Title
Sensorimotor Control of Human Singing

Permalink
https://escholarship.org/uc/item/6xq138dh

Author
Médé, Butovens

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, MERCED, CALIFORNIA

Sensorimotor Control of Human Singing

A dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Cognitive and Information Sciences

by

Butovens Médé

Committee in charge:

Professor Ramesh Balasubramaniam, Ph.D., Chair
Cognitive & Information Sciences
University of California, Merced

Professor Christopher T. Kello, Ph.D.
Cognitive & Information Sciences
University of California, Merced

Rick Dale, Ph.D.,
Department of Communication
University of California, Los Angeles

2021

Copyright

Butovens Médé, 2021

The dissertation of Butovens Médé is approved,
and it is acceptable

in quality and form for publication on microfilm and electronically:

Dr. Rick Dale

Dr. Christopher T. Kello

Dr. Ramesh Balasubramaniam

University of California, Merced

2021

This dissertation is dedicated to my family:

to Arol and Stéphanie who have always supported me along the way
to my mom, Marlène, my sister, Murielle, and my brother, Romain, who have fed the
fire that kept burning inside me.

This dissertation is also dedicated:

to all minorities in higher education.

Contents

List of Figures

List of Tables

Acknowledgments

My heartfelt thank you to my committee members - Ramesh Balasubramaniam, Chris Kello, and Rick Dale - for their support and mentorship during my graduate journey at UC Merced. The saying goes: *"It takes a village to raise a child"* , and I'll say: "It takes a great French speaking committee to make a (happy) PhD". I learned from each and everyone of them. They have all provided me with encouragements, thoughtful and unique lessons, and critical thoughts that have shaped the scientist and the person that I am.

Thank you to my lab mates past and present: Jessica Ross, Alex Pabst, Tim Shea, Chelsea Gordon, Dan Comstock, Brandon Batzloff, and Shannon Proksch who have made the lab a welcoming intellectual and social environment in which I could safely grow. This journey wouldn't have been the same without each and everyone of them.

Thank you to the entire CogSci department for the intellectual stimulation and support they provided me with over the years.

And special thanks to Sarah Depaoli from the Psychology department, and David Jennings, for the myriad of ways in which they have helped me. I am very grateful for those two.

BUTOVENS MÉDÉ

bmede@ucmerced.edu
University of California, Merced,
School of Social Sciences, Humanities, and Arts

EDUCATION

University of California Merced, Merced   *2014 - Present*
M.S./Ph.D. Cognitive and Information Sciences

The City College of New York, New York   *2014*
B.S. Cognitive Neuropsychology and Sound Design

LaGuardia Community College, Long Island City   *2012*
A.A. Music Recording Technology

ADDITIONAL TRAINING

Neuromatch Academy, An online school for Computational Neuroscience *2020*
JOSTR Online Course in Text Analysis with Python   *2020*
Ninth Annual JAGS and WinBUGS Workshop: Bayesian Modeling for Cognitive Science,
University of Amsterdam, Netherlands   *2020*
Summer School: Bayesian Statistical Analyses for the Human, Social and Cognitive Science,
University of Verona, Italy   *2019*
Data Carpentry Workshop: Data Analysis using Unix, Git and R,
University of California, Merced   *2019*
Institute for Computational & Mathematical Engineering (ICME) workshop:
Fundamentals of Data Science, Stanford University, Stanford   *2018*
Data Carpentry Workshop: Data management, analysis and visualization using the
OpenRefine, Python, and SQL, University of California, Merced   *2017*
Big Data Neuroscience Workshop, Indiana University, Bloomington   *2017*
Software Carpentry Workshop: Reproducible data management, analysis and visualization using
the shell, R, and git, University of California, Merced   *2017*
UC Music Experience Research Community Initiative (MERCI) 2016-17 Student Exchange,
University of California, San Francisco (Host: Julene Johnson)   *2016*
Certificate in Undergraduate Learning Outcomes Assessment: Pedagogy and Program Planning,
University of California, Merced   *2016*
APA Advanced Training Institute: Nonlinear Methods for Psychological Science,
University of Cincinnati, Cincinnati   *2016*
UC Retreat Workshop on Research in Music Experience and Communication,
Marconi, CA   *2016*
Data Analysis in Python Workshop, University of California, Merced
(Instructor: Bryan Kerster)   *2016*

UC Music Experience Research Community Initiative (MERCI) Student Exchange, Swartz Center for Computational Neuroscience (SCCN), University of California, San Diego (Hosts: Scott Makeig and John Iversen)   *2016*

UC MERCI Symposium/Workshop on Research on Music Experience and Communication, University of California, Los Angeles   *2015*

MathWorks Seminar: Data Analysis and Visualization with MATLAB, University of California, Merced   *2015*

Transcranial Magnetic Stimulation (TMS) Methods and Practice, Division of Biokinesiology and Physical Therapy, University of Southern California, Los Angeles   *2015*

Generalized Linear Model Workshop, University of California, Merced
(Instructor: Bodo Winter)   *2015*

CERTIFICATIONS

Adult CPR, Child CPR, and Infant CPR, FirstAidWeb, Inc. and National Health & Safety Foundation, Valid until February 2017 (renewed)   *2014*

EMPLOYMENT HISTORY

University of California Merced, Merced

SONA administrator (Coordinate UC Merced human subject pool)	*Aug. 2015 - Pres.*
Instructor of Record, Research Methods	*May 2021 - Jul. 2021*
Teaching Assistant, Research Methods (Tyler Marghetis)	*Jan. 2021 - May 2021*
Teaching Assistant, Ethics in Technology (Hanna Gunn)	*Aug. 2020 - Dec. 2020*
Teaching Assistant, Intro to Cog Neuro. (Kristina Backer)	*Jan. 2020 - May 2020*
Teaching Assistant, Applied Ethics (David Jennings)	*Aug. 2019 - Dec. 2019*
Teaching Assistant, Intro to Ethics (David Jennings)	*Jan. 2019 - May 2019*
Teaching Assistant, Mind, Brain & Computation (Chris Kello)	*Aug. 2018 - Dec. 2018*
Teaching Assistant, Introduction to Logic (Rolf Johansson)	*Jan. 2018 - May 2018*
Teaching Assistant, Introduction to Logic (Rolf Johansson)	*Aug. 2017 - Dec. 2017*
Teaching Assistant, Research Methods (Rick Dale)	*Jan. 2017 - May 2017*
Teaching Assistant, Intro to Cognitive Science (Rick Dale)	*Aug. 2016 - Dec. 2016*
Teaching Assistant, Applied Ethics (David Jennings)	*Jan. 2016 - May 2016*
Teaching Assistant, Ethics (David Jennings)	*Aug. 2015 - Dec. 2015*
Graduate Student Researcher (PI: R. Balasubramaniam)	
Effects of rhythm/ music processing on reading	*Jan. 2015 - May 2015*
Teaching Assistant, Perception & Action (R. Balasubramaniam)	*Aug. 2014 - Dec. 2014*

University of Cincinnati, Cincinnati

Research Assistant, Perceptual-Motor Dynamics Lab (Advisor, Michael Riley),	
NSF sponsored REU in Complex Psychological System	*Jun. 2013 - Aug 2013*

City College of New York, New York

Research Assistant, (Advisor, Jón G. Sigurjónsson),	
Effects of Concussions/ mTBIs on Full-Contact-Sport Athletes	*Jan. 2013 - Jul 2014*

RESEARCH ARTICLES

1. Proksch, S., Comstock, D. C., Médé, B., Pabst, A., & Balasubramaniam, R. (2020). Motor and predictive processes in auditory beat and rhythm perception. *Frontiers in Human Neuroscience*, 14, 375. doi: 10.3389/fnhum.2020.578546

2. Ramirez-Aristizabal, A. G.∗, Médé, B.∗, & Kello, C. T. (2018). Complexity matching in speech: Effects of speaking rate and naturalness. *Chaos, Solitons & Fractals, 111*, 175-179.

3. Balasubramaniam, R., Hove, M. J., & Médé, B. (2018). Factorization of Force and Timing in Sensorimotor Performance: Long?Range Correlation Properties of Two Different Task Goals. *Topics in cognitive science*, 10(1), 120-132. doi: 10.1111/tops.12301

4. Kello, C. T., Bella, S. D., Médé, B., & Balasubramaniam, R. (2017). Hierarchical temporal structure in music, speech and animal vocalizations: jazz is like a conversation, humpbacks sing like hermit thrushes. *Journal of The Royal Society Interface*, 14(135), 20170231.

5. Quatman-Yates, C. C., Bonnette, M. S., Hugentobler, J. A., Médé, B., Kiefer, A. W., Kurowski, B. G., & Riley, M. A. (2015). Postconcussion postural sway variability changes in youth: the benefit of structural variability analyses. *Pediatric physical therapy: the official publication of the Section on Pediatrics of the American Physical Therapy Association*, 27(4), 316.

POSTERS

1. Médé, B., Balasubramaniam, B., & Kello, C. T. (2019). *Jazz and Raga: A hierarchical temporal structure comparison*. Society for Music Perception and Cognition 2019, New York, NY

2. Médé, B., Balasubramaniam, B., & Kello, C. T. (2016). *Temporal event clustering in speech versus music*. 38th Annual Meeting of the Cognitive Science Society, Philadelphia, PA

3. Balasubramaniam, B., Hove, M. J, & Médé, B. (2015). *Factorization of pattern and energy in the motor sequence production: using long range correlations to understand individual differences*. 4th Bi-annual Meeting of the Society for Complex Systems in Cognitive Science, Pasadena, CA

4. Sigurjonsson, J., Florez, E., Iino, S., Rahman, M., Candolfi, M., Médé, B. & Tartter, V. (2015). *Possible effects of concussive and sub-concussive blows on cognition*. 14th FEPSAC Congress, Bern, Switzerland

5. Médé, B., Quatman-Yates, C., & Riley, M.A. (2013). *Use of Sample Entropy as a Tool to Assess a System?s Integrity: The Influence of Sport Related Concussion on Postural Balance*. REU in Complex Psychological System Symposium, Cincinnati, OH

PROFESSIONAL AFFILIATIONS

Cognitive Science Society, Cognitive Neuroscience Society, Society for Music Perception & Cognition, American Psychological Association

Abstract

Despite being a common activity, singing is a complex behavior. It involves processes such as: maintaining mental imagery of desired tones, activating a motor plan to make a vocal production, coordinating different parts of the vocal apparatus and respiratory system, and monitoring the auditory and kinesthetic output for accuracy. While singing can be achieved by a majority of people, some cannot do it accurately. In this dissertation, I investigate one of the posited causes of poor pitch matching and experimentally test it.

The inverse-model deficit of poor pitch matching suggests that poor singing results from the brain's inability to instantiate a precise motor plan with one's vocal apparatus when one tries to vocally reproduce a pitch. This hypothesis is motivated by previous work in sensorimotor control suggesting that in order to perform an action in an environment, one relies on two distinct internal models. These internal models are thought to be instantiated by the cerebellum. In this dissertation, I embed this hypothesis within the larger context of cerebellar cognition and the bayesian brain hypothesis.

First, I present multiple findings of the involvement of the cerebellum in cognitive tasks, including in musical production. This will provide some support for the role of the cerebellum in cognition, in addition to its role in motor control. This first part will also allow for an understanding of some of the neural underpinnings involved in implementing internal models.

Next, I present recent findings which could support the notion of internal models within the Bayesian brain framework. This will provide an understanding of these internal models at an algorithmic level (Marr, 1982).

Finally, I test one of the key predictions of the inverse-model deficit of poor pitch matching i.e. perturbing the cerebellum during singing task would lead to poorer pitch matching performance. Adult participants had to perform various singing assessment tasks before and after brain stimulations on different brain regions. I did not observe a decrease in pitch matching accuracy (relative to the individuals' baseline) after stimulation of the cerebellum. While these results were inconclusive, further research using TMS to investigate pitch accuracy should be undertaken so as to better understand the role of internal models in vocal production such as singing.

This dissertation, *Sensorimotor Control of Human Singing*, is submitted by Butovens Médé in 2021 in partial fulfillment of the degree Doctor of Philosophy in Cognitive and Information Sciences at the University of California, Merced, under the guidance of dissertation committee chair Ramesh Balasubramaniam.

Chapter 1

Introduction

1.1 Sensorimotor Prediction

A major theory of motor control in cognitive science posits that in order to enable the body to move around and act in the environment, the brain uses internal models (i.e., neural representations) of the external world (Wolpert et al., 1995, 1998; Kawato, 1999). Those models can be divided into two categories: Forward models and Inverse models. A forward model allows for the predictions of the sensory consequences of an action. In other words, when one takes an action or makes a movement, he or she anticipates the bodily experience associated with that action (or movement). As the movement is unfolding the brain compares the sensory anticipation (from the internal prediction) to the actual sensory information received from the world. The sensory prediction error (i.e., the difference between the internal prediction and the real world sensory information) can then be used to correct (or smoothen) the movement. As an example, imagine a person climbing down the stairs of a familiar place from the 2nd floor to the 1st floor. Because the person knows (and expects) each step of the stairs to be spaced equally, they can climb them down without looking down at their feet. At each step climbed down, the brain of that person anticipates and predicts the sensation that they should be feeling from the ball of their feet touching each stair step at regular interval. If, for an unexpected reason, one of those steps is higher than the other ones, the perceptual sensation the person will receive from the ball of one of their feet will be delayed compared to the prediction made by their internal (forward) model. Depending on how big the prediction error is, the person could either quickly readjust the movement of their foot (as well as their entire body position) and just stumble, or fail to correct that movement, which would lead them to fall. Thus, a forward model in the brain solve motor control problems that require the anticipation of the sensory consequences of an action in order to produce smooth movements.

The second type of internal model is the inverse model. This model can be seen as the inverse of a forward model (Ishikawa et al. 2016). It allows the planning of some actions for a desired perceptual outcome. In other words, given a certain perceptual state that one wants to achieve, what actions should be taken in order to reach that perceptual state (or bodily sensations). Thus, unlike a forward model which tries to predict the perceptual consequences of an action, here the final perceptual state is known and a plan (of actions) needs to be implemented in order to achieve it. For example, as an analogy, let's take a professional basketball player trying to score at the free throw line after being

fouled. It can be assumed that because that person has practiced free throws many times in their athlete career, they have a solid knowledge (or internal representation) of how it should feel like to unfold the series of movements that will enable them to score from that position (i.e., they know the perceptual desired state). However, at this moment in time they might be tired because of having run a lot during the game. This would add some noise to their motor movements. In addition, the game might be during an important championship final, which puts the person under more mental stress than usual. This could also add more noise to their motor movements. Finally, their vision might be blurry from the fatigue and lack of hydration before the game. From an inverse model perspective, the question becomes, what motor actions should be implemented in other for the player to achieve the desired perceptual state (i.e., bodily sensations leading them to scoring at the free throw line) despite the intrinsic and extrinsic noise affecting the body? It is believed that an inverse model allows the brain/body to solve this type of movement problems (among others) by instantiating a plan for those motor actions.

A body moving through an environment is facing these challenges on a constant basis. In motor control theories, these two types of models are believed to work together in order to enable precise and controlled movements. It is also believed that the cerebellum plays an important role in the cognitive and neural processes of these models (Ito, 2006, 2008; Tanaka et al, 2020). In addition, while there is accumulating evidence supporting the cerebellum to be the locus of the internal forward model formation (Ito, 2006, 2008; Tanaka et al, 2019, 2020) there appears to be less consensus to whether the cerebellum is also the locus of the internal inverse model formation (Yavari et al. 2016).

In the following section, we will see how one's performance in singing can be explained by a brain trying to optimize body movements using those models, and what it can teach us about sensorimotor control in general and the reason of poor singing performance in particular.

1.2   Singing

Singing is one of the most ubiquitous forms of musical activity (Hutchins et al., 2014). Despite being so common some people are unable to vocally match target pitches within a semitone while singing (Pfordresher, & Mantell, 2014). There have been a variety of causes being posited for poor pitch matching e.g., perceptual deficit, motor production deficit, memory deficit, sensorimotor integration (Pfordresher and Brown, 2007). While each of these bring an element of response to the problem, the question remains. What are the cognitive and neural processes underlying one's inability to vocally match pitches?

One of the most prevalent hypotheses has been the perceptual deficit hypothesis (i.e., poor pitch perception results in poor vocal pitch matching ability). Researchers such as Amir et al. (2003) and Nikjeh et al., (2009) have shown that pitch discrimination, or pitch perception threshold and vocal pitch matching accuracy are significantly correlated. However, the method used (i.e., different pitch ranges compared with sung tones that are different from the pitches of the perceptual task) renders the comparison problematic. Moreover, Hyde and Peretz (2004) demonstrated that most normal subjects have no difficulty discriminating between two or several pitches of a quarter of a semitone apart. Poor singing often being defined has being out of tune by half or a full semitone, it makes the perceptual deficit hypothesis for poor singing more arguable.

Adding to this body of work, Hutchins and Peretz (2012) have provided some compelling evidence showing that poor vocal pitch matching ability is generally not due to poor perception ability. They conducted a clever experiment that made use of a device called a 'slider'. The slider enabled participants to create a synthetic tone by pressing a horizontal touch-sensitive strip at a specific position. The use of this non-vocal alternative to pitch matching in their study has shown, for instance, that musicians and non-musicians were better at matching synthesized pitches when using the slider as opposed to using their own vocal apparatus. It has also demonstrated that even when one controls for timbre (i.e., where the sound quality of the target pitches is identical to that of the vocally produced ones), the slider "advantage" still remained. For instance, musicians and non-musicians, who were in general better at using their vocal apparatus to match pre-recorded pitches that they themselves produced for the experiment, were still less accurate in doing as compared to when using the slider (Hutchins and Peretz, 2012). This also strengthens the hypothesis that perceptual deficit problems may not be a main cause of poor singing ability. However, there seems to be more of a consensus about the role played by perceptual deficit in poor singers for people suffering from amusia (i.e., a congenital or acquired condition which makes people unable to recognize or produce musical tones; Peretz & Vuvan, 2017).

Other accounts of poor pitch matching have focused on its mechanistic aspect. Singing requires the coordination of core physical structures such as the lungs, the vocal folds, the vocal tracts (Ghazanfar & Rendall, 2008), which are regulated by a complex dedicated neural network (Brown et al., 2004; Zarate, 2013; Della Bella, 2016). The quality of the vocal output and its perception, therefore, relies on the efficient coordination of these respiratory, laryngeal and articulatory mechanisms (Sundberg, 1990; Sundberg 1994; Della Bella, 2016). These findings and similar ones have given rise to the motor deficit account of poor pitch matching. This account states that poor singers lack the vocal-motor control (e.g., vocal range, vocal stability, control of one's apparatus) necessary to reproduce pitches despite being able to accurately discriminate or perceive them; Joyner (1969) and Cleall (1970) (as cited by Hutchins & Peretz, 2012). Pfordresher and Brown (2007) also argued for a vocal motor control deficit but ruled out this hypothesis after finding no difference between poor and good singers in vocal ranges and sustainment of level tones.

Recently, a new hypothesis called the inverse model deficit hypothesis has started to emerge. This hypothesis was inspired by the internal model theory of motor control (Wolpert et al., 1995, 1998; Kawato, 1999). The inverse model deficit hypothesis states that poor vocal imitation of pitch is the result of an inefficient auditory vocal system representation (i.e., deficient internal model/inverse model). Within this view, the emphasis for the activity of singing is on the instantiation a motor plan (e.g., correct vocal apparatus movements) that will enable the desired perceptual outcome (e.g., correct proprioception sensations from the vocal apparatus, correct auditory feedback). Also, within this view, it can be assumed that well-trained singers have a very reliable internal representation of their auditory vocal system (which leads them to produce the desired pitches most of the time), and that non-musicians have a less reliable internal representation of their auditory vocal system. Guided by these assumptions, Pfordresher and Mantell (2014) have conducted an experiment that showed that there is a self-matching advantage when imitating recordings of oneself singing sequences of pitches (as opposed to imitating someone else's recordings of the same task). This self-matching advantage was taken as

evidence that one's intimate familiarity with a desired perceptual outcome (e.g., sound of own voice) helps in generating the actions (e.g., vocal productions) that lead to it. It also has provided some evidence that the ability of poor singers to imitate sung pitches, or a sung melody, is compromised when the pitches (or pitch pattern) is not one that they have produced in the past.

Taken together, it can be argued that most cases of poor singing may reflect the inability of poor singers to translate a novel perceptual stimulus representation into a series of actions that would give rise to the same perceptual stimulus. This points toward a deficit of the inverse model.

In this dissertation, I will first provide a foundational understanding of the cerebellum (Chapter 2). I will present some of the prevalent theories (past and present) regarding its function and show that it cannot be seen solely as a motor-control organ. I will then demonstrate how the cerebellum contributes to cognition in specific domains (e.g., language, music). This will show why the cerebellum is thought to be the locus of internal models, and it will provide an underlying framework to understand the role of the cerebellum in singing. Next, I will briefly introduce Bayesian inference (Chapter 3). I will show how Bayesian inference allows the integration of information under uncertainty and the use of prior knowledge. This will provide the tools necessary to understand the Bayesian model of cognition. I will give examples of how well-known psychological phenomena can be explained within this framework and, then I will present found evidence of how state estimation in motor control can be implemented from the point of view of an ideal Bayesian observer. This will provide an overall view of a Bayesian model of cognition that will allow us to understand how an activity such as singing can also be seen through the lens of Bayesian model of cognition. I will then present my experiment (Chapter 4), which investigated the internal model hypothesis in the context of singing. In this experiment, I used TMS to downregulate participants' cerebellum (among other brain regions). This was a pre-post test experiment where participants had to complete a singing task assessment before and after stimulations with TMS. The hypothesis was: perturbing the cerebellum using TMS will lead participants to do worse in their singing task assessment given the cerebellum's involvement in the instantiation of internal models. Finally, I will conclude (Chapter 5) with a general discussion and future directions.

Chapter 2

The Cerebellum

2.1 Introduction

Although a full understanding of the functional domain of the cerebellum has remained elusive, the understanding of its role has been extended in recent decades far beyond the realm of motor control. There has been an accumulation of evidence indicating that the cerebellum has a more important role in non-motor functions than was previously thought. The hypothesis that the cerebellum contributes to cognitive functions is largely based on what is known about its homogenous neural micro-circuitry, its anatomical connections, and its phylogenetic development (Dow, 1942; Diamond, 2000).

This chapter provides an overview of the theories about the role of the cerebellum in cognition: the internal model hypothesis, the universal cerebellar transform (UCT) theory, and the cerebellar cognitive affective syndrome (CCAS) theories. First, I will give a description of the cerebellar structure to lay the foundations for understanding how and why the particular microanatomy of the cerebellum has shaped and informed the theories of cerebellar functions. Next, I will focus on the role of the cerebellum in language and music.

2.2 Cerebellar Gross Anatomy

The cerebellum ("little brain") is a structure located ventrally and toward the posterior of the cerebral hemispheres. Despite accounting for only 10% of the brain mass, it contains more than 80% of the neurons in the brain (Azevedo et al., 2009). With only six types of cell, repeated throughout and arranged in an almost crystalline form, it is one of the most organized structures in the central nervous system (Miall, 2013).

There are various ways one can view the organization of the cerebellum. It can be divided by anatomical lobes, zones, or functions. Anatomical lobe division is along crosswise lines, whereas zone division is lengthwise. Both division schemes are relevant to functional division of the cerebellum. The anatomical lobes are the anterior lobe, the posterior lobes, and the flocculonodular lobe. These three lobes are separated by deep fissures going across the cerebellum from side to side: the primary fissure (or preclival fissure) and the posterolateral fissure (or postclival fissure). The anterior, posterior, and flocculonodular lobes can be divided into lobules (or folia) numbered from I to X (in Roman numerals). Lobules I–V correspond to the anterior lobes, VI–VII and part of

IX to the posterior lobes, and X and part of IX to the flocculonodular lobes (Apps & Garwicz, 2005; Schmahmann, 2019; see Figure 2.1).

Another way to view the organization of the cerebellum is by zones. The more central zone, in the midline of the cerebellum (from front to back and including all the lobes), is the vermis. On either side of the vermis is the intermediate zone or paravermis. The most lateral zones of the cerebellum, lateral to the paravermis, are the cerebellar hemispheres. Similar to the three lobes, these three zones can be further divided into parasagittal zones based on their anatomical connections.
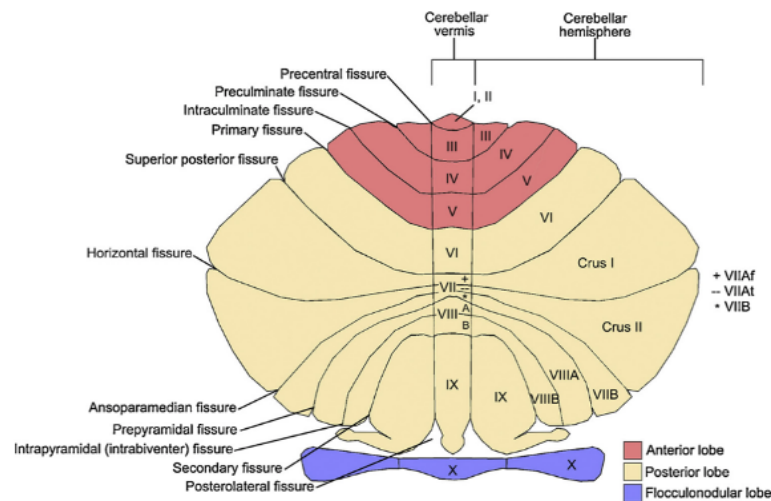
Figure 2.1: Flattened representation of the cerebellum depicting: the three major lobes and their lobules, the vermis and the cerebellar hemispheres. (Paravermis subdivision not shown here). (From "Cerebro-cerebellar circuits in autism spectrum disorder," by A. M. D'Mello, and C. J. Stoodley, 2015, Frontiers in neuroscience, (9), 408, p. 3. CC BY-NC-ND.)

The third way to map the cerebellum is by functions. The three main functional regions of the cerebellum (from the oldest to the newest) are the vestibulocerebellum (the functional analog of the flocculonodular), the spinocerebellum, and the cerebrocerebellum. The vestibulocerebellum is involved in controlling balance, head movements, and ocular reflexes such as the vestibulo-ocular reflex. It receives inputs mainly from the vestibular nuclei, but also from the auditory and visual sensory input (Broussard, 2014; Miall, 2013).

The spinocerebellum, which comprises the anterior vermis and the paravermis, is involved in the integration of sensory input with motor commands to produce adaptive motor coordination. It receives spinal proprioceptive information and auditory and visual input and projects its output back to the spinal cord via the deep cerebellar nuclei, cerebral cortex, and brain stem (Miall, 2013).

Lastly, the cerebrocerebellum, which is the largest functional subdivision of the human cerebellum and comprises the caudal vermis, paravermis, and lateral hemispheres, is involved in the planning and timing of movements and motor learning. It receives information exclusively from the cerebral cortex, via the pontine nuclei, mainly from regions such as the parietal lobe and the primary sensory lobe of the brain. Its output returns to the cerebral cortex via the ventrolateral part of the thalamus. It is the func-

tional region that is thought to be involved in the cognitive functions of the cerebellum (Broussard, 2014). This functional region will be discussed further in this manuscript.

