



# Developing Musical Imagery: Contributions from Pedagogy and Cognitive Science \*

*Sarah Gates*

NOTE: The examples for the (text-only) PDF version of this item are available online at:

<https://www.mtosmt.org/issues/mto.21.27.2/mto.21.27.2.gates.php>

KEYWORDS: musical imagery, pedagogy, cognition, expertise acquisition, memory

ABSTRACT: Research into the development of musical imagery ability has remained stagnant in both the fields of aural skills pedagogy and cognitive science. This article integrates scholarship from both disciplines to provide a way forward for both the study and practice of imagery development. Analysis of North American pedagogical practices provides a foundation for the types and functions of activities used to affect imagery ability, while newly designed measurement techniques in the cognitive sciences are shown to have promising implications for assessing change in imagery ability over time. Following consideration of insights from both fields, this article consolidates them by developing a model of imagery development. Framed through the lens of expertise acquisition and skilled memory performance, this model has implications for approaches to imagery in the aural skills classroom and for empirical studies of imagery development in music cognition.

DOI: 10.30535/mto.27.2.3

*Received January 2020*

Volume 27, Number 2, June 2021  
Copyright © 2021 Society for Music Theory

---

[1.1] Consider hearing internally the series of pitches shown in **Example 1**. This pattern, strongly situated in your mind's ear, is dependent on (and has been generated through) a host of different knowledge structures.<sup>(1)</sup> Take, for example, the key context. For those enculturated in Western art music, perhaps it manifests in C major as scale degrees  $\hat{1}-\hat{6}-\hat{4}-\hat{5}$  with a harmonization similar to that shown in **Example 2**. There is another candidate for key, however: F major, with scale degrees  $\hat{3}-\hat{3}-\hat{1}-\hat{2}$ . If you generated this pattern in F major, it may have a similar quality to the melody and harmonization shown in **Example 3**. The form your imagery first took reveals crucial aspects about the types of knowledge structures you possess from your own unique musical experiences and background. Viewed in this way, imagery is tightly coupled to past experience, long-term memory, and habit.<sup>(2)</sup>

[1.2] Nearly every musician has an intuitive sense about the importance of inner hearing for their craft. This imagining of musical content, or *musical imagery*, is a vital and seemingly ubiquitous activity for musicians of every discipline, be they performers imagining memorized pieces,

composers determining how to orchestrate their work, or theorists contemplating some aspect of a passage they are analyzing. Yet despite broad agreement about the importance of musical imagery, relatively little attention is typically directed toward it in pedagogical practice. Several pedagogues have noted this issue within the field of aural skills pedagogy (Klonoski 1998; Covington 2005), with some providing what they suggest are “more direct” methods of accessing and developing the skill rather than relying exclusively on traditional sight singing and dictation exercises (see Larson 1993; Klonoski 1998). Despite such efforts, these more direct methods have not been adopted in mainstream pedagogical practice, nor has their effectiveness for developing imagery been verified by experimental research.

[1.3] This article integrates scholarship from the fields of cognitive science and collegiate-level North American aural skills pedagogy to promote advancement in both the study and practice of imagery development. First, I examine the cognitive functions of the commonly discussed methodologies in North American aural skills scholarship for building imagery abilities. I explore research findings and practices in the cognitive sciences, with a focus on newly designed measurement techniques for auditory imagery. Of particular interest is the Bucknell Auditory Imagery Scale (hereafter BAIS), which has robust capabilities to predict outcomes on objective behavioral tasks on the basis of subjective imagery properties (Halpern 2015). To illustrate this, I present preliminary findings from a longitudinal pilot study using the BAIS to observe imagery development over the course of first-year aural skills. Following that, all previously noted pedagogical and cognitive insights are integrated using an adaptation of Ericsson and Kintsch’s (1995) *long-term working memory* model (LTWM), which reframes imagery development in the aural skills classroom as expertise acquisition. This model maps the aural skills methodologies discussed in the first section and the subjective measurement properties discussed in the second section onto a developmental trajectory to provide clearer, and importantly, more measurable outcomes for imagery expertise acquisition within the North American aural skills tradition. It is my hope that this proposed model will be used to reframe pedagogical approaches to imagery in the aural skills classroom and help to generate testable hypotheses for future empirical studies of imagery development in music cognition research.

### *Imagery Development in North American Music Pedagogy Scholarship: What is Imagery, and What Improves with Practice?*

[2.1] An effective approach to developing musical imagery requires focused pedagogical goals and techniques. Unfortunately, within the field of music pedagogy there is little consensus on what musical imagery is, let alone the best ways to develop it. This condition is perhaps most clearly demonstrated by the wide variety of ways in which scholars refer to musical imagery: “audiate/hear as” (Larson 1993), “internalization,” “inner singing/subvocalization,” and “inner hearing” (Klonoski 1998), “aural imagery,” “hearing eye,” and “mind singing” (Benward 1989), “musical thinking” and “auralize” (Karpinski 2000), “auditory imagery” (Covington 2005), “audiation” (Gordon 2012), “musical imagination” and “mind’s ear” (Adolphe 2013). While these terms may refer to slightly different aspects of related phenomena, they recognize the presence of some internal, introspectively available means of sound generation. This diversity of terminology also reflects the complex reality of musical imagery, namely that the phenomenon involves a host of interconnected cognitive skills.

[2.2] Pedagogical approaches to imagery development also vary widely. My analysis reveals that the diverse set of methodologies discussed by scholars can be divided into categories on the basis of four broad cognitive functions, which are listed below. The format of the list, which has been designed to facilitate discussion, does not reflect the fact that in practice significant overlaps between categories will occur.

1. Promoting the acquisition (long-term memory storage) of imaging content.
2. Encouraging semantic encoding of stored musical content in order to affect the *quality* of imagined sound.
3. Developing methods for cueing or generating imagery (including singing, subvocalization and other multimodal associations, like notational audiation).

4. Fostering expertise acquisition, including the freeing of “mental space” for multitasking and increased cognition during imaging, as well as increased metacognitive access to imagery.

[2.3] The first category, content acquisition, is by far the most prevalent within North American music pedagogy and forms the basis for many different methodologies. Most textbooks, remarkably, do not reference imagery development as a stated goal at all. Instead, they embrace a content-acquisition method implicitly, making frequent reference to the acquisition of common and meaningful patterns through exposure to Western tonal repertoire.<sup>(3)</sup> These schemata, which essentially are basic building blocks of musical information and sound structures, are meant to be stored in students’ long-term memory and relied upon as the basis for more advanced skills. The foundational importance of content acquisition is evidenced by practices associated with childhood music education, such as the focus on pattern acquisition in Gordon’s (2012) audiation model and the similar “child-development” model in the Kodály method (Choksy 1974, 12).

[2.4] The second category encourages semantic (meaning-based) encoding of musical content stored in long-term memory in order to affect the quality of internal sound generation. This is achieved primarily by developing connections between various statistically relevant schemata stored in memory. These connections help infuse internally generated musical content with meaning.<sup>(4)</sup> The most common forms of these techniques address tonal awareness, scale-degree hearing, functional hearing, and the like. Once these acquired musical structures are semantically encoded into categories and connected in memory, the internal quality of musical images is affected as larger memory networks become activated. Similarly, semantic encoding affects how external sounds are perceived, because these stored schemata become activated as images during engaged listening. These activated images therefore facilitate perception by acting as an internally generated perceptual prime, biasing incoming information and attention (Hubbard 2010, 319).

[2.5] This second category is distinct from the first in that implicit exposure to relevant musical schemata does not ensure semantic long-term encoding and retrieval in imagery. When connections between schemata are not explicitly built into memory, students may have the ability to form an internal image of a heard stimulus but only limited ability to imbue it with qualitative *feel*. Larson (1993) demonstrates this in his discussion of scale-degree function acquisition. Here he separates the processes of inner hearing and perceiving scale-degree function into two distinct but related phenomena. Larson uses the term *audiation* to refer to the inner hearing of internal sounds, and the term *hearing as* to refer to the process of providing meaning to a sound by subconsciously assigning it to a category (Larson 1993, 70). Developing this latter skill, Larson argues, requires an explicit focus on building functional scale-degree “hearing.”<sup>(5)</sup> Developing the sense of *quality* of imagined sound therefore relies on building large, interconnected networks of musical knowledge. Traditionally, these include gaining familiarity with Western tonal space, tightly associating scale degrees and their functional chord associations, and learning common patterns and their uses in tonal repertoire.<sup>(6)</sup> Once these relations have been deeply internalized (i.e., stored and connected in long-term memory), these sensitivities manifest themselves in imagined hearing, providing what we might perceive for example as scale-degree qualia, or “the threeness of  $\hat{3}$ ” (Karpinski 2000, 51).<sup>(7)</sup>

[2.6] The third pedagogical approach to image development enfolds a wide variety of tasks and techniques that serve as cues meant to promote internally generated sound. Most commonly, singing/subvocalization is addressed through sight singing, while notation mappings are encouraged through dictation exercises (see Johnson and Klonoski 2003; Klonoski 2006). Similarly, the acquisition of a solmization system can aid pitch cueing, as once the sound/symbol associations have been deeply learned, simply bringing to mind the symbol (e.g., thinking “ $\hat{7}$ ”) calls to mind aspects of the learned sound/imagistic content associated with it.<sup>(8)</sup> The explicit use of multimodal associations, such as imagined playing, is less prominent in traditional undergraduate curricula, in part perhaps because it is assumed that most students already possess pitch-motor mapping abilities on their primary instrument. That said, curricula with an emphasis on piano playing in the aural skills classroom do attempt to broaden the multimodal networks of sound to action mappings, especially for those who may lack experience with more tactile/motoric pitch-to-movement associations.<sup>(9)</sup> Other approaches, including the Dalcroze method, expressly employ

movement based strategies and metaphors to develop cross-modal connections that support imagery generation and that can act as cues for imagined sound (see [Urista 2016](#); [Godøy 2003, 2004](#)). While these methods encompass a wide range of different activities, they all function to support imagery cueing and maintenance.

