A Music Predictive Coding System: Chroma Directed Initial and Final Tonic Assignment

[1] This work seeks to develop a proposal about music perception: that all tonic phrases are perceived retrospectively (feedback model,) with their roles being determined by unfolding accidentals or chromatic notes of musical entities within the evolving musical continuum—and then they assign the role of the tonic functions in cadential phrases prospectively (feedforward model); We may refer this phenomenon to as "Chroma Directed Initial and Final Tonic Assignment." [2] To represent the syntactical underpinnings of this central claim, I devise a context-sensitive, grammar-based (CSG) analytical framework aimed at analyzing musical time slices on a moment-to-moment basis, with an emphasis on distinct musical phrasal durations, content, and their interrelationships using data-driven Phrase Structure Rules (PSR). [3] To demonstrate this concept, my work also incorporates statistical learning and corpus-driven probabilistic key-finding models to feed the mechanism of CSG with key-finding models' pitch class sets and chord structures at regular intervals. [4] Computational operations in CSG illustrate how these functions shift depending on different memory spans involved in musical perception. This framework investigates how memory duration and predictive mechanisms shape listeners' split-second musical perception, highlighting the interplay between prospective and retrospective processing.

Overall, by integrating statistical learning models and context-sensitive grammar, this work explores how sequential relationships within musical entities create predictive expectations, influencing listeners' interpretations of key and chord functions across varying memory spans. Before presenting the literature review of these interconnected arguments, I wish to demonstrate one key aspect and then revisit the related body of work. We will now explore how memory duration and predictive mechanisms shape listeners' moment-to-moment music perception and how this can be represented through CSG, thus clarifying the interaction between prospective and retrospective processing

The CSG framework analyzes musical time slices on a moment-to-moment basis, focusing on unique phrasal durations, content, and their interconnections. A Case of a probabilistic key-finding model is applied to the short musical example (8 seconds) shown in Figure 1, where the musical surface is analyzed every two seconds. It extracts the pitch class set, assigns it to a variable, and identifies the chord structure for this duration. We refer to this variable as "short-term memory." Every two seconds, the program clears the existing content of this set and fills it with new elements from the current musical surface structure, like a person who perceives external auditory stimuli but forgets them after a brief period.

Figure 1 : Short (8sec) musical example.

In this context, the representation of entities can be visualized as depicted in Figure 2. Here, the categories shown in blue can be considered as variables whose values are shown in red. With our limited two-second capacity, establishing relationships between these entities is challenging. Consequently, there is no distinct perception separating the first and the last G chord and constructing a relationship between other chordal entities.

Now, let us assume the program extends its capacity and detects the content of the musical surface every 4 seconds. From the perspective of statistical learning, entities that frequently occur in sequence create expectations, forming a predictive framework for musical processing. For example, when we hear a G Major chord at the beginning of a piece, we anticipate that the key might be G Major. Within the context of this key, we understand that G Major is often conclusive, C Major tends to lead to D Major, and D Major typically resolves back to G Major. As a reflection of this pattern, we can take these observations as syntactic tendencies within Western Classical Music (White, 2022.) Thus, the relationships between entities that populate the variables every 4 seconds can be constructed based on these predictive relationships, creating a structured model of musical expectancy and progression.

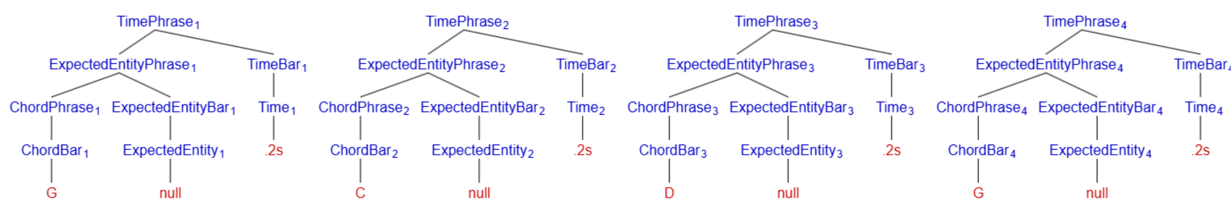

Figure 2 : 2-second-long memory and entity representation of 8-second-long music example.

