What Are the Truly Aural Skills?

Timothy Chenette

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.chenee.php

KEYWORDS: aural skills, pedagogy, cognition, perception, working memory, attentional control, curriculum, harmony

ABSTRACT: I argue that current models of aural skills instruction are too strongly linked to music theory curricula. I examine harmonic dictation as a case study, demonstrating that the system of roman-numeral/inversion-symbol labels can interfere with our ability to determine what exactly students are hearing and can distract students from more directly perceptual goals. A pilot study suggests that focusing on bass lines and schemata may make our harmonic dictation training more relevant to perception. I propose that a skill is “truly aural” to the extent that it engages working memory with minimal knowledge-based mediation. Finally, I consider the current state of aural skills instruction and suggest a number of curricular revisions. The more radical proposals call for redesigning aural skills classes to focus on perceptual skills and relocating knowledge-mediated listening to the music theory classroom. Other proposals take a more measured approach to integrating perceptual skills with otherwise traditional curricula.

DOI: 10.30535/mto.27.2.0

Received January 2020

[1.1] The standard approach to organizing an aural skills class is the result of filtering the content and order of a standard music theory curriculum through the two standard aural skills tasks of sight reading and dictation.

[1.2] While textbooks have made admirable strides towards including diverse activities from improvisation to keyboard exercises, virtually all are still focused primarily on sight reading and dictation. Chapters in Merritt 2016, for example, are based around “melodies for performance” and dictations; Horvit, Koozin, and Nelson 2013 is dedicated entirely to dictation; Rogers and Ottman 2019 is focused mainly on sight singing; and Karpinski and Kram 2017 is a sight-singing anthology, while the instructor’s manual for Karpinski 2017b gives answers only for its dictations. The overwhelming preeminence of these tasks is reflected in a recent core curriculum survey conducted by Murphy and McConville. In it, 99% of respondents reported engaging in sight singing and dictation in their aural skills courses; only two other, closely related tasks were reported by more than 50% of respondents, error detection (61%) and conducting (60.63%) (2017, 209).
The reliance on the content and order of a music theory curriculum is perhaps most explicit in *The Musician’s Guide to Aural Skills* (Murphy et al. 2016), every detail of which, from chapter organization to kinds of tasks, is modeled on the flagship text, *The Musician’s Guide to Theory and Analysis* (Clendinning and Marvin 2016). But of course, the sequence of topics in music theory textbooks is virtually always adopted by aural skills texts. The broad curriculum-level similarity between these two categories is illustrated with a few sample textbooks in Example 1. The parallels extend into smaller-level details; for example, Karpinski 2017b includes chapters about such music-theory topics as “Six-Four Figures” (Chapter 47).

From the perspective of perception and cognition, this reliance on music theory is pedagogically problematic. Though we often call these courses “aural skills,” the decision to base these curricula on music theory curricula leads us to prioritize the logical over the perceptual. This problem was recognized as early as the year 2000, when Edward Klonoski lamented in the *College Music Symposium* that cognitive research has had “little substantive influence” on the field of aural skills pedagogy. In particular, Klonoski argued that we do not yet know the best “perceptually-based learning hierarchy, with one perceptual skill serving as the building block for other skills” (2000). In response to this, we tend to substitute a logical learning hierarchy instead.

This means that aural skills curricula inherit and amplify a problem already inherent in music theory curricula: a lack of valid learning goals. As Gawboy 2013 points out, music theory pedagogy training often sidesteps what she calls “the most profound question facing every theory teacher”: what should students be able to do at the end of the course? Instead of learning goals focused on student action, we frequently substitute what she calls “content goals” focused on student learning. An example would be to offer an “introduction to species counterpoint and four-part voice leading” without identifying what it means to learn or know that material. In aural skills classes, the situation is similar. While sight singing and dictation are indeed focused on “doing” rather than “knowing,” they are in-class tasks and not longer-range learning goals. They are typically used as a lens through which to filter content derived from the music theory curriculum. Thus, the designs of aural skills curricula often seem to emerge not from the question, “what should students be able to do?,” but rather from the question, “what aspects of music theory will we adapt to our standard two modes of assessment?”

This article proposes a new model for aural skills curricula, one founded on learning goals related to perception and cognition rather than content goals derived from music theory. In section 2, I will demonstrate that, while music-theoretical systems typically have a close relationship with perception, when perception and music theory conflict we typically favor the latter. This behavior becomes particularly apparent when the music theoretical system at hand uses numbers. One such system, harmonic dictation—which typically requires roman numerals and inversion symbols—will serve as a case study in the ways that logical systems and physical scores can blind us to how people hear. In section 3, I will propose a new model for determining what fits in aural skills classes, focusing not on music theory curricula but on tasks that most directly exercise working memory and control of attention without a high degree of mediation through knowledge. This model does not exclude music theory from aural skills, but rather emphasizes the importance of fundamental aural skills in preparing students for a wider variety of subdisciplines and practices. The fourth and final section explores how this curricular shift might more fully acknowledge and embrace the central place of aural instruction for every aspect of music learning.