[2.7] The final category of expertise acquisition focuses on expanding and enriching the mental space required for imagery in order to improve multitasking, increased cognition and metacognitive awareness. The approaches of Edwin E. Gordon (*audiation*) and Gary Karpinski (*auralizing*) most clearly articulate this philosophy. Both approaches highlight the importance of the skills and goals discussed previously (i.e., content acquisition, semantic encoding, and cueing techniques), with further consideration given to automation and the freeing of mental resources during imaging. While each term involves imagery as a central feature, both scholars note that their particular concept encompasses more than imagery alone.<sup>(10)</sup>

[2.8] Gordon's term "audiation" refers to conscious awareness of meaning in a wide variety of musical situations and activities. For example, Gordon notes that a jazz drummer can audiate a melody while soloing, and a conductor can audiate patterns while conducting. The importance of conscious awareness to audiation is also evident since "Fine musicians know when they are audiating: it occurs when ears become more important than fingers and arms" ([Gordon 2012](#), 6). This suggests that musicians who can audiate not only have the capacity to engage in their craft, but also possess sufficient cognitive fluency to multitask. While playing and/or hearing music, they may consciously engage with music through predicting, completing, and understanding what they are hearing and/or playing. Or they may choose to imagine something else entirely. Gordon's types and stages of audiation more directly reference expertise acquisition by proposing a developmental trajectory in which abilities build off one another and increase in complexity over time. Advanced musicians, for example, are able to recognize learned musical patterns while actively listening, and can multitask even under severe real-time processing constraints, as when improvising ([Gordon 2012](#), 19–23). Gordon's audiation skill, once fully acquired, is an all-encompassing, imagery-dependent musical ability that can be deployed concurrently during various types of activities.

[2.9] Karpinski's term *auralize* has some similar features to Gordon's *audiation*, but focuses more narrowly on conscious comprehension of musical meaning, and only implicitly refers to expertise and multitasking ability. Karpinski defines "thinking in music," as follows: "Music listeners who understand what they hear are thinking in music. Music readers who understand and auralize what they read are thinking in music" (2000, 4). In this way, "musical understanding" is tied to the ability to actively make sense of musical structures. These structures, once deeply learned and automated through expertise, free up mental space for actively reflecting on internal experiences, providing a more conscious form of understanding.<sup>(11)</sup> Framed in this manner, to possess musical understanding is to consciously grasp functionality—to be aware of what musical structures are and how they function. Such understanding is most often associated with the ability to identify and apply a theoretical label to a given event (e.g., a metric grouping, set of scale degrees, Roman numeral, etc.). The labeling ability is itself not the *goal* of acquired expertise; it is simply representative of it. If sound-to-meaning mappings have been deeply ingrained, then the application of the label (e.g.,  $\hat{7}$ ) stems from the internal sense of musical meaning, rather than from a method of calculation (e.g., with interval distances). This internal sense becomes linked with several other types of musical knowledge and skills, so that it can, similarly to Gordon, be fluently employed during many different activities (e.g., conducting, performance, notation reading, etc.).

[2.10] Taken together, Gordon's and Karpinski's conceptions of imagery development represent the acquired expertise acquisition resulting from mastery of the categories discussed previously, including content acquisition, semantic encoding, and imagery cueing techniques. Beyond this, their approaches highlight important markers of expertise acquisition, including expanded mental resources for metacognition and the presence of reciprocal interconnections among the many skills that broadly encompass imagery ability.<sup>(12)</sup> As these skills are acquired, related, and automated through practice, more mental space and attentional resources become available for further cognition and metacognitive reflection. Our musical actions become permeable and observable

from the inside and are infused with more sonic life and meaning. In turn, we can better hone and access our individual, internal musical worlds.

## *Musical Imagery in the Cognitive Sciences: Research Trends and Measurement Techniques*

[3.1] The understanding of musical imagery is more standardized in the cognitive sciences. There, it is considered a music-specific application of the broader ability of “auditory imagery,” defined as “the introspective persistence of an auditory experience, including one constructed from components drawn from long-term memory, in the absence of direct sensory instigation of that experience” (Hubbard 2010, 302).<sup>(13)</sup> Experimental researchers have struggled to develop accurate measurement techniques, since imagery cannot be directly observed. The advent of neurological imaging (EEG, MEG, fMRI) has greatly aided in the experimental observation of imagery. These techniques, however, are still relatively indirect, as they depend on behavioral task constructions that must actively recruit imagery as opposed to other forms of representations (Hubbard 2010, 302).<sup>(14)</sup> While cognitive science has yet to explore any developmental aspects of musical imagery, some experimental findings can be enlisted to support assertions made about musical imagery in the pedagogical domain.

[3.2] Research on musical imagery has confirmed many pedagogues’ claims concerning imagery. This includes:

- The important role of imagery in accurate singing (Pfordresher and Halpern 2013; Pfordresher and Mantell 2014; Pfordresher, Halpern, and Greenspon 2015; Greenspon, Pfordresher, and Halpern 2017)
- The involvement of multimodal support networks and subvocalization in particular during imagery (Davidson-Kelly et al. 2015; Smith, Reisberg, and Wilson 1992; Kalakoski 2001; Hubbard 2013)
- Imagery as improving auditory pitch acuity and expectation (Janata and Paroo 2006; Navarro and Janata 2010)
- The importance of imagery for performance preparation and memorization (Keller 2012; Sainilan 2014; Highben and Palmer 2004; Brown and Palmer 2012)
- The use of imagery during notational audiation (Brodsky et al. 2003, 2008)
- The tight coupling of tonal pitch (scale degree) imagery with long-term representations of the tonal hierarchy in musicians trained within Western traditions (Vuvan and Schmuckler 2011)

[3.3] These findings support traditional pedagogical practices that focus on tonal pitch imagery and rely on a large support network of multimodal skills—primarily singing and piano proficiency—to build, access, and maintain imagery. While helpful, few studies to date have explored imagery ability between differing populations or examined expertise-related imagery acquisition from a developmental perspective.<sup>(15)</sup> Another confound of current findings is that standardized measurement tools have yet to be adopted, making the comparison of results and exploration of implications difficult. Recently, a more reliable psychometric has been developed by cognitive psychologist Andrea Halpern (2015) that shows promising results for imagery measurement. The upcoming section will introduce and detail the subjective/objective measurement approach used in the cognitive sciences and discuss how it facilitates the observation of differences in imaging ability between individuals and groups.

## *Subjective/Objective Measurements Paradigms: Detecting Effects of Training and Individual Differences in Imagery*

[4.1] Research in the comparatively larger field of visual imagery has traditionally employed a dual approach to measurement. This methodology, which has been adapted to the auditory realm, entails using a subjective measure in conjunction with an objective measure for increased predictive power and accuracy. The subjective measurement typically takes the form of a psychometric questionnaire in which participants are prompted to generate some form of imagery

from a verbal description and then rate its properties on a Likert scale. The predictive validity and reliability of these types of subjective measures has historically been inconsistent within the visual domain, leading many scholars to conclude that they have poor predictive validity in general (McAvinue and Robertson 2007, 196). More recently, the development of questionnaires has been informed by theoretical scholarship on imagery ability, resulting in more robust relationships between subjective and objective measures. Questionnaires now often divide subjective imagery scales into subtypes, such as spatial imagery and object imagery, which have been found to correlate with specific objective tasks involving those imagery processes (e.g., mental rotation and degraded image tasks, respectively; see McAvinue and Robertson 2007, 204–205). Therefore, the more accurately a subjective measure captures the imagery properties relevant to completing a specific objective task, the better that measure is for predicting objective performance.

[4.2] The Bucknell Auditory Imagery Scale (or BAIS), a reliable, verified subjective measure for auditory imagery, was developed by cognitive psychologist Andrea Halpern (2015). This measure is designed to capture two distinct types of processing in auditory imagery: generation, the ability to bring an image to mind, and transformation, the ability to make changes to a generated image. These processes are measured using relevant qualitative features of imagery, specifically *vividness* (for generation) and *control* (for transformation). The BAIS measures auditory imagery broadly, and includes environmental, speech and musical items. The two subscales of vividness and control have been shown to have several behavioral and neural correlates (Halpern 2015), providing predictive power for various objective tasks that rely to varying extents on generation and transformation (Halpern and Overy 2019).

#### Measuring Image Generation: The BAIS Vividness Subscale

[4.3] The vividness rating scale measures image generation ability by prompting individuals to subjectively judge how life-like their imagery is. In the questionnaire, a verbal cue is presented to subjects that asks them to construct an auditory scene; they are then instructed to rate the vividness of their auditory imagery on a scale from 1 (no image present at all) to 7 (as vivid as real sound); see **Example 4**. The vividness subscale on the BAIS has been highly correlated with several behavioral and neurological measures: high vividness scores have been shown to predict accurate singing of individual pitches (Pfordresher and Halpern 2013) and longer melodic sequences (Greenspon, Pfordresher, and Halpern 2017), as well as memory for previously heard melodies (Herholz, Halpern, and Zatorre 2012). Vividness has also predicted the explicit use of imagery strategies for scale-degree imagery in the Pitch Arrow Imagery Task (Gelding, Thompson, and Johnson 2015)<sup>(16)</sup> and levels of involuntary musical imagery or earworms experienced by individuals (Floridou et al. 2015). Recently, the vividness subscale was shown to predict the ability to perceive expressive timing patterns (Colley, Keller, and Halpern 2018). Performance on the vividness subscale has also been correlated with increased brain activation during the encoding of imagined melodies (Herholz, Halpern, and Zatorre 2012), increased activity during melody reversal (Zatorre, Halpern, and Bouffard 2009), and increased grey matter volume in the brain's supplementary motor area (SMA) implicated in subvocalization (Lima et al. 2015).

#### Measuring Image Transformation: The BAIS Control Subscale

[4.4] Auditory imagery control measures subjects' capabilities for manipulating and making changes to their musical imagery. In the control subscale, a verbal cue instructs subjects to construct an auditory scene, after which they are prompted with another cue to make some sort of change to the generated image. They are then asked to rate how easy it was to alter their auditory imagery on a scale from 1 (no image present at all) to 7 (extremely easy to make the change); see **Example 5**. Compared to the vividness subscale, the control subscale has shown slightly less predictive power, which is likely due to the fact that imagery control is more complex and relies on a wider range of related cognitive mechanisms, such as working memory (WM). This subscale has, nevertheless, been shown to predict better performance on the Pitch Imagery Arrow Task (Gelding, Thompson, and Johnson 2015). It has also been correlated with lower error rates in the imitation of pitch sequences (Greenspon, Pfordresher, and Halpern 2017) and was predictive of the ability for

musically untrained subjects to synchronize with expressive music (Colley, Keller, and Halpern 2018).