Figure 3 illustrates an activity of a 4-second memory span that resets every 4 seconds. Based on observational adequacy derived from statistical learning, I assign a conclusive role to the G Major chord in the first tree, which triggers a predictive process, leading me to expect that the next chord will most often be D Major and occasionally C Major. However, before I can fully interpret this progression, the memory resets, leaving me with a data-free blank slate once again.

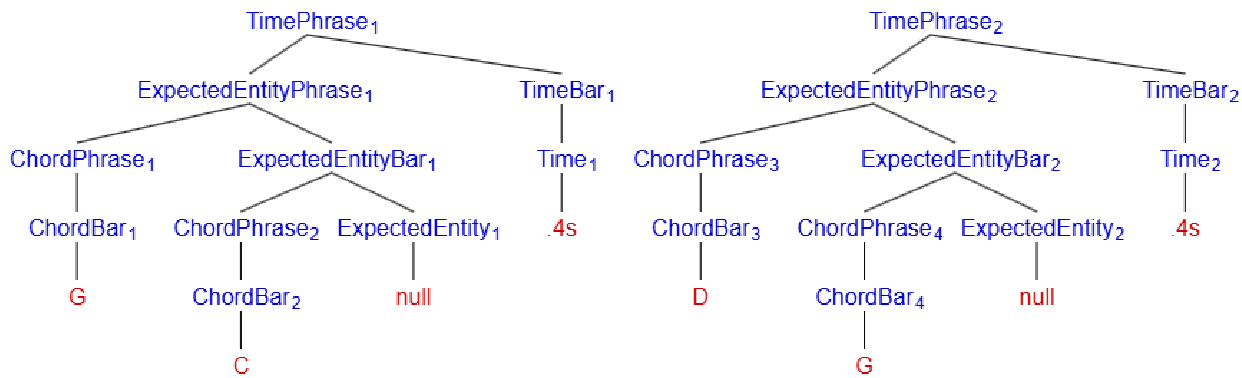

Figure 3 : 4-second-long memory and entity representation of the 8-second-long music example.

When I hear the C Major chord and then the D Major chord, my mind resets every 4 seconds, preventing me from finding enough entities to form meaningful relationships. Thus, at that moment, D Major becomes the conclusive chord, and the key is perceived as D Major. Statistically, if a G Major chord follows D Major, it suggests that the next chord might commonly be A Major. However, before I can establish any connections, the memory resets once more. By increasing the program's capacity to 6 seconds, there is a more insightful understanding of tonal perception compared to the limited 4-second memory span.

Figure 4 illustrates what happens when we have a 6-second memory span that resets every 6 seconds. Observational adequacy from statistical learning leads me to assign a conclusive role to the G Major chord in the initial cycle. When I hear the C Major chord, it triggers a predictive process that makes me expect the next chord to most often be D Major, and occasionally G Major. When the expected D Major follows C Major, my brain is relieved from additional cognitive effort, but it continues to function as an expectation-driven mechanism, wondering what comes next—until the memory resets once again.

Upon hearing the G Major chord in the second cycle of Figure 4, no meaningful comparison can be established due to the resetting of memory. Additionally, this G Major chord lasts for about 2 seconds, followed by a rest, further complicating any ongoing relationships within the musical context.