**Hearing Harmony: A Case Study in Why Logical Systems May Be Perceptually Problematic**

While many labeling and theoretical systems taught in music theory classes are useful tools for modeling perception, they are never a perfect representation of it. Nowhere is this more apparent than when our systems use numbers, whose apparent logic and clarity often tempt us to consider them, at least in some way, more substantial than perception. For example, Karpinski has drawn attention to the problems associated with interval-identification training, pointing out that “a preponderance of experimental evidence shows little connection between the ability to
identify intervals acontextually and the ability to do so in a tonal context” (2000, 52). The music-theoretical terms “perfect fourth,” “major third,” etc., imply clear categories with clear divisions among them, and indeed this might be a useful perspective when reading scores. The evidence cited by Karpinski, however, suggests that our brains perceive tones primarily with regard to their context—especially their relation to a key center—and not their abstract relationships to immediately adjacent pitches. The minor third between the fifth and seventh of a V7 chord, for example, will sound far different from the minor third at the foundation of a minor tonic chord, at least to listeners who are enculturated to common-practice music. When we ask students to identify these as the same, we are prioritizing music theory content goals—“learn intervals”—rather than perception-based learning goals.

[2.2] Logical, number-based systems also distort another perceptual task: harmonic identification. Virtually every aural skills textbook dedicates one or more sections to developing this skill, relying on the standard labels derived from music theory instruction: roman numerals paired with inversion symbols (henceforth “RN/IS labels”). For aural skills classes based on a music theory curriculum in which these symbols are virtually omnipresent, this system makes sense. But it is not clear that this system is the right one for perception-based learning goals. In this section, I will demonstrate some of the ways that the RN/IS system prioritizes music theory knowledge over perception. I will then present preliminary results from a research study on hearing harmony that suggests alternative pathways forward.

[2.3] It is important to recognize that root, quality, and bass—the specific aspects conveyed by RN/IS—can indeed correlate to important aspects of chord perception. Example 2 presents harmonic dictation #38.6 from Karpinski 2017a. In this example, there is a sense of moving through the phrase as the passage goes through a traditional cycle of tonic, pre-dominant (IV6), and dominant (V) functions before returning to tonic. The roman numerals—through their associations with harmonic functions—roughly capture this functional succession, which is important to an educated musician’s understanding of tension, release, and the notion of “progression” often attributed to tonal music. The major quality of each chord might explain in part the affective quality of the passage (cheerful?), though it would be hard to disentangle the effects of individual chord qualities from the influence of the overall key/mode. The change of inversion symbol where the initial tonic chord moves from to position reflects its transition from lesser to greater stability.

[2.4] Nevertheless, the elements specifically delineated by a RN/IS label are not those most directly available to perception. Harmonic dictation involves multiple simultaneous “voices,” and it is well established that humans can attend closely to only one object at a time. In an ideal situation, a listener would hear each chord as a Gestalt rather than as a collection of separate notes; Karpinski 2000 describes this kind of hearing as the intended endpoint for training in harmonic hearing. Yet this goal is complicated by the fact that any given chord (say, the subdominant) can be instantiated in different keys, inversions, spacings, voicings, timbres, registers, and textures. This makes it all but certain that most beginning students—especially those lacking training on a chord-producing instrument—will choose a voice to focus on. Ideally this would be the bass, since that is the single voice that allows a listener to infer the greatest amount of harmonic information. Yet the scale degree of a chord’s bass note is only indirectly implied by a RN/IS label; for example, although scale degree 4 in the bass is crucial to the identity of the ii6 chord, there is no number 4 anywhere in its RN/IS label. Each of the elements directly represented by the RN/IS label—root, quality, and inversion—requires hearing every single note in the progression correctly, potentially distracting from the perceptual goal of improving one’s ability to follow the bass.

[2.5] Forging connections between bass notes and their resultant chords’ likely RN/IS labels is a worthy goal—but it is a music-theoretical goal and not a directly perceptual goal. For music theorists with long experience pondering the connections between bass lines and harmony, concrete information directly available from the audio stimulus (the bass) and calculated conceptual information (root, quality, inversion) are joined in an RN/IS label. For beginners learning to control their perception, the labels may seem to focus on the conceptual rather than the concrete and can be an obstacle and a distraction if our goal is for them to listen to the bass. The
difference between these two classes of listeners is not necessarily how well they hear harmony, but how comfortable they are with the labeling system we use for harmony—again, an important focus of music theory classes.