### Individual Differences and Effects of Expertise

[4.5] The BAIS represents a compelling potential resource for accurate measurement of imagery for developmental purposes; however, it has only shown to loosely correlate with musical training (Halpern 2015). This raises several important questions. First, is the measure adequate for detecting changes in musical imagery brought about through formal instruction? Secondly, and perhaps more importantly, is imagery a learnable skill, or a trait with a biologically determined capacity? (17) There is some research suggesting that musicians have an imagery advantage compared to untrained populations. For example, using both a musical and a non-musical auditory imagery task, Aleman et al. (2000) found that musicians outperformed nonmusicians on auditory imagery tasks, but not on a visual imagery task. Differences between musicians and nonmusicians have also been observed in the characteristics of their involuntary auditory imagery, or earworms. Those with musical training have been shown to experience more involuntary imagery on average than nonmusicians (Beaty et al. 2013), and they also experience earworms with more fidelity and complexity (e.g., music with multiple parts; see Williamson and Jilka 2014). Musicians have also reported earworms that lead to experiencing visual and motor sensations related to the act of playing (e.g., imagined piano playing, staff reading, etc.; see Williamson and Jilka 2014, 661). This research suggests that musical training has the potential to affect the quality, complexity, and functioning of auditory imagery in individuals. Yet until developmental approaches can be used to study musical imagery acquisition, it will remain impossible to determine the extent to which these differences result from pre-existing differences versus formal training.

#### Measuring Imagery Development in the Aural Skills Classroom: A Pilot Study

[4.6] To examine directly the impact of training on music imagery acquisition, I conducted a pilot study using the subjective/objective measurement paradigm. Using the BAIS and a previously verified objective-task correlate of the BAIS—a melodic reversal task from Zatorre, Halpern, and Bouffard (2009)—I measured first-year university students' imagery abilities in the fall and again in the spring, near the end of their first year of aural skills training. As a control measure for comparison, I also tested students who were exempted from the first-year aural skills sequence but who were still enrolled in all other required music courses. These students were included to help determine whether observed changes in imaging ability resulted from aural skills training (in the experimental group) or from more generalized musical training (in the control group). No measurable change was observed in auditory imagery scores on either the BAIS or the melodic reversal task for those enrolled in first-year aural skills. Those exempted from aural skills, however, had higher scores overall, and showed improvement on both the melodic reversal and BAIS scores between the pre- and post-tests.

[4.7] These tentative results could indicate several things. First, a generalized auditory imagery measure like the BAIS may not be suitable for musical imagery measurement in the aural skills classroom. If this is the case, however, the methodology could easily be adjusted, as I will demonstrate shortly. Second, improvements in imagery ability may not be observable within the relatively small timeframe of one year of aural skills training. Last, these results support the observation about the reciprocal nature of imagery development made by Gordon, Karpinski, and Rogers. To wit: students whose imagery abilities are already developed enough to permit their exemption from first-year aural skills may already be able to actively recruit their imagery in other musical endeavors, resulting in improved imagery ability over the seven months of first year, general training. Those with more limited imaging abilities are less likely to use imagery in complex musical tasks, resulting in less growth over the same time period. While more testing is needed to investigate these claims, these preliminary findings will inform the model for imagery development proposed below.

### *Bridging the Divide: Expertise and Skilled Memory Performance*

[5.1] While research has yet to establish whether and how musical imagery can be developed and improved, another field, that of expertise acquisition, offers profound insights on how to reframe the processes, goals, and outcomes of advanced aural skills training. Anders Ericsson's (2018) work on memory expertise, including the LTWM model (Ericsson and Kintsch 1995) allows for a greater understanding of what developing imagery skill may entail. The LTWM model evolved out of extensive research into expert performance in domains such as chess and memorization<sup>(18)</sup> to explain how experts were able to bypass known memory constraints.<sup>(19)</sup> The field of expertise acquisition has since grown to explore many other domains, such as memorized musical performance, medical diagnosis, and sports. Ericsson's expert performance and LTWM model proposes that acquiring an expert level of performance in a given discipline entails developing integrated domain-specific memory skill and domain-specific expertise.

[5.2] The LTWM model claims that every expert, through many hours of deliberate practice, acquires a set of knowledge structures and related set of retrieval cues that forms the basis of their acquired memory skill (Ericsson and Kintsch 1995). This memory skill is comprised of three components. The first component is the large body of domain-specific knowledge that experts acquire; specifically, the deeply encoded knowledge structures stored in semantic (long-term) memory. The second component is a set of retrieval cues that are explicitly associated with knowledge in the form of a stable hierarchical structure called a retrieval structure (see **Example 6**).<sup>(20)</sup> The last component of acquired LTWM skill is that the process of encoding and retrieving information using a retrieval structure is cultivated and rapidly sped up with deliberate practice, eventually making the rate of information storage and retrieval from LTM comparable to that of WM (Ericsson 1985, 194).<sup>(21)</sup>

[5.3] Another important feature of acquired memory skills, including LTWM, is that the mechanisms are extremely domain-specific and are tailored to meet the demands imposed by a specific task or set of tasks on storage and retrieval of information (Ericsson and Roring 2007). Therefore, the specificities of an expert's acquired memory skill (schema, retrieval structure) will be drastically different based on their domain of expertise and will also present with a set of processing benefits and deficits. For example, while some domains have a high demand for retrieval accuracy (e.g., memorized musical performance or delivering a script/speech), others put more of a premium on the encoding and processing of new information so as to adjust potential future actions (e.g., chess, sports, medical diagnosis).

[5.4] Recent neurological imaging work on the acquisition of expertise has verified much of Ericsson and Kintsch's LTWM model, including the developmental trajectory implied by said model. Guida et al. (2012) completed a meta-analysis of experimental neurological data comparing novices to experts in various domains. Their findings reveal a developmental trajectory of functional brain reorganization gained with expertise that supports Ericsson's claims. They also suggest a gradual two-step process in the acquisition of expertise and related brain changes (Guida et al. 2012, 221–44). The first stage of expertise acquisition in Ericsson's model includes chunk creation and storage of relevant information in long-term memory. Once chunks become stored in LTM, they then become bound with practice; that is, relationships among stimuli are formed (i.e., building associated knowledge) and then become hierarchically organized.<sup>(22)</sup> As novices gain relevant knowledge structures in LTM that are more efficiently processed, working memory becomes less burdened, resulting in a reduction of brain activity during more skilled performance (Guida et al. 2012, 235–236). The second phase of Ericsson's LTWM model is the speeding up of memory skill with practice, resulting in long-term retrieval being nearly equal in speed to that of working memory. This second phase, which occurs with expertise acquisition, results in functional reorganization of the brain (Guida et al. 2012, 236). Such functional reorganization indicates that the retrieval structures in LTWM are fully acquired, allowing experts to actively use part of LTM as WM (or LTWM).

[5.5] It is important to note that Ericsson's work—his deliberate practice research in particular—has recently been criticized by scholars investigating factors important for expertise acquisition. A series of meta-analytical studies has revealed that deliberate practice only accounts for approximately 20–30% of the variance in performance, indicating that other factors—notably non-

environmental ones, such as starting age, intelligence, personality, and genetics—are also needed to explain expert performance (see Hambrick et al. 2014a; Macnamara, Hambrick, and Oswald 2014).<sup>(23)</sup> To account for this explanatory gap, researchers in Hambrick's lab have proposed a multifactorial gene-environment interaction framework in order to examine the contribution of multiple factors to expertise (Ullén, Hambrick, and Mosing 2016; Hambrick et al. 2018). Their model positions individual differences in expertise as stemming from an interaction of genetic and environmental factors, including both domain-general traits and domain-specific knowledge, which can influence expertise both indirectly and directly (Hambrick et al. 2018, 291). From this perspective, deliberate practice is required for maximizing potential and acquiring domain-specific skills; however, it is not the sole, nor potentially the most important factor in expertise acquisition.  
(24)

[5.6] What remains uncertain is the extent to which non-environmental factors play a role in expertise after the initial stages of learning. While there is strong evidence to suggest the importance of domain-general abilities such as general intelligence and working memory at the beginning stages of learning, the role of the same abilities at expert levels remains unclear (Hambrick, Burgoyne, and Oswald 2019). Similarly, evidence supporting the circumvention-of-limits hypothesis as promoted by the LTWM model—which holds that that cognitive limitations may be overcome by extensive practice—remains mixed.<sup>(25)</sup> This condition suggests that the degree to which experts may be able to circumvent limitations imposed by domain-general abilities may be contingent on the domain itself, or even on the given task within that domain. Hambrick, Altmann and Burgoyne recommend that future work in this area build theoretical frameworks capable of making testable predictions about when cognitive ability factors should affect performance and when they should not (2018, 320). Doing so would entail using current theories and methodologies to identify potential predictors of expertise and the cognitive mechanisms underlying performance—especially as these may differ across domains and different types of tasks (Hambrick, Burgoyne, and Oswald 2019). Part of the difficulty in teasing apart contributions to expertise lies in the challenges of expertise studies in general.<sup>(26)</sup> The task is further complicated by the fragmentary nature of methodologies, tasks, and domains studied. Therefore, it is of the utmost importance, moving forward, that researchers more clearly define the expertise they are examining, understand the components of the tasks they are analyzing, and theorize the various contributing factors—both environmental and nonenvironmental—that contribute to said expertise.

[5.7] This project represents the beginning phase of such research in musical imagery. By building off existing frameworks such as LTWM, it is possible to conceptualize those domain-specific mechanisms that may be susceptible to change with training interventions, while also theorizing the contributing roles for domain-general abilities (e.g., WM capacity).<sup>(27)</sup> I will speak more to these issues in the conclusion and future directions section. For the time being, I will discuss those domain-specific features of imagery expertise that may be susceptible to change. As I will show, the LTWM framework allows for a modelling of imagery expertise according to current pedagogical practices.

### *Rethinking Musical Imagery Development in Pedagogy and Cognitive Science*

[6.1] By integrating scholarship from the fields of North American aural skills pedagogy, musical imagery, and expertise acquisition, it is possible to construct a more nuanced and insightful model of imagery development in higher education. Reframed using the LTWM model, I position the development of musical imagery as the acquisition of expert memory skill. This entails the development of an integrated network of skills, including content formation and connections in LT (semantic) memory, retrieval structures for that information, and increased memory skill (LTWM) for speeding up retrieval, encoding and storage. It is further possible to use imagery processes (generation and transformation) and their relevant properties (vividness and control) to determine both what imagery subcomponents might be affected through expertise acquisition and what related neurological changes occur with this expertise as proposed by Guida et al. (2012).