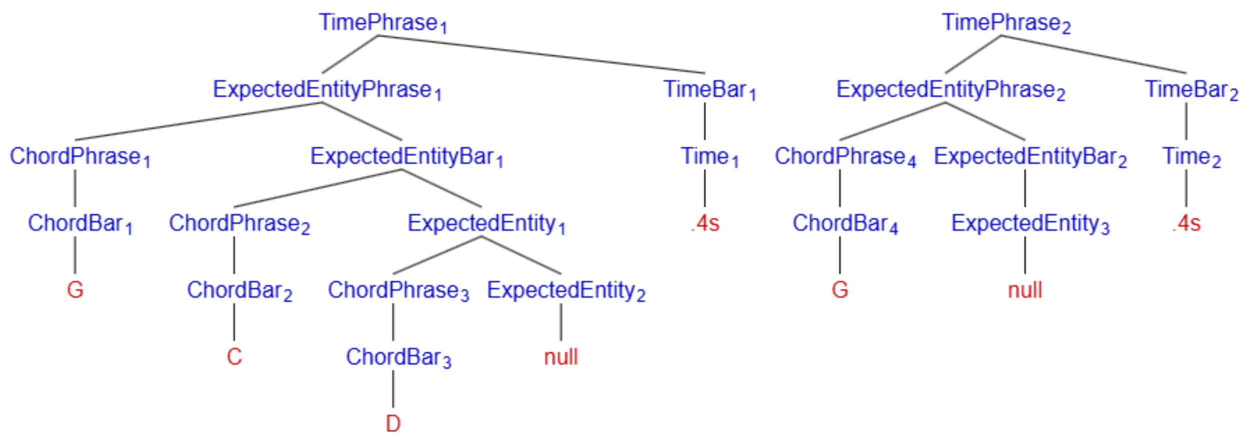

Figure 4 : 6-second-long memory and entity representation of the 8-second-long music example.

This is indeed a crucial point. Should I ask whether a hypothetical person with a 4-second memory perceives the third chord as conclusive, thus assigning the key as D Major? However, when the memory extends to 6 seconds, its function shifts—it becomes a conclusive and expected entity following C Major, not as part of the D Major key, but as part of the G Major key. I would genuinely appreciate being corrected if there is an alternative explanation that I have not considered.

This idea highlights three important points: First, the expected chord function depends on the memory duration and the relationships among entities within that span. Second, music interpretation involves both prospective (forward-looking) and retrospective (reflective) processes. Third, at any given moment, an entity's auditory function can either maintain our expectations or shift our prospective understanding, which may have an impact on the retrospective understanding. By extending the memory duration to 8 seconds, we can see another important dimension of this process.

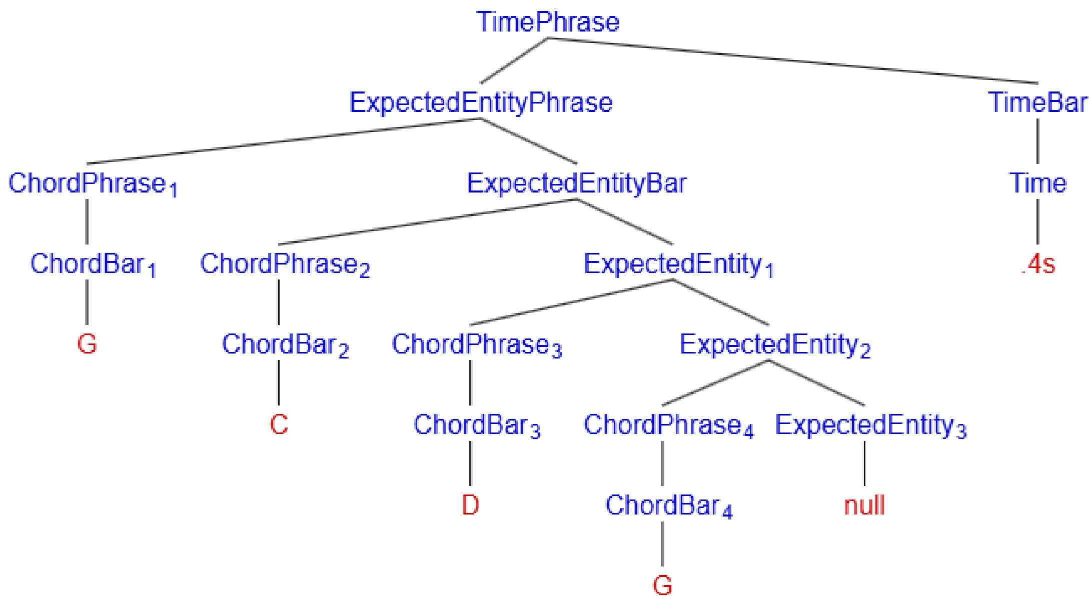

Figure 5 : 8-Second-long memory and entity representation of the 8-second-long music example.

As observational adequacy, we hear G Major as a conclusive and highly stable chord. However, an important question arises: how does it differ from the first conclusive chord we hear, or from subsequent conclusive chords of the same level (subdominant, dominant, tonic), particularly when other elements appear in between? Addressing these questions will require a deeper exploration of data-driven PSRs later in this work, leading to tree structures that are depicted quite differently from what has been presented so far. However, even this demonstration might suffice to say that harmonic functions shift depending on different memory spans involved in musical perception.