2.3 Cerebellar Cortical Structure and Microanatomy

The cerebellar cortex and the deep cerebellar nuclei (the unique output structures of the cerebellum) constitute two major parts of the cerebellum. The cerebellar cortex, which includes most of the cerebellar neurons, is composed of three layered sheets of neurons "folded like an accordion" (Broussard, 2014). The surface, middle, and deep layers of the cerebellar cortex, called the molecular layer, the Purkinje layer, and the granular layer, respectively, contain six major cell types. The principal type of cell is the Purkinje cell. They extend through the entirety of the cerebellar cortex and compose the Purkinje layer. They are some of the largest cells in the brain and are central to the cerebellar circuitry. In the granule layer, below the Purkinje-cell layer, are the granule cells (some of the smallest cells in the brain), Golgi cells, and the unipolar brush cells. In the molecular layer, above the Purkinje layer, are the Basket cells and the outer Stellate cells (Glickstein & Voogd, 2010; Broussard, 2014; see Figure 2.2).
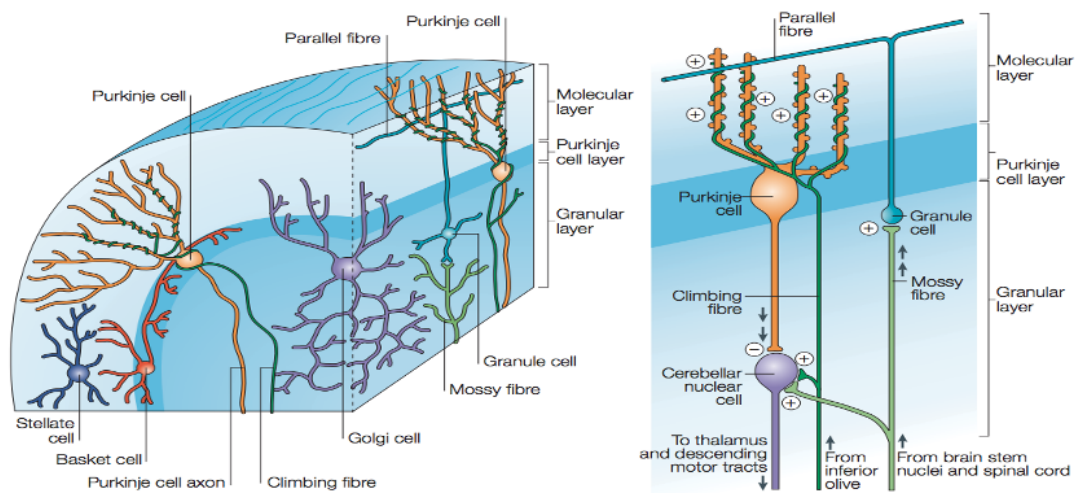
Figure 2.2: Simplified cytoarchitecture of cerebellar cortex. The Mossy fibers and the Climbing fibers are the main cerebellar inputs. The Mossy fibers are coming from the brain stem nuclei and spinal cord. They project to the Granule cells (which branch into parallel fibers) and Golgi cells. The Climbing fibers are coming from the inferior olive and project to the Purkinje cells and Golgi cells as well. Both Mossy and Climbing fibers have (+) projections. Granule cells (parallel fibers) make synaptic excitatory (+) connections whereas the Golgi, Basket, Stellate, and P-cells make inhibitory (-) synaptic connections. (From "Anatomical and physiological foundations of cerebellar information processing," by R. Apps, and M. Garwicz, 2011, Nature Reviews Neuroscience, 6(4), p. 298. Copyright 2021 by Springer Nature. Reproduced with permission.)

The Purkinje cells are inhibitory (GABAergic) neurons whose axons terminate in the deep cerebellar nuclei, in the cerebellar white matter, and in the vestibular nuclei. They form the main output element of the cerebellar cortex and are the only neurons

in the cerebellum whose axons project outside of it. It is estimated that several million nerve fibers exit the cerebellum (Broussard, 2014). They can be excited by mossy fibers through the granule cells and parallel fibers and by climbing fibers directly. They can also emit action potentials "spontaneously" (i.e., in the absence of intrinsic synaptic input).

The granule cells, which are excitatory, glutamatergic neurons, are innervated by excitatory mossy fiber inputs, and they also receive inhibitory inputs from Golgi cells. Their projections, which do not leave the cerebellum, extend from the granular layer through the P-cell layer and into the molecular layer, where they make a "T" junction to terminate on the P-cell dendrites. Granule cell axons are called parallel fibers. The P-cell dendritic arbors and the parallel fibers form a rectangular lattice that has been thought to be an efficient device for processing and/or storing information (Apps & Garwicz, 2005).

The Golgi cells are inhibitory interneurons located at the top of the granular layer and whose projections terminate on granule cell dendrites. They form an inhibitory feedback loop with the granule cells. In the molecular layer are the Basket and outer Stellate cells, which are inhibitory GABAergic interneurons. The Basket cells' terminals surround the soma and axon hillocks of the P-cells, which allows them to strongly inhibit and modulate the P-cells firing patterns. The Stellate cells synapse with the P-cells' dendrites, which also allows for some inhibition of the P-cells (albeit to a lesser degree than the Basket cells). The Basket and Stellate cells form a link between the P-cells and the granular cells, and they also receive inputs from the climbing fibers (Glickstein & Voogd, 2010; Broussard, 2014; see Figure 2.2).

Lastly, the mossy and climbing fibers form the two major inputs of the cerebellum. The mossy fibers arise from various extra-cerebellar sites such as the pontine nuclei, the vestibular nuclei, and the spinal cord. They make excitatory contact with the deep cerebellar nuclei (directly and via a granule cells–parallel fiber–Purkinje cells cortical loop) and with the granule cells (Miall, 2013; see Figure 2.2, right panel). The climbing fibers, which originate solely from the inferior olivary nucleus within the brain stem, receive input from various areas such as the vestibular system, spinal cord, and from much of the cranial and cortical descending information (Miall, 2013). They send collaterals to the deep cerebellar nuclei before entering the cerebellar cortex and the climbing fibers terminals reach the primary Purkinje cells' dendrites and form excitatory synapses with proximal Basket, Stellate, and Golgi cells (Miall, 2013).

2.4 Theories of the Role of the Cerebellum: Motor and Non-Motor Functions

Most functional analyses of the cerebellum have revolved around its role in the co-ordination and the control of movements. Over the past 30 years, however, clinical, neurophysiological, neuroanatomical and neuroimaging studies have provided substantial evidence for the involvement of the cerebellum in higher-level processing due to its unique structure. As mentioned previously, the cytoarchitecture (i.e., the distribution of cells in cortical layers and sublayers) of the cerebellum is very stereotyped and uniform across all of its subdivisions. However, this quasi-uniformity in the composition of the cortical layers is contrasted with the functional diversity and specificity of cerebellar

connections to the cerebrum (Schmahmann, 2004; Broussard, 2014; Sokolov, Miall, & Ivry, 2017). This has led researchers and theorists to seek to identify the nature of a transform (or computational mechanism) that could originate from that homogenous synaptic organization and yet support that functional heterogeneity (Sokolov, Miall, & Ivry, 2017; Schmahmann, 2019).

One early theory about the role of the cerebellum is the theory of motor learning proposed by Marr (1969) and developed by Albus (1971). This theory has substantially contributed to the creation of the discipline of computational neuroscience (Strata, 2009). According to this theory, the cerebellar cortex has the task of learning motor skills for movements and posture. It does so through the interaction of mossy fibers, climbing fibers, and Purkinje cells. This theory assumes that each action has a specific representation, in terms of elemental movements, that is expressed as a sequence of firing patterns in the inferior olive (Marr, 1969). The olivary cells, which have a one-to-one connection with Purkinje cells via climbing fibers, respond to cerebral instruction for an elemental movement. When those olivary cells fire, they send signals to the corresponding Purkinje cells. The Purkinje cells also receive signals from mossy fibers, via the granule cells and their parallel fibers axons, that are thought to provide information about the context in which the olivary cells fired (Marr, 1969). When there are simultaneous activations of the climbing fibers and parallel fibers at the Purkinje cell level, the parallel fibers and Purkinje cell synapses are reinforced through long-term potentiation (Marr, 1969). It is believed that, during action rehearsal, the Purkinje cells learn to recognize the contexts in which the olivary cells' firing patterns occurred. This recognition leads those Purkinje cells to be able to fire and cause the next elemental movement with only the occurrence of the context after the action has been learned

Extending Marr's theory and modifying it slightly, Albus (1971) proposed that, in order for the learning process to be stable and to generate appropriate motor command signals, pattern recognition of olivary cell bursts (or climbing fiber bursts) and pattern storage must be accomplished principally by weakening parallel fibers and Purkinje cell synapses through long-term depression. While Marr's (1969) theory assumes that no other synapse type is modifiable, Albus (1971) states that not only parallel fiber synapses are adjustable on Purkinje cell dendrites, but also on both Stellate and Basket cell dendrites. The plasticity assumption in the Marr–Albus model is also supported by Ito and Kano (1982).

The motor learning theory as proposed by Marr (1969) and Albus (1971) builds upon Eccles et al.'s "cellular beam hypothesis" (as cited in Ito, 2006; D'Angelo et al., 2011; Bower, 2010), which saw the cerebellum as a collection of individual beams (referring to the parallel fibers of the granule cells connecting to Purkinje cells) that propagate excitatory signals to sequentially activate the Purkinje cells. The excitatory parallel activity along the beams sharpened by lateral inhibition adjacent to those beams was thought to be an important computational mechanism in the cerebellar cortical circuitry (Bower, 2010; Broussard, 2014). The "cellular beam hypothesis" tried to provide support for the role of inhibition in the cerebellum, which was believed to modulate excitatory effects on neuronal outputs by "sculpting," blocking, or cancelling them (Bower, 2010; Braitenberg, 1997, 2002).

Braitenberg (as cited in Braitenberg, 1997) originally saw parallel fibers as generators of time delays regulating the timing sequence of the activation of different muscles involved in one movement. Assuming a fixed point of origin, the excitatory signals

along those parallel fibers could represent precise timing solely based on distance, and it was believed that a system would learn skilled movements by using the precise timing information from task-relevant parallel fiber-Purkinje cell synapses to generate precise temporal patterns.

However, it was shown that, given the length of parallel fibers (a few millimeters) and the relatively constant rate at which signals propagate along them (about 0.5 m/s), the delays generated would be an order of magnitude lower than the known time for movement completion (about 200 ms to 300 ms; Braitenberg, 1997). This led to the abandonment of the original timing hypothesis. Nonetheless, the original timing hypothesis, as posited by Braitenberg (1997) in 1967, set the ground for subsequent theories about how the cerebellum microphysiology may be well suited for temporal representation (Ivry & Spencer, 2004; Spencer & Ivry, 2013) in sensory motor timing (Flament & Hore, 1996; Manto, 2009), perceptual timing (Rao et al., 1997; de Zeeuw et al., 1998; Xu et al., 2006), and in sensorimotor learning tasks (Yeo & Hesslow, 1998; Brach et al., 2009).

Central to the idea of cerebellar-dependent motor learning developed by Marr and Albus is the use of error signals to "instruct" or supervise the motor-learning scheme. A more complete version of the Marr–Albus hypothesis is the Marr–Albus–Ito hypothesis (Ito, 2006), which posits that motor error feedback activates the climbing fiber inputs from the inferior olive. That activation induces changes at parallel fiber-Purkinje cell synapses through long-term depression that alter Purkinje cells' firing pattern responses to mossy fiber inputs. The learned Purkinje cell firing patterns are then reflected in behavioral modifications (Popa et al., 2012, 2016).

Closely related to the error-driven learning mechanism hypothesis is the comparator hypothesis of olivary function originally proposed by Oscarsson (as cited in de Zeeuw et al., 1998). This theory states that the olivocerebellar system compares intended movements with performed movements to provide error detection (de Zeeuw et al., 1998). Due to its afferent connections with the sensory system in the body, the cerebellum can use that error detection mechanism to make appropriate adjustment during movements (Tesche & Karhu, 2000).

Some hypotheses that have built upon the cerebellum-as-a-comparator theory include the internal models for motor control and trajectory planning (Kawato, 1999; Wolpert et al., 1995, 1998; Ebner, 2008) and Ito's (2006, 2008) internal models for motor control and mental representations in the cerebral cortex. According to this concept of internal model (i.e., a neural representation of the external world), the cerebellum forms (through a learning process) two different types of internal model: a forward model, which predicts the sensory consequences of motor commands; and an inverse model, which plans motor commands to achieve or maintain a desired goal (Kawato, 1999). The forward model needs a copy of the motor commands, termed an efference copy, and sensory feedback from afferent signals. These afferent signals give information about the motor apparatus' current state (Shadmehr & Krakauer, 2008; Ishikawa et al., 2016). It is assumed that when a movement is initiated, the neural activities of efference copy inputs arise shortly after those of the primary motor cortex neurons (Ishikawa et al., 2016). Forward and inverse models can be seen as inversion of each other in terms of information flow (Ishikawa et al., 2016).

Taking into consideration past theories about the cerebellum's function and the growing body of evidence about its involvement in cognition, Sokolov, Miall, and

Ivry (2017) have proposed a model that tries to explain how the cerebellum uses its most prominent motor-control computational features (i.e., prediction- and error-based learning) to instantiate or facilitate cognitive behaviors (see Figure 2.3). This model applies the forward model of motor commands that predicts the consequences of an action to a forward model that predicts the changes in perceptual or mental states. Both are error-based learning-driven models.
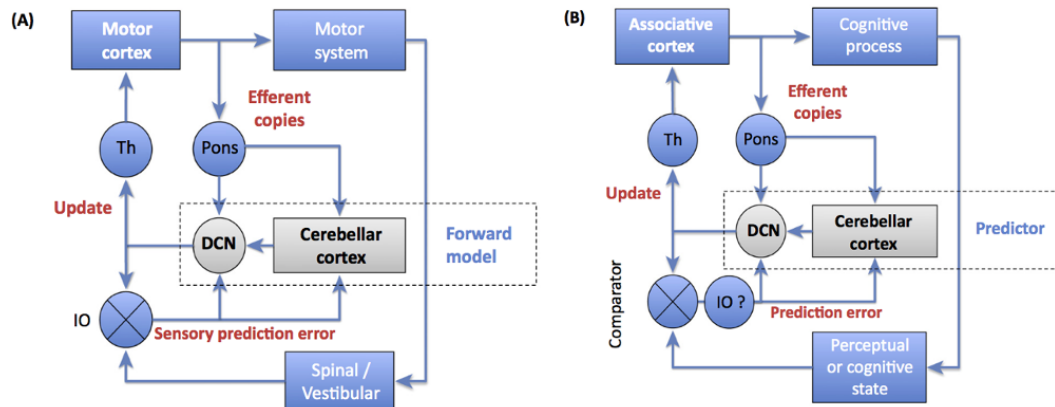
Figure 2.3: Schematic of hypothesized cerebellar mechanisms for predictions of motor movements and cognitive processes. (A) Motor cortex sends motor commands down the corticospinal track. Efferent copies of those motor commands are sent down the cortico-ponto-cerebellar track to predict the sensory consequences of the motor commands. The sensory differences between the actual motor commands and the expected sensory outcomes from the motor commands are the sensory prediction errors. These predictions are believed to reach the cerebellum via the climbing fibers of the inferior olive (IO). (B) The cognitive cerebellum is thought to rely on similar mechanisms and thought to predict changes in perceptual or mental states. Note: Th = thalamus. (From "The cerebellum: adaptive prediction for movement and cognition," by A. A. Sokolov, R. C. Miall, and R. B. Ivry, 2017, Trends in cognitive sciences, 21(5), p. 319. Copyright 2021 by Elsevier. Reproduced with permission.)

In parallel, Schmahmann (1991, 2019) has proposed a set of theories to unify most of the other theories into a general framework: the universal cerebellar transform (UCT) and the dysmetria of thought (DoT) theories. The UCT theory posits that the homogenous cytoarchitecture of the cerebellar cortex provides the anatomical and physiological substrate for consistent cerebellar computations across sensorimotor and associative processes (Schmahmann, 1991). According to this theory, the cerebellum automatically maintains behavior around a homeostatic baseline, without conscious awareness, by integrating multiple internal representations with external stimuli and appropriate responses. This transform or computation is informed by implicit learning and acts as an oscillation dampener that seeks to optimize performance according to context (Schmahmann, 1991, 2019).

The DoT theory, which is based on the UCT, posits that, depending on the location of a cerebellar lesion, the resulting behaviors will either lead to gait ataxia, limb dysmetria, or dysarthric speech if the lesion occurs in a cerebellar motor region, or lead to what is

known as the cerebellar cognitive affective syndrome (CCAS) if the lesion occurs in a cognitive limbic region of the cerebellum (Schmahmann, 1991, 2004, 2019; Schmahmann & Sherman, 1998). The CCAS comprises a cluster of cognitive and affective symptoms that can be categorized into four deficit types: executive function impairments, such as in the case of planning, set-shifting, abstract reasoning, and working memory; spatial cognition deficits, such as in the case of visuospatial organization and memory; behavioral-affective changes, such as blunting of affect or disinhibited and inappropriate behavior; and language impairments, such as agrammatism, dysprosodia, and anomia (Schmahmann & Sherman, 1998; Marien & Manto, 2017; Hoche et al., 2018). Thus, according to the DoT theory, the symptoms of cerebellar dysfunctions seen in motor behaviors (because of a compromised cerebellar motor region) have analogs in cognitive behaviors (if a cerebellar cognitive limbic region is compromised) due to the "universal" nature of the cerebellar computation (see Figure 2.4).
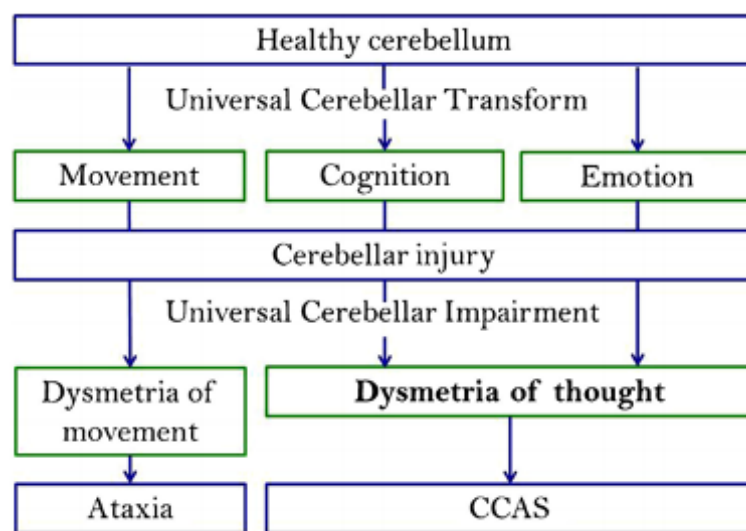
Figure 2.4: Schematic of the dysmetria of thought theory. (From "Metalinguistic deficits in patients with cerebellar dysfunction: empirical support for the dysmetria of thought theory," by X. Guell, F. Hoche, and J. D. Schmahmann, 2015, The Cerebellum, 14(1), p. 51. Copyright 2021 by Springer Nature. Reproduced with permission.)

2.5 The Cerebellum and Language

In the course of the brain's phylogenetic evolution, its enlargement was accompanied by a concomitant enlargement of the newest parts of the cerebellum (e.g., the neocerebellum) and the association areas in the cerebral cortex (Leiner et al, 1991). This led to new neural connections that descend (via enlarged structures in the brainstem) from the cerebral cortex to the new areas of the lateral cerebellum (Leiner et al., 1991). Given that the cerebellum also projects to the neocortex via the thalamus, it has been shown that the connections between the cerebellum and the cerebral cortex are arranged in closed loops (Kelly & Strick, 2003; Sokolov, Miall, & Ivry, 2017; see Figure 2.5).
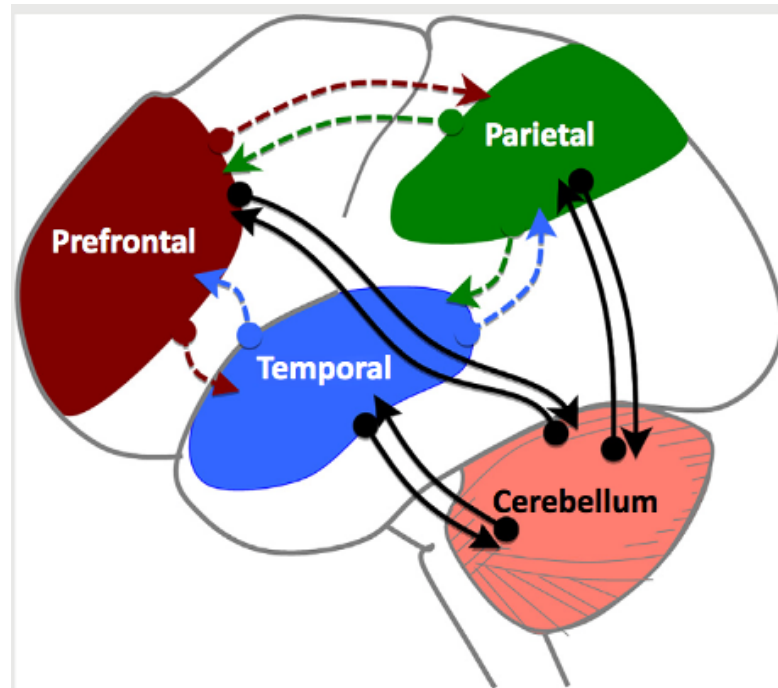
Figure 2.5: Cerebellar indirect modulation of cognitive cortical networks. The cerebellum has closed-loop afferent and efferent connections (black arrows) to specific cortical areas involved in the cognitive networks. Complex cognitive processes involves communication between regions in cerebral cortex (colored dashed arrows). By acting uniquely on these single cortical brain areas of the cognitive network which also interact with each other, the cerebellum is believed to also have an indirect influence on the cognitive network. (From "The cerebellum: adaptive prediction for movement and cognition," by A. A. Sokolov, R. C. Miall, and R. B. Ivry, 2017, Trends in cognitive sciences, 21(5), p. 315. Copyright 2021 by Elsevier. Reproduced with permission.)

Some of those closed loops have motor functions, such as the loops connecting regions of the vermis and areas of the cerebellar hemispheres (i.e., lobules IV, V, VI, VII, and VIII) with the primary motor cortex. Others do not, such as the indirect cerebellar projections that connect parts of the cerebellum with the posterior parietal cortex (a cortical region known to be active during nearly all cognitive tasks; Broussard, 2014; Kelly & Strick, 2003). The cerebrocerebellar loops linking the dentate nucleus (in the deep cerebellar nuclei) with Walker's area 46 in monkeys (the monkey analog to the human dorsolateral prefrontal cortex) are also closed loops that do not have a motor function (Broussard, 2014; Kelly & Strick, 2003).

The dorsolateral prefrontal cortex is a region known for its importance in higher cognitive functions. It is believed that the closed loops between the neocerebellum and the frontal lobe are not limited to frontal motor areas, but include other areas of the frontal cortex such as Broca's area (Leiner et al., 1991; Murdoch, 2010). Taken together, these anatomical findings suggest not only that the cerebellum has the necessary anatomical connections to be involved in cognitive functions, but also that these series of distinct neural loops formed by reciprocal connectivity may enable the cerebellum to facilitate cognitive and linguistic function in the same way it enhances motor functions (Leiner et al., 1991; Murdoch, 2010).

While there is a great deal of evidence showing that by having to coordinate the muscles of the larynx, throat, tongue, and lips, the cerebellum contributes to the motor processes that produce fluent speech in humans, its involvement in the cognitive processes of "thinking of the word to be expressed" is not yet as established. Nonetheless, recent advancements in neuroimaging techniques and imaging, lesion, clinical, and behavioral studies have increased the amount of converging evidence and the understanding of the role of the cerebellum in non-motor linguistic processing (Marien & Manto, 2018).

For instance, a neuroimaging meta-analysis by Stoodley, Valera, and Schmahmann (2011) showed that verb generation engaged the right cerebellar lobules VI–Crus I and a second cluster in lobules VIIB–VIIIA (see Figure 2.1 for anatomy of cerebellum). The verb generation task required participants to internally (or covertly) generate a verb from a given noun (e.g., "beer" $->$ "drink"). Another neuroimaging meta-analysis by Keren-Happuch et al. (2014) revealed that cerebellar peak activation coordinates were found bilaterally in lobules VI, midline lobule VIII, left Crus 1, and right Crus 2 in expressive and receptive language tasks. They concluded that their results added to the increasing body of research describing lateralized activation in the posterior regions of the cerebellum for language tasks.

It is worth noting that these regions are often active during higher order functions (Price, 2012). In a study on verbal working memory using fMRI-guided TMS, Desmond et al. (2005) tested whether disruption of the right superior cerebellum (hemispheric lobule VI/Crus I) impaired verbal working memory. They administered a single-pulse TMS to participants immediately after the simultaneous presentation of six letters during the encoding phase of some of the trials. The encoding phase was followed by a maintenance period and then by a retrieval period in which four probe letters were presented, only one of which had already appeared during the encoding phase. Participants were asked to press one of four buttons (oriented in a single row) to indicate which probe letter had appeared in the initial encoding period. The authors found that TMS on the right superior cerebellum did not increase accuracy; however, it did significantly improve reaction time for the trials that were accurate. They concluded that

the observed effects indicated cerebellar involvement in verbal working memory.

Like Moberget and Ivry (2016), one might ask, then, to what extent these findings can be explained by a generalization of the internal model hypothesis to cognition. It is possible that in the context of the cognitive aspect of language, cerebellar internal models contribute to the timing of language processing, the adaptability of linguistic representation (e.g., semantic, syntactic, lexical), and the predictive capabilities during language processing (e.g., inferring the next linguistic states; Moberget & Ivry, 2016).

Several studies seem to show support for these themes. For instance, it has been shown that patients with cerebellar disorder are impaired in speech perception tasks when they need to use temporal cues as opposed to spectral ones (Ackermann et al., 1997, 2004). In an fMRI study, Guediche et al. (2015) used distorted speech in a word-recognition task to investigate whether adaption in speech perception also involves the cerebellum. They found co-activation of the right Crus I of the cerebellum and other cerebral cortical regions during the adaptive changes in speech perception. The results were consistent with known cerebellar contributions to sensorimotor adaptation (some of which will be discussed in a later section).

Using a visual world paradigm and repetitive TMS, Lesage et al. (2012) demonstrated that demonstrated that stimulating participants' right cerebellum (a region implicated in language) disrupted the their predictive language processing, but only when the task involved making a prediction. Participants' performance in the task remained unaffected, despite the stimulations, when no prediction was involved. This provided some evidence that the computational operations used by the cerebellum for making predictions during motor control might also be in use during language processing.