[6.2] **Example 7** illustrates my proposed mapping of the commonly employed areas of imagery development in aural skills pedagogy—content acquisition, imagery quality, methods of image generation, and Gordon and Karpinski’s notions of imagery expertise—onto the three stages of the LTWM model. The result is a developmental trajectory of imagery expertise, in which each of the pedagogical activities has a distinct function within an expertise framework. The focus on content acquisition entails the formation of relevant and accessible chunks in long-term memory. Developing imagery quality through tonal awareness, for example, aids in semantic encoding of relevant schemata by forming associations between items stored in long-term memory, which in turn boosts their qualitative content and meaningfulness. Cultivating different ways of producing imagery, such as singing, subvocalizing, or solmization skills, provides cues for retrieval and imagery maintenance within a retrieval structure. The final category discussed by Karpinski and Gordon reflects the acquired memory skill. This expertise acquisition frees up attentional resources for reasons already discussed: unburdening WM during performance facilitates multitasking, increased cognition, and metacognition. Each of these LTWM functions (chunk formation, semantic encoding, retrieval cue formation, and the speeding up of memory skill) maps onto stages of brain reorganization proposed by Guida et al. (2012). A gradual reduction in brain activity occurs first, as structures and retrieval cues are acquired and made more efficient. The reduction is followed by functional reorganization with LTWM skill acquisition (expertise acquired).

[6.3] I propose that each of these expertise stages reflects the development of different imaging processes (generation and transformation) and their relevant subjective properties (vividness and control) as described in the BAIS (Halpern 2015). Both the acquisition and semantic encoding of content chunks as well as the development of more diverse and reliable ways of cuing imagery increase image generation ability. As musical structures (schemata) become increasingly semantically encoded in long-term memory, they pull more associated knowledge with them when retrieved, providing a boost to the quality (vividness) of internally generated images. For example, the more an individual associates scale degrees with their functional harmonic usages in LTM, the more likely those harmonic implications will be automatically cued when imagining or perceiving that scale degree, affording traditional notions of Western functional scale-degree qualia. Perceived vividness may also increase as images become more readily available in their schematic forms (as chunks). As the association between the retrieval cue and image becomes more stable, more rapid access will ensue. Conversely, the acquisition of LTWM helps to aid in image transformation ability, or control of musical imagery. Through aid of increased attentional resources, we gain the ability to more easily and flexibly cue, control, and modify generated musical images, essentially gaining LTWM skill. These hypotheses are supported by current research in musical imagery using the BAIS, which shows that perceived vividness is more related to the formation of a detailed sound image (i.e., more related to long-term memory), whereas imagery control is more related to working memory ability and action planning (Colley, Keller, and Halpern 2018).

[6.4] The new framework allows us to hypothesize how a musical task that involves skilled imagery use is modeled. Take, for example, the exercise of sight singing. Let us assume that one begins preparation for this task by silently reading a score excerpt. The retrieval structure of a hypothetical expert using movable-*do* solfège might look something like that given in **Example 8**. The explicit set of retrieval cues is represented hierarchically. The lowest level of retrieval cue takes the form of the solfège symbols situated within the frame/grid of the  $\frac{6}{8}$  compound meter.<sup>(28)</sup> The higher-level retrieval cues involve chunking and abstraction, notably the explicit recognition of tonic and dominant functions in each bar, and of how these are expanded with arpeggiations, neighbor tones, and passing tones. The resulting “Encoded Information” retrieved using these cues is meant to represent the musical image.<sup>(29)</sup>

[6.5] Retrieval structures may also vary as a function of experts’ intentions, goals, and the specific tasks to be accomplished. Say, for example, that instead of audiating Example 8 with the intention of sight singing it in solfège, one is planning to perform the excerpt at sight on their primary instrument. While one might opt for an initial silent reading of the score using solfège to firmly grasp the pitch structure, it is likely that a performer would rely more on fluent pitch-motor mappings to generate the intended performance. The retrieval structure would instead include the

motor movements at the lowest level, rather than solfège. It is highly likely, moreover, that the internally sounding image would differ in quality.

[6.6] Personally, when I imagine this excerpt in solfège, the imagined pitches have a more vivid, robust, and centered quality to them. In contrast, when I imagine performing this excerpt on my primary instrument, the pitch structure is less vivid but the available tempo is much faster. I am also much more consciously aware of concurrent motor imagery, which affects the quality of the internally generated sound.<sup>(30)</sup> This thought experiment demonstrates that there are slightly different retrieval structures for these tasks. Each activates different associated knowledge structures, resulting in perceptible differences in the activated imagistic content.

[6.7] The LTWM model also offers a new perspective on the earlier-reported pilot study results with regard to including Gordon, Karpinski, and Roger's claims about the reciprocal nature of imagery development. Generating and maintaining a detailed and vivid image in working memory takes an extraordinary amount of mental resources (Halpern and Overy 2019, 394). Those who lack facile access to information stored in long-term memory or those who have yet to explicitly and deeply encode Western tonal schemata in long-term memory—i.e., many students enrolled in first-year aural skills—are likely to employ imagery strategies that are inefficient and costly in terms of processing. Those who already possess imagery expertise—i.e., many of those exempted from first-year aural skills—incur fewer processing costs to using imagery strategies in their musical tasks and will be able to continue developing and refining their acquired imagery expertise in a reciprocal manner.

### *Pedagogical and Curricular Implications*

[7.1] While the expertise model proposed here is more descriptive than prescriptive for pedagogical purposes, I believe certain prescriptive claims can be made regarding approaches to imagery in the classroom. It may be that a domain-general auditory imagery psychometric like the BAIS may not be suitable for use in musical imagery measurement. Still, I argue that the BAIS's subjective methodology can easily be paired with commonly used objective measures, such as sight singing and dictation, for a more robust and direct focus on imagery acquisition.<sup>(31)</sup> The vividness subscale can be used to discern to what extent a score is being represented imagistically before sight singing. A sample question like that used in **Example 9** could be given to a student to quickly complete before they sight sing the excerpt. The metric, which helps capture the vividness of students' musical images, is extended to include information about image consistency across the excerpt. Further customization may make it possible to inquire, for example, into the vividness of different musical parameters, for example rhythm, pitch, and dynamics. Similarly, the control subscale can be adapted to measure various control-related skills, such as the ability to make changes to the given material in imagery; see **Example 10**. This type of subscale allows for the assessment of expertise-related skills, such as the speed of cue-to-imagery mappings, and the availability of musical structures in long-term memory. These types of reflective questions will not only allow instructors to assess the state of various imagery skill components, but encourages metacognitive awareness in students which, with practice, may help them become more efficient at making adjustments to their own learning approaches and practice routines.

[7.2] Along with employing an array of subjective measurement paradigms, there are other facets of imagery skill acquisition that are important to keep in mind. For example, imagery generation (vividness) is likely to develop faster than transformation (control) because content acquisition, the first stage in the LTWM model, is a required first step before expertise acquisition. As such, the acquisition of imagery transformation ability will likely take longer than the typical two- or three-year aural skills sequence permits. Rather than implying that work on control skills should be avoided, this natural constraint simply makes it likely that we may not see the benefits of this work directly within the timeframe of most undergraduate music programs. As such, more effort ought to be placed on curricular designs that give more thought to how learned skills and material contribute to post-course expertise acquisition. In this way, we would expect to find more growth on vividness tasks within the span of the course, and comparatively less on control.

[7.3] In order to facilitate the acquisition of control ability, I urge pedagogues to be explicit about the strategies necessary to complete complex musical tasks involving imagery. This will help to ease the initial learning curve of acquiring both new skills and strategies for tasks with heavy working memory demands. Pedagogues should also actively encourage metacognitive reflection to help develop control skills. The more students can monitor and adjust their individual strategies, the more efficiently they will be able to make progress in the development of the intricate, reciprocal skill set that is musical imagery.

[7.4] Another important takeaway is that we cannot assume that relevant musical structures will be stored in and fluently accessible from LTM from superficial exposure alone. I am referring here to quick passes through sight singing and dictation exercises. In order to ensure that relevant musical structures are stored and are accessible *as sound* in LTM, I propose that more emphasis be placed on memorization through rote learning, imitation, and singing, especially early on in development, to build a foundation of musical vocabulary. However, this cannot be done out of context in isolation. While the pedagogical approaches examined here separate “content acquisition” and “imagery quality” into two categories, within the LTWM expertise framework, chunk formation and semantic encoding in LTM should be completed simultaneously. To build efficient LTWM, there must also be a focus on retrieval efficiency (i.e., what to recall and when, see [Ericsson 2014](#)). Pattern-based approaches seem to be the most effective for building meaningful chunks in memory, but in order to ensure meaningful encoding and retrieval efficiency, these should always be taught with regard to their usage and context. For example, with reference to how they are used in repertoire, and/or explicitly tied to their typical temporal orderings (i.e., when/where they are used, and to what other chunks they are associated). Therefore, the more that relevant musical schema can be meaningfully encoded into LTM (i.e., associated with other knowledge structures for efficient retrieval in specific contexts), the faster it will become integrated in expert memory skill.<sup>(32)</sup>

[7.5] Last, I believe that the expertise framework provided here allows for a more precise understanding of imagery development as it is conceived in North American collegiate-level pedagogy. Operating from a memory skill perspective allows for a functional view of imagery development; such a viewpoint could help free us from the current state in which imagery is entangled in memory for schemata rooted in the Western art-music tradition. For example, despite the notion some instructors might harbor about students possessing pitch imagery in the absence of “tonal awareness,” (scale-degree function), it is in fact impossible to have meaning-neutral imagery.<sup>(33)</sup> This student “difficulty” they are supposedly observing is not due to a lack of imagery ability; instead, it can be understood as resulting from differences in semantic encoding of scale degrees in which their common-practice functional harmonic contexts have not been associated in LTM.<sup>(34)</sup> All incoming students will possess some imagery for scale degrees; however, what form their imagery takes will greatly depend on their past experience, long-term memory (including enculturation), and habit.

[7.6] As we work to broaden our approach to aural skills instruction, we can relinquish these stricter notions and incorporate other understandings of both semantic encoding and retrieval structures in imagery expertise, especially ones that are more stylistically and disciplinarily inclusive.<sup>(35)</sup> Within cognitive science research, I believe that giving careful attention to participant populations, expertise context and goals, interactions with memory, and task demands for different musical imagery expertise will help to provide a more nuanced scientific understanding of this expertise moving forward.<sup>(36)</sup> I will conclude by suggesting some such avenues for future research. It is my hope that this future research continues to shed light on malleable domain-specific imagery skills and the best ways to foster them through teaching. By better understanding the mechanisms underlying musical imagery expertise across a broad range of musical disciplines, we can begin to reshape our instruction in undergraduate core in ways that are more beneficial and more equitable.

## *Conclusions and Future Directions*

[8.1] This article provides a foundation for future research in musical imagery development. Reframing pedagogical practices for imagery development within the framework of LTWM allows

us to theorize imagery development as a form of memory skill and more clearly define the features of this expertise that may be susceptible to change (i.e., schematic content acquisition, retrieval structures, imagery maintenance processes and efficiency). Adopting an expertise perspective also allows for general predictions about imagery processes and properties that may change as a result, which may help to inform hypotheses in future longitudinal studies. Such studies may consider, for example, examining change in different imagery processes (generation and transformation) over time. Given the model proposed here, we would likely expect vividness tasks to improve before control type skills.