Probabilistic mechanisms within key-finding models, which operate based on specific musical slices, can determine not only the overall key of a piece but also identify local key changes that occur at specific moments, referred to as musical discontinuities in this work. It will be important for us to see not just each slice in isolation but also how these slices interact and combine within the relative flow of musical time, offering a nuanced understanding of music's structure and its cognitive implications. At the same time, when we discuss statistical models of key findings, it is closely related to corpus studies. So, the question then is: what are the general harmonic tendencies of these music slices in corpora? While context-sensitive grammar provides a theoretical framework to describe complex musical building block dependencies, statistical learning offers practical and flexible methods to approximate these dependencies through corpus-directed PSRs.

In summary, I have made 4 points so far; first, the primary thesis assertion: that all tonic phrases are understood retrospectively (using a feedback model), with their roles shaped by the unfolding accidentals or chromatic notes of musical building blocks within the developing musical context—while they subsequently assign the tonic functions of cadential phrases in a forward-looking, feedforward manner —a concept I refer to as the " Chroma Directed Initial and Final Tonic Assignment." Secondly, to provide discrete musical building blocks to CSG, this approach utilizes statistical learning and corpus-driven probabilistic key-finding models to use pitch class sets and chord structures at consistent intervals. Thirdly, the CSG framework will allow me to illustrate this hypothesis with an analytical framework for analyzing musical time slices moment-to-moment, focusing on distinct phrasal durations, content, and their interrelationships through data-driven Phrase Structure Rules. Finally, computational operations in CSG illustrate how these functions shift depending on different memory spans involved in musical perception. In this thesis, I will follow a bottom-up strategy to find evidence for data-driven PSRs of CSG and demonstrate the central claim of this work. Thus, in the literature review and beyond, I will first examine what is the role of memory span in music perception and incorporate it into key finding models in the second chapter. In the next to last chapter, I will present the analytic notation of CSG with its constituent structures. In the final chapter, I will demonstrate the “Chroma Directed Initial and Final Tonic Assignment.”

The processing of music listening is a dynamic interaction between sensory memory, short-term memory, and long-term memory. Sensory memory captures immediate auditory input, short-term memory allows for the manipulation and retention of musical information, and long-term memory enables the storage and retrieval of complex musical structures and emotional associations. This multi-tiered memory system underscores the cognitive richness of music listening and highlights the potential for enhancing memory through musical engagement. (Saffran et al., 2001; Ginsborg, 2002; Xu, 2009; Strait & Kraus, 2011; Zendel & Alain, 2012; Swaminathan & Schellenberg, 2018; Zanto et al., 2022.) As a prerequisite for a minimally viable listening process, it is essential to understand the roles and functions of different memory capacities in music listening. Therefore, in the first section of the present thesis, I will introduce the basics of memory types and elaborate on their impact on music listening and what some melodic harmonic led memory discontinuities as a surprise effect. This part will give us memory-based data-driven rules that are the first working mechanism in CSG I devices in the third chapter. So, it will be important to look at the theories of key finding models in the second chapter to integrate our explorations in human memory types with the models.

The literature on key-finding models illustrates a rich interplay between statistical analysis, cognitive psychology, and music theory. The contributions of Krumhansl, Temperley, Quinn, and White have laid a robust foundation for understanding how musical keys are perceived and processed, with ongoing research continuing to contribute to these models for broader applications across diverse musical genres. (Krumhansl&Kessler, 1982; Krumhansl, 1990; Petri&Krumhansl, 2003, Temperley, 2004, Schmuckler&Tomovski, 2005; Temperley&Marvin, 2007; Quinn, 2010; Vuvan&Olivia et al., 2014; White&Quinn, 2016, 2018; White, 2018a, 2018b, 2022) In the present work, I aim to include capacity-related functions of memory types in key-finding models with a web application to present their main differences in terms of processing music in human memory with feasible and manageable chunks. Then, I will particularly focus on Ian Quinn and Chris White's studies of corpus data-driven key-finding methods, particularly "Yale Classical Archive Corpus" (YCAC.) In YCAC, I will investigate two points, the minimum time constraints of musical slices as entities and general sequential harmonic tendencies as relationships of entities with each other. The first point will communicate with the minimum capacity of human memory to make meaning from the auditory stimulus and the second point will provide a general music technical data-driven consistencies that is another working mechanism of CSG in the next section. The first chapter's Memory base and the second chapter's music-technical data-driven consistencies will provide the constituent structure of the CSG.