In proposing an alternative to the internal model hypothesis, Guell et al. (2015) and Schmahmann (2019) argued for an account of cerebellar cognitive and linguistic function through the lenses of the DoT theory. In a study of linguistic capabilities in patients with cerebellar damage (Guell et al., 2015), patients were given the Oral Sentence Production Test to assess their basic semantic and syntactical abilities. No difference in scores was found between the patients and healthy control participants. The patients were also given the Test of Language Competence (Expanded Edition) to assess their metalinguistic abilities. In all its subtests, the patients with cerebellar damage scored significantly lower than healthy control participants.

From these results, Guell et al. 2015 concluded that patients with cerebellar damage exhibited metalinguistic deficits that are more specifically linked to social communication (e.g., metaphor, ambiguity, inference), despite retaining most of their grammar and semantic abilities. They postulated that these types of cognitive deficits were in line with the DoT, as they were akin to motor deficits. Their explanation was that, in the case of movement, cerebellar injury causes dysmetria but not weakness (i.e., disruption of modulation of movement but not disruption of generation of movement); and in the case of language, it causes metalinguistic deficits but not aphasia (i.e., disruption of modulation of language but not disruption of generation of language; Guell et al., 2015; Schmahmann, 2019). It is important to emphasize that the DoT theory does not go against the internal model hypothesis, but integrates it within its framework.

2.6   The Cerebellum and Music

As with language, making music (e.g composing, performing) is a complex cognitive function. The process of creating music using one's voice or an instrument requires many cognitive skills, such as executive function, attention, working memory, sequence learning, auditory discrimination, auditory memory, mental imagery, perception of speed, and knowledge of temporal order. It also involves motor skill learning, eye–hand coordination, and, for musical performance, the ability to synchronize. Thus, it is very likely that the process of making music also involves parts of the cerebellum.

Hutchinson et al., (2003) have found that the volume of the cerebellum is positively correlated with lifelong intensity of music practice and that musicians have larger cerebellar volumes than non-musicians. In addition, Abdul-Kareem et al. (2011) found that musicians had greater volumes of right cerebellar white matter and right superior and middle cerebellar peduncles (the nerve tracts permitting communication between the cerebellum and the other parts of the central nervous system) than non-musicians. Interestingly, Baer et al. (2015) showed that cerebellar regional volumes in early-trained musicians were smaller than those in late-trained musicians, and no difference in cerebellar regional volumes was found between late-trained musicians and non-musicians. They acknowledged that their findings might go against the more typical view, which associates expertise with greater volumes of brain regions, but they also noted that in other fields, expertise has been associated with reduced cerebellar volumes (see Nigmatullina et al., 2015). Even though correlation does not imply causation, it is worth noting these changes.

Using positron emission tomography (PET), Brown et al. (2004) found that part of the posterior cerebellum (the quadrangular lobule VI) played a role in singing repetitions of novel melodies, singing harmonization with novel melodies, and vocalizing monotonically.

Moreover, in a series of four neuroimaging and neurological studies involving professional musicians and non-musicians, Parsons (2001) provided support for a non-motor role of the cerebellum in a variety of sensory and cognitive musical tasks. The first study had pianists play music from memory and then play scales while in a PET scanner in order to localize brain areas specifically supporting music. Different regions of the cerebellum were activated when playing music compared to when playing scales. The second study found bilateral activation of the cerebellum (primarily lateral hemispheres) when musicians sight-read a score while listening for specific melodic, harmonic, or rhythmic errors in its performance. These results parallel that of another study in the language domain showing implication of the cerebellum in error detection (Fiez et al., 1992).

The third study involved musicians and non-musicians discriminating pairs of rhythms with respect to pattern, tempo, meter, or duration. There were significant differences between the two groups in terms of the strength of activations and locations, but all participants exhibited cerebellar activations. It was noted that these activations were probably unrelated to motor activity, as there was no overt motor activity and, little, to no activity in neocortical motor areas. In the last study, patients suffering from cerebellar degeneration exhibited impaired pitch discrimination (Parsons, 2001). These last results echo other studies demonstrating impaired musical ability in patients with cerebellar disorders (Tölgyesi & Evers, 2014) or impaired pitch discrimination ability

after downregulating cerebellar TMS (Lega et al., 2016). Similar to the language studies discussed above, these findings not only add to a growing body of literature providing supporting evidence for the role of the cerebellum in cognitive functions, but they also add to the music cognition literature.

It may also be asked to what extent a generalization of the internal model hypothesis can explain aspects of music processing in the brain. To answer that question, it may be useful to recapitulate what was discussed in the case of language processing. The previous section presented evidence supporting the involvement of the cerebellum in some cognitive aspects of language processing. It was also shown that these could potentially be explained by an internal model hypothesis account. In addition, a meta-analysis by Keren-Happuch et al. (2014), which looked at PET and fMRI studies of cerebellar contribution to higher cognition, revealed that some cerebellar cortical regions used during language processing overlap with some of those used during music processing (including the right cerebellar hemisphere; see Figure 2.6). Lastly, there is a large body of evidence supporting the use of internal models in motor behaviors such as speech production (Houde & Nagarajan, 2011; Hickok, 2012; Parrell et al., 2019a; Parrell et al., 2019b).
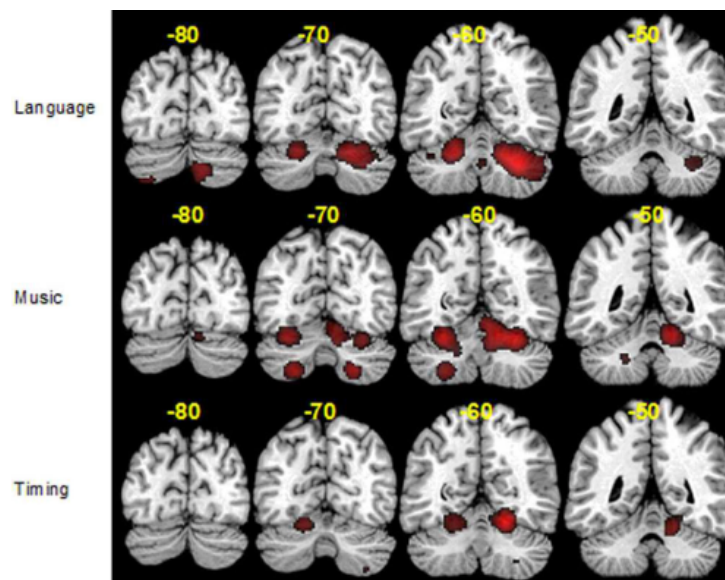
Figure 2.6: Segment of activation likelihood estimation maps for language, music, and timing at four different slices (y = -80, -70, -60, -50) mapped onto coronal sections of the Colin27 brain. (Original figure also showed activation for emotion, executive function, and working memory.) (From "A meta-analysis of cerebellar contributions to higher cognition from PET and fMRI studies," by E. Keren-Happuch, S. H. A. Chen, M. H. R. Ho, and J. E. Desmond, 2014, Human brain mapping, 35(2), p. 598. Copyright 2021 by John Wiley and Sons. Reproduced with permission.)

Singing is a musical activity that has similar motor implementations to and shares multiple cognitive aspects with speech production. It involves precise control of the vocal apparatus, an accurate control of the fundamental frequency of the voice (Jones & Keough, 2008), coordination between cognitive planning and motor execution (Pruitt & Pfordresher, 2015), and processing of the auditory feedback that helps the precise

control of the fundamental frequency (Pruitt & Pfordresher, 2015). Thus, it is reasonable to hypothesize that some aspects of music processing may also be explained by a generalization of the internal model hypothesis.

One place to start investigating this hypothesis could be in the domain of sensorimotor adaptation. In a frequency-altered feedback study comparing trained singers and non-singers, in which participants sang the syllable "ta" while hearing a frequency-altered version of their singing (shifted down by 100 cents), Jones and Keough (2008) showed that trained singers compensated for the difference in pitch to a lesser degree than non-singers. However, prolonged exposure to the frequency-altered feedback led the trained singers to compensate as much as the non-singers. In addition, they demonstrated that when participants subsequently received their unaltered feedback, the singers' fundamental frequency remained higher than their pre-test baseline fundamental frequency. The authors concluded that trained singers might rely more on internal models to maintain accurate vocal control as opposed to using a real-time auditory feedback.

Studies using a similar frequency-altered feedback paradigm have indicated that trained singers can voluntarily ignore altered feedback to produce a pitch that remains identical to their fundamental frequency (by presumably using an internal model) or use it to produce the target pitch as instructed (Keough & Jones, 2009; Zarate & Zatorre, 2005, 2008). These results were followed by another similar study that showed that singers' internal models seem to be more sensitive to subtle discrepancies in auditory feedback compared to those of non-singers (Keough & Jones, 2009). While these studies provide supporting evidence for the use of internal models in a musical activity, it is still unclear whether these internal models participate in a cognitive aspect of music processing (as opposed to motor control) that makes use of the cerebellum.

Other researchers have examined singing and the use of internal models in sensorimotor prediction. One hypothesis about poor singing, or more specifically poor pitch matching (defined as the inability to match a target pitch within one semitone), is the inverse model deficit hypothesis (Pfordresher & Mantell, 2014). According to this hypothesis, poor vocal imitation of pitch is the result of an inefficient auditory–vocal system representation that leads to difficulties instantiating a motor plan to reach a perceptual goal (Pfordresher & Brown, 2007; Pfordresher & Mantell, 2014).

Testing a similar hypothesis (i.e poor pitch matching being the result of poor motor control), as well as others hypotheses, Hutchins and Peretz (2012) devised a series of studies using a specifically designed device called a slider. The slider creates a vocal tone that mimics the human voice on a continuous spectrum. When pressed at a specific position on its touch-sensitive strip, the slider plays a tone that increases or decreases in pitch as the finger is slid up or down the strip. This allowed the participants (musicians and non-musicians) to use a different motor mechanism to reproduce a pitch. In addition, they had an experimental condition that made use of a synthesized voice. The main findings from this series of studies were that the singers and non-singers were better at matching pitches when using the slider as opposed to their own vocal apparatus; and when using their vocal apparatus, the non-singers were better at matching recordings of their own voices as opposed to recordings of a synthesized one at the same pitch.

Taking these findings together, the authors concluded that these patterns of results suggested that poor singing can mostly be attributed to poor motor control and timbral-translation error (Hutchins & Peretz, 2012). While the first cause does not provide any new element that would enable researchers to determine if the cerebellum is involved in

music processing beyond motor control, the second one leaves open the possibility of a generalization of the internal model hypothesis for music processing. This possibility, however, would need to be empirically tested.

The paucity of data about cerebellar disorder studies that involve pitch production in the context of music makes it difficult to assess the plausibility of the DoT theory within the context discussed in this section.

2.7   Conclusion

The main aims of this chapter were to provide a foundational understanding of the cerebellum structure, a foundational understanding of its well-established known functions, and to present more recent theories regarding its contribution to non-motor domains. The chapter has provided evidence that the cerebellum cannot be seen solely as a motor-control "machine", and a more holistic and realistic view of its role (which includes some aspects of cognition) should be taken into consideration to fully understand the extent of its functional domains. In addition, we particularly addressed the role of the cerebellum in vocalizations such as in speech and in music. The work of Lega et al., (2016) underlined the involvement of the cerebellum in pitch discrimination, which has some impact on the accuracy of pitch production (Pruitt & Pfordresher, 2015). The work of Brown et al.(2004) provided evidence of the involvement of the cerebellum in vocalizing singing tones. Jones and Keough (2008) have provided some evidence that expert singers might rely on internal models in order to keep the accuracy of their vocal productions when receiving inaccurate feedback of their own voice. Our hypothesis was that if the cerebellum is involved in the instantiation of internal models during singing, perturbing it using TMS, would lead healthy participants to do worse at the singing task because of the resulting compromised internal model. So far we have seen evidence that motor and cognitive aspects of singing involve the cerebellum, and that some aspect of singing may rely on internal models. In the following chapter we will see how those internal models can be seen through the framework of Bayesian cognitive models.

Chapter 3

The Bayesian Brain

3.1 Introduction

The human brain does not have direct access to the external environment. As a result, it relies on senses to rebuild the world internally. However, the sensory inputs received and the causes eliciting them do not have a one-to-one relation. A sensory input can have many external causes, and different sensory inputs can have the same external cause. Given this challenge, how can the brain reliably determine the sources of its sensory input? One theory was that the brain constructs percepts from the bottom up (Gibson & Carmichael, 1966; Marr, 1982). According to this theory, the brain's representation of the environment starts from a two-dimensional array (i.e., the retina), from which features of a scene are extracted (e.g., edges, shapes, texture), resulting in a model in which a scene is visualized in a continuous 3D map. Opposing that theory was the top-down theory of perception (Gregory, 1980), which posits that higher cognitive information acquired from experience or prior knowledge shapes how the brain perceives.

In recent years, however, an accumulation of compelling evidence in theoretical neuroscience has led to a paradigm shift. The brain is no longer seen as a passive sensory-information recipient, but as an active processor of sensory information that tries to make reliable causal inferences about the origins of its sensations. Two approaches that are used to investigate the problem of perception in terms of causal inference are Bayesian inference and predictive coding (Friston, 2010; Howhy, 2013; Clark, 2013).

This chapter will introduce core concepts in Bayesian inference. It will then consider the supporting evidence for the so-called "Bayesian brain" i.e. the Bayesian model of cognition. Finally, it will examine how predictive coding compares to Bayesian inference.

3.2 Beliefs as Probability Distributions

In contrast to frequentist statistics, which focuses on obtaining point estimates (e.g., mean, standard deviation or variance) of the observed data, Bayesian statistics assigns probabilities to the possible causes of the observed data, and thus so does Bayesian inference. Despite there being only a unique value for the true state of the environment (or one true cause of an individual's observations), the Bayesian observer's model represents the causes as probability density functions.
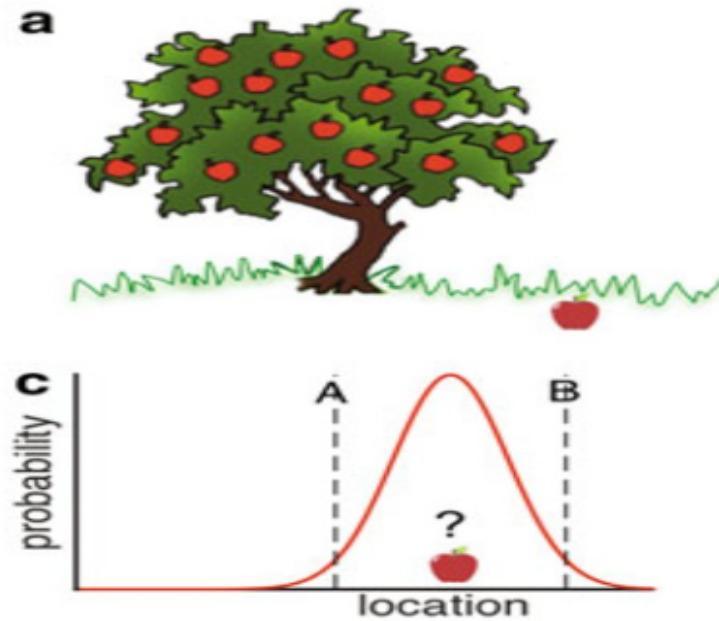
Figure 3.1: (a) An apple that has fallen from a tree but whose whereabout is unknown, (c) beliefs about the location of the apple represented as a probability distribution with the distribution's central tendency representing the most likely location (From "Bayesian Models in Cognitive Neuroscience: A Tutorial," by J. O'Reilly and R. Mars, 2015, An introduction to model-based cognitive neuroscience, p. 182. Copyright 2021 by Springer Nature. Reproduced with permission.)

Consider the example used in O'Reilly et al. (2012) in which an observer tries to estimate the width of a pen by looking at it. The observer could estimate the width either by using a single number (e.g., "This pen is 6 mm wide") or by using a distribution of possible values with their degree of belief about each value (e.g., "There is a 95% chance that this pen is between 4.5 mm and 7 mm wide"). This probability density function (PDF) represents the observer's uncertainty about a quantity for which there is a single true value that is unknown (the true width of the pen).

This point can be further illustrated with the scenario used in O'Reilly and Mars (2015): "Isaac Newton is foraging for apples in his garden when he sees an apple fall from a tree into long grass". Similarly to the previous example, Newton could estimate the location of the fallen apple using a single set of location coordinates, or he could represent its location with a PDF of the possible coordinates with his degree of belief (or uncertainty) for each location (see Figure 3.1).

In both examples, the representation of the uncertainty in the PDF is in direct relation to the level of noise in the estimation process. The noise may be intrinsic to the observer, as in the case of stochasticity in neural activity (Knill & Pouget, 2004; Ma et al., 2006; Ma et al., 2008) or the central nervous system (Faisal et al., 2008); or it may be extrinsic to the system, as when an observer tries to visually estimate the width of a pen while someone is writing with it or the location of a fallen apple at dusk. Less uncertainty (i.e., less noise) is represented by a narrower PDF, and more uncertainty (i.e., more noise) is represented by a wider PDF.

In more general statistical terms, all unknown estimated parameters can incorporate (un)certainty that can be defined by a probability distribution, and each parameter is believed to have its own probability distribution that captures the (un)certainty about its value (Van de Schoot & Depaoli, 2014). Thus, one of the key features of a Bayesian system is the ability to take a quantity that has an unknown true value and represent it as a probabilistic distribution (O'Reilly et al., 2012).

3.3   Integration of Information Using Uncertainty

Another key feature of the Bayesian approach resides in how a Bayesian system uses the level of uncertainty from its different sources of information. In a PDF that represents a belief about the state of the world (e.g. width of a pen, location of an apple), the variance of the PDF (i.e., its width) represents the degree of confidence about the state of the world (i.e. the wider the variance, the less certain one is about that world's state, and vice versa). Bayesian systems use this degree of confidence (or uncertainty) to weight different sources of information relative to their level of precision (Berniker & Kording, 2011; O'Reilly & Mars, 2015).

Continuing with the apple example from O'Reilly and Mars (2015), imagine that Newton not only saw the apple falling from the tree but also heard it. He now has two sources of information that can be represented as PDFs and which reflect his degree of belief about the location of the apple. To obtain the best estimate of where the apple fell in the long grass, Newton (as a Bayesian system) could choose either to rely on (or favor) one of his two senses or to combine the location estimates from both senses and use their levels of precision as weights (precision weighting). The Bayesian approach would favor the latter (i.e., considering both measurements at once weighted by their relative precisions). The measurement (or observations) with the lowest variance would be given more weight. For instance, if Newton's vision was not clear because of the time of day (e.g., dusk) or because of the weather (e.g., fog) and yielded an unreliable estimate of the apple's location, then the auditory evidence should be given more weight; inversely, if Newton's hearing was impaired or was not as sensitive because of some loud ambient noise, then vision should be given more weight.

Thus, in this example, a Bayesian system would optimally combine both sources of information by considering their levels of precision and estimating a (posterior) PDF that represents the location of the apple given vision and hearing measurements or PDFs (see Figure 3.2). The highest point of the PDF would be where the two distributions overlap, and its mode would be closer to the mode of the distribution with the highest level of precision (i.e., lowest variance).

3.4   Bayes' Theorem

The previous example demonstrated that a Bayesian approach can combine independent and parallel sources of information (e.g., vision, hearing). Prior beliefs (also known as priors in the Bayesian framework) can also be represented by PDFs and therefore they possess same mathematical properties as if they were other sensory modalities. A prior provides information that can be weighted and combined with current observations (or sensory evidence), depending on how reliable it is relative to the current sensory evidence.
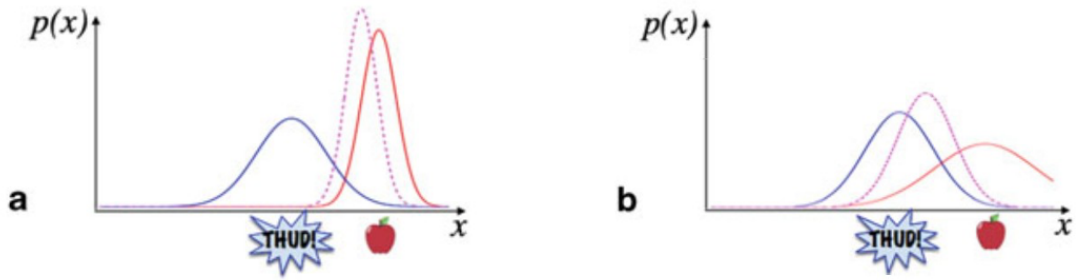
Figure 3.2: Multisensory integration. The red probability distribution represent beliefs about the fallen apple location based on visual information. The blue probability distribution represents beliefs about the fallen apple location based on auditory location. The purple distribution represents the beliefs about the fallen apple location based on the combination of both visual and auditory information. (From "Bayesian Models in Cognitive Neuroscience: A Tutorial," by J. O'Reilly and R. Mars, 2015, An introduction to model-based cognitive neuroscience, p. 183. Copyright 2021 by Springer Nature. Reproduced with permission.)

The result of this combination will be a posterior distribution that takes current observations and the prior into account. This combination is embodied in a theorem called Bayes' theorem. In its general form, Bayes' theorem can be expressed as follows:

$$P(h|d) = \frac{P(d|h)P(h)}{P(d)} \tag{3.1}$$

where: $P(h|d)$ is the probability of the hypothesis given the data, $P(d|h)$ is probability of the data given the hypothesis, $P(h)$ is the prior probability of the hypothesis and $P(d)$ is the probability of the data. In Bayesian terminology $P(h|d)$ is the posterior probability (and represents the posterior distribution or PDF); $P(d|h)$ is the likelihood (and represents the likelihood distribution or PDF); and $P(d)$ is the marginal probability, which can be obtained by summing over the hypotheses (marginalization process). The marginal probability can also be rewritten:

$$P(h|d) = \sum_{h' \in H} P(d|h')P(h') \tag{3.2}$$

where H is the hypothesis space (i.e. the set of all hypotheses considered). The summation of all the elements in the denominator ensures that the resulting probabilities are normalized to sum to one (Griffiths & Yuille, 2006).

Bayes' rule can also be written:

$$P(h|d) \propto P(d|h)P(h)P(d) \tag{3.3}$$

This formulation, which is derived from Equation (3.1) makes it clear that the posterior is directly proportional to the product of the prior and the likelihood.

This is a model of how prior beliefs should be updated when taking new evidence into consideration.In Bayesian inference models of cognition, the prior, the evidence and the posterior (or hypothesis) form, respectively, one's current knowledge, one's

new pieces of information and one's conclusion based on the combination of first two elements (Griffiths et al. 2008).

3.5   Bayesian Optimality

The optimization problem in Bayes' rule involves maximizing the posterior PDF P(h/d) through maximum likelihood estimation and determining the maximum a posteriori probability estimate. Maximum likelihood estimation maximizes the probability of collecting this type of data given the parameters to be estimated (e.g., the state of the world). In the absence of prior knowledge, maximizing the likelihood PDF is equivalent to maximizing the posterior PDF. However, given an informative prior PDF (i.e., some knowledge about the parameters or the state of the word), then Bayes' rule should be applied to estimate the posterior. The most probable estimate is found at the peak of the posterior distribution (i.e., the maximum a posteriori; Wolpert, 2007).

Newton's apple example can now be formalized. The optimal integration of two Gaussian priors with some degree of uncertainty due to measurement noise (i.e., the variance) is given by the weighted average:

$$\hat{S} = \hat{S}_1 \frac{r_1}{r_1 + r_2} + \hat{S}_2 \frac{r_2}{r_1 + r_2} \qquad (3.4)$$

where r_i is the reciprocal of the variance, and S_1 and S_2 are the means of the Gaussians. Thus, the posterior mean is the sum of the priors' means, each weighted by their additive relative precision (Penny, 2015).

This relation can be seen in Figure 3.2a and Figure 3.2b. In the first scenario, the mean of the posterior distribution (magenta dotted line) is closer to the mean of the red distribution, because the visual information was more precise. In the second scenario, the mean of the posterior distribution is closer to the mean of the blue distribution, because the auditory information was more precise. In both scenarios, the posterior distribution has a lower variance than both the distributions from which it derived. This is one of the direct consequences of Bayesian inference: As one increases the amount of information, one decreases the amount of uncertainty. Equation (3.4) shows how two independent sources of information are optimally combined using uncertainty.

While Bayes' formula helps to make optimal inferences about the state of the world in the face of uncertainty, it can also help to choose an action based on these inferences. Bayesian decision theory is one way to accomplish this feat, as its essence is to minimize expected loss (or maximize expected utility; Wolpert, 2007).

In this context, a loss function that quantifies the value of taking each possible action for each possible state θ of the world, $L(action, \theta)$, is added to specify the cost of estimating incorrectly:

$$\sum_{\theta} L(action, \theta) P(\theta|S) \qquad (3.5)$$

where σ denotes a summation over all possible states. The action can then be chosen that has the smallest expected loss (Wolpert, 2007).

Having reviewed the statistical underpinning of the Bayesian approach, this chapter now examines some of the evidence supporting the account that human observers behave as optimal Bayesian observers.

25

3.6 Behavioral Model

3.6.1 Multisensory Integration

Ernst and Banks (2002) conducted one of the seminal studies showing that humans might behave in a Bayes-optimal way. In their experiment, Ernst and Banks were interested in the problem of integrating two sources of information from different modalities: tactile (haptic) and visual. Participants were asked to judge which of two sequentially presented ridges was the tallest. The ridges were presented through a dot stereogram. To determine which ridge was the tallest, participants used vision alone, touch alone, or the two simultaneously. In addition, the researchers could vary the level of noise in the visual modality (i.e., the variance in the visual information) by manipulating the levels of noise in the dot stereogram. They observed stronger reliance on vision to make the estimations of the ridges' heights when the levels of visual noise were low, and stronger reliance on touch to make those same estimations when the levels of visual noise were high. Thus, the results of this experiment closely matched Bayesian behavior.

Other studies have shown the role of Bayesian inference in multisensory integration. Alais and Burr (2004) reported near-optimal combination of visual and auditory space cues in the context of object localization. They showed not only that, in this context, visual and auditory information are combined and weighted by their precisions, but also that this combination led to an improvement in bimodal spatial location over unimodal spatial location. The authors also mentioned that the visual stimuli in the experiment had to be severely blurred (making visual localization very poor) for audition to outweigh vision in perceptual localization. This might suggest that there is substantial sensory dominance of vision over audition in multisensory integration during spatial localization of audio-visual stimuli. However, they also noted that the location of the auditory stimuli was defined using only a single cue (i.e., interaural time difference). Thus, the sensory dominance of vision in their study might have been stronger than it would be in ecologically valid contexts.

Battaglia et al. (2003) found comparable results in a similarly designed spatial localization study. However, their study showed that vision tended to dominate much more than what one might expect from an optimal multimodal combination of information. It is worth noting, too, that the dominance of vision over audition in spatial localization is the underlying basis of the ventriloquist effect.

Precision-based sensorimotor integration has also been shown to occur between vision and proprioception in visuo-proprioceptive estimation tasks (van Beers et al., 1996; van Beers et al., 1999). Moreover, while proprioceptive information is given greater weight than one would expect, Sexton (2019) showed that, if a force field is applied to the hand doing the reaching during these tasks, the "system" interprets it as somatosensory unreliability and increases the weight given to vision in the estimation.