[8.2] I will conclude by proposing several avenues of future research. One of the most pressing issues, in my estimation, is the need to more fully explain and assess the various contributing factors to imagery expertise. To address this, more theoretical work on musical imagery expertise is needed to understand the underlying cognitive mechanisms and processes important for performance in various tasks. This includes theorizing about the types of processes involved in auditory imagery in a manner similar to that already conducted in visual imagery research.<sup>(37)</sup> While the BAIS may show some promise in measuring auditory imagery as a domain-general trait, it may not be ideal for measuring domain-specific changes in musical imagery expertise. As such, further theorizing about imagery processes—particularly as they pertain to musical imagery expertise—may lead to the development of more accurate measures and protocols.<sup>(38)</sup> Such would also allow for a more nuanced understanding of those abilities that are more domain-general (e.g., the BAIS as general auditory imagery measure and WM capacity), those that are more domain specific (e.g., imagery vividness for musical schema, use of various cueing and maintenance techniques), and interactions between the two types.

[8.3] I propose that cognitive task analyses would be extremely useful in examining the cognitive mechanisms and processes that underlie various types of imagery tasks musicians regularly engage in (as recommended by [Hambrick, Burgoyne, and Oswald 2019](#)). For example, the musical imagery skill for notational audiation may differ from that required for improvisation in meaningful ways. Theoretical work in this area would allow for the creation of testable hypotheses regarding what underlying processes are common across tasks, which processes are fixed and unchangeable (i.e., domain-general), and which processes differ with expertise in those given areas.<sup>(39)</sup> This would not only greatly contribute to our understanding of musical imagery expertise, but to theories of expertise acquisition in general. Similarly, in better understanding the contributing factors and task demands for different types of musical imagery expertise, it may be possible to design more effective training procedures to maximize growth in areas that are, in fact, malleable ([Hambrick et al. 2018, 292](#)).

[8.4] Last, I believe it vital to more thoroughly examine the relationship between memory and musical imagery. According to the expertise perspective, the ability to cue and maintain information in WM relies heavily on the structure and organization of information stored in LTM. This claim is supported by prior research, which has already demonstrated important roles for and interactions between WM and LTM for perceived imagery vividness ([Baddeley and Andrade 2000](#)).<sup>(40)</sup> Because music, unlike vision, has very few fixed objects, the schematic representations of musical content are likely much more diverse between populations, given that the majority of listeners develop long-term representations solely through passive exposure. We could hypothesize, then, that imagery ability might vary in part as a function of the musical structures (content, organization, and retrieval mechanisms) stored in LTM, either acquired through statistical learning (i.e., “in the wild”), or cultivated more deliberately through training. One such avenue would be to examine interactions between imagery and long-term memory representations such as those in stylistic familiarity or enculturation.<sup>(41)</sup>

[8.5] It is my hope that research into musical imagery development will continue on multiple fronts. As I have argued, such sustained collaboration between the fields of cognitive science, music theory and music theory pedagogy will benefit not only students and instructors in the aural skills classroom, but also the academic community more broadly.

---

Sarah Gates  
Northwestern University  
70 Arts Circle Drive  
Evanston, IL 60208  
[sarahgates2015@u.northwestern.edu](mailto:sarahgates2015@u.northwestern.edu)

---

## Works Cited

- Adolphe, Bruce. 2013. *The Mind's Ear: Exercises for Improving the Musical Imagination for Performers, Composers, and Listeners*. 2nd ed. Oxford University Press.
- Aleman, André, Mark R. Nieuwenstein, Koen B. E. Böcker, and Edward H. F. de Haan. 2000. "Music Training and Mental Imagery Ability." *Neuropsychologia* 38 (12): 1664–68. [https://doi.org/10.1016/S0028-3932\(00\)00079-8](https://doi.org/10.1016/S0028-3932(00)00079-8).
- Arnett, Jeffrey J. 2008. "The Neglected 95%: Why American Psychology Needs to Become Less American." *American Psychologist* 63 (7): 602–14. <https://doi.org/10.1037/0003-066X.63.7.602>.
- Arthur, Claire. 2018. "A Perceptual Study of Scale-Degree Qualia in Context." *Music Perception* 35 (3): 295–314. <https://doi.org/10.1525/mp.2018.35.3.295>.
- Baddeley, Alan D., and Jackie Andrade. 2000. "Working Memory and the Vividness of Imagery." *Journal of Experimental Psychology: General* 129 (1): 126–45. <https://doi.org/10.1037/0096-3445.129.1.126>.
- Baddeley, Alan D., and Robert Logie. 1992. "Auditory Imagery and Working Memory." In *Auditory Imagery*, ed. Daniel Reisberg, 179–97. Lawrence Erlbaum Associates.
- Beaty, Roger E., Chris J. Burgin, Emily C. Nusbaum, Thomas R. Kwapil, Donald A. Hodges, and Paul J. Silvia. 2013. "Music to the Inner Ears: Exploring Individual Differences in Musical Imagery." *Consciousness and Cognition* 22 (4): 1163–73. <https://doi.org/10.1016/j.concog.2013.07.006>.
- Benward, Bruce. 1989. *Advanced Sightsinging and Ear Training: Strategies and Applications*. Wm. C. Brown Publishers.
- Bland, Leland D. 1984. *Sight Singing Through Melodic Analysis: A Guide to the Study of Sight Singing and an Aid to Ear Training Instruction*. Scarecrow Press.
- Brodsky, Warren, Avishai Henik, Bat-Sheva Rubinstein, and Moshe Zorman. 2003. "Auditory Imagery from Musical Notation in Expert Musicians." *Perception & Psychophysics* 65 (4): 602–12. <https://doi.org/10.3758/BF03194586>.
- Brodsky, Warren, Yoav Kessler, Bat-Sheva Rubinstein, Jane Ginsborg, and Avishai Henik. 2008. "The Mental Representation of Music Notation: Notational Audiation." *Journal of Experimental Psychology: Human Perception and Performance* 34 (2): 427–45. <https://doi.org/10.1037/0096-1523.34.2.427>.
- Brown, Rachel M., and Caroline Palmer. 2012. "Auditory-Motor Learning Influences Auditory Memory for Music." *Memory & Cognition* 40 (4): 567–78. <https://doi.org/10.3758/s13421-011-0177-x>.
- Burgoyne, Alexander P., Lauren Julius Harris, and David Z. Hambrick. 2019. "Predicting Piano Skill Acquisition in Beginners: The Role of General Intelligence, Music Aptitude, and Mindset." *Intelligence* 76. <https://doi.org/10.1016/j.intell.2019.101383>.
- Burgoyne, Alexander P., Giovanni Sala, Fernand Gobet, Brooke N. Macnamara, Guillermo Campitelli, and David Z. Hambrick. 2016. "The Relationship between Cognitive Ability and Chess Skill: A Comprehensive Meta-Analysis." *Intelligence* 59: 72–83. <https://doi.org/10.1016/j.intell.2016.08.002>.

- Byros, Vasili. 2009. "Foundations of Tonality as Situated Cognition, 1730–1830: An Enquiry into the Culture and Cognition of Eighteenth-Century Tonality with Beethoven's 'Eroica' Symphony as a Case Study." PhD diss., Yale University. <http://search.proquest.com/docview/305040356/?pq-origsite=primo>.
- . 2012. "Meyer's Anvil: Revisiting the Schema Concept." *Music Analysis* 31 (3): 273–346. <https://doi.org/10.1111/j.1468-2249.2012.00344.x>.
- Choksy, Lois. 1974. *Kodály Method: Comprehensive Music Education from Infant to Adult*. Prentice-Hall.
- Cleland, Kent D. 2015. *Developing Musicianship Through Aural Skills*. 2nd ed. Routledge. <https://doi.org/10.4324/9780203738474>.
- Colley, Ian D., Peter E. Keller, and Andrea R. Halpern. 2018. "Working Memory and Auditory Imagery Predict Sensorimotor Synchronisation with Expressively Timed Music." *Quarterly Journal of Experimental Psychology* 71 (8): 1781–96. <https://doi.org/10.1080/17470218.2017.1366531>.
- Covington, Kate. 2005. "The Mind's Ear: I Hear Music and No One Is Performing." *College Music Symposium* 45: 25–41. <https://www.jstor.org/stable/40374518>.
- Cox, Arnie. 2016. *Music and Embodied Cognition: Listening, Moving, Feeling, and Thinking*. Musical Meaning and Interpretation. Indiana University Press.
- Davidson-Kelly, Kirsteen, Rebecca S. Schaefer, Nikki Moran, and Katie Overy. 2015. "'Total Inner Memory': Deliberate Uses of Multimodal Musical Imagery during Performance Preparation." *Psychomusicology: Music, Mind, and Brain, Musical Imagery* 25 (1): 83–92. <https://doi.org/10.1037/pmu0000091>.
- Ericsson, K. Anders. 1985. "Memory Skill." *Canadian Journal of Psychology/Revue Canadienne de Psychologie* 39 (2): 188–231. <https://doi.org/10.1037/h0080059>.
- . 2014. "Why Expert Performance Is Special and Cannot Be Extrapolated from Studies of Performance in the General Population: A Response to Criticisms." *Intelligence* 45 (July–August): 81–103. <https://doi.org/10.1016/j.intell.2013.12.001>.
- . 2016. "Summing Up Hours of Any Type of Practice Versus Identifying Optimal Practice Activities: Commentary on Macnamara, Moreau, & Hambrick (2016)." *Perspectives on Psychological Science* 11 (3): 351–54. <https://doi.org/10.1177/1745691616635600>.
- . 2018. "Superior Working Memory in Experts." In *The Cambridge Handbook of Expertise and Expert Performance*, ed. K. Anders Ericsson, Robert Hoffman, Aaron Kozbelt, and A. Mark Williams, 2nd ed., 696–714. Cambridge University Press. <https://doi.org/10.1017/9781316480748.036>.
- Ericsson, K. Anders, and Walter Kintsch. 1995. "Long-Term Working Memory." *Psychological Review* 102 (2): 211–45. <https://doi.org/10.1037/0033-295X.102.2.211>.
- Ericsson, K. Anders, and Jerad H. Moxley. 2014. "Experts' Superior Memory: From Accumulation of Chunks to Building Memory Skills That Mediate Improved Performance and Learning." In *The SAGE Handbook of Applied Memory*, 404–20. London: SAGE Publications Ltd. <https://doi.org/10.4135/9781446294703>.
- Ericsson, K. Anders, and Roy W. Roring. 2007. "Memory as A Fully Integrated Aspect of Skilled and Expert Performance." In *Psychology of Learning and Motivation* 48: 351–80. Elsevier. [https://doi.org/10.1016/S0079-7421\(07\)48009-4](https://doi.org/10.1016/S0079-7421(07)48009-4).
- Floridou, Georgia A., Victoria J. Williamson, Lauren Stewart, and Daniel Müllensiefen. 2015. "The Involuntary Musical Imagery Scale (IMIS)." *Psychomusicology: Music, Mind, and Brain* 25 (1): 28–36. <https://doi.org/10.1037/pmu0000067>.
- Gelding, Rebecca W., William Forde Thompson, and Blake W. Johnson. 2015. "The Pitch Imagery Arrow Task: Effects of Musical Training, Vividness, and Mental Control." *PLoS ONE* 10 (3):

<https://doi.org/10.1371/journal.pone.0121809>.