[2.6] The RN/IS system is particularly problematic from the standpoint of perception when tasked with distinguishing similarly functioning chords that have the same bass note—that is, chords whose RN/IS appear highly divergent despite sounding almost the same. After all, “aurally identify the functional areas of a phrase” is clearly an important learning objective, as these are a crucial aspect of how common-practice music conveys a sense of tension and release and likely affect how a performer would “shape” the phrase. “Distinguish aurally between different varieties of pre-dominants” (such as IV and ii$^6$) is less clearly important, as this distinction may be perceived more as one of color than of function. Example 3 presents harmonic dictation #42.9 from Karpinski 2017b’s chapter on “The Supertonic Chord.” Hearing that the phrase moves to a pre-dominant chord—or, alternatively, a “chord with scale degree 4 in the bass”) at the downbeat of m. 2—is likely important for an educated listener. Yet is the distinction between pre-dominant chords—the supertonic in first inversion vs. root-position subdominant—particularly meaningful in this context? Example 4 presents the same dictation with ii$^6$ replaced by iv; distinguishing these two chords is a primary goal of Karpinski’s chapter. Is there an important learning goal that is served by focusing on this distinction?

[2.7] An instructor who wishes to focus on the ii$^6$ vs. iv distinction might object that the supertonic is more intense than the subdominant and thus worth distinguishing. Note, though, that intensity changes in other non-RN/IS domains such as orchestration, spacing, timbre, and ornamentation may have similar effects. Example 5 adds an ornament to the iv chord on the downbeat of m. 2 with an accented dissonance; Example 6 sharply alters the dynamics at this point in the phrase. Compared to Example 4 with its plain iv chord, Examples 3 (with its ii$^6$), 5 (with its accented dissonance and more active rhythm/counterpoint), and 6 (with its drop to subito piano) contribute similarly to a rise in intensity anticipating the arrival of dominant function, but because the types of shifts in Examples 5 and 6 are not captured in the roman numeral analysis provided in the textbook, they are unlikely to be assessed by the instructor. The chord substitution of Example 4, on the other hand, changes the chord symbol in every way: the root is different (4 vs. 2), the quality is different (minor vs. diminished), and the inversion is different (root position vs. first inversion), and these distinctions are what are graded in RN/IS harmonic dictation. I encourage readers to listen to the audio attached to each example, recorded by a professional pianist instructed only “to play what is written,” and note which examples seem the most significantly different and which appear to have elicited the most expressive differences in performance.

[2.8] It is well known that the distinctions between similar chords with different RN/IS labels can be extremely difficult to hear. Example 7 presents the opening of harmonic dictation #73.7 from Karpinski 2017a. The third chord is V$^4$, a chord often confused with the rather similar vii$^6$ and V$^1$. To distinguish among these three chords in the key of F minor, one must accurately hear all the voices of the chord (to determine whether it includes Bb, C, or both), detect the bass, and add all of the notes up to arrive logically at the correct answer. This task is relatively straightforward when one is looking at a physical score, as is common in music theory classes. But this kind of perfectly accurate hearing seems unlikely for most listeners attending to an ephemeral and complex audio signal, particularly if the passage is played quickly, as it is in Karpinski’s accompanying recordings.

[2.9] Some instructors might object that since we teach these chord symbols in music theory, we should also use them in aural skills, even when we may be uncomfortable with the priorities they imply in certain situations. Indeed, when a curriculum or textbook is based on content goals rather than learning goals, the only basis for deciding whether or not a system is effective is whether or not it fits the content. In other words, if the content goal driving our curriculum is “learn harmony through roman numerals and inversion symbols,” then the kind of dictation described above is probably inevitable. But thinking about what we want students to be able to do might inspire different kinds of activities and assessment.
Setting aside habits and assumptions derived from music theory classes and looking at scores, how might listeners perceive a passage and its harmonies? Certain musicians do seem to have extremely accurate harmonic hearing. In particular, I have found that highly experienced pianists are often able to imagine playing a heard passage, even one scored for winds, strings, or voice. (Needed to say, aural skills instructors often fall into this category.) In this method, the perceived stimulus is mediated through norms of physical motion internalized through many years of practice, typically begun during the period in child development when the brain is still highly plastic.