So far this chapter has presented cases where a Bayesian system integrates multiple sources of information to reach the most optimal estimation. It has yet to provide evidence of how, in this framework, new observations are incorporated into existing ones using prior beliefs to constrain the estimates given by each new observation. This is the topic of the next section.

3.6.2   Perception and Ambiguity: Use of Priors

Expectations play an important role in shaping the way people perceive the world. This idea has its origins in the work of Helmholtz, who viewed perception as "unconscious inference" (as cited in Howhy, 2013), and can be traced back to Kant (1908), who believed that "a priori" knowledge may contribute to one's "sensibility" (i.e., the faculty that provides sensory representations). One of the earliest demonstrations of how perception might be influenced by prior knowledge, and thus be inferential in this regard, can be seen in the case of binocular rivalry. In (conventional) binocular rivalry, incompatible images presented to the two eyes result in a perceptual alternation between the images, rather than in their combination (Denison et al., 2011). This suggests that perception is not solely driven by sensory stimuli, but also includes some assumptions or expectations.

In 1928, Diaz-Caneja showed that the perception of alternating coherent percepts, in binocular rivalry, could also occur even when the images presented to each eye were incompatible halves (Diaz-Caneja, 1928, translated by Alais et al., 2000). This again underlines the role of expectations in perception. These assumptions or expectations can be conceptualized as Bayesian priors (Kok et al., 2013). More recently, work on visual object perception using a Bayesian approach has shown that much of the ambiguity, weakness, or noise in the sensory inputs in visual processing can be resolved with prior knowledge (de Lange et al., 2018; Kersten, 2004).

Perceptual illusions are testimonies to the use of priors in perception. In the Müller-Lyer illusion (Figure 3.3a), two vertical lines of the same length, one with outward-pointing fins on the ends and the other with inward-pointing fins, are perceived to be of different lengths (the line with inward-pointing fins appears to be shorter). It has been suggested that this illusion is due to strong priors learned from depth perception and a lifelong exposure to Western architectural features (Gregory, 1997, 2005). The line with outward-pointing fins cues the visual system to perceive it as closer to the viewer, as if they were looking at the outside corner of a house, for instance; whereas the vertical line with inward-pointing fins cues the visual system to perceive it as farther away, as if the viewer were looking at the inside corner of a room (Gregory, 1997).

Another type of illusion driven by acquired priors is the Ames room illusion (Figure 3.3c). In this illusion, the viewer's powerful assumptions about parallel floors and ceilings, as well as orthogonal walls, make a trapezoidal room look cubic, and the people standing in it appear to be of different sizes when seen monovisually through a peephole. It is thought that these types of priors come from implicitly learning and exploiting statistical regularities in sensory inputs.

Moreover, it has been proposed that these priors can broadly be divided into two categories depending on the extent to which their effects on perception can be generalized across different environments: structural and contextual priors (Series & Seitz, 2013). Structural priors are those that reflect the lifelong learning of statistical regularities in the natural environment or those that are thought to be innate. They seem to apply to a broad range of sensory inputs. The Müller-Lyer illusion can be seen as the consequence of structural priors. Another type of illusion resulting from structural priors is the light-from-above illusion (Figure 3.3b), which makes the viewer assume the presence of an overhead light (Ramachandran, 1988; Morgenstern, 2011) and see objects as having concave or convex features depending on where the "shadow" appears in relation to these objects.
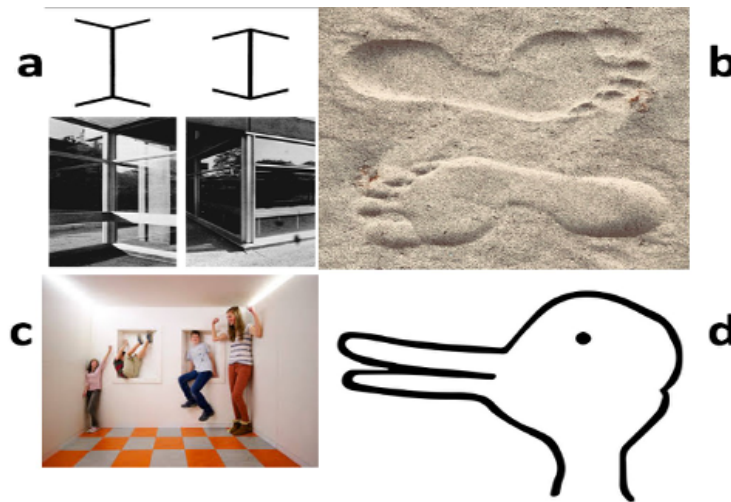
Figure 3.3: (a) Müller-Lyer illusion with potential causes for this illusion underneath; from Gregory (2005). (b) Light-from-above illusion: The bottom footprint seems to be convex, whereas the top one seems to be concave; these are identical footprints, with the bottom one turned upside-down. (c) Ames room illusion. (d) Rabbit–duck bistable figure. ((a) & (d) From "The Medawar Lecture 2001: Knowledge for Vision: Vision for Knowledge. Philosophical Transactions," by R. Gregory, 2015, Biological Sciences, 360(1458), p1239, p1244. Copyright 2021 by The Royal Society (U.K.). Reproduced with permission.)

Contextual priors, on the other hand, affect perception only in specific spatial or temporal circumstances (Series & Seitz, 2013). They can be implicitly or explicitly manipulated and rapidly induced through different means, ranging from sensory cues (Stein & Peelen, 2015; Pinto et al., 2015) to context exposure (Bar, 2004; Oliva & Torralba, 2007) and even instructions (Sterzer et al., 2008). For instance, using the inherently ambiguous duck–rabbit bistable figure (Figure 3.3d), Brugger and Brugger (1993) showed that people were more likely to recognize the figure as a rabbit on Easter Day, whereas on a random October day, the figure was more likely to be recognized as a bird.

While most of the examples above relate to vision, it is important to note that the use of priors to disambiguate or inform sensory stimuli is not restricted to vision. Aspects of speech recognition seem to depend on one's expectations about the probabilities that words will appear based on their frequency of occurrence in the language (Norris et al., 2016). Given that the ability to recognize words relies in part on the individual having learned about their regularities of occurrence, achieving this feat is an example of the use of structural priors in spoken language (Series & Seitz, 2013). An example of the use of contextual priors in spoken language is when an individual is in a noisy environment and is trying to work out the unheard spoken words of an interlocutor by considering the topic of the conversation (Series & Seitz, 2013).

In addition to disambiguation, the use of priors can even be seen, albeit more subtly, when the sensory inputs are unambiguous, clear, and strong. In this context, priors have been shown to have some role in the efficiency of perception (de Lange, 2018; Pinto et al., 2015). For instance, Stein and Peelen (2015) demonstrated that giving valid information about the category of an object led to improved performance of the visual

localization of that object, relative to having no information at all or being misinformed. Wyart et al. (2012) found that cueing participants about the source locations of sensory inputs increased their detection by suppressing internal noise, and giving participants information about the apparition probability of sensory inputs influenced the participants' baseline level of detection.

3.6.3   Sensorimotor Integration

So far, this study has discussed the integration of priors in the Bayesian framework for the estimation of perceptual inputs that were assumed to have static properties. However, the world is not static. It is dynamic and continually changing. Thus, its properties are also dynamic. Those changes in the state of the world (and of one's perception of it) are often brought about by the actions of people, which are themselves dynamic. These actions necessitate the estimation of parameters that evolve over time. For instance, the control of one's body requires an estimation of its current state (i.e., its configuration), which continuously changes as it moves, and must be updated as accurately as possible to continue generating appropriate motor commands (Wolpert, 2007; Franklin, 2011).

Despite having the ability to perform skillful actions, the body, and more precisely the sensorimotor control system, faces multiple challenges during this estimation task. For instance, given the large number of degrees of freedom in the body, there is an infinite number of ways to achieve the same task. How does the sensorimotor control system select the specific muscles and joints to coordinate, as well as select a path and force to achieve the goal? This is known as the redundancy problem (Franklin, 2011). There is intrinsic noise limiting one's ability to act precisely and adding variability to one's movements (Harris & Wolpert, 1998; van Beers et al., 2004). In addition, delays in the reception of afferent sensory information and the transmission of efferent commands (Merton & Morton, 1980) not only oblige the system to find a way to cope with out-of-date sensory information, but also make its control of the body difficult.

Moreover, combined noise and delays within the system as well as the ambiguity in sensory inputs from the world (Yuille & Kersten, 2006) can give rise to uncertainty about the current state of the body and make parameter estimation difficult. There are also changes in the body as it ages that modify properties of the motor system on multiple timescales (i.e., non-stationarity of the sensorimotor system problem; Lindle et al., 1997; Cantone et al., 2019). Lastly, the motor system initiates motor commands that undergo nonlinear transformation when converted into forces, trajectories, or movements (Franklin, 2011).

To compensate for sensorimotor delays and reduce the uncertainty in the system's time- varying state estimation due to these challenges, the system needs a model of the body that anticipates these dynamics. It has been proposed that the sensorimotor system uses inferential probabilistic models that combine an internal forward (predictive) model with the sensory feedback of a motor command to make an optimal estimate of the current state of the body (Wolpert, 2007; Wolpert & Flanagan, 2001, 2016).

In motor control theory, it is common practice to model optimal state estimation which relies on a forward predictive model and sensory feedback using a Kalman filter (Kalman, 1960; Todorov, 2004; Wolpert, 2007). This technique is considered an equivalent to the application of Bayesian integration for the state estimation of a time-varying system (Wolpert, 2007; Berniker & Kording, 2011).
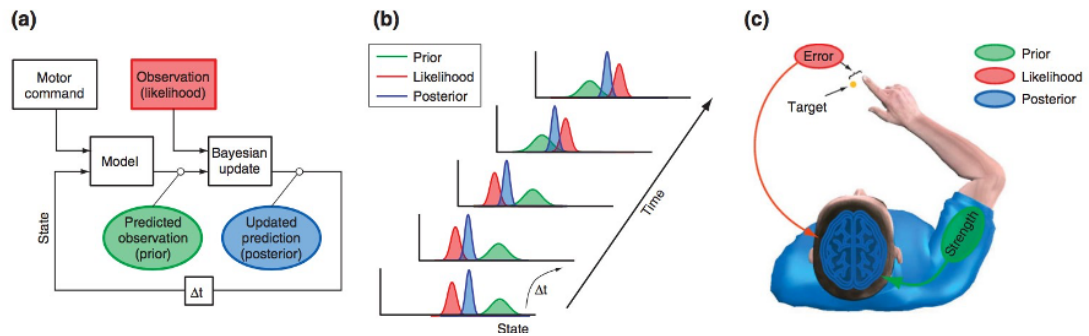
Figure 3.4: (a) World's state estimation process during a motor command according to modern control theory principles. Predictions about the current world's state (posterior in blue) are obtained by combining observations of the world's state (likelihood in red) with expectation of the world's state (prior in green) which is based on a model of the world and a motor command. Bayesian belief updating about the current state of the world (posterior in blue) is done with a Kalman filter. (b) Bayesian belief updating. Distributions representing the expected state of the world (prior in green), the observation of the world (likelihood in red) and the prediction about the current state of the world (posterior in blue). The Bayesian belief updating which combines the prior and likelihood to get the posterior is reimplemented at each time step with the posterior at one time step becoming the prior in the following time step. (c) Motor movements and adaptation framed as Bayesian belief updating. Individual's prior knowledge about their muscle properties (e.g. strength) is combined with their observed movement errors to reassess and create updated beliefs about their muscle properties. (From "Bayesian approaches to sensory integration for motor control," by M. Berniker, and K. Kording, 2011, Wiley interdisciplinary reviews. Cognitive science, 2(4), p. 423. Copyright 2021 by John Wiley and Sons. Reproduced with permission.)

Here, the predictive model, which reflects the distribution over the previous state estimates (given the best estimate of the current distribution), acts as a prior; and the sensory evidence, which reflects the most likely sensory feedback to be received given the body's state, acts as the likelihood. When combined, they form a posterior that represents the updated prediction of the body's current state in the world. That newly computed state estimate is then used in the model to define the future prior (Berniker & Kording, 2011; see Figure 3.4).

Several empirical studies support this model. For instance, McIntyre et al. (2001) showed that the brain uses its prior knowledge of gravity combined with sensory information when making motor predictions. Wolpert et al. (1995) developed a computational model that uses a Kalman filter to integrate a forward model and proprioceptive feedback to determine a hand's position. They demonstrated that it closely modeled the bias and variance in the data of participants, who were tasked with estimating one of their hands' location at the end of a movement in the dark and under external forces. The results of this state estimation study showed that proprioceptive information and forward model are likely to be integrated in a Bayes-optimal way. Other studies examining posture (Kuo, 1995) and head rotation (Merfeld et al., 1999) also support that model.

Having made the case that the way one perceives and interacts with the world is not

solely determined by one's sensory inputs, but that it is strongly influenced by one's expectations of them, this chapter now turns to the potential neural implementation of Bayesian inference.

3.7 From Bayesian Inference to Bayesian Predictive Coding

While Bayesian inference prescribes the optimal prediction a system can make given the data, it does not specify the underlying neural representation; it only describes the end-results of computations (i.e., behaviors). Predictive coding, on the other hand, aims to provide an account of the actual neural responses or computations (Aitchison & Lengyel, 2017). In its simplest form, predictive coding represents the prediction error rather than the raw inputs. That prediction error can either be the difference or the ratio between sensory inputs and predictions (Spratling, 2017):

$$Prediction error = input - prediction \qquad (3.6)$$

Efficiency is assumed to be a computational advantage of such representation, as messages are relayed only when there is a discrepancy between the incoming signal and its prediction (Mead, 1990). More advanced versions of predictive coding suggest that the brain creates a generative model, the goal of which is to extract and encode patterns or statistical regularities from the world that are believed to be possible causes of sensory inputs (Mumford, 1992; Friston, 2005). Once encoded, these statistical regularities are used as parameters of the generative model to make more accurate causal inferences (or hypotheses) about the origin of new sensory inputs (Rao & Ballard, 1999; Friston, 2005).

The process is often thought of as a processing hierarchy in which the hypotheses at the upper level constrain the generation of new hypotheses at the next level down, and so on until the lowest level is reached. At each level, the difference is calculated between the predictive top-down signals and the bottom-up sensory inputs from the lower level (i.e., the prediction error). The prediction error is then propagated back to the level above, where it is used to revise the initial hypothesis. If the prediction error cannot be minimized at the level at which it is being processed, it is relayed to the next level above. The higher in the system the prediction error is relayed to, the more substantial the revision of the hypothesis. This routine is repeated simultaneously throughout the hierarchy, and the content of perception arises as successive prediction error minimizations take place and a winning hypothesis is selected.

This model requires that at least two functionally distinct classes of neurons (or units) be present at each level: one that represents the hypothesis (or the prediction) and tries to best explain away the inputs received, and one that represents the prediction error and encodes for the difference between the hypothesis and the inputs (Kok & de Lange, 2015; Keller, 2018). When the hypothesis signal matches the information carried by the input signal (i.e., prediction error), the prediction error response decreases. These sensory input suppressions have the effect of reducing the firing rate. One advantage of such a coding scheme is that potentially relevant but unexpected stimuli are made more salient (Spratling, 2012). Another advantage of the prediction-error suppression is the

simultaneous enhancement of the hypothesis, which will lead to a potentially accurate expectation and representation of the next sensory input (Kok et al., 2012).

There are different proposals for the neural architecture underlying this predictive coding scheme (Spratling, 2017), such as Rao and Ballard's algorithm (Rao & Ballard, 1999), the predictive coding/bias competition divisive input modulation (PC/BC-DIM) model (Spratling, 2008), and the free energy principle (Friston, 2005). They differ primarily in how the error is computed (subtraction or division) and how the different classes of neurons are connected (Spratling, 2017; Aitchison & Lengyel, 2017).

In the classical predictive coding schemes (Mumford, 1992; Rao & Ballard, 1999; Friston, 2005), the "hypothesis" neurons have feedback projections and the prediction-error neurons have feedforward projections. Because feedforward connections mostly begin at layers II/III and feedback connections come from layers V/VI (Felleman & van Essen, 1991), it is expected that the hypothesis neurons reside in layers II/III and the prediction-error neurons reside in layers V/VI (Heilbron & Chait, 2018). On the other hand, in the PC/BC-DIM scheme (Spratling, 2008), the hypothesis neurons have reciprocal connections between regions, and the prediction-error neurons are processed intracortically (Spratling, 2008). This leads to a different location for hypothesis and prediction-error neurons compared to the classic predictive coding schemes (see Figure 3.5). Another consequence of this different neural architecture is that the calculation of the residual error is performed by connections within each cortical region, as opposed to between cortical regions (Spratling, 2010).

3.8   Empirical Evidence for Predictive Coding

Some findings from well-known classical studies can be understood in terms of, and provide evidence for, predictive coding. For instance, Rao and Ballard (1999) showed that the end- stopping (Hubel & Wiesel, 1965) classical visual phenomenon may be the result of predictions from higher-level representation inhibiting the surrounding area of extra-classical receptive fields (ECRF). The stimulus presented in these fields gives a clue about the content of the surrounding regions, which is accurately predicted by higher-order cortical regions and thus suppressed.

In line with this, cases of decreased neural response when a stimulus is predictable have also been shown in learned association (Egner et al., 2010) or after the presentation of the same stimulus (Ulanovsky et al., 2003). These could be explained by a common principle of sensory processing where the exposure or the presentation of sustained stimuli lead to an adaptation and a reduction of the neural responses over time (Keller et al. 2017). These examples are reminiscent of the prediction-error suppression scheme in predictive coding. In a study on mice, Keller et al. (2017) demonstrated that the reduction in neuronal responses due to the aforementioned principle could be reversed in contrast adaptation if the stimulus (re)acquired behavioral relevance. If stimulus predictability by top-down processes leads to error suppression, a violation of that predictability should lead to an increased prediction-error response.
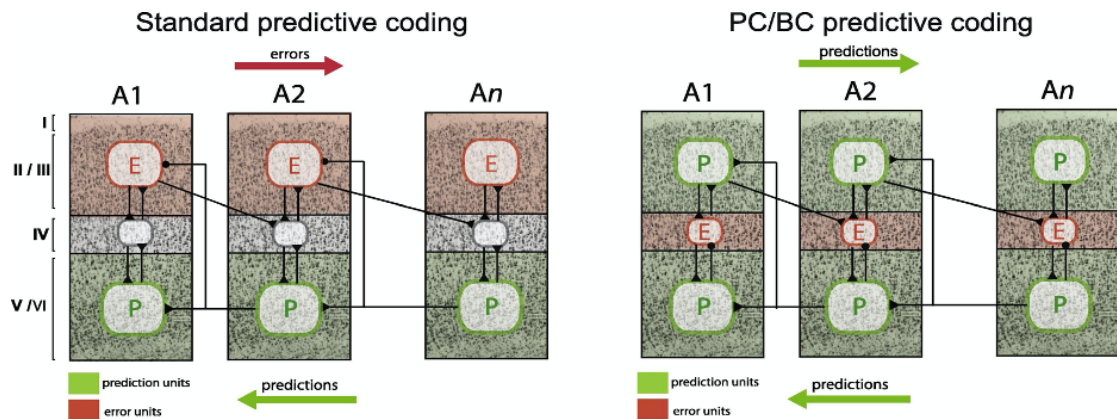
Figure 3.5: Error neurons and expectation neurons in the auditory cortex arranged according to two different formulations of predictive coding (PC) (i.e. Standard predictive coding, PC/BC predictive coding). Cortical columns of the primary (A1), secondary (A2), and higher order (An) auditory areas are represented. (Left) standard PC: errors and predictions flow in opposite directions (upward and downward respectively). Error units are in superficial layers (II/III) and predictions units are in deep layers (V/VI) with predictions units at higher auditory areas being able to suppress errors units at lower auditory areas via polysynaptic top-down inhibitory connections (black units in layer IV). (Right) Biased competition PC: predictions flow bidirectionally (upward and downward) with prediction unit being both in superficial and deep layers. Error is computed in layer IV and can be suppressed via inhibition by prediction units that are within the same cortical columns as the error unit itself. Black arrows represent excitatory top-down connections. (From "Great expectations: is there evidence for predictive coding in auditory cortex?," by M. Heilbron, and M. Chait, 2018, Neuroscience, 389, p. 55. CC BY-NC-ND.)

Studies on novel events or oddballs in a stream of repeated events (Squires et al., 1975; Lieder et al., 2013) provide evidence that the mechanisms underlying the mismatch negativity can be studied within the predictive-coding framework. This effect can be seen as the result of the inability of the brain to predict a stochastic event, which leads to a large prediction error every time the event is presented.

Similarly, Eliades and Wang (2008) showed that there is increased cortical activity in marmosets' auditory cortex when they received a frequency-shifted feedback of their self-generated vocalizations. They noted that the average neural activity for the population of neurons they had studied was strongly inhibited during normal vocal production. This suggests that self-generated vocalizations, which are generally suppressed in the auditory cortex, became unexpected and salient as soon as there was an unexpected pitch-shift in the vocalization.

3.9 Conclusion

This chapter started with the concept of Bayesian inference. It showed how our beliefs could be represented as probability distributions that take into account our degrees of uncertainty. We saw that those beliefs could be combined with new information (also represented as probability distributions) to form new beliefs. This forms the basis of the Bayesian optimal agent model. We also saw that this updating process which consists of taking some information (e.g. prior beliefs), and updating that information by combining it in a systematic fashion with new information could also be applied in the behavioral context and used to explain multisensory integration. Indeed, studies like the seminal study of Ernst and Banks (2002) showed that there is evidence that we integrate multimodal sensory information following Bayesian optimality. In addition to perception, this concept of Bayesian optimality could be used to explain sensorimotor integration i.e. coupling of the motor system and the sensory system to instantiate voluntary movements or actions, which seems to rely on a forward predictive model and sensory feedback (Wolpert, 2007). Last, after looking at evidence to provide higher-level explanation of behavioral mechanisms with a Bayesian model of cognition, we looked for evidence of lower level implementation of this Bayesian model of cognition in the brain. Combining these pieces of evidence with those we found about the cerebellum, we've seen that: the cerebellum is involved not only in motor behavior, but also in cognition. We've seen that an activity like singing seems to recruit the cerebellum on both of these aspects (i.e. behavioral and cognitive). In addition, singers were shown to rely on internal models for accurate vocal production, and those internal models could be explained within a Bayesian cognitive model framework. While it would be very informative to know if there is evidence that the cytoarchitecture of the cerebellum implements predictive coding, here we will restrict ourselves to look at higher level explanations for inaccurate singing and see if our experiment provides evidence that poor singing is the result of an inefficient internal model implementation that is thought to be instantiated by the cerebellum.

Chapter 4

The Experiment

4.1 Introduction

In this project we intend to test the inverse-model-deficit hypothesis as a possible reason for pitch inaccuracy. In order to do so we will use neuroscience tools and methods. The goal of this research project is to shed lights on the cognitive processes believed to be involved in vocal-pitch-matching accuracy.

4.2 Materials and Methods

4.2.1 Participants

16 students were recruited from the University of California, Merced, student population. All participants were dominantly right-handed and screened for atypical hearing, amusia, and contraindications for TMS including increased risk for seizure, unstable medical problems, metal in the body other than dental fillings, neurological or psychiatric illness, history of syncope, and head or spinal cord surgery or abnormalities (Huang et al., 2005). Participants were asked to remove all metal jewelry before TMS. The age range was 18-34 years old (M = 21.63, SD=4.5). 30% of the sample was male and 70% was female. The experimental protocol was carried out in accordance with the Declaration of Helsinki, reviewed by the University of California, Merced, institutional review board, and all participants gave informed consent before testing.

4.2.2 Task

Participants took the Seattle Singing Accuracy Protocol (SSAP) (Demorest & Pfordresher, 2015) which is an online pitch measurement test hosted by Northwestern university. The test consisted of the following tasks:

1. Pitch matching to a vocal model: This subtask required participants to sing back a single vocal target pitch in 10 trials (5 pitches x 2 trials per pitch). The pitches within each series of 5 different pitches were randomized.

2. Pitch matching to piano tones: This subtask required participants to sing back a single piano timbre in 10 trials (5 pitches x 2 trials per pitch). The piano timbres within each series of 5 different timbres were randomized

3. Imitative singing of 4-note vocal patterns: This subtask required participants to sing back a 4-note pattern based on the same vocal pitch used in the first imitation subtask. There were 6 trials in total (6 x 4-pitch patterns, 3 ascending and 3 descending, in the range of a 5th based on the participant's comfort pitch)

4. Familiar song singing: This subtask required participants to sing a familiar song a cappella from a predetermined list, first with the lyrics then with a neutral syllable (i.e., "doo")

5. Simple pitch discrimination: Adaptive pitch discrimination test

6. Questionnaire responses: Survey about participants' musical background.

4.2.3   Procedure

This experiment was a within subject design, pretest-posttest with four different randomized conditions. The conditions were stimulation of the left cerebellum (region of interest), stimulation of the right cerebellum (region of interest), sham stimulation (i.e., placebo, passive control) with coiled turned away from the motor cortex and stimulation of the visual cortex (active control). Below is an example of the entire experiment for one participant:
Condition 1: SSAP pre $->$ TMS right cerebellum $->$ SSAP post
Condition 2: SSAP pre $->$ sham TMS $->$ SSAP post
Condition 3: SSAP pre $->$ TMS V1 $->$ SSAP post
Condition 4: SSAP pre $->$ TMS left cerebellum $->$ SSAP post
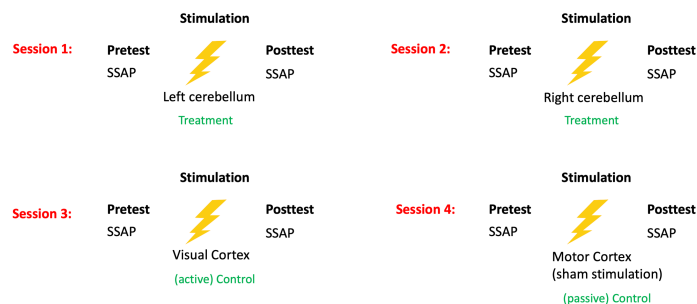
Figure 4.1: Example of the entire study for 1 participant

All conditions were done with one-week interval in between to respect the TMS safety guidelines established by Rossi et al. (2009).

Below is a description of the material used and how the stimulations were conducted:

4.2.4   Material

The SSAP was given on a Macbook (Retina, 13-inch, Mid 2014) with processor 3 GHz Dual-Core Intel Core i7 and 8 GB 1600 MHz DDR3 Memory . Participants heard the

tones on a MHX 5 pair of headphones. They were instructed to adjust the volume to a level that they deemed comfortable to complete the task. Their voices were recorded using a Senheiser microphone. The fundamental singing frequencies of the recordings were analyzed in SSAP via the Yin algorithm (de Cheveigné & Kawahara, 2002).