The exploration of context-sensitive grammar about music perception reveals that harmonic functions are not static but rather dynamic elements that shift depending on the listener's memory spans and contextual factors. Constructing context-sensitive grammar in music involves recognizing the interplay between harmonic functions and the cognitive capacities of listeners. The literature indicates that memory spans, prior musical experience, and contextual factors significantly influence how harmonic functions are perceived and processed. Experienced listeners demonstrate a greater ability to navigate these shifts due to their extensive cognitive frameworks, while everyday listeners may rely more heavily on immediate contextual cues. This nuanced understanding of music perception highlights the importance of both cognitive and contextual factors in shaping how harmonic functions are experienced. (Dienes, 2004; Schmuckler & Tomovski, 2005; DePape et al., 2012; Gordon et al., 2014; Eerola et al., 2015; Ding et al., 2018; Saarikallio et al. 2020; Lad et al., 2021; Corrigall et al. 2013.) In chapter 3, after reviewing why a context-sensitive grammar might be a useful tool for providing some explanatory adequacy for the observations of corpus-driven musical entities and harmonic tendencies, I will introduce memory and melody-harmony technical tendencies base production rules of CSG to provide several depth in not only analysis of music structures but also why they are perceived in the way of elucidation of human memory and key finding studies.

In the last chapter, I will demonstrate the "Chroma Directed Initial and Final Tonic Assignment" which proposes that all tonic phrases are understood retrospectively through a feedback model, where their roles are shaped by changes in musical elements within the unfolding musical context. Subsequently, their tonic functions in cadential phrases are determined prospectively using a feedforward model. The duality of feedback and feedforward models provides a comprehensive understanding of tonic phrases in music. The retrospective feedback model allows for the interpretation of tonic phrases within their musical context, while the prospective feedforward model facilitates the anticipation of tonic functions in cadential phrases. This synthesis of models underscores the complexity of musical perception and the dynamic interplay of various musical elements that shape listener experiences. (Nada&Diedrichsen, 2012; Supèr et.al, 2010, 2011; Claire, 2018; Verosky, 2021.)

Chris White's "Feedback and Feedforward Models of Musical Key" (White, 2018) will be the main reference to support this point. According to White, there are two distinct approaches to understanding musical key perception: feedforward and feedback models. He defines feedforward models as those that determine key based on immediate surface features, such as pitch distributions or melodic intervals, without considering other musical parameters. In contrast, feedback models identify the key in conjunction with other musical elements like harmony and meter, creating an interdependent analytical process on a moment-to-moment basis.

In the present work, I will create a software program that harmonic accidentals will provide feedback to previous harmonic entities if they are retrospectively tonic phrases or not and this assignment will determine the role of the prospective harmonic entities that create an expectation with a musical closure, cadence or ending of the musical continuum. This program takes its basis from the CSG of the previous chapter and offers a potential music-predictive coding system in music-predictive information processing studies.

Another important characteristic of White's feedback model is the role of the musical meter in feedback models. In the present work, I will propose that a musical meter is just like a key that may not be given but constructed moment-to-moment durational changes with its content. Thus, with a similar strategy to key perception, the current split moment of the time will always be compared with the next musical time slice in an incremental way. If the last time slice is proportionate to previous accumulative chunks of time slices with values of basic or compound meters, or duple or triple meter, our expectation from the beat stays constant, if not current durations of the music will be divided by a coefficient that divides whole durations to equal discrete units whose value is adjustable. Then, I will look at Fast Fourier Transform amplitude levels in each time slice and depending on the output we will assign the meter of the current cursor time of the music.

Overall, my goal is to illustrate how integrating feedback and feedforward processes provides a more holistic view of musical key perception, capturing both retrospective and prospective functions. By examining these dynamic interactions within a context-sensitive grammar framework, I hope to contribute to a deeper understanding of how memory, expectancy, and evolving musical content shape listeners' experiences of tonal structures.

[References Chapter1]

Ginsborg, Jane. "Book Review: Music and Memory: An Introduction." *Musicae Scientiae* 6, no. 1 (2002): 101-105. https://doi.org/10.1177/102986490200600107.