Since the background of extensive keyboard training does not apply to most students, how might a non-pianist effectively hear a passage such as Example 7? Such a student might hear that there is a less-structurally-important chord in the second half of m.1 harmonizing a bass scale degree 2 and then guess at its identity based on their music theory knowledge. Those who know the most likely three options might choose at random among V₈, vii°₆, and V°F, those who do not will often guess a root-position ii chord. By the standards of music theory class, where ii is usually classified as a predominant that must progress to V, the latter group of students are “wrong;” but is their perception any worse than the first category? After all, the perception “less-structurally-important chord over a bass scale degree 2” has a more direct logical connection—through shared number—to the roman numeral ii than to any of the more likely chords. If we mark this wrong, we are using music theory knowledge as a barrier to the student’s ability to effectively communicate what they are hearing and grading the student’s understanding of music theory at least as much as their aural perception.

There is no published research on exactly how people aurally identify chords, but a pilot study I recently conducted with a research team may be helpful in framing the issue of how perception interacts with harmony. The core of the study, an anonymous online questionnaire, asked participants (N = 74) to identify in their own words the chords of the three recordings represented in Examples 8, 9, and 10: Harmonic dictation #42.6 from the sound recordings accompanying Karpinski 2017b, the repeating four-chord progression in the song “Halo” by Beyoncé, and the opening two measures of a recording of Mozart’s Piano Sonata K. 332, movement 2. They were then asked to reflect on the process of identifying these chords. The participants’ answers suggested several important aspects of how they perceived harmony.

Examples 11–13 demonstrate that study participants who were particularly successful at all three harmonic identification tasks had significantly different backgrounds than other participants. 10 of our 74 respondents identified every chord correctly in all three excerpts; we call this the “Correct” group. Comparing these 10 respondents to the 16 participants who got an average score of 80–99% across the three excerpts (“Nearly correct”), the highest performers were significantly more likely to report some variety of absolute pitch, to report piano as their primary instrument, and to report a higher number of years of formal training. Though the correlation of these characteristics with highly accurate chord identification does not firmly establish causation, it is possible that some combination of these factors may be necessary for the highest degree of accuracy. Those in the 80–99% group, in contrast, have profiles more similar to the remaining 46 participants who scored less than 80%, except for the prevalence of heightened tonal memory.

If the “Correct” group is truly a distinct population, this would suggest that a primary goal in aural skills classes should be to move less-successful students into the “Nearly correct” group. Notably, when presented with the Karpinski dictation, which featured a ii°₆, only half (eight) of those in this “Nearly correct” group labeled it correctly; seven labeled it “IV” and one described it merely as “predominant.” Thus the level of precision we often demand—for example, distinguishing iv from ii°₆ and vii°₆ from V°F and V°F, as described above—may only or primarily be available to people with certain backgrounds. These backgrounds, in turn, may include abilities that we cannot inculcate in four years of post-secondary instruction. While more research would be beneficial to further test this hypothesis, demanding a high level of accuracy may serve primarily to reward those who have these characteristics and penalize those who do not.
Another behavior that manifested in our data was pattern matching, particularly in responses to the four-chord progression in the song “Halo.” This progression is very similar to the more common “Axis” four-chord loop shown in Example 14 (Richards 2017), and quite a few participants appeared to make this connection. Several wrote the roman numerals of the Axis progression instead of the correct progression, often adding one of the common names for the progression. Two participants noticed the “Halo” progression’s resemblance to a stock progression but incorrectly labeled it the I–vi–IV–V “doo-wop progression.” These correct and incorrect responses indicate the importance of recognizing familiar patterns and the limitations of such recognition.

The clearest result of our study by far is that hearing the bass is a crucial aspect of hearing harmonically. This is apparent in respondents’ reflections on how they identified the chords: of the 10 “Correct” respondents, five mentioned listening for the bass or outer voices first, and two others mentioned the bass as part of the process. (The remaining three had immediate access to higher-level hearing, such as an immediate, apparently unmediated perception of roman numerals.) Of the 16 “Nearly correct” respondents, all but one mentioned listening for the bass or outer voices. When participants were asked to choose from a list of all the elements and strategies they use in identifying chords, “bass lines” was chosen by 88% of all participants, more than 10 percentage points ahead of any other element or strategy. The importance of the bass was also reflected in other aspects of the data, in particular the fact that incorrect answers nevertheless often still matched the correct bass line.

These findings align with two important principles derived from research on perception. The first concerns pattern matching. As Huron 2006 points out,

In a stable environment, the most frequently occurring events of the past are the most likely events to occur in the future. Thus, a simple yet optimum inductive strategy is to expect the most frequent past event. Acquiring such knowledge through exposure is referred to as statistical learning (360; italics in original).