Single Pulse

Single pulses of TMS were delivered with a figure-of-eight coil (Magstim Rapid2). The coil was placed at the optimal position for eliciting MEPs from the contralateral FDI muscle. The optimal position was tracked and saved using an ANT Visor 2 motion capture based neuro-navigation system in order to ensure identical placement of the coil throughout the experiment. The handle of the coil pointed backward and was perpendicular to the presumed direction of the central sulcus, about 45 degrees to the midsagittal line (see Figure 4.2). The resting motor threshold (RMT) was defined as the lowest intensity that produced MEPs of 50 μV in at least five out of 10 trials with the muscles relaxed (Rossini et al., 1994). The active motor threshold (AMT) was defined as the lowest intensity that produced MEPs of 200 μV in at least five out of 10 trials when the subject made a 10% of maximum contraction using visual feedback (Rothwell, 1997).

Continuous Theta Burst Stimulation (cTBS)

Following Koch et al.'s (2008) procedure, a MagStim Super Rapid magnetic stimulator (Magstim Company, Whitland, Wales, UK), connected with a figure-of-eight coil with a diameter of 90 mm was used to deliver rTMS over the scalp site corresponding to the lateral cerebellum. The magnetic stimulus had a biphasic waveform with a pulse width of about 300 ls. During the first phase of the stimulus, the current in the centre of the coil flowed toward the handle. Three-pulse bursts at 50 Hz repeated every 200 ms for 40 s (equivalent to "continuous theta burst stimulation, cTBS" in Huang et al. (2005) were delivered at 80% AMT (600 pulses).

Cerebellar Stimulation

After eliciting MEPs from the contralateral FDI muscle to determine the adequate motor threshold to be used for the cerebellar stimulation, cTBS was applied via a coil positioned 1 cm inferior and 3 cm lateral to the inion (referred to as 3L1I) on both the contralateral side, and ipsilateral side of the head corresponding to the participant's dominant hand (see Figure 4.2)

Visual Cortex Stimulation

To determine the phosphene threshold, the coil was placed in a vertical position (handle pointing rightward) on the inion–nasion line, with its inferior limit 1 cm above the inion (see Figure 4.2). Stimulation was initially applied at 40% of maximum stimulator output and increased in 5% increments until the subject reported phosphenes. The threshold was then finely determined, by changing the intensity in 1% increments. cTBS was then applied to the visual cortex.

Coil position for stimulation

Motor Threshold   Cerebellar Stimulation   Visual Cortex Stimulation
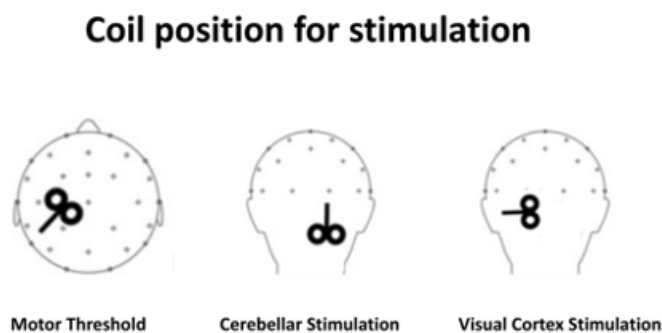
Figure 4.2: Coil position for the different stimulation sites

4.2.5 Pitch Scoring

The scoring was done automatically by the SSAP. The three imitation singing tasks were scored acoustically. A percentage accuracy score was given based on mean deviation scores from target pitches for all items in a block. To be considered accurate, the sung pitch needed to be within +/- 50 CENTS of the target pitch (see Figures 4.3 & 4.4 for examples). The accuracy of 'Familiar song' was scored by comparing the histogram of the fundamental frequency that is sung to the 'ideal' histogram and then computing the percent match.

4.3 Analysis

For this analysis, we decided to divide the models in three parts: One for each vocal imitation subtask. For the first model used the SSAP data coming from the first pitch-matching subtests (i.e., Pitch matching to a vocal model). The second pitch matching subtask (i.e., Pitch matching to piano tones) will be analyzed with a slightly different model that takes into consideration the different piano notes to match. The last pitch imitation subtask (i.e., Imitative singing of 4-note vocal pattern) had about 1.5% of missing data. We conducted a missing data analysis which combines multiple-imputations and Bayesian multilevel model for that data of this subtask

In addition it is relevant to note that we removed 6 participants from the analysis for the first two subtasks. Four of those students were removed because they left the study before completion of the four conditions, and two were removed from the study analysis because the system had failed to record their singing data for two (or more) of the conditions. Thus, the remaining population sample that was analyzed was 10 for subtask 1 and 2. The criteria to keep or exclude participants from the analysis of subtask 1 and 2 was that participants had to have completed all the conditions. We also decided not to include the "Familiar song singing" data in the analysis based on a recommendation from the team of researchers who built the SSAP. The measures from this subtask might be too coarse for research purposes according to them.

The analysis was done using R version 4.0.2 (2020-06-22), and brms_2.15.0. We implemented the WAMBS checklist steps (van de Schoot & Depaoli, 2020).

39
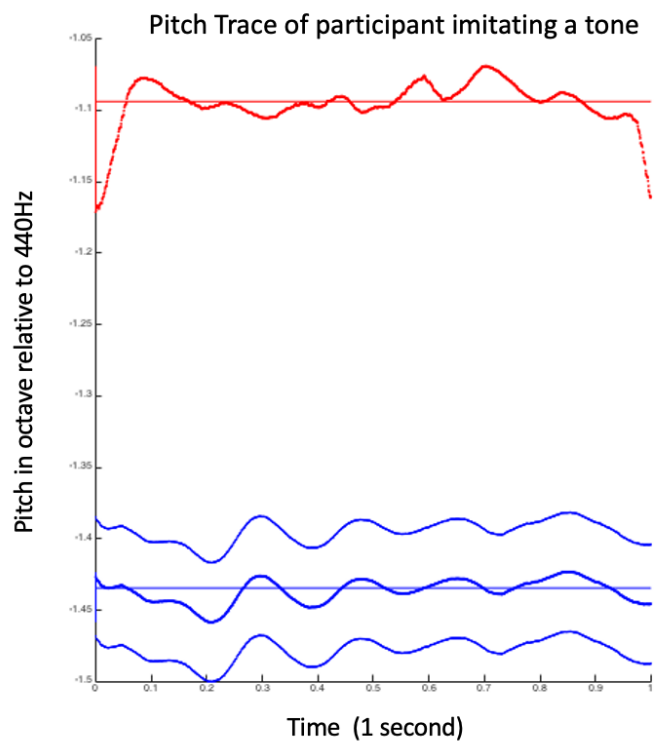
Figure 4.3: SSAP pitch traces of the fundamental frequency of target pitch (blue) and the fundamental frequency of sung pitch (red) for male participant. Middle blue trace is target pitch, upper blue trace is target pitch +50 cents, lower blue trace is target pitch –50 cents. Here participant is not within +/- 50 cents of target pitch, thus the participant pitch matching is considered inaccurate.
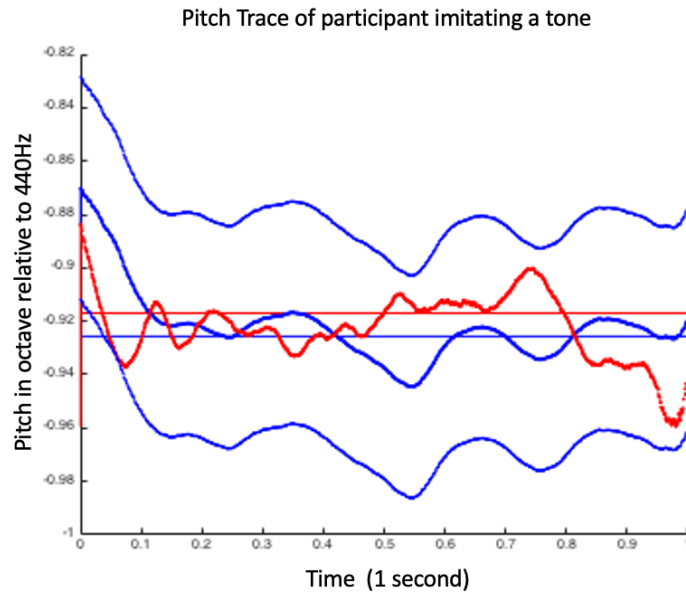
Figure 4.4: SSAP pitch traces of the fundamental frequency of target pitch (blue) and the fundamental frequency of sung pitch (red) for female participant. Middle blue trace is target pitch, upper blue trace is target pitch +50 cents, lower blue trace is target pitch –50 cents. Here participant is within +/- 50 cents of target pitch, thus the participant pitch matching is considered accurate

4.3.1   Subtask 1: Pitch matching to a vocal model

The model used for this analysis was a Bayesian hierarchical multilevel model with random intercept and random slope. The dependent variable was the deviation score from the target pitch (in cents), the predictors were: the stimulation condition (with four levels i.e., left cerebellum, right cerebellum, sham, visual cortex) and the test time point (with two levels i.e., pre, post). A random intercept and slope were fitted for participants in both those predictors. A random intercept was also added for notes/intervals items that each participant had to reproduce. The predictors were sum-zero coded -0.5, 0.5 with 'sham' being the reference level in 'condition' and 'pre' being the reference level in 'test timepoint'.

Most parameters in this model have weakly informative prior distributions. The baseline parameter β_0 was given a normal prior distribution with a fairly wide variance relative to the measure of interest (i.e., deviation score in cents) . Given the contrast coding, β_0 here is the grand mean across conditions and time points. β_1 and β_2 are the group deflection parameter (i.e., parameter indicating how much the central tendency value of a predictor deviates from the grand mean β_0) are given a normal prior distribution with mean of zero as well. The mean of zero for these parameters was chosen following Kruschke's (2014) analysis so as to constrain the deflection parameters to a zero-sum (i.e., deflection parameters need to sum to zero in this model). γ_0, γ_1 , γ_2 the subject adjustments for the intercept, and the slopes for the two predictors were also given multivariate normal distributions with mean zero and a variance covariance matrix Σ.

41

Within that variance covariance matrix the standard deviations σ's for each of the adjustment parameter were half-normally distributed with mean zero, and a moderately restricted scale parameter. ζ_0 is the 'note' or 'interval' adjustment for the intercept. It is recommended to have this adjustment in the model as each participant has to imitate the same series of notes or intervals. The correlation parameter ρ of the variance covariance matrix was given an η of 2. This assumes the correlation between the intercept and the slopes to be most likely zero, but it allows correlations that are non-zero (with correlations that are furtherer away from zero being less likely). A prior predictive check was done and gave plausible values, thus we decided to keep those parameter settings for our model.

Model

$$Deviation_{ij} \sim N(\mu_{ij}, \sigma)$$
$$u_{ij} = (\beta_0 + \gamma_{0j} + \zeta_{0i}) + (\beta_1 + \gamma_{1j}) \cdot x_{1ij} + (\beta_2 + \gamma_{2j}) \cdot x_{2ij} + \beta_3 \cdot (x_{1ij} \cdot x_{2ij})$$
$$\beta_0 \sim N(0, 20)$$
$$\beta_1 \sim N(0, 10)$$
$$\beta_2 \sim N(0, 10)$$
$$\sigma \sim N_+(0, 50)$$

$$\begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{pmatrix} \sim N_3 \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\gamma 0}^2 & \rho_\gamma \sigma_{\gamma 0} \sigma_{\gamma 1} & \rho_\gamma \sigma_{\gamma 0} \sigma_{\gamma 2} \\ \rho_\gamma \sigma_{\gamma 0} \sigma_{\gamma 1} & \sigma_{\gamma 1}^2 & \rho_\gamma \sigma_{\gamma 1} \sigma_{\gamma 2} \\ \rho_\gamma \sigma_{\gamma 0} \sigma_{\gamma 2} & \rho_\gamma \sigma_{\gamma 1} \sigma_{\gamma 2} & \sigma_{\gamma 2}^2 \end{pmatrix} \right)$$

$$\zeta_{0i} \sim N(0, \sigma_{\zeta 0})$$
$$\sigma_{\gamma 0} \sim N_+(0, 8)$$
$$\sigma_{\gamma 1} \sim N_+(0, 5)$$
$$\sigma_{\gamma 2} \sim N_+(0, 5)$$
$$\sigma_{\zeta 0} \sim N_+(0, 15)$$
$$\rho \sim LKJ(2)$$

i = indexes observations
j = indexes subjects

Descriptive plots

Absolute value pitch to target difference by condition and time points
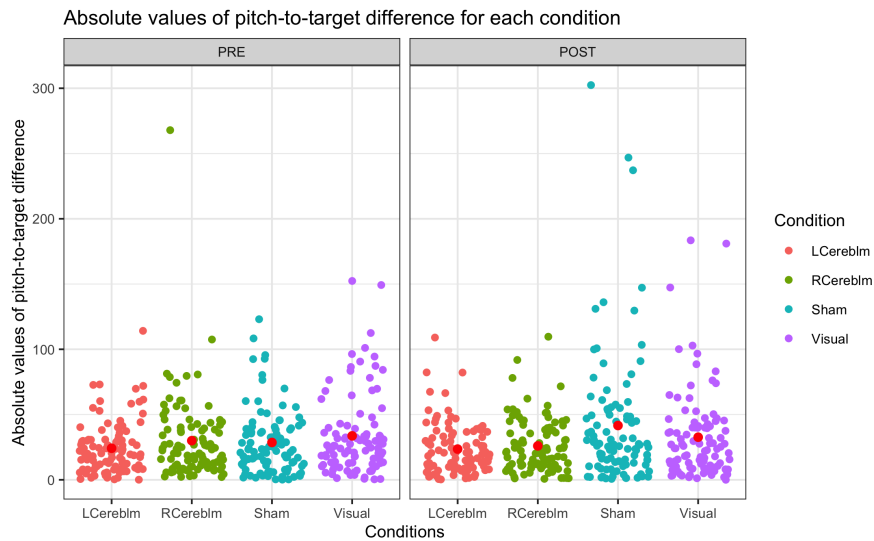
Figure 4.5: Figure shows the absolute deviation values from pitch sung to target pitches. It seems to indicate that most vocal pitch imitations are below +/-100 cents regardless of the conditions or time points. There are however some people who sung more than +/-200 cents of the target pitches.

Mean absolute value pitch to target difference by condition and time points
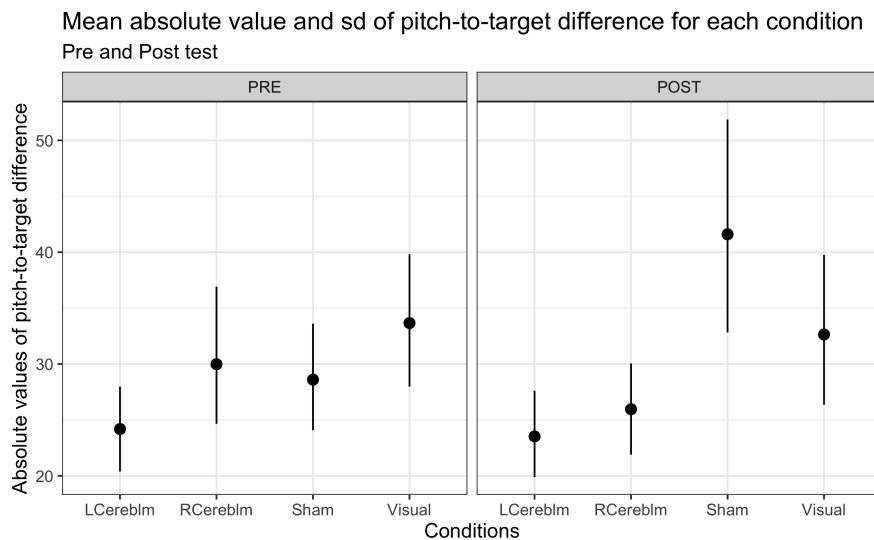
Figure 4.6: Figure shows the mean deviation from the target pitches in each condition, pre and post. Some means are higher than other (e.g., Post-sham stimulation, Pre and Post visual stimulations) but that higher means seem to be driven by what appears to be outlier sung pitches.

Absolute value pitch to target difference by condition and participants (pre and post)
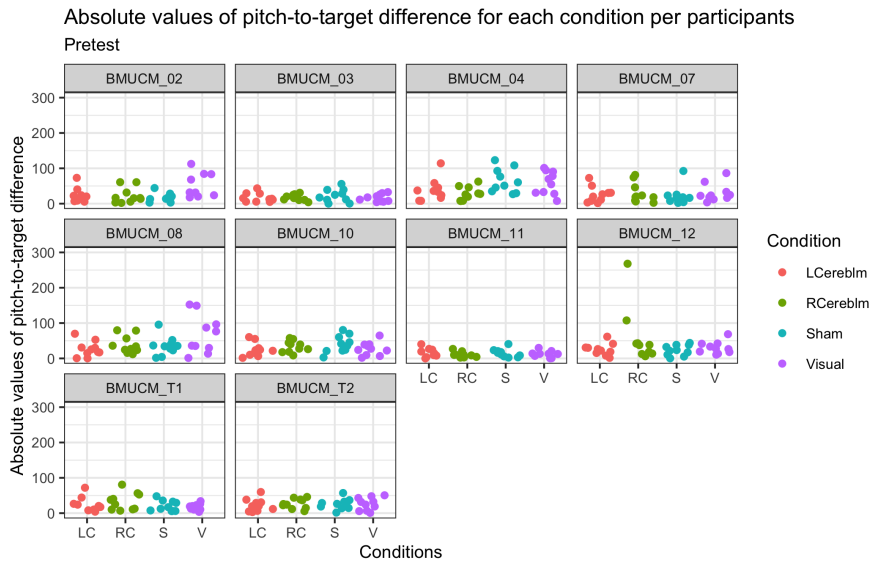
Figure 4.7: Figure shows the absolute deviation from the target pitches in each condition, pretest. Some participants (e.g., 04 and 08) were less accurate overall (i.e., deviation score above 50 cents) in several conditions (e.g., sham and visual) prior to stimulation. Other participants (e.g., 11) were fairly accurate (i.e., deviation score below 50 cents) in all conditions prior to stimulation.
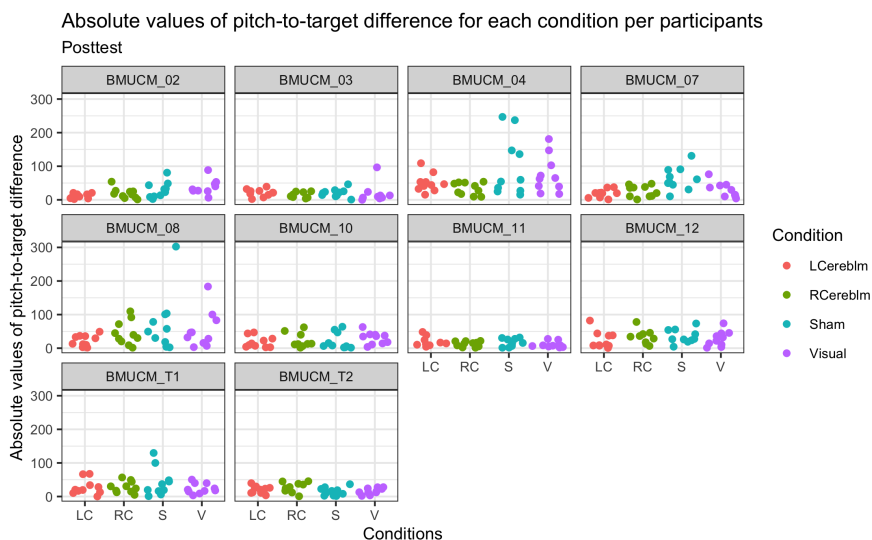
Figure 4.8: Figure shows the absolute deviation from the target pitches in each condition, posttest. Participants seem to be showing more variability post stimulations in some conditions (e.g., sham and visual) and less in others (e.g., left cerebellum.)

44

Results

We ran the Bayesian multilevel analysis with 4 chains for the MCMC, 8000 iterations and 2000 warm-up iterations. No thinning was applied. We assessed convergence with the diagnostic plots, Potential Reduction Scale Factor, and Effective Sample Size. The caterpillar plots showed that the chains perfectly blended with each other, the variance of the chains remained stable and not too large for all the parameters estimated. The PSRF (i.e., Rhat) was below 1.05 and the effective sample sizes were relatively high for all parameters estimated. Thus, these diagnostic measures did not seem to indicate absence of convergence. The prior predictive check gave reasonable values. The result of posterior predictive check indicated some level of misfit. This needs to be taken into account when interpreting the results.
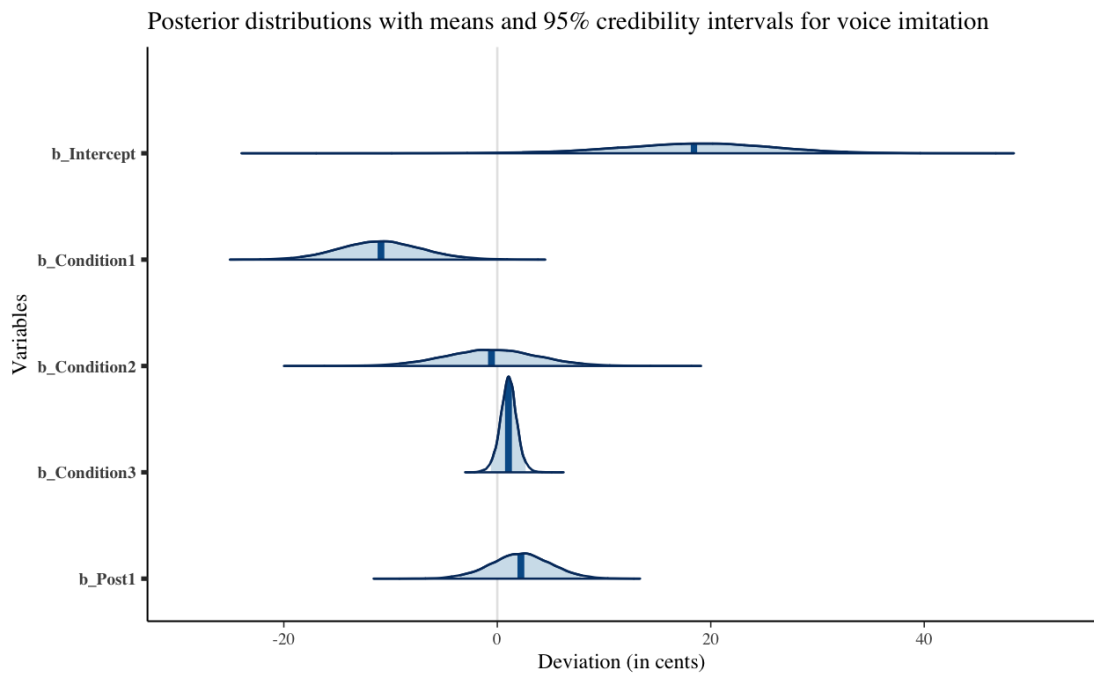
Figure 4.9: Figure shows the posterior distributions of mean deviation scores for the intercept (i.e., grand mean) and the four other conditions (i.e., Condition1 = Left Cerebellum, Condition2 = Right Cerebellum, Condition3 = Visual, Post = Posttest) in the vocal model imitation task. (Posterior distributions of interactions not shown)

Based on the model, the posterior estimate of grand mean deviation score (i.e., collapsed across conditions and time points) was 18.43 cents, 95% HDI [2.37, 32.78]. The posterior estimate of the mean difference from the grand mean for the left cerebellum condition (i.e., Condition1 in our analysis) was -10.90 cents, 95% HDI [-18.41, -3.04]. This seems to indicate that there is some main effect of left cerebellum stimulation. However, because it is a main effect (resulting from the sum-coding step), it is averaging across time points (i.e., Pre and Post). The posterior estimate of the mean difference from the grand mean for the right cerebellum condition, visual condition and the Post timepoint were -0.55 cents, 95% HDI [-9.19, 8.05], 1.04 cents, 95% HDI [-0.60, 2.66], and 2.20 cents 95% HDI[-3.50, 7.80] respectively.
The posterior estimate of the interaction coefficient for the left cerebellum condition and

Table 4.1: Summary results from Bayesian Analysis of vocal model

Summary table Bayesian analysis for voice imitation

Pop_Level_Effects	estimate	est_error	l_95_CI	u_95_CI	PSRF
Intercept	18.43	7.63	2.37	32.78	1.00
Condition1	-10.90	3.89	-18.41	-3.04	1.00
Condition2	-0.55	4.38	-9.19	8.05	1.00
Condition3	1.04	0.81	-0.60	2.66	1.00
Post	2.20	2.87	-3.50	7.80	1.01
Condition1:Post	-4.70	5.56	-15.53	6.07	1.00
Condition2:Post	-7.00	5.52	-17.77	3.89	1.00
Condition3:Post	-0.37	0.96	-2.26	1.50	1.01

the Post timepoint was -4.70, 95% HDI [-15.53, 6.07]. The one for the right cerebellum condition and the Post timepoint was -7.00, 95% HDI [-17.77, 3.89]. The last one, for the visual condition and the Post timepoint was -0.37, 95% HDI [-2.26, 1.50]. Aside from the posterior distribution for the intercept (i.e., the grand mean) and the posterior distribution for the left cerebellum condition level, all the other posterior distributions of the mean estimates have estimates that include 0. This was also the case for the correlations between our variables and their levels. Thus, in the majority of the cases, we do not have enough evidence to say that there is an effect of conditions or time points on our dependent variable (i.e., deviation score). In addition, while we can say that there seems to be some evidence that the left cerebellum condition may have an impact on singing accuracy (and the resulting deviation from target scores), that difference from the grand mean is collapsed across time points (i.e., Pretest, Posttest). Thus, it may be possible that by chance, participants ended up being fairly accurate overall while in this condition.

Discussion

This study aimed to investigate the possible involvement of the cerebellum in one's pitch production accuracy (Pfordresher et al., 2014). Our hypothesis was that by stimulating the cerebellum of participants, they would become less accurate at vocally matching pitches. This task in particular looked at participants' ability to imitate pitches of similar timbre as their own voice. Contrary to our hypothesis, TMSing the cerebellum didn't seem to have a substantial effect on the pitch matching performance of participants. Data from this experiment seems to lack strong evidence to support this hypothesis when it comes to pitch matching to a vocal model. It showed the opposite prediction for one of the condition (i.e., left cerebellum stimulation) where the variance in the deviation scores seemed to decrease. Several elements from the present study could explain the obtained results. One could be related to the experimental protocol. It is difficult to assess proper

stimulation of the cerebellum. Unlike stimulations of the motor cortex or visual cortex which elicit responses such as twitches or phosphenes, respectively, the stimulation of the cerebellum doesn't elicit any responses. In addition, the cerebellum is 'hidden' behind thick neck muscles, which may dampen the strength of the stimulations. Studies have shown that the depth of stimulation has an impact on the stimulation effectiveness (Harwick, Lesage & Miall, 2014). Adding to this, there is an upper limit to maximum stimulation power output that a researcher can administer when using the cTBS protocol (Huang et al. 2005), and that upper limit is maintained regardless of participants' active motor threshold. Thus, it could be possible that participants who had high active motor threshold were understimulated given the cTBS stimulation cap. Another element that could explain the results is the small sample size. This study had to remove participants from the analysis. This removal was due to attrition or equipment failure. Thus, the study might be underpowered, and the chances of detecting a true effect were reduced (Button et al, 2013).