Xu, Fang. "Short-Term Memory in EFL Listening Comprehension." *Asian Social Science* 4, no. 4 (2009). https://doi.org/10.5539/ass.v4n4p103.

Zanto, Theodore P., Vinith Johnson, Avery E. Ostrand, and Adam Gazzaley. "How Musical Rhythm Training Improves Short-Term Memory for Faces." *Proceedings of the National Academy of Sciences* 119, no. 41 (2022). https://doi.org/10.1073/pnas.2201655119.

Zendel, Benjamin R., and Claude Alain. "Musicians Experience Less Age-Related Decline in Central Auditory Processing." *Psychology and Aging* 27, no. 2 (2012): 410-417. https://doi.org/10.1037/a0024816.

Swaminathan, Swathi, and E. G. Schellenberg. "Musical Competence Is Predicted by Music Training, Cognitive Abilities, and Personality." *Scientific Reports* 8, no. 1 (2018). https://doi.org/10.1038/s41598-018-27571-2.

Saffran, Jenny R., Michelle M. Loman, and Rachel R. Robertson. "Infant Long-Term Memory for Music." *Annals of the New York Academy of Sciences* 930, no. 1 (2001): 397-400. https://doi.org/10.1111/j.1749-6632.2001.tb05751.x.

Strait, Dana L., and Nina Kraus. "Playing Music for a Smarter Ear: Cognitive, Perceptual and Neurobiological Evidence." *Music Perception* 29, no. 2 (2011): 133-146. https://doi.org/10.1525/mp.2011.29.2.133.

Pritz, Christian O., Eyal Itskovits, Eduard Bokman, Rotem Ruach, Vladimir Gritsenko, Tal Nelken, Mai Menasherof, Aharon Azulay, and Alon Zaslaver. "Principles for Coding Associative Memories in a Compact Neural Network." 2020. https://doi.org/10.1101/2020.06.20.162818.

[References Chapter2]

Kania, Dariusz, and Paulina Kania. Archives of Acoustics. 2019. https://doi.org/10.24425/aoa.2019.129260.

Krumhansl, C. L. The Cognitive Foundations of Musical Pitch. Oxford: Oxford University Press, 1990.

Krumhansl, C. L., and E. J. Kessler. "Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys." Psychological Review 89 (1982): 334–368.

Quinn, Ian. "What's 'Key for Key': A Theoretically Naive Key-Finding Model for Bach Chorales." Zeitschrift der Gesellschaft für Musiktheorie 7, no. 2 (2010): 151–163.

Schmuckler, Mark A., and Robert Tomovski. "Perceptual Tests of an Algorithm for Musical Key-Finding." Journal of Experimental Psychology: Human Perception & Performance 31, no. 5 (2005): 1124-1149. https://doi.org/10.1037/0096-1523.31.5.1124.

Temperley, David. "Bayesian Models of Musical Structure and Cognition." Musicae Scientiae 8, no. 2 (2004): 175-205. https://doi.org/10.1177/102986490400800204.

Temperley, David, and Elizabeth West Marvin. "Pitch-Class Distribution and the Identification of Key." Music Perception 25, no. 3 (2008): 193–212.

Toiviainen, Petri, and Carol L. Krumhansl. "Measuring and Modeling Real-Time Responses to Music: The Dynamics of Tonality Induction." Perception 32, no. 6 (2003): 741-766. https://doi.org/10.1068/p3312.

Vuvan, Dominique T., Olivia M. Podolak, and Mark A. Schmuckler. "Memory for Musical Tones: The Impact of Tonality and the Creation of False Memories." Frontiers in Psychology 5 (2014). https://doi.org/10.3389/fpsyg.2014.00582.

White, Christopher W. M., and Ian Quinn. "The Yale-Classical Archives Corpus." Empirical Musicology Review 11, no. 1 (2016): 50-58. https://doi.org/10.18061/emr.v11i1.4958.

——————White, Christopher W. M., and Ian Quinn. “Chord Context and Harmonic Function in Tonal Music.” *Music Theory Spectrum* 40, no. 2 (2018): 314–37. https://www.jstor.org/stable/90026032.

White, Christopher W. “Feedback and Feedforward Models of Musical Key.” Music Theory Online 24, no. 2 (2018a).