- Gjerdingen, Robert, and Janet Bourne. 2015. "Schema Theory as a Construction Grammar." *Music Theory Online* 21 (2). <https://doi.org/10.30535/mt.21.2.3>.
- Godøy, Rolf Inge. 2003. "Motor-Mimetic Music Cognition." *Leonardo* 36 (4): 317–19. <https://doi.org/10.1162/002409403322258781>.
- Godøy, Rolf Inge. 2004. "Gestural Imagery in the Service of Musical Imagery." In *esture-Based Communication in Human-Computer Interaction*, ed. Antonio Camurri and Gualtiero Volpe, 55–62. Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-540-24598-8\\_5](https://doi.org/10.1007/978-3-540-24598-8_5).
- Gordon, Edwin, E. 2012. *Learning Sequences in Music: Skill, Content, and Patterns*. GIA Publications.
- Gottschalk, Arthur, and Phillip Kloeckner. 1998. *Functional Hearing: A Contextual Method for Ear Training*. Ardsley House Publishers.
- Graybill, Roger. 2018. "Activating Aural Imagery through Keyboard Harmony." In *The Norton Guide to Teaching Music Theory*, ed. Rachel Lumsden and Jeffrey Swinkin, 182–97. W.W. Norton.
- Greenspon, Emma B., Peter Q. Pfordresher, and Andrea R. Halpern. 2017. "Pitch Imitation Ability in Mental Transformations of Melodies." *Music Perception* 34 (5): 585–604. <https://doi.org/10.1525/mp.2017.34.5.585>.
- Guida, Alessandro, Fernand Gobet, Hubert Tardieu, and Serge Nicolas. 2012. "How Chunks, Long-Term Working Memory and Templates Offer a Cognitive Explanation for Neuroimaging Data on Expertise Acquisition: A Two-Stage Framework." *Brain and Cognition* 79 (3): 221–44. <https://doi.org/10.1016/j.bandc.2012.01.010>.
- Hall, Anne Carothers. 2004. *Studying Rhythm*. 3rd ed. Pearson.
- Halpern, Andrea R. 2015. "Differences in Auditory Imagery Self-Report Predict Neural and Behavioral Outcomes." *Psychomusicology: Music, Mind, and Brain, Musical Imagery* 25 (1): 37–47. <https://doi.org/10.1037/pmu0000081>.
- Halpern, Andrea R, and Katie Overy. 2019. "Voluntary Auditory Imagery and Music Pedagogy." In *The Oxford Handbook of Sound and Imagination*, ed. Mark Grimshaw-Aagaar, Mads Walther-Hansen, and Martin Knakkegaard, vol. 2, 390–407. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190460242.013.49>.
- Hambrick, David Z., Erik M. Altmann, and Alexander P. Burgoyne. 2018. "A Knowledge Activation Approach to Testing the Circumvention-of-Limits Hypothesis." *The American Journal of Psychology* 131 (3): 307. <https://doi.org/10.5406/amerjpsyc.131.3.0307>.
- Hambrick, David Z., Erik M. Altmann, Frederick L. Oswald, Elizabeth J. Meinz, Fernand Gobet, and Guillermo Campitelli. 2014b. "Accounting for Expert Performance: The Devil Is in the Details." *Intelligence* 45: 112–14. <https://doi.org/10.1016/j.intell.2014.01.007>.
- Hambrick, David Z., Alexander P. Burgoyne, Brooke N. Macnamara, and Fredrik Ullén. 2018. "Toward a Multifactorial Model of Expertise: Beyond Born versus Made." *Annals of the New York Academy of Sciences* 1423 (1): 284–95. <https://doi.org/10.1111/nyas.13586>.
- Hambrick, David Z., Alexander P. Burgoyne, and Frederick L. Oswald. 2019. "Domain-General Models of Expertise: The Role of Cognitive Ability." In *The Oxford Handbook of Expertise*, ed. by Paul Ward, Jan Maarten Schraagen, Julie Gore, and Emilie M. Roth, 55–84. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198795872.013.3>.
- Hambrick, David Z., Julie C. Libarkin, Heather L. Petcovic, Kathleen M. Baker, Joe Elkins, Caitlin N. Callahan, Sheldon P. Turner, Tara A. Rench, and Nicole D. LaDue. 2012. "A Test of the Circumvention-of-Limits Hypothesis in Scientific Problem Solving: The Case of Geological Bedrock

- Mapping." *Journal of Experimental Psychology: General* 141 (3): 397–403.  
<https://doi.org/10.1037/a0025927>.
- Hambrick, David Z., Frederick L. Oswald, Erik M. Altmann, Elizabeth J. Meinz, Fernand Gobet, and Guillermo Campitelli. 2014a. "Deliberate Practice: Is That All It Takes to Become an Expert?" *Intelligence* 45: 34–45. <https://doi.org/10.1016/j.intell.2013.04.001>.
- Hambrick, David Z., and Elliot M. Tucker-Drob. 2015. "The Genetics of Music Accomplishment: Evidence for Gene–Environment Correlation and Interaction." *Psychonomic Bulletin & Review* 22 (1): 112–20. <https://doi.org/10.3758/s13423-014-0671-9>.
- Hansberry, Benjamin. 2017. "What Are Scale-Degree Qualia?" *Music Theory Spectrum* 39 (2): 182–99. <https://doi.org/10.1093/mts/mtx014>.
- Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. "The Weirdest People in the World?" *Behavioral and Brain Sciences* 33 (2–3): 61–83. <https://doi.org/10.1017/S0140525X0999152X>.
- Herholz, Sibylle C., Andrea R. Halpern, and Robert J. Zatorre. 2012. "Neuronal Correlates of Perception, Imagery, and Memory for Familiar Tunes." *Journal of Cognitive Neuroscience* 24 (6): 1382–97. [https://doi.org/10.1162/jocn\\_a\\_00216](https://doi.org/10.1162/jocn_a_00216).
- Highben, Zebulon, and Caroline Palmer. 2004. "Effects of Auditory and Motor Mental Practice in Memorized Piano Performance." *Bulletin of the Council for Research in Music Education* 159: 58–65.
- Hubbard, Timothy L. 2010. "Auditory Imagery: Empirical Findings." *Psychological Bulletin* 136 (2): 302–29. <https://doi.org/10.1037/a0018436>.
- . 2013. "Auditory Imagery Contains More Than Audition." In *Multisensory Imagery*, ed. Simon Lacey and Rebecca Lawson, 221–47. Springer. [https://doi.org/10.1007/978-1-4614-5879-1\\_12](https://doi.org/10.1007/978-1-4614-5879-1_12).
- Janata, Petr, and Kaivon Paroo. 2006. "Acuity of Auditory Images in Pitch and Time." *Perception & Psychophysics* 68 (5): 829–44. <https://doi.org/10.3758/BF03193705>.
- Johnson, Eric A., and Edward Klonoski. 2003. "Connecting the Inner Ear and the Voice." *The Choral Journal* 44 (3): 35–40. <https://www.jstor.org/stable/23554524>.
- Jones, Evan, Matthew R. Shaftel, and Juan Chattah. 2013. *Aural Skills in Context: A Comprehensive Approach to Sight Singing, Ear Training, Keyboard Harmony, and Improvisation*. 1st ed. Oxford University Press.
- Kalakoski, Virpi. 2001. "Musical Imagery and Working Memory." In *Musical Imagery*, ed. Rolf Inge Godøy and Harald Jorgensen, 43–56. Taylor & Francis.
- Karpinski, Gary S. 2000. *Aural Skills Acquisition: The Development of Listening, Reading, and Performing Skills in College-Level Musicians*. 1st ed. Oxford University Press.
- . 2017. *Manual for Ear Training and Sight Singing*. 2nd ed. W.W. Norton.
- Keller, Peter E. 2012. "Mental Imagery in Music Performance: Underlying Mechanisms and Potential Benefits." *Annals of the New York Academy of Sciences* 1252 (1): 206–13. <https://doi.org/10.1111/j.1749-6632.2011.06439.x>.
- Klonoski, Edward. 1998. "Teaching Pitch Internalization Processes." *Journal of Music Theory Pedagogy* 12: 81–96.
- . 2006. "Improving Dictation as an Aural-Skills Instructional Tool." *Music Educators Journal* 93 (1): 54–59. <https://doi.org/10.1177/002743210609300124>.
- Kosslyn, Stephen Michael. 1994. *Image and Brain: The Resolution of the Imagery Debate*. MIT Press. <https://doi.org/10.7551/mitpress/3653.001.0001>.

