract intentions. *Proceedings of the National Academy of Sciences*, *110*(15), 6217–6222. doi:10.1073/pnas.1212218110/-/DCSupplemental/pnas.201212218SI.pdf

Spitsyna, G., Warren, J. E., Scott, S. K., Turkheimer, F. E., & Wise, R. J. (2006). Converging language streams in the human temporal lobe. *The Journal of Neuroscience*, *26*(28), 7328–7336. doi:10.1523/JNEUROSCI.0559-06.2006

Springer, S. P., & Gazzaniga, M. S. (1975). Dichotic testing of partial and complete split brain subjects. *Neuropsychologia*, *13*(3), 341–346.

Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, *253*(5026), 1380–1386.

Staeren, N., Renvall, H., De Martino, F., Goebel, R., & Formisano, E. (2009). Sound categories are represented as distributed patterns in the human auditory cortex. *Current Biology*, *19*(6), 498–502.

Studdert-Kennedy, M. (1963). Reaction Time to Synthetic Stop Consonants and Vowels at Phoneme Centers and at Phoneme Boundaries. *The Journal of the Acoustical Society of America*, *35*(11), 1900. doi:10.1121/1.2142747

Tagaris, G. A., Kim, S.-G., Strupp, J. P., Andersen, P., Uğurbil, K., & Georgopoulos, A. P. (1997). Mental rotation studied by functional magnetic resonance imaging at high field (4 Tesla): Performance and cortical activation. *Journal of Cognitive Neuroscience*, *9*(4), 419–432.

Takahashi, E., Ohki, K., & Kim, D.-S. (2013). Dissociation and convergence of the dorsal and ventral visual working memory streams in the human prefrontal cortex. *Neuroimage*, *65*(C), 488–498. doi:10.1016/j.neuroimage.2012.10.002

Tallal, P., Miller, S., & Fitch, R. H. (1993). Neurobiological basis of speech: a case for the preeminence of temporal processing. *Annals of the New York Academy of Sciences*, *682*(1), 27–47.

Tanaka, K., Saito, H., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology*, *66*, 170–189.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, *311*(5761), 670–674.

Tsunada, J., Lee, J. H., & Cohen, Y. E. (2011). Representation of speech categories in the primate auditory cortex. *Journal of Neurophysiology*, *105*(6), 2634–2646. doi:10.1152/jn.00037.2011

Ungerleider, L. G., & Haxby, J. V. (1994). 'What'and 'where'in the human brain. *Current Opinion in Neurobiology*, *4*(2), 157–165.

Vallar, G. (1998). Spatial hemineglect in humans. *Trends in Cognitive Sciences*, *2*(3), 87–97.

Van Essen, D. C., & Maunsell, J. H. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in Neurosciences*, *6*, 370–375.

Vouloumanos, A., Kiehl, K., Werker, J., & Liddle, P. (2001). Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *Cognitive Neuroscience, Journal of*, *13*(7), 994–1005.

Wada, J., & Rasmussen, T. (1960). Intracarotid injection of sodium amytal for the lateralization

of cerebral speech dominance: experimental and clinical observations. *Journal of Neurosurgery*, 266–282.

Warren, J. D., & Griffiths, T. D. (2003). Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. *The Journal of Neuroscience*, *23*(13), 5799–5804.

Warren, J. E., Wise, R. J. S., & Warren, J. D. (2005). Sounds do-able: auditory–motor transformations and the posterior temporal plane. *Trends in Neurosciences*, *28*(12), 636–643. doi:10.1016/j.tins.2005.09.010

Weeks, R. A., Aziz-Sultan, A., Bushara, K. O., Tian, B., Wessinger, C. M., Dang, N., et al. (1999). A PET study of human auditory spatial processing. *Neuroscience Letters*, *262*, 155–158.

Wernicke, C. (1874). *Der aphasische Symptomencomplex*.

Westbury, C. F., Zatorre, R. J., & Evans, A. C. (1999). Quantifying variability in the planum temporale: a probability map. *Cerebral Cortex*, *9*(4), 392–405.

Wheaton, L. A., & Hallett, M. (2007). Ideomotor apraxia: A review. *Journal of the Neurological Sciences*, *260*(1-2), 1–10. doi:10.1016/j.jns.2007.04.014

Whorf, B. L. (1956). *Language, thought, and reality*. Oxford, England: Technology Press of MIT.

Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research*, *156*, 249–268. doi:10.1016/S0079-6123(06)56013-3.3d

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*(7), 701–702. doi:10.1038/nn1263

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780–7785.

Wise, S. P., di Pellegrino, G., & Boussaoud, D. (1996). The premotor cortex and nonstandard sensorimotor mapping. *Canadian Journal of Physiology and Pharmacology*, *74*(4), 469–482.

Wolmetz, M., Poeppel, D., & Rapp, B. (2011). What does the right hemisphere know about phoneme categories? *Journal of Cognitive Neuroscience*, *23*(3), 552–569.

Worsley, K. J., Liao, C. H., Aston, J., Petre, V., Duncan, G. H., Morales, F., & Evans, A. C. (2002). A General Statistical Analysis for fMRI Data. *Neuroimage*, *15*(1), 1–15. doi:10.1006/nimg.2001.0933

Yoo, L., & Fujinaga, I. (1999). A comparative latency study of hardware and software pitch-trackers. *Proceedings of the 1999 ICMC. International Computer Music Association*.

Zacks, J. (2008). Neuroimaging studies of mental rotation: a meta-analysis and review. *Cognitive Neuroscience, Journal of*, *20*(1), 1–19.

Zarahn, E., Aguirre, G. K., & D'Esposito, M. (1997). Empirical analyses of BOLD fMRI statistics. *Neuroimage*, *5*(3), 179–197.

Zarate, J. M., & Zatorre, R. J. (2008). Experience-dependent neural substrates involved in vocal pitch regulation during singing. *Neuroimage*, *40*(4), 1871–1887. doi:10.1016/j.neuroimage.2008.01.026

Zarate, J. M., Wood, S., & Zatorre, R. J. (2010). Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers. *Neuropsychologia*, *48*(2), 607–618. doi:10.1016/j.neuropsychologia.2009.10.025

Zatorre, R. J. (1983). Category-boundary effects and speeded sorting with a harmonic musical-interval continuum: Evidence for dual processing. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(5), 739.

Zatorre, R. J. (1985). Discrimination and recognition of tonal melodies after unilateral cerebral excisions. *Neuropsychologia*, *23*(1), 31–41.

Zatorre, R. J. (1989). Perceptual asymmetry on the dichotic fused words test and cerebral speech lateralization determined by the carotid sodium amytal test. *Neuropsychologia*, *27*(10), 1207–1219.

Zatorre, R. J., & Belin, P. (2001). Spectral and Temporal Processing in Human Auditory Cortex. *Cerebral Cortex*, *11*, 946–953.

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1087–1104. doi:10.1016/S1364-6613(00)01816-7

Zatorre, R. J., & Halpern, A. R. (1979). Identification, discrimination, and selective adaptation of simultaneous musical intervals. *Perception & Psychophysics*, *26*(5), 384–395. doi:10.3758/BF03204164

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002a). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, *6*(1), 37–46.

Zatorre, R. J., Bouffard, M., & Belin, P. (2004). Sensitivity to auditory object features in human temporal neocortex. *The Journal of Neuroscience*, *24*(14), 3637–3642. doi:10.1523/JNEUROSCI.5458-03.2004

Zatorre, R. J., Bouffard, M., Ahad, P., & Belin, P. (2002b). Where is “where” in the human

auditory cortex? *Nature Neuroscience*, *5*(9), 905–909. doi:10.1038/nn904

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience*, *8*(7), 547–558. doi:10.1038/nrn2152

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, *256*(5058), 846–849.

Zatorre, R. J., Meyer, E., Gjedde, A., & Evans, A. C. (1996). PET studies of phonetic processing of speech: review, replication, and reanalysis. *Cerebral Cortex*, *6*(1), 21–30.

Zeki, S. M. (1971). Cortical projections from two prestriate areas in the monkey. *Brain Research*, *34*, 19–35.

Zeki, S. M. (1976). Colour coding in the superior temporal sulcus of the rhesus monkey [proceedings]. *The Journal of Physiology*, *263*(1), 169P.

Zeki, S. M. (1978). Functional specialisation in the visual cortex of the rhesus monkey. *Nature*, *274*, 423–428.

Zevin, J. D., & McCandliss, B. D. (2005). Dishabituation of the BOLD response to speech sounds. *Behav Brain Funct*, *1*(4).

Zimmer, H. D. (2008). Visual and spatial working memory: from boxes to networks. *Neuroscience and Biobehavioral Reviews*, *32*(8), 1373–1395.