——————White, Christopher W., and Ian Quinn. “Chord Content and Harmonic Function in Tonal Music.” Music Theory Spectrum (2018b): 314–350.

——————White, Christopher W. The Music in the Data: Corpus Analysis, Music Analysis, and Tonal Traditions. Routledge, 2022.

[References Chapter3]

Corrigall, Kathleen A., E. G. Schellenberg, and Nicole M. Misura. "Music Training, Cognition, and Personality." Frontiers in Psychology 4 (2013). https://doi.org/10.3389/fpsyg.2013.00222.

DePape, Anne-Marie, Geoffrey B. Hall, Barbara Tillmann, and Laurel J. Trainor. "Auditory Processing in High-Functioning Adolescents With Autism Spectrum Disorder." PLOS One 7, no. 9 (2012): e44084. https://doi.org/10.1371/journal.pone.0044084.

Ding, Yongsheng, Kathleen Gray, Alexander D. Forrence, Xiaoqin Wang, and Juan Huang. "A Behavioral Study on Tonal Working Memory in Musicians and Non-Musicians." PLOS One 13, no. 8 (2018): e0201765. https://doi.org/10.1371/journal.pone.0201765.

Dienes, Zoltán. "Can Musical Transformations Be Implicitly Learned?" Cognitive Science 28, no. 4 (2004): 531-558. https://doi.org/10.1016/j.cogsci.2004.03.003.

Eerola, Tuomas, Henna-Riikka Peltola, and Jonna K. Vuoskoski. "Attitudes Toward Sad Music Are Related to Both Preferential and Contextual Strategies." Psychomusicology: Music, Mind, and Brain 25, no. 2 (2015): 116-123. https://doi.org/10.1037/pmu0000096.

Gordon, Reyna L., Carolyn M. Shivers, Elizabeth A. Wieland, Sonja A. Kotz, Paul J. Yoder, and J. D. McAuley. "Musical Rhythm Discrimination Explains Individual Differences in Grammar Skills in Children." Developmental Science 18, no. 4 (2014): 635-644. https://doi.org/10.1111/desc.12230.

Lad, Meher, Alexander J. Billig, Sukhbinder Kumar, and Timothy D. Griffiths. "A Specific Relationship Between Musical Sophistication and Auditory Working Memory." 2021. https://doi.org/10.1101/2021.07.08.451659.

Saarikallio, Suvi, William M. Randall, and Margarida Baltazar. "Music Listening for Supporting Adolescents' Sense of Agency in Daily Life." Frontiers in Psychology 10 (2020). https://doi.org/10.3389/fpsyg.2019.02911.

Schmuckler, Mark A., and Robert Tomovski. "Perceptual Tests of an Algorithm for Musical Key-Finding." Journal of Experimental Psychology: Human Perception & Performance 31, no. 5 (2005): 1124-1149. https://doi.org/10.1037/0096-1523.31.5.1124.

[References Chapter4]

Verosky, Niels. "Interpreting the Tonal Hierarchy Through Corpus Analysis." Psychomusicology: Music, Mind, and Brain 31, no. 2 (2021): 96-106. https://doi.org/10.1037/pmu0000276.

Yousif, Nada, and Jörn Diedrichsen. "Structural Learning in Feedforward and Feedback Control." Journal of Neurophysiology 108, no. 9 (2012): 2373-2382. https://doi.org/10.1152/jn.00315.2012.

Arthur, Claire. "A Perceptual Study of Scale-Degree Qualia in Context." Music Perception 35, no. 3 (2018): 295-314. https://doi.org/10.1525/mp.2018.35.3.295.

Supèr, Hans, and August Romeo. "Feedback Enhances Feedforward Figure-Ground Segmentation by Changing Firing Mode." PLOS One 6, no. 6 (2011): e21641. https://doi.org/10.1371/journal.pone.0021641.

Supèr, Hans, August Romeo, and Matthias Keil. "Feed-Forward Segmentation of Figure-Ground and Assignment of Border-Ownership." PLOS One 5, no. 5 (2010): e10705. https://doi.org/10.1371/journal.pone.0010705.

White, Christopher W. "Feedback and Feedforward Models of Musical Key." Music Theory Online 24, no. 2 (2018).