- Kozhevnikov, Maria, Stephen Kosslyn, and Jennifer Shephard. 2005. "Spatial versus Object Visualizers: A New Characterization of Visual Cognitive Style." *Memory & Cognition* 33 (4): 710–26. <https://doi.org/10.3758/BF03195337>.
- Larson, Steve. 1993. "Scale-Degree Function: A Theory of Expressive Meaning and Its Application to Aural Skills Pedagogy." *Journal of Music Theory Pedagogy* 7: 69–84.
- Lima, César F., Nadine Lavan, Samuel Evans, Zarinah Agnew, Andrea R. Halpern, Pradheep Shanmugalingam, Sophie Meekings, et al. 2015. "Feel the Noise: Relating Individual Differences in Auditory Imagery to the Structure and Function of Sensorimotor Systems." *Cerebral Cortex* 25 (11): 4638–50. <https://doi.org/10.1093/cercor/bhv134>.
- Macnamara, Brooke N., David Z. Hambrick, and David Moreau. 2016. "How Important Is Deliberate Practice? Reply to Ericsson (2016)." *Perspectives on Psychological Science* 11 (3): 355–58. <https://doi.org/10.1177/1745691616635614>.
- Macnamara, Brooke N., David Z. Hambrick, and Frederick L. Oswald. 2014. "Deliberate Practice and Performance in Music, Games, Sports, Education, and Professions: A Meta-Analysis." *Psychological Science* 25 (8): 1608–18. <https://doi.org/10.1177/0956797614535810>.
- McAvinue, Laura P., and Ian H. Robertson. 2007. "Measuring Visual Imagery Ability: A Review." *Imagination, Cognition and Personality* 26 (3): 191–211. <https://doi.org/10.2190/3515-8169-24J8-7157>.
- Meinz, Elizabeth J. and David Z. Hambrick. 2010. "Deliberate Practice Is Necessary but Not Sufficient to Explain Individual Differences in Piano Sight-Reading Skill: The Role of Working Memory Capacity." *Psychological Science* 21(7): 914–19. <https://doi.org/10.1177/095679761037393>
- Morrow, Daniel G., William E. Menard, Elizabeth A. L. Stine-Morrow, Thomas Teller, and David Bryant. 2001. "The Influence of Expertise and Task Factors on Age Differences in Pilot Communication." *Psychology and Aging* 16 (1): 31–46. <https://doi.org/10.1037/0882-7974.16.1.31>.
- Murphy, Paul, Joel Phillips, Elizabeth West Marvin, and Jane Piper Clendinning. 2016a. *The Musician's Guide to Aural Skills: Ear Training*. 3rd ed. W.W. Norton.
- . 2016b. *The Musician's Guide to Aural Skills: Sight-Singing*. 3rd ed. W.W. Norton.
- Navarro Cebrian, Ana, and Petr Janata. 2010. "Influences of Multiple Memory Systems on Auditory Mental Image Acuity." *The Journal of the Acoustical Society of America* 127 (5): 3189–3202. <https://doi.org/10.1121/1.3372729>.
- Paivio, Allan. 2007. *Mind and Its Evolution*. Routledge.
- Pearson, Joel, and Stephen M. Kosslyn. 2015. "The Heterogeneity of Mental Representation: Ending the Imagery Debate." *Proceedings of the National Academy of Sciences of the United States of America* 112 (33): 10089–92. <https://doi.org/10.1073/pnas.1504933112>.
- Pfordresher, Peter Q., and Andrea R. Halpern. 2013. "Auditory Imagery and the Poor-Pitch Singer." *Psychonomic Bulletin & Review* 20 (4): 747–53. <https://doi.org/10.3758/s13423-013-0401-8>.
- Pfordresher, Peter Q., Andrea R. Halpern, and Emma B. Greenspon. 2015. "A Mechanism for Sensorimotor Translation in Singing: The Multi-Modal Imagery Association (MMIA) Model." *Music Perception* 32 (3): 242–53. <https://doi.org/10.1525/mp.2015.32.3.242>.
- Pfordresher, Peter Q., and James T. Mantell. 2014. "Singing with Yourself: Evidence for an Inverse Modeling Account of Poor-Pitch Singing." *Cognitive Psychology* 70: 31–57. <https://doi.org/10.1016/j.cogpsych.2013.12.005>.
- Pruitt, Tim A., Andrea R. Halpern, and Peter Q. Pfordresher. 2019. "Covert Singing in Anticipatory Auditory Imagery." *Psychophysiology* 56 (3). <https://doi.org/10.1111/psyp.13297>.

- Rogers, Michael. 2004. *Teaching Approaches in Music Theory: An Overview of Pedagogical Philosophies*. 2nd ed. Southern Illinois University Press.
- Saintilan, Nicole. 2014. "The Use of Imagery During the Performance of Memorized Music." *Psychomusicology; Music, Mind, and Brain* 24 (4): 309–15. <https://psycnet.apa.org/doi/10.1037/pmu0000080>.
- Sánchez-Kisielewska, Olga. 2017. "The Rule of the Octave in First-Year Undergraduate Theory: Teaching in the Twenty-First Century with Eighteenth-Century Strategies." *Journal of Music Theory Pedagogy* 31 (22): 113–34. <https://jmt.pappstate.edu/rule-octave-first-year-undergraduate-theory-teaching-twenty-first-century-eighteenth-century>.
- Smith, David, Daniel Reisberg, and Meg Wilson. 1992. "Subvocalization and Auditory Imagery: Interactions Between the Inner Ear and the Inner Voice." In *Auditory Imagery*, ed. Daniel Reisberg, 1st ed., 95–119. Psychology Press.
- Sohn, Young Woo, and Stephanie M. Doane. 2003. "Roles of Working Memory Capacity and Long-Term Working Memory Skill in Complex Task Performance." *Memory & Cognition* 31 (3): 458–66. <https://doi.org/10.3758/BF03194403>.
- . 2004. "Memory Processes of Flight Situation Awareness: Interactive Roles of Working Memory Capacity, Long-Term Working Memory, and Expertise." *Human Factors* 46 (3): 461–75. <https://doi.org/10.1518/hfes.46.3.461.50392>.
- Tacka, Philip, and Micheal Houlahan. 2004. *Sound Thinking: Music for Sight-Singing and Ear Training*. Vol. 1. Boosey & Hawkes.
- Ullén, Fredrik, David Zachary Hambrick, and Miriam Anna Mosing. 2016. "Rethinking Expertise: A Multifactorial Gene–Environment Interaction Model of Expert Performance." *Psychological Bulletin* 142 (4): 427–46. <https://doi.org/10.1037/bul0000033>.
- Urista, Diane. 2016. *The Moving Body in the Aural Skills Classroom: A Eurythmics Based Approach*. Oxford University Press.
- Vuvan, Dominique T., and Mark A. Schmuckler. 2011. "Tonal Hierarchy Representations in Auditory Imagery." *Memory & Cognition* 39 (3): 477–90. <https://doi.org/10.3758/s13421-010-0032-5>.
- Williamson, Victoria J., and Sagar R. Jilka. 2014. "Experiencing Earworms: An Interview Study of Involuntary Musical Imagery." *Psychology of Music* 42 (5): 653–70. <https://doi.org/10.1177/0305735613483848>.
- Zatorre, Robert J., Andrea R. Halpern, and Marc Bouffard. 2009. "Mental Reversal of Imagined Melodies: A Role for the Posterior Parietal Cortex." *Journal of Cognitive Neuroscience* 22 (4): 775–89. <https://doi.org/10.1162/jocn.2009.21239>.

---

#### Footnotes

\* An earlier version of this research was presented as a conference paper titled "Developing Auditory Imagery: Contributions from Aural Skills Pedagogy and Cognitive Science" at the 41st Annual Meeting of the Society for Music Theory special session "Rethinking Aural Skills Instruction Through Cognitive Research." The present article represents a substantial expansion of this previous work with additional original research. I would like to thank Vivian Luong, Miriam Piilonen, Alyssa Barna, Lena Console and Anjni Amin for their feedback on earlier drafts of this article. I would also like to thank the anonymous reviewers and editorial team at *Music Theory Online* for their time and feedback.

[Return to text](#)

1. This can include, but is not limited to, musical staff visualization, typical key or usage, motor imagery of a familiar instrument, subvocalization, etc.

[Return to text](#)

2. Perhaps you are a bass player, or simply someone who frequently attends to basslines. You would probably be more prone to hearing Example 1 as a bass pattern. Or perhaps you have a jazz background and your generation of this pattern was more chord based and less key based, recalling patterns (schemata) you may have learned in the past.

[Return to text](#)

3. Many texts make direct reference to common patterns (Jones, Shaftel, and Chattah 2013; Tacka and Houlahan 2004; Hall 2004; Murphy et al. 2016a and 2016b; Bland 1984), while others simply imply this approach given the content and ordering of the instructional material (Karpinski 2017; Cleland 2015; Gottschalk and Kloeckner 1998).

[Return to text](#)

4. I am taking “meaning” here to refer primarily to how a particular musical schema is used in relation to other material (essentially, its function).

[Return to text](#)

5. Tonal awareness, as Larson clearly emphasizes, is not the result of a conscious effort (i.e., a secondary cognitive process layered on top of imagery) to assign meaning to sound. It is a learned sensitivity to tonal relationships, which become active automatically during listening and imagining, shaping the perception of musical material as goal directed (i.e., meaningful). His pedagogical approach, however, does rely on metaphor to facilitate this internalization of functional relationships within the frame of musical forces (gravity, magnetism, and inertia) (Larson 1993, 69).

[Return to text](#)

6. Such learned associational networks can include other meaning and movement-related information outside of traditional notions of function, such as imagined playing or singing, or the addition of movement metaphors, which can add an internal qualitative sensation of liveliness and/or action to imagined sound. Other scholars refer to this *hearing as* as the internal experience of musical actions as the results of an agent, such as in the mimetic hypothesis proposed by Cox (2016, 73).

[Return to text](#)

7. Research has shown that both trained musicians and nonmusicians demonstrate consistency in their scale-degree qualia ratings; however, musicians provide more consistent ratings than their nonmusician counterparts. Arthur 2018 proposes that these rating differences may arise due to increased statistical learning in musician populations, or from the conceptualization of scale degrees, as suggested by Hansberry 2017. Hansberry’s claim supports some of the notions of imagery control discussed below. For a quantitative example of imagery control over scale degrees stemming from conceptualization in trained musicians, see Vuvan and Schmuckler 2011.

[Return to text](#)

8. Sound-to-solfège mappings can be thought of as a form of dual-coding entailing cooperative activity between verbal and nonverbal systems (Paivio 2007).

[Return to text](#)

9. This is especially true for vocalists and percussionists who focus more on non-pitched instruments. For a detailed review of a multi-modal keyboard approach, see Graybill (2018).

[Return to text](#)

10. Gordon specifically distinguishes audiation from musical imagery (a vivid or figurative picture), noting that audiation is a “more profound process” because it involves “assimilation and comprehension” (2012, 4). Karpinski does not explicitly make such a distinction between auralizing and musical imagery. He does, however, in a similar manner to Gordon, distinguish “musical understanding” (comprehension) from “musical memory” (the ability to repeat what has been heard) (Karpinski 2000, 78). Both scholars therefore emphasize that skilled imagery performance involves more than “mere imitation or repetition” in that it entails contemplation, comprehension

and multitasking. This is similar to Larson's (1993) separation of *audiation* and *hearing as*, discussed above.

[Return to text](#)

11. I see this relating to but differing from Larson's (1993) emphasis on scale-degree function in that both Gordon and Karpinski specifically make reference to the ability to either consciously understand or multitask, whereas Larson proposes *hearing as* as a primarily subconscious phenomenon.

[Return to text](#)

12. Gordon discusses this reciprocal nature of audiation, see stages of audiation (2012, 18). Also see Karpinski (2000, 3). Rogers refers to this as the inseparable association between thinking and listening (2004, 8).