Other elements from this study that could explain those results could be related to the model building and analysis themselves. The posterior predictive check of the model revealed difference between the model predicted values and the observed data. This could indicate potential failing of the model to account for some aspects of the data if those differences appear to be systematic. Therefore, these results should be interpreted with caution. In addition, a number of potentially impactful choices were made in building the model and before running the analysis (e.g., inclusion and exclusion criteria for participants in the dataset, variables to include or exclude from the model, prior distribution parameters to choose). While these decisions are faced by all scientists, making one or two different decisions could lead to different results. To understand the impact of those decisions in our results, we could incorporate a multiverse analysis (Steegen, 2016).

4.3.2   Subtask 2: Pitch matching to piano tones

The model used to analyze task 2 (i.e., pitch matching to piano tone) was similar to the model used for task 1, but the priors were made broader so as to account for the increased difficulty that participants might have when imitating a pitch that is of different timbre (Pfordresher, 2014). The contrast coding remain identical as the previous model.

Model

$$Deviation_{ij} \sim N(\mu_{ij}, \sigma)$$

$$u_{ij} = (\beta_0 + \gamma_{0j} + \zeta_{0i}) + (\beta_1 + \gamma_{1j}) \cdot x_{1ij} + (\beta_2 + \gamma_{2j}) \cdot x_{2ij} + \beta_3 \cdot (x_{1ij} \cdot x_{2ij})$$

$$\beta_0 \sim N(0, 40)$$

$$\beta_1 \sim N(0, 20)$$

$$\beta_2 \sim N(0, 20)$$

$$\sigma \sim N_+(0, 100)$$

$$\begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{pmatrix} \sim N_3 \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\gamma 0}^2 & \rho_\gamma \sigma_{\gamma 0} \sigma_{\gamma 1} & \rho_\gamma \sigma_{\gamma 0} \sigma_{\gamma 2} \\ \rho_\gamma \sigma_{\gamma 0} \sigma_{\gamma 1} & \sigma_{\gamma 1}^2 & \rho_\gamma \sigma_{\gamma 1} \sigma_{\gamma 2} \\ \rho_\gamma \sigma_{\gamma 0} \sigma_{\gamma 2} & \rho_\gamma \sigma_{\gamma 1} \sigma_{\gamma 2} & \sigma_{\gamma 2}^2 \end{pmatrix} \right)$$

$$\zeta_{0i} \sim N(0, \sigma_{\zeta 0})$$

$$\sigma_{\gamma 0} \sim N_+(0, 16)$$

$$\sigma_{\gamma 1} \sim N_+(0, 10)$$

$$\sigma_{\gamma 2} \sim N_+(0, 10)$$

$$\sigma_{\zeta 0} \sim N_+(0, 30)$$

$$\rho \sim LKJ(2)$$

i = indexes observations
j = indexes subjects

Descriptive plots

Absolute value pitch to target difference by condition and time points

Figure 4.10: Figure shows the absolute deviation values from pitch sung to target piano pitches. It seems to indicate that most vocal pitch imitations are below +/-200 cents regardless of the conditions or time points. There are however extreme outliers i.e., people who sung more than +/-600 cents of the target pitches.

Mean absolute value pitch to target difference by condition and time points
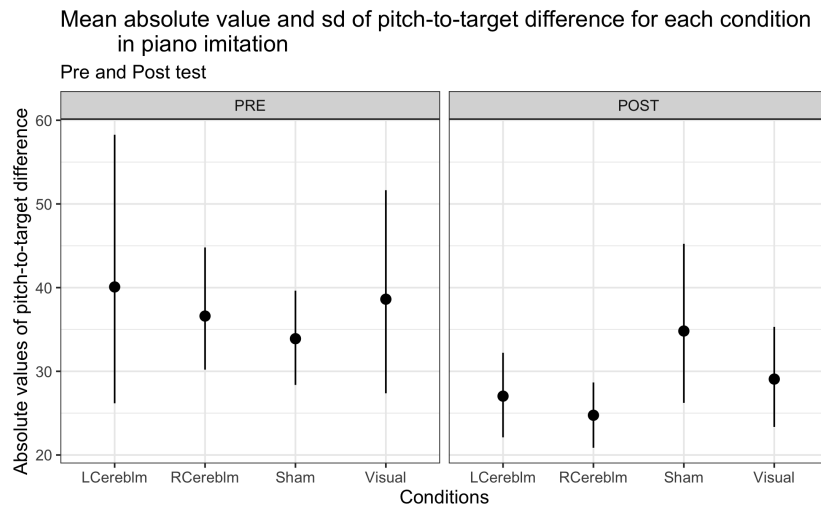
Figure 4.11: Figure shows the mean deviation from the target pitches in each condition, pre and post. Pre left-cerebellar stimulation condition shows a higher mean deviation. This mean is probably driven by the outlier sung pitches of a participant. Overall all mean absolute pitch deviations are below +/- 50 cents which indicates that participants are on average fairly accurate in imitating a vocal model.

Absolute value pitch to target difference by condition and participants (pre and post separate)
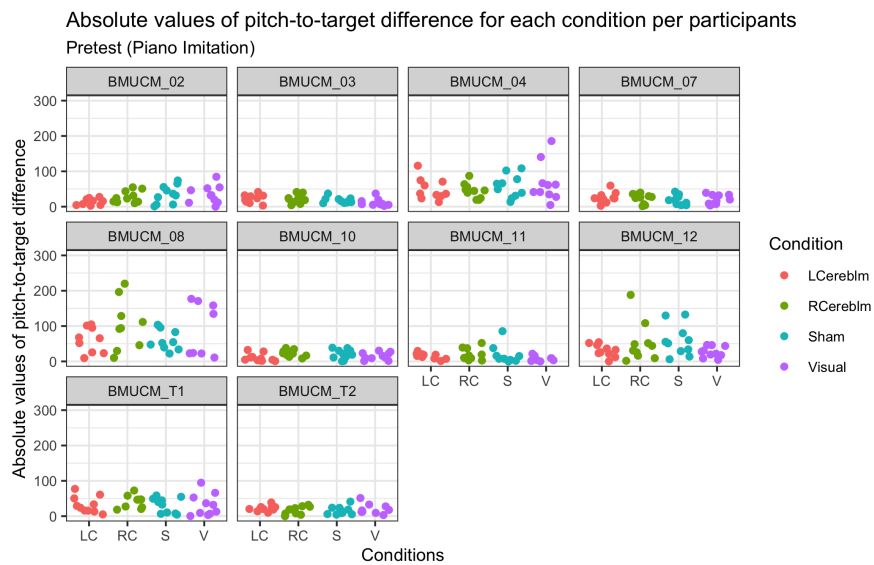
Figure 4.12: Figure shows the absolute deviation from the target pitches in each condition per participants, pretest. Pre stimulations, some participants show more variability in their absolute deviation scores (e.g., 04, 08, 12) while others are fairly accurate (e.g., 10, T2) regardless of conditions.
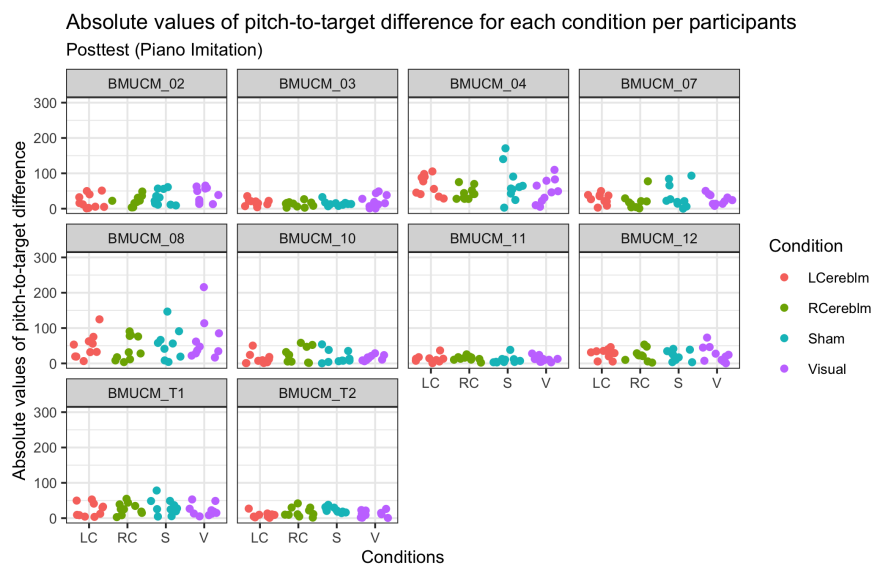
Figure 4.13: Figure shows the absolute deviation from the target pitches in each condition per participants, posttest. It appears that participants who were accurate pre stimulations remained accurate post stimulations (e.g., 03, 11, T2) regardless of conditions. Others (e.g., 04, 08) also showed the same level of variability.

Results
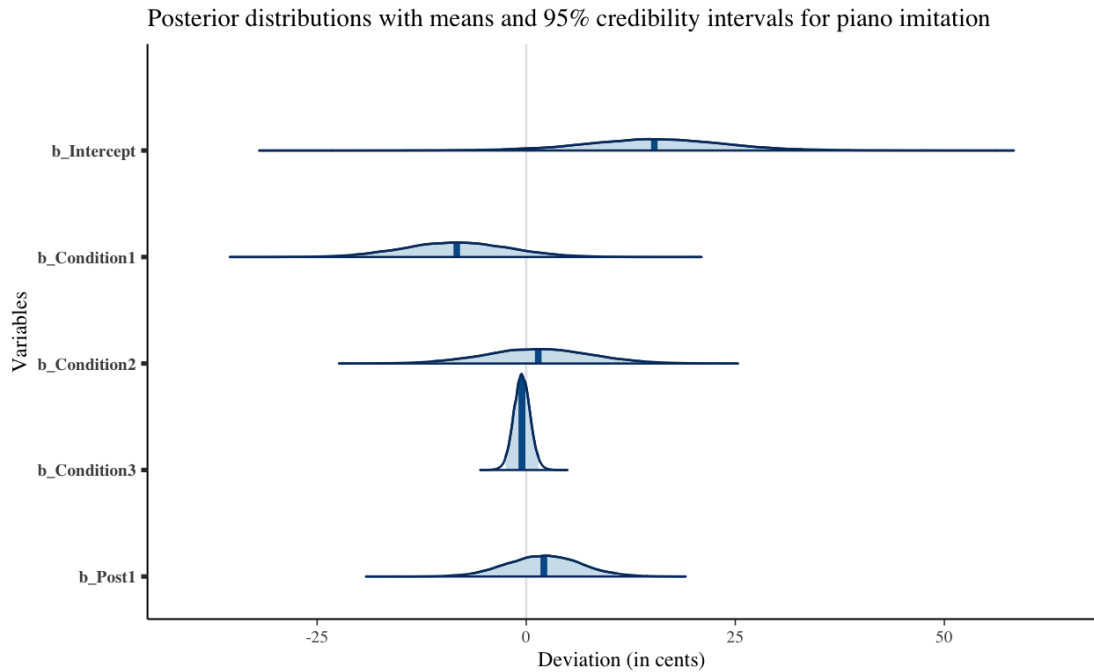
Figure 4.14: Figure shows the posterior distributions of mean deviation scores for the intercept (i.e., grand mean) and the four other conditions (i.e., Condition1 = Left Cerebellum, Condition2 = Right Cerebellum, Condition3 = Visual, Post = Posttest) in the piano tone imitation task. (Posterior distributions of interactions not shown)

Table 4.2: Summary results from Bayesian Analysis of piano tone model

Summary table Bayesian analysis for piano imitation

Pop_Level_Effects	estimate	est_error	l_95_CI	u_95_CI	PSRF
Intercept	15.11	8.88	-2.97	32.24	1
Condition1	-8.37	6.33	-20.78	4.05	1
Condition2	1.52	6.38	-11.12	14.07	1
Condition3	-0.51	1.00	-2.49	1.45	1
Post	2.12	4.45	-6.54	10.85	1
Condition1:Post	15.28	10.36	-5.03	35.79	1
Condition2:Post	-11.03	10.32	-31.60	9.15	1
Condition3:Post	0.38	1.71	-2.94	3.73	1

The results indicated that the posterior estimate of the grand mean deviation score (i.e., collapsed across conditions and time points) was 15.11 cents, 95% HDI [-2.97, 32.24].

The posterior estimates of the mean difference from the grand mean for the left cerebellum condition, the right cerebellum condition and the (post) timepoint variable (i.e., in our Pre-Post condition) were -8.37 cents, 95% HDI [-20.78, -4.05], 1.52 cents, 95% HDI [-11.12, 14.07] and 2.12 cents, 95% HDI [-6.54, 10.85] respectively. The posterior estimate of the interaction coefficient for the left cerebellum condition and the Post timepoint was -15.28, 95% HDI [-5.03, 35.79]. The one for the right cerebellum condition and the Post timepoint was -11.03, 95% HDI [-31.60, 9.15]. The last one, for the visual condition and the Post timepoint was 0.38, 95% HDI [-2.94, 3.73]. All credibility interval of the posterior estimates included 0. This was also the case for the correlations between our variables and their levels. Thus, this suggests that we do not have enough evidence to say that there is an effect of conditions or time points on our dependent variable (i.e., deviation score).

Discussion

This task looked at participants' ability to imitate pitches of a piano tone. Imitating pitches of a different timbre than one's own voice has been shown to be more difficult than imitating a sung pitch (Pfordresher, 2014). Our hypothesis remained the same i.e., stimulating participants' cerebellum will perturb their auditory-vocal system representation which would make them less accurate at vocally imitating pitches in general, and piano tones in particular for this task. Similarly to the previous task, stimulations of the cerebellum didn't not seem to impact participants' ability to imitating a piano tones. Participants who were accurate before stimulations, remained accurate (i.e., absolute deviation scores below +/-50 cents) after stimulations regardless of stimulations sites, and those who were less accurate (i.e., absolute deviation scores above +/-50 cents) also maintained their level of inaccuracy. As the same sample was used, under the same experimental protocol and with similar model of analysis, elements that could explain the obtained results are identical to the previous task.

4.3.3 Subtask 3: Imitative singing of 4-note vocal patterns

Descriptive plots

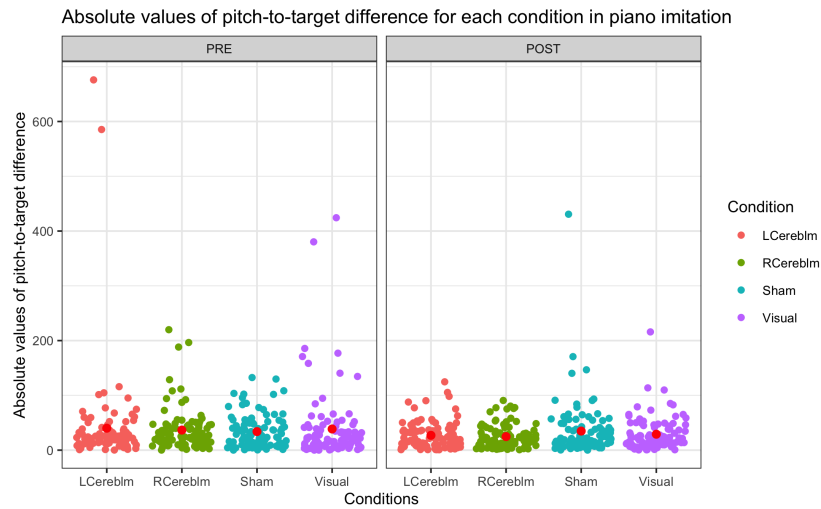Absolute value pitch to target difference by condition and time points
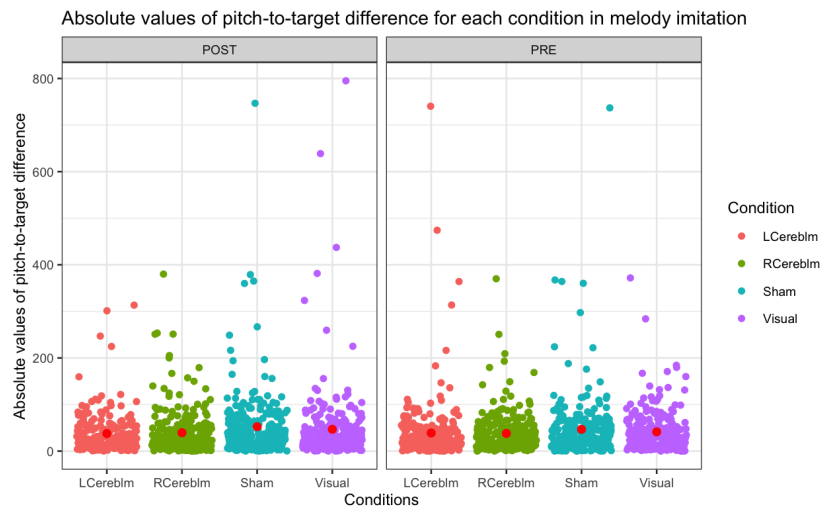
Figure 4.15: Figure shows the absolute deviation values from pitch sung to target vocal melody. It seems to indicate that most vocal pitch imitations are below +/-200 cents regardless of the conditions or time points. Some are below +/-400 and some condition have people who sung more than +/-600 cents of the target pitches.

Although participants completed all conditions of this subtask, some of the measurements were missing. Thus, we first checked amount of missingness in the data (see Figure 4.16). We used the R software version 4.0.2 (2020-06-22), the VIM package (version 6.1.0) and the naniar package (version 0.6.0). The aggr function in VIM and the vis_miss function in naniar were used to check missing data patterns and missing data proportion respectively. The missingness here was believed to be due to a piece of software failing to record some of the participants singing. Because the data was recorded on an online platform, and saved at another university, it was not possible for researchers to double-check right after each session that each recording was properly saved. Those problems seemed to have happened randomly. We checked our missingness by doing a MCAR little's test and using the mcar_test function in naniar.

We ran the analysis of the missing data pattern, missing data proportion, and the MCAR little's test. Missing data patterns was at 2, and only 1.5% of the total data set for this subtask was missing. The variable with the missingness was 'Values' which had 7.39% (Result are shown in Figure 4.15). Little's test was significant (T-statistic = 211.95, df = 5, p-value < 0.001) meaning that the data was not missing completely at random. We attempted to identify auxiliary variables through a path analysis, but none could be identified. Given the hierarchical structure of our data and our small sample size, we decided to analyze our data using a Bayesian hierarchical model. Also given the nature of the analysis and that our data was not normally distributed (data was leptokurtic), we decided to handle missing data using multiple imputation. Some of the practical issues were, as mentioned above, the non-normality of the data, the hierarchical nature of the data and the data violating the MCAR assumption.

53
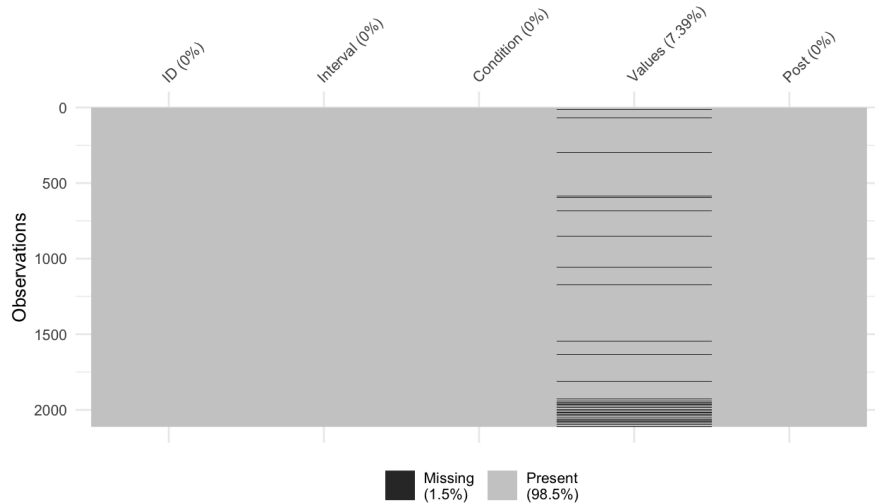
Figure 4.16: Figure shows missing data proportion for Subtask 3

We kept in mind that multiple imputation is a method that assumes a MAR (i.e., missing at random) assumption, and if this is not the case then our results may be biased.

The software we used to run the multiple imputation was R version 4.0.2 (2020-06-22) with the mice package version 3.30.0 (to create the imputed data set), and the brms package version 2.15.0 to fit each imputed data set and pooled the result.

We decided to impute 20 data sets with 50 iterations. The chains seems to have converged (although there seemed to be some recurrent values coming periodically). We increased the number of iterations but we were not able to remove the periodicity observed in some of the chains.
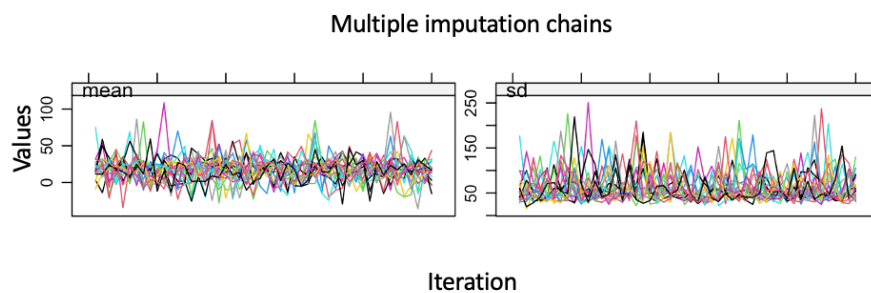
Figure 4.17: Figure shows MCMC Chains from imputed data sets

We checked that imputed values were also plausible by plotting the imputed values against the data. The plot showed that the imputed values were all plausible (see Figure 4.17, 4.18 and 4.19).

Our imputation model was a Bayesian hierarchical mixed effect model with Condition and Session time (i.e., pre or post) as fixed effects, and participants and intervals as random effect. Participants had random intercept and a slope in Condition in our model. We also included an interaction between Condition and Session time (full model specification identical to subtask 1).

54
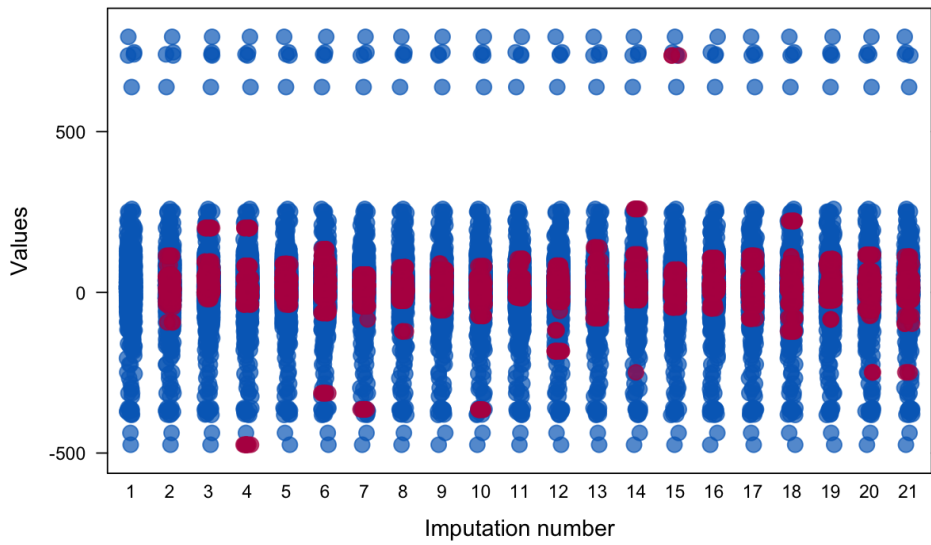
Figure 4.18: Figure shows imputed values (red) against complete data (blue). Imputed values were plausible and remained within boundaries of the real data values
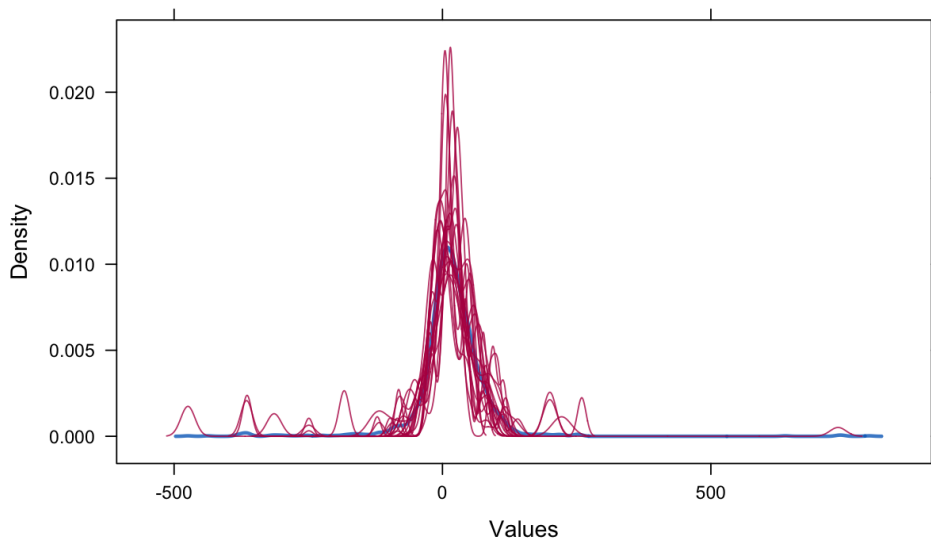
Figure 4.19: Figure shows imputed values (red) against complete (blue) data in density plot forms

We intended to sum-code both predictors [-0.5,0.5] with 'pre' being the reference level for the 'Session time' predictor and sham/placebo being the reference level for the 'Condition' predictor, but the imputation function from mice did not seem to allow dataset with such coding. Thus, we dummy coded our variables with sham being the reference condition within the Condition variable and pretest being the reference time point within the Session time variable. The pooling procedure necessary in multiple imputation could be done in two different ways with Bayesian analysis (to our knowledge). We decided to use the most common way which is the "two-step process" recommended by Gelman et al. (2013). In this process, imputation happens before model fitting (as in frequentist statistics). After the data set is imputed multiple times, one can use a software to analyze each imputed data set. The same Bayesian model is ran on all the imputed data sets and then draws from the different models are just pooled, combining results from multiple imputed Bayesian models. The two step process bypass the need for Rubin's rules. One thing to be aware of is that the PSRF may indicate non-convergence. This is usually a "false positive" which can be due to some differences between the data in the different imputed datasets as well as the chains being from different data sets. It is recommended when possible to inspect the PSRF individually.

Results

Posterior distributions with means and 95% credibility intervals for melody imitation imputed dataset
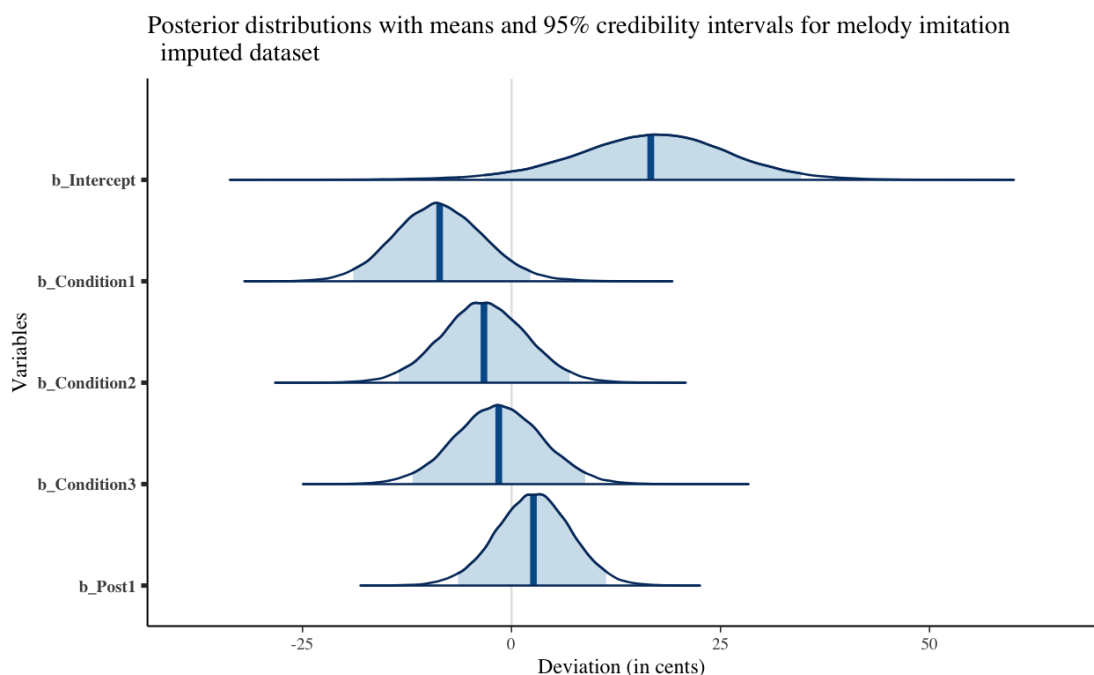
Figure 4.20: Figure shows the posterior distributions of mean deviation scores for the intercept (i.e., mean sham) and the four other conditions (i.e., Condition1 = Left Cerebellum, Condition2 = Right Cerebellum, Condition3 = Visual, Post = Posttest) in the vocal melody imitation task. (Posterior distributions of interactions not shown)

The results from this Bayesian analysis with imputed data reveal that there doesn't seem to be any evidence of decreased pitch matching accuracy when participants have their cerebellum downregulated. In Bayesian terms, the results are inconclusive.

Table 4.3: Summary results from Bayesian Analysis of vocal melody model

Summary table Bayesian analysis for piano imitation

Pop_Level_Effects	estimate	est_error	l_95_CI	u_95_CI	PSRF
Intercept	15.11	8.88	-2.97	32.24	1
Condition1	-8.37	6.33	-20.78	4.05	1
Condition2	1.52	6.38	-11.12	14.07	1
Condition3	-0.51	1.00	-2.49	1.45	1
Post	2.12	4.45	-6.54	10.85	1
Condition1:Post	15.28	10.36	-5.03	35.79	1
Condition2:Post	-11.03	10.32	-31.60	9.15	1
Condition3:Post	0.38	1.71	-2.94	3.73	1

For instance, all variable estimates have a credible interval which includes zero (see Table 4.3). The intercept estimate is 15.11 cents, 95% HDI [-2.97, 32.24]. This means that its estimate of the average deviation score, pre-test in the sham condition (which was the default reference level for the variable Condition) can be anywhere between a positive and a negative number. The 'Pre' session time estimate is 2.12 cents, 95% [-6.54, 10.85]. The same is true for the other predictors and the interactions. In addition, if we take a look at the estimated errors, some are much bigger than the estimated themselves. This underlines the uncertainty in the those estimate. However, the PRSF are all below one which implies that convergence was not an issue in this model for the various predictors. We also conducted an analysis with the complete data set by doing a classic listwise deletion (i.e., all observations with missing data were removed). The analysis gave parameter estimates slightly different from the multiple imputation analysis. Some parameters had their effect size flip sign (i.e., becoming negative while they were positive in the multiple imputation analysis or vice versa). This might be expected given the uncertainty in those and the fact that the results had credible intervals which included zero. The complete data set analysis also yielded slightly bigger estimate error overall.

Discussion

This task looked at participants' ability to imitate pitches of sung melodies. Similarly to previous tasks within this study, this task aimed to investigate the involvement of the cerebellum in one's pitch production accuracy (Pfordresher et al., 2014). In addition, given that our data set had a small sample size and had missing data, we aimed to use a missing data analysis technique in order to compensate for the lack of power. As with the previous tasks our analysis remained inconclusive. Elements that could explain the obtained results remain identical to the previous tasks. In addition, predictive mean matching algorithm was used with multiple imputation, but another method could have

been more appropriate such as the *2l.pan* or the *2lonly.norm*. While both are for imputing univariate missing data, the first one uses a two-level normal model with homogeneous within group variances, whereas the second imputes the data at level 2 using Bayesian linear regression analysis (van Buuren, 2018).

Chapter 5

General Discussion/Conclusion

5.1   Introduction

The voice is considered to be one of the oldest musical instruments and as such singing is one of the oldest forms of music. It is a behavior that starts in infancy, develops through the norms and characteristics of a culture, and can be found in all cultures. Yet, despite the universal nature of this musical activity, it has been found, in western societies, that it was not uncommon to find people who could not vocally reproduce a pitch within accurate range (by western standards). At the onset of this dissertation, I presented the inverse-model-deficit hypothesis of poor pitch matching which had for goal to provide an explanation about the cognitive mechanism responsible for poor pitch matching ability. This hypothesis states that poor singing results from a compromised internal model and an inability to instantiate a plan of action with one's vocal apparatus that would lead to the correct pitch being sung.

Along with investigating the inverse-model-deficit hypothesis of poor pitch matching, this dissertation addressed the questions of whether the cerebellum is the locus of the internal inverse-model formation and if those internal models could be considered Bayesian. Chapter 2 introduced the cerebellum, its structure and functions. Chapter 3 presented an account of the Bayesian brain. Chapter 4 tested the inverse-model-deficit hypothesis. This chapter will summarize the contributions, important observations and points across chapters.

5.2   Cerebellum: the seat of inverse model formation?

Across the three chapters presented in this dissertation, first I focused on establishing the cerebellum as being the locus of formation of internal models which are necessary for singing; second, I focused on providing evidence that these internal models could be explained under the Bayesian brain hypothesis; and third I focused on experimentally testing the assumed mechanisms responsible for inaccurate pitch matching.

Singing is an activity that engages the coordinated control of a respiratory system (e.g., the lungs), the vocal folds (which creates the sound source) and the vocal tracts (which acts as a source filter) (Sundberg, 2000). It necessitates the reliance on auditory and sensorimotor feedback (Pfordresher & Larrouy-Maestri, 2015; Dalla Bella et al. 2011), and makes use of acquired mental representations to drive the vocal production

(Pfordresher, 2014). In Chapter 2 we observed that, in addition to supporting, controlling, coordinating, and learning gross and fine movements (Marr 1969; Albus, 1971; Ito, 2008) the cerebellum was involved in music cognition (Hutchinson et al., 2003; Brown et al. 2004; Parsons, 2001; Keren-Happuch et al., 2014). Moreover, we observed that its neural architecture and its parallel loop connections to the cerebral cortex (e.g., motor and associative cortices) could support internal predictive models of actions and, potentially, of perceptual states (Sokolov, Miall, and Ivry, 2017).

The proposed neural mechanisms for the formation of these internal models are thought to take advantage of the Cerebro-Pontocerebello-dentato-thalamocortical pathway (Ito, 2006, 2008; Sokolov, Miall, and Ivry, 2017). The mechanisms are also thought to rest on a unifying algorithmic principle coined the universal transform i.e., the homogeneous local circuitry and single neural computation of the cerebellum that give rise to a diversity of cerebellar functions (Schmahmann, 1991, 2018, 2019; Diedrichsen et al., 2019; Tanaka et al., 2020; Welniarz et al.,2021). In particular, some have proposed that the forward model, which predicts the consequence of an action, and plays a role in online movement control and motor learning (e.g., singing) (Sokolov et al., 2017) seems to rely on computations of predicted state estimate (by Purkinje cells) from a current state estimate relayed by the mossy fibers. Both state estimates are then filtered and integrated to the deep cerebellar nuclei for optimal estimation and future prediction (Tanaka et al. 2020).

We can see here the similarities with the use of a Kalman filter to update a state estimate within the optimal feedback control (OFC) framework (Kording & Wolpert, 2006; Berniker & Kording, 2011). We have observed in chapter 3 that optimal state estimation of movements involving a forward predictive model and sensory feedback could be modeled using a Kalman filter. The optimal Kalman filter can be seen as "a Bayesian technique for continuously varying problems" (Kording & Wolpert, 2006).

It is important to note that the neural underpinnings of the forward model seems to have gathered all the attention. There seems to be less data regarding the neural underpinnings of the inverse model and its locus of formation in the sensorimotor literature. Few studies having investigated this topic include studies on eye movements which posited that the burst-tonic (BT) neurons (i.e., a group of premotor neurons) represent the output for the eye's inverse model (Belknap & McCrea, 1998; Green, 2007; Ghasia et al., 2008; de Xivry & Ethier, 2008). However, those neural correlates of the inverse model were restricted only to its output. The notion of prediction, which is usually associated with a forward model, appears to be more prevalent than the notion of instantiation, which is usually associated with an inverse model even in music cognition.

One of the aims of Chapter 4, in addition to test the inverse-model deficit hypotheses, was to address this gap in literature about the locus of formation of the inverse model. If participants had indeed been less accurate at pitch matching after downregulating and targeted cerebellar stimulations, that would have provided some evidence that: the cerebellum is involved in poor pitch matching, and that this involvement may be related to the formation of an inverse model at the cerebellar level. While we were not able to determine a causal link between the cerebellum, the formation of an inverse model and poor singing ability, this research question needs to be explored further in order to fill this gap in the sensorimotor control, the sensorimotor prediction, and the music cognition literature.

5.3 Measurement and Quantification of singing accuracy

In this dissertation, fairly novel methods and analysis techniques to experimental studies in music cognition were used and combined to study and estimate singing accuracy. Most studies in music cognition which investigate singing make use of fMRI (Brown et al., 2008, Zarate 2013), PET (Parsons, 2001; Brown et al, 2004) and EEG but only the context of perception (not vocal production) (Gordon et al., 2010, 2011).

While some studies have used TMS to investigate singing perception (Leveque et al, 2013; Royal et al. 2015), cortical representation of motor functions in singing (Sparing et al, 2006), speech and song arrest (Stewart et al. 2001), this is to our knowledge the first study using TMS to investigate singing accuracy.

In addition, given the hierarchical nature of our data, the small sample size, and the presence of missing information we made choice to use and combine Bayesian hierarchical modeling with multiple imputation. Both of these techniques have gained in popularity recently. While the foundation of Bayesian inference was explained in chapter 3, the multiple imputation technique which was used to guesstimate the missing deviation scores of some participants (given what they had previously sung) was not.

Multiple imputation is a statistical approach to handling missing data. It's mathematical underpinning relies on Bayesian estimation principles (Enders, 2010). It involves three distinct phases: the imputation phase, the analysis phase and the pooling phase. The imputation phase consists of creating multiple copies of the dataset and replace each missing values within those copies by likely values sampled from the posterior predictive distribution of the observed data (Huque, 2018). Then each (now complete) data set is analyzed with a specific statistical model (specified by the researcher). Finally, the pooling phase combines everything into a single complete data set result. Because the method relies on Bayesian estimation principles, the imputed values all take into account the uncertainty in their estimation. Thus, in this study, the imputed deviation scores for each of the missing sung pitch took into account the uncertainty associated with that missingness given what participants had already produced. The combination of Bayesian hierarchical model and multiple imputation is also an active area of research.

5.4 Limitations and Future Directions

5.4.1 Limitations

The results drawn from the study in this dissertation were inconclusive at determining if poor pitch matching is indeed caused by a compromised auditory-vocal system representation. However, they opened up new avenues to improve future studies looking at pitch accuracy using cognitive neuroscience methodologies.

One of the main limitations of this study was the sample size. Data collection had to be stopped short because of the covid-19 pandemic which started in the beginning of 2020 and lasted though my finalizing of this dissertation. Restrictions put in place during this period severely restricted access to laboratories and forbade behavioral experiments and behavioral-data collection. The intended number of participants for this study was originally 30. In addition, like a number of experimental designs that rely on testing the

same participants several times over a period of time (e.g., longitudinal studies, within-subject design studies with more than two conditions), the study suffered from attrition. While nothing can be done to prevent the impact of a pandemic, slight modifications to this study's design could increase the power of future similar studies while decreasing the burden on participants. Graham et al. (2001) have shown that planned missing data designs could give very similar results (in terms of power) as designs with complete data. The idea of a planned missing data design is to split the sample of participants into subgroups and each sub groups is then assigned to do a subsets of the experimental conditions instead of being in all the condition. For instance, this study had four pretest-posttest conditions (i.e., sham, visual, left cerebellum, right cerebellum), which required participants to do the same task (i.e., SSAP) eight times. A planned missing data design would have participants be in two of those conditions out of four (e.g., sham & right cerebellum, or visual & left cerebellum). Graham et al. (2001) also showed that in most of the cases the planned missing data designs was more powerful than complete data designs with identical number of observations. It is believed that this is due to the improved quality of the data that one's get by decreasing the demand on participants.

Another potential limitation was the TMS power output required to effectively inhibit cerebellar regions might not have been met for some participants in reason of their high motor thresholds. This could have led to insufficient cerebellar stimulations needed to perturb the instantiation of an internal model of singing. Studies like the ones by Hartwick et al. (2014) and Pauly et al. (2019) have shown that the geometry of coils had an impact on the depth of the tissues reached by TMS, and that the brand of the coil could have an impact on the efficacy of cerebellar stimulations. One possible way to ensure proper stimulation could be to combine cerebellar TMS with EEG. While this method is usually used to investigate cerebellar-cortical connectivity (Fernandez et al. 2020), it could provide a useful way to assess cerebellar stimulation.

In addition, while the accuracy and validity of the SSAP's scoring of vocal pitch imitation is very close to that of expert judges for coarse grained categorical measure, the automated analysis of singing accuracy may not be as robust and accurate to measure pitch deviation scores (Pfordresher & Demorest, 2020). To this day it remains very difficult to automate pitch accuracy scoring that rivals that of human expert judges. Being able to do the same experiment in a conservatory would be ideal. Last but not least, the fact that the SSAP was hosted online and was not run locally in this experiment led to multiple instances where some participants had to wait longer (30s-1:30 mins) in between trials for the content to load. This made the experiment longer and probably less engaging for participants who had to take it eight times. Having a local version of the accuracy software will solve that issue.

5.4.2 Future Directions

Once the aforementioned limitations have been addressed, it would be a worthwhile endeavor to keep investigate pitch accuracy using non invasive brain stimulation techniques. The presented study appears to have been one of the first of its kinds. Other singing related studies in music cognition that have used TMS investigated: motor cortex excitability in music perception (Royal et al. 2015), representation of the larynx motor areas in singing perception (Lévêque et al., 2013), and auditory feedback control of vocal production (Liu et al, 2020).

Another interesting future direction would be to look at cross-cultural aspects of pitch production. In pitch perception/production study involving the Tsimane' (i.e., residents of the Bolivian Amazon) and US residents, Jacoby et al. (2019) asked participants to listen to tones and musical intervals, and reproduce them by singing. While the US residents sang back notes (or intervals) that were out of their range by taking into account the chroma of those notes (i.e., notes were sung some octaves below or above the target tones to match their range), the Tsimane' completely ignored those chroma. One of their conclusions was that octave equivalence might be different across cultures and that it may be dependent on pitch representations gained during enculturation. This provides some evidence that pitch production (whether dimmed accurate or not) may relies on cultural specific internal models. In addition, this is a reminder that pitch matching accuracy should always be defined within the contexts and norms of where the test is being given. Western 'international' music theory standards will not hold in other systems.

Last, but not least, one could also look at the effect of facilitatory cerebellar stimulations on pitch accuracy. It is interesting to note that in a sensorimotor study on patients with cerebellar ataxia, Farzan et al. (2013) have shown that there can be improvement in speech (as well as limb coordination, and gait) following daily cerebellar 1Hz rTMS sessions for 3 weeks. Because 1Hz rTMS is an inhibitory type of stimulation, it was hypothesized that the improvements seen in those patients could be the results of a decrease in cerebellar brain inhibition (CBI), which allowed for enhanced activation and potentiation of the dentate nuclei. In turn, the increased activation of the dentate nuclei was believed to free resources along the cerebello-thalamo-cortical (CTC) pathways, which enhanced the performance in motor and non-motor functions. Similarly, cTBS on the right cerebellum of patients who suffer from speech motor disorders (as well as movement disorders) due to spinocerebellar ataxia has been shown to improve auditory-motor integration for vocal pitch regulation (Lin et al., 2021). In this frequency-altered feedback (FAF) study in which the fundamental frequency (F0) of participants is altered while they produce vocalizations, Lin et al. (2021) demonstrated that spinocerebellar patients decreased the magnitude of their, usually, abnormal compensatory vocal responses in these types of tasks after cTBS on the right cerebellum, but not after a sham stimulation. Akin to Farzan et al. (2013) study, it was hypothesized that downregulation of CBI might have led to increased activity in the prefrontal cortex and allowed for better top-down inhibitory control required in compensatory vocal productions during pitch perturbations (Lin et al., 2021). Following these lines of evidence on the suppression of cerebellar cortical excitability using TMS and its impact on sensorimotor control of the vocal apparatus during speech, one may conjecture that if we were to use facilitatory cerebellar stimulations (e.g., iTBS, 10Hz rTMS) during a pitch-matching accuracy task, we would see an increase in CBI that may lead to a net result of worse vocal control and worse pitch imitation outcomes. However, it is important to take into account that: (1) motor control improvements following inhibitory cerebellar stimulations were mostly seen in patients with cerebellar disorders, and (2) induced inhibitory CBI can result in either motor performance improvement or cognitive performance impairment depending on: participants (healthy vs clinical population), type of coils (figure-of-eight vs double-cone), stimulation intensity, and the task (Oliveri et al., 2017). Future work investigating the effects of excitatory cerebellar stimulations on pitch-matching accuracy in a healthy population would not only help increase our understanding of the neural and cognitive mechanisms involved in poor pitch matching, but would also provide a better

understanding of sensorimotor control in human vocalizations.

5.5   Conclusion

This dissertation focused on testing the inverse-model-deficit hypothesis of poor pitch matching. I have presented two review chapters and one study that focused on establishing this hypothesis within a broader literature and testing it. This dissertation has also made used of the most recent statistical analysis for multilevel data and missing data to quantify results on singing accuracy tasks. Future work should focus on ascertaining the locus of formation of inverse models and their implications in singing and human vocalizations in general. Future work should also focus on expanding singing accuracy studies in cross-cultural contexts.

References

Abdul-Kareem, I. A., Stancak, A., Parkes, L. M., Al-Ameen, M., AlGhamdi, J., Aldhafeeri, F. M., ... & Sluming, V. (2011). Plasticity of the superior and middle cerebellar peduncles in musicians revealed by quantitative analysis of volume and number of streamlines based on diffusion tensor tractography. *The Cerebellum*, *10*(3), 611. doi.org/10.1007/s12311-011-0274-1

Ackermann, H., Gräber, S., Hertrich, I., & Daum, I. (1997). Categorical speech perception in cerebellar disorders. *Brain and language*, *60*(2), 323-331.

Ackermann, H., Mathiak, K., & Ivry, R. B. (2004). Temporal organization of "internal speech" as a basis for cerebellar modulation of cognitive functions. *Behavioral and cognitive neuroscience reviews*, *3*(1), 14-22.

Aitchison, L., & Lengyel, M. (2017). With or without you: predictive coding and Bayesian inference in the brain. *Current opinion in neurobiology*, *46*, 219-227.

Alais, D., O'Shea, R. P., Mesana-Alais, C., & Wilson, I. G. (2000). On binocular alternation. *Perception*, *29*(12), 1437-1445.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current biology*, *14*(3), 257-262.

Albus, J. S. (1971). A theory of cerebellar function. *Mathematical biosciences*, *10*(1-2), 25-61.

Amir, O., Amir, N., & Kishon-Rabin, L. (2003). The effect of superior auditory skills on vocal accuracy. *The Journal of the Acoustical Society of America*, *113*(2), 1102-1108.

Apps, R., & Garwicz, M. (2005). Anatomical and physiological foundations of cerebellar information processing. *Nature Reviews Neuroscience*, *6*(4), 297-311.

Baer, L. H., Park, M. T. M., Bailey, J. A., Chakravarty, M. M., Li, K. Z. H., & Penhune, V. B. (2015). Regional cerebellar volumes are related to early musical training and finger tapping performance. *Neuroimage*, *109*, 130-139.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617-629.

Barton, R. A., & Harvey, P. H. (2000). Mosaic evolution of brain structure in mammals. *Nature*, *405*(6790), 1055-1058. doi: 10.1038/35016580

Barton, R. A., & Venditti, C. (2014). Rapid evolution of the cerebellum in humans and other great apes. *Current Biology*, *24*(20), 2440-2444. doi: 10.1016/j.cub.2014.08.056.

Baumann, O., Borra, R. J., Bower, J. M., Cullen, K. E., Habas, C., Ivry, R. B., ... & Sokolov, A. A. (2015). Consensus paper: the role of the cerebellum in perceptual processes. *The Cerebellum*, *14*(2), 197-220.

Belknap, D. B., & McCrea, R. A. (1988). Anatomical connections of the prepositus and abducens nuclei in the squirrel monkey. *Journal of Comparative Neurology*, *268*(1), 13-28.

Berniker, M., & Kording, K. (2011). Bayesian approaches to sensory integration for motor control. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(4), 419-428.

Bower, J. M. (2010). Model-founded explorations of the roles of molecular layer inhibition in regulating purkinje cell responses in cerebellar cortex: more trouble for the beam hypothesis. *Frontiers in cellular neuroscience*, *4*, 27.

Bracha, V., Zbarska, S., Parker, K., Carrel, A., Zenitsky, G., & Bloedel, J. R. (2009). The cerebellum and eye-blink conditioning: learning versus network performance hypotheses. *Neuroscience*, *162*(3), 787-796.

Braitenberg, V., Heck, D., & Sultan, F. (1997). The detection and generation of sequences as a key to cerebellar function: experiments and theory. *Behavioral and Brain Sciences*, *20*(2), 229-245. doi: 10.1017/S0140525X9700143X

Braitenberg, V. (2002). In defense of the cerebellum. *Annals of the New York Academy of Sciences*, *978*(1), 175-183. doi: 10.1111/j.1749-6632.2002.tb07565.x

Broussard, D. M. (2013). *The cerebellum: learning movement, language, and social skills*. John Wiley & Sons.

Brown, S., Martinez, M. J., Hodges, D. A., Fox, P. T., & Parsons, L. M. (2004). The song system of the human brain. *Cognitive Brain Research*, *20*(3), 363-375.

Brugger, P., & Brugger, S. (1993). The Easter bunny in October: Is it disguised as a duck?. *Perceptual and motor skills*, *76*(2), 577-578. doi: 10.2466/pms.1993.76.2.577

Brunoni, A. R., Nitsche, M. A., Bolognini, N., Bikson, M., Wagner, T., Merabet, L., ... & Fregni, F. (2012). Clinical research with transcranial direct current stimulation (tDCS): challenges and future directions. *Brain stimulation*, *5*(3), 175-195.

Busan, P., Zanon, M., Vinciati, F., Monti, F., Pizzolato, G., & Battaglini, P. P. (2012). Transcranial magnetic stimulation and preparation of visually-guided reaching movements. *Frontiers in neuroengineering*, *5*, 18. doi: 10.3389/fneng.2012.00018

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature reviews neuroscience*, *14*(5), 365-376.

Cantone, M., Lanza, G., Vinciguerra, L., Puglisi, V., Ricceri, R., Fisicaro, F., ... & Pennisi, M. (2019). Age, height, and sex on motor evoked potentials: Translational data from a large Italian cohort in a clinical environment. *Frontiers in human neuroscience*, *13*, 185. doi.org/10.3389/fnhum.2019.00185

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, *36*(3), 181-204.

Colombo, M., & Seriès, P. (2012). Bayes in the brain—on Bayesian modelling in neuroscience. The British journal for the philosophy of science, 63(3), 697-723.

Dalla Bella, S. (2016). Vocal performance in occasional singers. In G. Welch, D. Howard & J. Nix (Eds.), *The Oxford Handbook of Singing*, Oxford University Press, Oxford, UK.

D'angelo, E., Mazzarello, P., Prestori, F., Mapelli, J., Solinas, S., Lombardo, P., ... & Congi, L. (2011). The cerebellar network: from structure to function and dynamics. *Brain research reviews*, *66*(1-2), 5-15.

De Cheveigné, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, *111*(4), 1917-1930.

De Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception?. *Trends in cognitive sciences*, *22*(9), 764-779.

Demorest, S. M., & Pfordresher, P. Q. (2015). Seattle Singing Accuracy Protocol – SSAP [Measurement instrument]. https://ssap.music.northwestern.edu/

Denison, R. N., Piazza, E. A., & Silver, M. A. (2011). Predictive context influences perceptual selection during binocular rivalry. *Frontiers in human neuroscience*, *5*, 166. doi: 10.3389/fnhum.2011.00166

Desmond, J. E., Chen, S. A., & Shieh, P. B. (2005). Cerebellar transcranial magnetic stimulation impairs verbal working memory. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *58*(4), 553-560.

de Xivry, J. J. O., & Ethier, V. (2008). Neural correlates of internal models. *Journal of Neuroscience*, *28*(32), 7931-7932.

De Zeeuw, C. I., Hoogenraad, C. C., Koekkoek, S. K. E., Ruigrok, T. J., Galjart, N., & Simpson, J. I. (1998). Microcircuitry and function of the inferior olive. *Trends in neurosciences*, *21*(9), 391-400.

Diamond, A. (2000). Close interrelation of motor development and cognitive development and of the cerebellum and prefrontal cortex. *Child development*, *71*(1), 44-56.

Diedrichsen, J., King, M., Hernandez-Castillo, C., Sereno, M., & Ivry, R. B. (2019). Universal transform or multiple functionality? Understanding the contribution of the human cerebellum across task domains. *Neuron*, *102*(5), 918-928. doi.org/10.1016/j.neuron.2019.04.021

Dow, R. S. (1942) The evolution and anatomy of the cerebellum. *Biological Reviews of the Cambridge Philosophical Society*, 17, 179-220.

Ebner, T. J., & Pasalar, S. (2008). Cerebellum predicts the future motor state. *The Cerebellum*, *7*(4), 583-588.

Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *Journal of Neuroscience*, *30*(49), 16601-16608.

Eliades, S. J., & Wang, X. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature*, *453*(7198), 1102-1106.

Enders, C. K. (2010). *Applied missing data analysis*. Guilford press.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429-433.

Faisal, A. A., Selen, L. P., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature reviews neuroscience*, *9*(4), 292-303.

Farzan, F., Wu, Y., Manor, B., Anastasio, E. M., Lough, M., Novak, V., ... & Pascual-Leone, A. (2013). Cerebellar TMS in treatment of a patient with cerebellar ataxia: evidence from clinical, biomechanics and neurophysiological assessments. *The Cerebellum*, *12*(5), 707-712. doi: 10.1007/s12311-013-0485-8

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

Fernandez, L., Rogasch, N. C., Do, M., Clark, G., Major, B. P., Teo, W. P., ... & Enticott, P. G. (2020). Cerebral cortical activity following non-invasive cerebellar stimulation—A systematic review of combined TMS and EEG studies. *The Cerebellum*, *19*(2), 309-335.

Fiez, J. A., Petersen, S. E., Cheney, M. K., & Raichle, M. E. (1992). Impaired non-motor learning and error detection associated with cerebellar damage: A single case study. *Brain*, *115*(1), 155-178.

Flament, D., & Hore, J. (1986). Movement and electromyographic disorders associated with cerebellar dysmetria. *Journal of neurophysiology*, *55*(6), 1221-1233.

Franklin, D. W., & Wolpert, D. M. (2011). Computational mechanisms of sensorimotor control. *Neuron*, *72*(3), 425-442.

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences*, *360*(1456), 815-836.

Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature reviews neuroscience*, *11*(2), 127-138.

Gao, W., Chen, G., Reinert, K. C., & Ebner, T. J. (2006). Cerebellar cortical molecular layer inhibition is organized in parasagittal zones. *Journal of Neuroscience*, *26*(32), 8377-8387.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 2013. *Bayesian data analysis*. 3rd ed. Boca Raton, Florida: Chapman & Hall/CRC.

Ghasia, F. F., Meng, H., & Angelaki, D. E. (2008). Neural correlates of forward and inverse models for eye movements: evidence from three-dimensional kinematics. *Journal of Neuroscience*, *28*(19), 5082-5087.

Ghazanfar, A. A., & Rendall, D. (2008). Evolution of human vocal production. *Current biology*, *18*(11), R457-R460.

Gibson, J. J., & Carmichael, L. (1966). *The senses considered as perceptual systems* (Vol. 2, No. 1, pp. 44-73). Boston: Houghton Mifflin.

Glickstein, M., Oberdick, J., & Voogd, J. (2010). Evolution of the Cerebellum.

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological methods*, *11*(4), 323.

Green, A. M., Meng, H., & Angelaki, D. E. (2007). A reevaluation of the inverse dynamic model for eye movements. *Journal of Neuroscience*, *27*(6), 1346-1355.

Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *290*(1038), 181-197.

Gregory, R. L. (2005). The Medawar lecture 2001 knowledge for vision: Vision for knowledge. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1458), 1231-1251.

Gregory, R. L. (1997). Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *352*(1358), 1121-1127.

Griffiths, T. L., & Yuille, A. L. (2006). Technical introduction: A primer on probabilistic inference. UCLA. Department of Statistics Papers no. 2006010103. UCLA, Los Angeles, CA.

Griffiths, T., Kemp, C., & B Tenenbaum, J. (2008). Bayesian models of cognition.

Grimaldi, G., Argyropoulos, G. P., Boehringer, A., Celnik, P., Edwards, M. J., Ferrucci, R., ... & Ziemann, U. (2014). Non-invasive cerebellar stimulation—a consensus paper. *The Cerebellum*, *13*(1), 121-138. doi.10.1007/s12311-013-0514-7

Guediche, S., Holt, L. L., Laurent, P., Lim, S. J., & Fiez, J. A. (2015). Evidence for cerebellar contributions to adaptive plasticity in speech perception. *Cerebral Cortex*, *25*(7), 1867-1877.

Guell, X., Hoche, F., & Schmahmann, J. D. (2015). Metalinguistic deficits in patients with cerebellar dysfunction: empirical support for the dysmetria of thought theory. *The Cerebellum*, *14*(1), 50-58.

Hardwick, R. M., Lesage, E., & Miall, R. C. (2014). Cerebellar transcranial magnetic stimulation: the role of coil geometry and tissue depth. *Brain stimulation*, *7*(5), 643-649.

Harris, C. M., & Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature*, *394*(6695), 780-784.

Heilbron, M., & Chait, M. (2018). Great expectations: is there evidence for predictive coding in auditory cortex?. *Neuroscience*, *389*, 54-73. doi.org/10.1016/j. neuroscience.2017.07.061

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature reviews neuroscience*, *13*(2), 135-145.

Hoche, F., Guell, X., Vangel, M. G., Sherman, J. C., & Schmahmann, J. D. (2018). The cerebellar cognitive affective/Schmahmann syndrome scale. *Brain*, *141*(1), 248-270. doi: 10.1093/brain/awx317

Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in human neuroscience*, *5*, 82.

Hohwy, J. (2013). *The predictive mind*. Oxford University Press.

Huang, Y. Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron*, *45*(2), 201-206.

Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of neurophysiology*, *28*(2), 229-289.

Huque, M. H., Carlin, J. B., Simpson, J. A., & Lee, K. J. (2018). A comparison of multiple imputation methods for missing data in longitudinal studies. *BMC medical research methodology*, *18*(1), 1-16.

Hutchins, S., Larrouy-Maestri, P., & Peretz, I. (2014). Singing ability is rooted in vocal-motor control of pitch. *Attention, Perception, & Psychophysics*, *76*(8), 2522-2530. Doi: 10.3758/s13414-014-0732-1.

Hutchins, S. M., & Peretz, I. (2012). A frog in your throat or in your ear? Searching for the causes of poor singing. *Journal of Experimental Psychology: General*, *141*(1), 76.

Hutchinson, S., Lee, L. H. L., Gaab, N., & Schlaug, G. (2003). Cerebellar volume of musicians. *Cerebral cortex*, *13*(9), 943-949.

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological science*, *15*(5), 356-360.

Ishikawa, T., Tomatsu, S., Izawa, J., & Kakei, S. (2016). The cerebro-cerebellum: could it be loci of forward models?. *Neuroscience research*, *104*, 72-79.

Ito, M. & Kano, M. (1982) Long-lasting depression of parallel fiber-Purkinje cell transmission induced by conjunctive stimulation of parallel fibers and climbing fibers in the cerebellar cortex. *Neuroscience. Letters.* 33(3), 253-258. doi: 10.1016/0304-3940(82)90380-9

Ito, M. (2006). Cerebellar circuitry as a neuronal machine. *Progress in neurobiology*, *78*(3-5), 272-303.

Ito, M. (2008). Control of mental activities by internal models in the cerebellum. *Nature Reviews Neuroscience*, *9*(4), 304-313.

Ivry, R. B., & Spencer, R. M. (2004). The neural representation of time. *Current opinion in neurobiology*, *14*(2), 225-232.

Jacoby, N., Undurraga, E. A., McPherson, M. J., Valdés, J., Ossandón, T., & McDermott, J. H. (2019). Universal and non-universal features of musical pitch perception revealed by singing. *Current Biology*, *29*(19), 3229-3243.

Jones, J. A., & Keough, D. (2008). Auditory-motor mapping for pitch control in singers and nonsingers. *Experimental brain research*, *190*(3), 279-287.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems [J]. *Journal of basic Engineering*, *82*(1), 35-45.

Kant, I. (1908). Critique of pure reason. 1781. *Modern Classical Philosophers, Cambridge, MA: Houghton Mifflin*, 370-456.

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current opinion in neurobiology*, *9*(6), 718-727.

Kelly, R. M., & Strick, P. L. (2003). Cerebellar loops with motor cortex and prefrontal cortex of a nonhuman primate. *Journal of neuroscience*, *23*(23), 8432-8444.

Keough, D., & Jones, J. A. (2009). The sensitivity of auditory-motor representations to subtle changes in auditory feedback while singing. *The Journal of the Acoustical Society of America*, *126*(2), 837-846. doi: 10.1121/1.3158600

Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive processing: a canonical cortical computation. *Neuron*, *100*(2), 424-435. doi:10.1016/j.neuron.2018.10.003

Keller, A. J., Houlton, R., Kampa, B. M., Lesica, N. A., Mrsic-Flogel, T. D., Keller, G. B., & Helmchen, F. (2017). Stimulus relevance modulates contrast adaptation in visual cortex. *Elife*, *6*, e21589.

Keren-Happuch, E., Chen, S. H. A., Ho, M. H. R., & Desmond, J. E. (2014). A meta-analysis of cerebellar contributions to higher cognition from PET and fMRI studies. *Human brain mapping*, *35*(2), 593.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annu. Rev. Psychol.*, *55*, 271-304.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, *27*(12), 712-719.

Kok, P., Jehee, J. F., & De Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron*, *75*(2), 265-270.

Kok, P., Brouwer, G. J., van Gerven, M. A., & de Lange, F. P. (2013). Prior expectations bias sensory representations in visual cortex. *Journal of Neuroscience*, *33*(41), 16275-16284.

Kok, P., & de Lange, F. P. (2015). Predictive coding in sensory cortex. In *An introduction to model-based cognitive neuroscience* (pp. 221-244). Springer, New York, NY.

Körding, K. P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in cognitive sciences*, *10*(7), 319-326.

Kruschke, J. (2014). Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan.

Kuo, A. D. (1995). An optimal control model for analyzing human postural balance. *IEEE transactions on biomedical engineering*, *42*(1), 87-101.

Lega, C., Vecchi, T., D'Angelo, E., & Cattaneo, Z. (2016). A TMS investigation on the role of the cerebellum in pitch and timbre discrimination. *Cerebellum & ataxias*, *3*(1), 1-7.

Leiner, H. C., Leiner, A. L., & Dow, R. S. (1991). The human cerebro-cerebellar system: its computing, cognitive, and language skills. *Behavioural brain research*, *44*(2), 113-128.

Lesage, E., Morgan, B. E., Olson, A. C., Meyer, A. S., & Miall, R. C. (2012). Cerebellar rTMS disrupts predictive language processing. *Current Biology*, *22*(18), R794-R795.

Lévêque, Y., Muggleton, N., Stewart, L., & Schön, D. (2013). Involvement of the larynx motor area in singing-voice perception: A TMS study. *Frontiers in Psychology*, *4*, 418.

Lieder, F., Stephan, K. E., Daunizeau, J., Garrido, M. I., & Friston, K. J. (2013). A neurocomputational model of the mismatch negativity. *PLoS computational biology*, *9*(11), e1003288.

Lin, Q., Chang, Y., Liu, P., Jones, J. A., Chen, X., Peng, D., ... & Liu, H. (2021). Cerebellar Continuous Theta Burst Stimulation Facilitates Auditory–Vocal Integration in Spinocerebellar Ataxia. *Cerebral Cortex*.

Lindle, R. S., Metter, E. J., Lynch, N. A., Fleg, J. L., Fozard, J. L., Tobin, J., ... & Hurley, B. F. (1997). Age and gender comparisons of muscle strength in 654 women and men aged 20–93 yr. *Journal of applied physiology*, *83*(5), 1581-1587.

Liu, D., Dai, G., Liu, C., Guo, Z., Xu, Z., Jones, J. A., ... & Liu, H. (2020). Top–down inhibitory mechanisms underlying auditory–motor integration for voice control: evidence by TMS. *Cerebral Cortex*, *30*(8), 4515-4527.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature neuroscience*, *9*(11), 1432-1438.

Ma, W. J., Beck, J. M., & Pouget, A. (2008). Spiking networks for Bayesian inference and choice. *Current opinion in neurobiology*, *18*(2), 217-222.

Manto, M., Bower, J. M., Conforto, A. B., Delgado-García, J. M., Da Guarda, S. N. F., Gerwig, M., ... & Timmann, D. (2012). Consensus paper: roles of the cerebellum in motor control—the diversity of ideas on cerebellar involvement in movement. *The Cerebellum*, *11*(2), 457-487.

Mariën, P., & Manto, M. (2018). Cerebellum as a master-piece for linguistic predictability. doi.org/10.1007/s12311-017-0894-1

Marr, D. (1969). A theory of cerebellar cortex. *The Journal of Physiology*, 202(2), 437-470. https://doi.org/10.1113/jphysiol.1969.sp008820

Marr, D (1982). Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. W. H. Freeman and Company.

McIntyre, J., Zago, M., Berthoz, A., & Lacquaniti, F. (2001). Does the brain model Newton's laws?. *Nature neuroscience*, *4*(7), 693-694.

Mead, C. (1990). Neuromorphic electronic systems. *Proceedings of the IEEE*, *78*(10), 1629-1636.

Merfeld, D. M., Zupan, L., & Peterka, R. J. (1999). Humans use internal models to estimate gravity and linear acceleration. *Nature*, *398*(6728), 615-618.

Merton, P. A., & Morton, H. B. (1980). Stimulation of the cerebral cortex in the intact human subject. *Nature*, *285*(5762), 227-227.

Miall, R. C. (2013). Cerebellum: Anatomy and function. In *Neuroscience in the 21st Century: From Basic to Clinical* (pp. 1149-1167). Springer. doi 10.1007/978-1-4614-1997-6_38

Moberget, T., & Ivry, R. B. (2016). Cerebellar contributions to motor control and language comprehension: searching for common computational principles. *Annals of the New York Academy of Sciences*, *1369*(1), 154.

Molinari, M., Leggio, M. G., Solida, A., Ciorra, R., Misciagna, S., Silveri, M. C., & Petrosini, L. (1997). Cerebellum and procedural learning: evidence from focal cerebellar lesions. *Brain: a journal of neurology*, *120*(10), 1753-1762.

Morgenstern, Y., Murray, R. F., & Harris, L. R. (2011). The human visual system's assumption that light comes from above is weak. *Proceedings of the National Academy of Sciences*, *108*(30), 12551-12553.

Mumford, D. (1992) On the computational architecture of the neocortex. II. The role of cortico- cortical loops. *Biological Cybernetics* 66(3), 241–251

Murdoch, B. E. (2010). The cerebellum and language: historical perspective and review. *Cortex*, *46*(7), 858-868.

Nigmatullina, Y., Hellyer, P. J., Nachev, P., Sharp, D. J., & Seemungal, B. M. (2015). The neuroanatomical correlates of training-related perceptuo-reflex uncoupling in dancers. *Cerebral Cortex*, *25*(2), 554-562. doi.org/10.1093/cercor/bht266.

Nikjeh, D. A., Lister, J. J., & Frisch, S. A. (2009). The relationship between pitch discrimination and vocal production: Comparison of vocal and instrumental musicians. *The Journal of the Acoustical Society of America*, *125*(1), 328-338.

Norris, D., McQueen, J. M., & Cutler, A. (2016). Prediction, Bayesian inference and feedback in speech recognition. *Language, cognition and neuroscience*, *31*(1), 4-18.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in cognitive sciences*, *11*(12), 520-527.

Oliveri, M., Torriero, S., Koch, G., Salerno, S., Petrosini, L., & Caltagirone, C. (2007). The role of transcranial magnetic stimulation in the study of cerebellar cognitive function. *The Cerebellum*, *6*(1), 95-101.

O'Reilly, J. X., Jbabdi, S., & Behrens, T. E. (2012). How can a Bayesian approach inform neuroscience?. *European Journal of Neuroscience*, *35*(7), 1169-1179.

O'Reilly, J. X., & Mars, R. B. (2015). Bayesian models in cognitive neuroscience: a tutorial. *An introduction to model-based cognitive neuroscience*, 179-197.

Parrell, B., Lammert, A. C., Ciccarelli, G., & Quatieri, T. F. (2019a). Current models of speech motor control: a control-theoretic overview of architectures and properties. *The Journal of the Acoustical Society of America*, *145*(3), 1456-1481. doi.org/10.1121/1.5092807

Parrell, B., Ramanarayanan, V., Nagarajan, S., & Houde, J. (2019b). The FACTS model of speech motor control: Fusing state estimation and task-based control. *PLoS computational biology*, *15*(9), e1007321. doi.org/10.1101/543728

Parsons, L. M. (2001). Exploring the functional neuroanatomy of music performance, perception, and comprehension. *Annals of the New York Academy of Sciences*, *930*(1), 211-231.

Pauly, M. G., Steinmeier, A., Bolte, C., Hamami, F., Tzvi, E., Münchau, A., ... & Weissbach, A. (2021). Cerebellar rTMS and PAS effectively induce cerebellar plasticity. *Scientific Reports*, *11*(1), 1-13.

Penny, W. (2012). Bayesian models of brain and behaviour. *International Scholarly Research Notices*, *2012*.

Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. *Annu. Rev. Psychol.*, *56*, 89-114.

Peretz, I., & Vuvan, D. T. (2017). Prevalence of congenital amusia. *European Journal of Human Genetics*, *25*(5), 625-630.

Perry, D. W., Zatorre, R. J., Petrides, M., Alivisatos, B., Meyer, E., & Evans, A. C. (1999). Localization of cerebral activity during simple singing. *Neuroreport*, *10*(18), 3979-3984.

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of" tone deafness". *Music Perception*, *25*(2), 95-115.

Pfordresher, P. Q., & Mantell, J. T. (2014). Singing with yourself: Evidence for an inverse modeling account of poor-pitch singing. *Cognitive psychology*, *70*, 31-57.

Pfordresher, P. Q., & Demorest, S. M. (2020). Construction and validation of the Seattle Singing Accuracy Protocol (SSAP): An automated online measure of singing accuracy. In *The Routledge Companion to Interdisciplinary Studies in Singing* (pp. 322-333). Routledge.

Pinto, Y., van Gaal, S., de Lange, F. P., Lamme, V. A., & Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. *Journal of Vision*, *15*(8), 13-13.

Popa, L. S., Hewitt, A. L., & Ebner, T. J. (2012). Predictive and feedback performance errors are signaled in the simple spike discharge of individual Purkinje cells. *Journal of Neuroscience*, *32*(44), 15345-15358.

Popa, L. S., Streng, M. L., Hewitt, A. L., & Ebner, T. J. (2016). The errors of our ways: understanding error representations in cerebellar-dependent motor learning. *The Cerebellum*, *15*(2), 93-103.

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, *62*(2), 816-847.

Pruitt, T. A., & Pfordresher, P. Q. (2015). The role of auditory feedback in speech and song. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(1), 152. doi.org/10.1037/a0038285

Ramachandran, V.S. (1988) Perceiving shape from shading. *Scientific American*. 259, 76–83.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, *2*(1), 79-87.

Rao, S. M., Harrington, D. L., Haaland, K. Y., Bobholz, J. A., Cox, R. W., & Binder, J. R. (1997). Distributed neural systems underlying the timing of movements. *Journal of Neuroscience*, *17*(14), 5528-5535.

Rossi, S., Hallett, M., Rossini, P. M., Pascual-Leone, A., & Safety of TMS Consensus Group. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clinical neurophysiology*, *120*(12), 2008-2039.

Rossini, P. M., Barker, A. T., Berardelli, A., Caramia, M. D., Caruso, G., Cracco, R. Q., ... & Tomberg, C. (1994). Non-invasive electrical and magnetic stimulation of the brain, spinal cord and roots: basic principles and procedures for routine clinical application. Report of an IFCN committee. *Electroencephalography and clinical neurophysiology*, *91*(2), 79-92.

Rothwell, J. C. (1997). Techniques and mechanisms of action of transcranial stimulation of the human motor cortex. *Journal of neuroscience methods*, *74*(2), 113-122.

Royal, I., Lidji, P., Théoret, H., Russo, F. A., & Peretz, I. (2015). Excitability of the motor system: A transcranial magnetic stimulation study on singing and speaking. *Neuropsychologia*, *75*, 525-532.

Schmahmann, J. D. (1991). An emerging concept: the cerebellar contribution to higher function. *Archives of neurology*, *48*(11), 1178-1187.

Schmahmann, J. D., & Sherman, J. C. (1998). The cerebellar cognitive affective syndrome. *Brain: a journal of neurology*, *121*(4), 561-579.

Schmahmann, J. D. (2004). Disorders of the cerebellum: ataxia, dysmetria of thought, and the cerebellar cognitive affective syndrome. *The Journal of neuropsychiatry and clinical neurosciences*, *16*(3), 367-378.

Schmahmann, J. D., Guell, X., Stoodley, C. J., & Halko, M. A. (2019). The theory and neuroscience of cerebellar cognition. *Annual review of neuroscience*, *42*, 337-364.

Seriès, P., & Seitz, A. (2013). Learning what to expect (in visual perception). *Frontiers in human neuroscience*, *7*, 668.

Sexton, B. M., Liu, Y., & Block, H. J. (2019). Increase in weighting of vision vs. proprioception associated with force field adaptation. *Scientific reports*, *9*(1), 1-13. doi.org/10.1038/s41598-019-46625-7

Shadmehr, R., & Krakauer, J. W. (2008). A computational neuroanatomy for motor control. *Experimental brain research*, *185*(3), 359-381.

Sokolov, A. A., Miall, R. C., & Ivry, R. B. (2017). The cerebellum: adaptive prediction for movement and cognition. *Trends in cognitive sciences*, *21*(5), 313-332.

Spencer, R. M., & Ivry, R. B. (2013). Cerebellum and timing. *Handbook of the cerebellum and cerebellar disorders*, 1201-1219.

Spratling, M. W. (2008). Predictive coding as a model of biased competition in visual attention. *Vision research*, *48*(12), 1391-1408.

Spratling, M. W. (2010). Predictive coding as a model of response properties in cortical area V1. *Journal of neuroscience*, *30*(9), 3531-3543.

Spratling, M. W. (2012). Predictive coding as a model of the V1 saliency map hypothesis. *Neural Networks*, *26*, 7-28.

Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and cognition*, *112*, 92-97.

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and clinical neurophysiology*, *38*(4), 387-401.

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702-712.

Stein, T., & Peelen, M. V. (2015). Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. *Journal of Experimental Psychology: General*, *144*(6), 1089-1104.

Sterzer, P., Frith, C., & Petrovic, P. (2008). Believing is seeing: expectations alter visual awareness. *Current Biology*, *18*(16), R697-R698. doi: 10.1016/j.cub.2008.06.021

Stoodley, C. J., & Schmahmann, J. D. (2009). Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies. *Neuroimage*, *44*(2), 489-501.

Strata, P. (2009). David Marr's theory of cerebellar learning: 40 years later. *The Journal of physiology*, *587*(Pt 23), 5519.

Sundberg, J. (1990). What's so special about singers?. *Journal of voice*, *4*(2), 107-119.

Sundberg, J. (1994). Perceptual aspects of singing. *Journal of voice*, *8*(2), 106-122.

Tanaka, H., Ishikawa, T., Lee, J., & Kakei, S. (2020). The cerebro-cerebellum as a locus of forward model: a review. *Frontiers in systems neuroscience*, *14*, 19.

Tesche, C. D., & Karhu, J. J. (2000). Anticipatory cerebellar responses during somatosensory omission in man. *Human brain mapping*, *9*(3), 119-142.

Todorov, E. (2004). Optimality principles in sensorimotor control. *Nature neuroscience*, *7*(9), 907-915.

Tölgyesi, B., & Evers, S. (2014). The impact of cerebellar disorders on musical ability. *Journal of the neurological sciences*, *343*(1-2), 76-81.

Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature neuroscience*, *6*(4), 391-398.

van Beers, R. J., Sittig, A. C., & van der Gon Denier, J. J. (1996). How humans combine simultaneous proprioceptive and visual position information. *Experimental brain research*, *111*(2), 253-261.

van Beers, R. J., Sittig, A. C., & Gon, J. J. D. V. D. (1999). Integration of proprioceptive and visual position-information: An experimentally supported model. *Journal of neurophysiology*, *81*(3), 1355-1364.

van Beers, R. J., Haggard, P., & Wolpert, D. M. (2004). The role of execution noise in movement variability. *Journal of neurophysiology*, *91*(2), 1050-1063.

van Buuren, S. (2018). *Flexible Imputation of Missing Data*. Chapman and Hall/CRC. https://stefvanbuuren.name/fimd/sec-multioutcome.html.

van de Schoot, R., & Depaoli, S. (2014). Bayesian analyses: Where to start and what to report. *The European Health Psychologist*, *16*(2), 75-84.

Van de Schoot, R., Veen, D., Smeets, L., Winter, S. D., & Depaoli, S. (2020). A tutorial on using the WAMBS checklist to avoid the misuse of Bayesian statistics. *Small Sample Size Solutions: A Guide for Applied Researchers and Practitioners; van de Schoot, R., Miocevic, M., Eds*, 30-49.

Von Helmholtz, H. (1867). *Handbuch der physiologischen Optik: mit 213 in den Text eingedruckten Holzschnitten und 11 Tafeln* (Vol. 9). Voss. (English tranl. 1924 JPC Southall as Treatise on Physiological Optics)

Welniarz, Q., Worbe, Y., & Gallea, C. (2021). The forward model: a unifying theory for the role of the cerebellum in motor control and sense of agency. *Frontiers in Systems Neuroscience*, *15*.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*(5232), 1880-1882.

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in cognitive sciences*, *2*(9), 338-347.

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current biology*, *11*(18), R729-R732.

Wolpert, D. M. (2007). Probabilistic models in human sensorimotor control. *Human movement science*, *26*(4), 511-524.

Wolpert, D. M., & Flanagan, J. R. (2016). Computations underlying sensorimotor learning. *Current opinion in neurobiology*, *37*, 7-11. doi: 10.1016/j.conb.2015.12.003

Wyart, V., Nobre, A. C., & Summerfield, C. (2012). Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proceedings of the National Academy of Sciences*, *109*(9), 3593-3598.

Xu, D., Liu, T., Ashe, J., & Bushara, K. O. (2006). Role of the olivo-cerebellar system in timing. *Journal of Neuroscience*, *26*(22), 5990-5995.

Yavari, F., Mahdavi, S., Towhidkhah, F., Ahmadi-Pajouh, M. A., Ekhtiari, H., & Darainy, M. (2016). Cerebellum as a forward but not inverse model in visuomotor adaptation task: a tDCS-based and modeling study. *Experimental brain research*, *234*(4), 997-1012.

Yeo, C. H., & Hesslow, G. (1998). Cerebellum and conditioned reflexes. *Trends in cognitive sciences*, *2*(9), 322-330.

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis?. *Trends in cognitive sciences*, *10*(7), 301-308.

Zarate, J. M., & Zatorre, R. J. (2005). Neural substrates governing audiovocal integration for vocal pitch regulation in singing. *Annals of the New York Academy of Sciences*, *1060*(1), 404-408.

Zarate, J. M., & Zatorre, R. J. (2008). Experience-dependent neural substrates involved in vocal pitch regulation during singing. *Neuroimage*, *40*(4), 1871-1887.

Zarate, J. M. (2013). The neural control of singing. *Frontiers in human neuroscience*, *7*, 237.

Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *Journal of neuroscience*, *14*(4), 1908-1919.