[Return to text](#)

13. This definition generally includes any type of internal sound generation, including speech. Within the field, auditory imagery is used to refer to the whole host of internal sound generation skills and contents, while the term musical imagery is typically reserved for the internal generation of musical content.

[Return to text](#)

14. Other forms of representation, such as propositional or language/symbolic-based representation, do not contain imagistic properties (see Pearson and Kosslyn 2015, 10089).

[Return to text](#)

15. "Developmental," as used here means tracking the change of imagery ability over the course aging, or through some training intervention.

[Return to text](#)

16. This task involves the imagining of scale degrees prompted through arrows (up and/or down) indicating which direction participants must imagine a scale moving. The task measures imagined pitch accuracy compared to a presented target pitch (correct/incorrect), which increases with difficulty as participants get higher accuracy scores.

[Return to text](#)

17. Gordon's position is that the audiation skill is in part biologically determined, yet still malleable (Gordon 2012, 3). This claim is in line with recent approaches to expertise acquisition that posit important roles for both genetics and environmental factors; see Hambrick, Burgoyne, Macnamara, and Ullén (2018) for a review.

[Return to text](#)

18. This includes, for example, memory experts like those who have learned to recall the numerical values of pi for several hundred digits (see Ericsson and Kintsch 1995).

[Return to text](#)

19. These constraints are primarily related to the functionality and limitations of working memory and long-term memory. Information held in working memory is temporarily stored and therefore accessible very quickly. Due to its temporary nature, it is known to be susceptible to interference effects such as retroactive interference, a process by which information in working memory is overwritten by new incoming information and/or processing. Long-term memory is not temporary, and therefore is not prone to these sorts of effects, however processing speeds of long-term memory for retrieval and encoding are extremely slow (over 5–10 seconds). Ericsson and colleagues have found that certain memory experts are not susceptible to interference effects while performing tasks that require working memory, suggesting they instead use long-term memory at comparable retrieval and encoding speeds to working memory (see Ericsson and Kintsch 1995; Ericsson and Roring 2007).

[Return to text](#)

20. These retrieval cues are presumed to be organized hierarchically, both spatially and sequentially in memory, and are used to efficiently recall any encoded information association with them. For example, such a structure was used by a subject called “SF” to memorize and recall a series of 30 digits. SF was reported to use a mnemonic coding scheme at retrieval cue level 1 (e.g., digits 3596 encoded as 3 mins 59.6 seconds), along with a spatial encoding scheme (relative positions of such cues into groups) at level 2 in the retrieval structure (Ericsson and Kintsch 1995, 217).

[Return to text](#)

21. This theory helps to explain how experts circumvent known limitations of working memory, namely retrospective interference (the wiping of information in working memory) and capacity constraints like Miller’s magic number (see Ericsson and Kintsch 1995, 215).

[Return to text](#)

22. This refers to the formation of a stable retrieval structure.

[Return to text](#)

23. This has since sparked a large debate within the expertise community, centering between Ericsson and those associated with the Hambrick lab (see Ericsson 2014; Hambrick et al. 2014b; Ericsson 2016; Macnamara, Hambrick, and Moreau 2016). Unfortunately, this back forth exchange has been rather heated and has not led to any collaboration between the parties involved.

[Return to text](#)

24. Within the domain of music, working memory capacity has been found to be an important predictor of success in sight reading both at low and high levels of deliberate practice (Meinz and Hambrick 2010). Burgoyne, Harris and Hambrick (2019) found that general intelligence was a stronger predictor for beginner piano skill acquisition than music aptitude and mindset. Evidence for the gene-environment model also comes from the musical domain. Hambrick and Tucker-Drob’s (2015) analysis of musical practice and accomplishment in a database of 800 pairs of twins found that genetic impact on musical accomplishment was maximized by deliberate practice.

[Return to text](#)

25. Some scholarship supports the circumvention-of-limits hypothesis (Sohn and Doane 2003, 2004; Hambrick et al. 2012), while others do not (Morrow et al. 2001, Meinz and Hambrick 2010).

[Return to text](#)

26. For example, the study of pre-existing groups (e.g., comparing “experts” versus “novices”) in quasi-experimental designs presents many confounds because participants are not assigned to groups randomly (Burgoyne et al. 2016). Conducting fully experimental studies with randomized groups is often characterized as unfeasible, as it would take an inordinate amount of time to train one group to expert levels of performance. While methodologies have been developed to test aspects of expertise in experimental ways, these methodologies also have limitations; see the knowledge-activation approach in Hambrick, Altman and Burgoyne (2018).

[Return to text](#)

27. While individuals may not be able to completely “overcome” constraints imposed by domain-general traits, such as working memory, it is still important to understand those domain-specific mechanisms that aid performance improvements with increasing expertise. In this way, working memory can be thought of as demarcating the possible upper limit of performance in certain tasks, while deliberate practice (cultivating domain-specific skills and knowledge) allows for a maximizing of this potential.

[Return to text](#)

28. It is important to note that in the context of sight singing, the notation itself is also a retrieval cue. The interpretation of these presented cues, namely solfège, is the explicit cue used to generate the internal sound structure while reading the notation.

[Return to text](#)

29. The thought process I engage in during this process proceeds something like the following. At the first reading, I note the higher-level structures of each bar (tonic and dominant alternations). I then use this higher-level scheme to cue solfège in context of this structure, i.e., “sol–do–mi” arpeggio in tonic, “sol–fa” in a dominant context, etc. By the time I am ready to sing the excerpt (after one or two scans of the notation), all I need to do is cue the solfège, and the contextual harmonic information occurs simultaneously with the syllable retrieval cue, affecting my inner hearing of the scale-degree qualia as “sol in tonic context.” Technically, the tonic function shifts to include a motion to vii<sup>0</sup>/V (as PD) in measure 3.4 in the original harmonized version of the concerto; however given the information available purely from the melody, it is more likely that a blank read of this score would prompt a purely tonic reading. With more passes, the PD function might be more perceptually evident as other knowledge structures become activated (e.g., antecedent function, therefore HC in measure 4).

[Return to text](#)

30. My primary instrument is saxophone, so the image takes on a more breathy and flowing quality.

[Return to text](#)

31. This application is similar to more recent use of trial-by-trial vividness ratings which have shown to predict performance on a behavioral task. See Pruitt, Halpern and Pfordresher (2019) where low vividness ratings were associated with inaccurate singing on a given trial.

[Return to text](#)

32. Some approaches already emphasize this type of learning, such as eighteenth-century schema theory-based methods, which draw analogies between the learning of meaningful musical patterns and linguistic collocations (see Gjerdingen and Bourne 2015; Sánchez-Kisielewska, 2017). Similarly, there may indeed be a desirable level of difficulty for efficient encoding, meaning that the task should not be easy (e.g., just listening), or else encoding might not occur. A moderate level of difficult where students are stretched a little out of their comfort zone seems to be ideal (see Halpern and Overy 2019, 394).

[Return to text](#)

33. Cognitive findings indicate that imagery is not meaning- or interpretation-neutral (see Hubbard 2010), and that musical imagery in particular is highly related to LTM and stored schema (see Halpern and Overy 2019, 399–401).

[Return to text](#)

34. This can be thought of as a form of situated cognition (Byros 2009; 2012) whereby certain perceptions are afforded by and conditional on specific interactions between humans and an environment (e.g., eighteenth-century scale degree and chord associations through the lens of schema theory).

[Return to text](#)

35. This will likely entail moving away from a reliance on Western notation as the primary source of retrieval structures and functional tonality as the prevailing means for semantically encoding pitch, both of which are quite particular to music theoretic expertise for a rather narrow range of repertoire (i.e., primarily the Western canon). skills.

[Return to text](#)

36. This includes expanding both qualitative and quantitative research to examine more diverse participant populations, including cross cultural studies. Currently, psychological studies are predominantly conducted at universities (with undergraduates), in wealthy countries, limiting their scope to approximately 5% of the world’s population (Arnett 2008). This WEIRD (Western, Educated, Industrialized, Rich and Democratic) bias for participant pools in the behavioral sciences is extremely problematic and calls into question the validity of research findings more broadly (Henrich, Heine, and Norenzayan 2010).

[Return to text](#)

37. See for example Kosslyn's (1994) theoretical work on visual imagery processes, where general vividness and control processes were supplanted by spatial and object imagery measures (McAvinue and Robertson 2007). This work has allowed research to examine interactions between imagery and cognitive style (Kozhenikov, Kosslyn, and Shephard 2005).

[Return to text](#)

38. We may consider in addition to general vividness and control properties others such as schematic (reduced) versus veridical (or high definition) imagery, imagery for specific content or musical features, imagery cueing efficiency, and imagery durability (i.e., how long an image lasts for before decaying, and how well it can be maintained through interference).

[Return to text](#)

39. For example, verbalization and graphic sketching as mediating factors in imagery maintenance and cuing for music theoretic expertise, versus procedural generation (motor activation) of imagery in fluent improvisors. We might hypothesize that domain-general abilities such as processing speed may be more important for imagery expertise in improvisers than it would be for in domains where there are fewer real-time processing constraints like music theoretic imagery expertise.

[Return to text](#)

40. A fair amount of work has been done on working memory and auditory imagery, although comparatively little has been done on long-term memory and imagery, or their interactions with working memory (for a review, see Kalakoski 2001; Baddeley and Logie 1992).

[Return to text](#)

41. Conversely, it would also be beneficial to explore the relation in the opposite direction, i.e., inquiring into whether imagery ability (measured as a domain-general trait) impacts the acquisition of statistically relevant schema in a novel musical environment using an implicit learning paradigm.

[Return to text](#)

---

## Copyright Statement

Copyright © 2021 by the Society for Music Theory. All rights reserved.

[1] Copyrights for individual items published in *Music Theory Online (MTO)* are held by their authors. Items appearing in *MTO* may be saved and stored in electronic or paper form, and may be shared among individuals for purposes of scholarly research or discussion, but may *not* be republished in any form, electronic or print, without prior, written permission from the author(s), and advance notification of the editors of *MTO*.

[2] Any redistributed form of items published in *MTO* must include the following information in a form appropriate to the medium in which the items are to appear:

This item appeared in *Music Theory Online* in [VOLUME #, ISSUE #] on [DAY/MONTH/YEAR]. It was authored by [FULL NAME, EMAIL ADDRESS], with whose written permission it is reprinted here.

[3] Libraries may archive issues of *MTO* in electronic or paper form for public access so long as each issue is stored in its entirety, and no access fee is charged. Exceptions to these requirements must be approved in writing by the editors of *MTO*, who will act in accordance with the decisions of the Society for Music Theory.

This document and all portions thereof are protected by U.S. and international copyright laws. Material contained herein may be copied and/or distributed for research purposes only.