Pattern matching encompasses not just stock four-chord progressions in popular music but also cadential formulas in Classical and Romantic music and contrapuntal-harmonic schemas such as those presented in Gjerdingen 2007. The second principle is that humans can only focus on one stream of information. Justin London has pointed out that the human attentional requirement to focus on a single primary stimulus rules out the possibility of hypermeter from a perceptual perspective (2012, 67). Similarly, in our study, the frequency with which participants mentioned listening to the bass suggests that many people do not hear harmonies as simultaneous Gestalts. Instead, they are likely focusing on individual voices that provide the quickest route into harmonic understanding (typically the bass).

These observations suggest a way to untangle the two groups of skills involved in harmonic dictation as currently practiced—one grounded in perception, the other in logic. In terms of harmonic perception, the primary attributes that people seem to attend to are stock patterns and the bass. To turn these perceptions into roman numerals, however, requires a certain amount of logic and knowledge; specifically, the construction of the patterns, stylistically appropriate norms of harmonic progression, and common associations between bass scale degrees and RN/IS labels. This knowledge is, of course, important to music theory study. Therefore, as currently practiced, harmonic dictation in aural skills classes requires both perceptual skills and logic/knowledge, meaning that success in music theory is necessary for success in aural skills. This is not true the other way around, which again lays bare our current, implicit priorities. One could, however, imagine an aural skills class that focuses on perceptual skills alone. Students in such classes might be asked for bass-line dictation alone, to reproduce the bass vocally, and to describe cadences; they might also do extensive practice listening for (and distinguishing among) common patterns.

This relatively in-depth exploration of harmonic hearing is intended as a specific case to demonstrate how shifting our focus towards perception might alter what we do in aural skills classes more generally. It is not possible here to examine every aural skills task—melodic and rhythmic dictation, sight reading, improvisation, listening for form, etc. The goal is rather to
recognize that how we approach all of these activities would be different if we undertook seriously this change in focus. Some activities might be less affected; certain kinds of simple dictation, for example, require fairly little knowledge-based mediation beyond understanding of notation. That being said, we should make sure that every activity we incorporate in our teaching is based in perception and on learning goals, or else we will be grading our students’ understanding of music theory at least as much as their perceptual skills—in effect doubling the penalty we extract for faulty knowledge structures, since music-theoretical understanding is already graded in music theory courses.

If we are convinced that the prioritization of logic over perception in aural skills classes is a problem, then inverting that relationship must take us beyond activity planning and into the realm of curriculum (re)design. After all, we teach students harmonic dictation at least in part because harmony is a central focus of music theory. The issue of curricular change is examined in the next two sections, which examine respectively how we might prioritize the “truly aural skills,” and then how we might reflect this foundation in our curricula.

**Determining the Truly Aural Skills**

[3.1] If we want to reorient aural skills curricula towards cognition and perception rather than logic and systems, then we need a way to determine: which are the truly aural skills?

[3.2] I propose that what makes a skill truly aural is the degree to which it directly engages working memory. Working memory, the cognitive system that coordinates the short-term storage and manipulation of information, is strongly correlated with the ability to control one’s attention. All listening and performing skills, and presumably all music classes, engage working memory and attention, but they do so with greater and lesser degrees of mediation through knowledge-based information. This principle acknowledges the ways that knowledge and perception are intimately intertwined: it is probably impossible to have one without the other. In addition, it allows us to distinguish skills which are “more fundamentally aural” from those which are “more mediated.”

[3.3] As working memory is a complex subject, my discussion here will be limited to a few of its most pertinent aspects. First, working memory is of limited capacity. Most aural skills instructors are familiar with George Miller’s famous claim that humans can store between five and nine chunks of information for immediate use; modern researchers tend to place that limit lower, perhaps between three and four (Cowan, Chen, and Rouder 2004, 639). Second, working memory and the ability to control one’s focus of attention are strongly linked—and attentional focus is an important aspect of aural skills training (for example, training attention on bass lines). (The reason I refer to “working memory” rather than “short-term memory” is that manipulation of information appears to interfere with short-term storage capacity, suggesting that manipulation and storage share a mechanism.) Finally, working memory capacity may be fixed by adulthood (Melby-Lervåg and Hulme 2013). When we work with our students on perceptual skills, then, our focus should not be on increasing the capacity of their memory (except through established mechanisms such as chunking) but rather on helping them make their use of memory, manipulation, and focus more efficient and flexible.

[3.4] Decoupling aural skills training from its foundation on music theory concepts and ordering in favor of focusing on foundational perceptual skills yields a much broader view of what might be included in aural skills classes. Examples include evaluating ensemble balance like a conductor or chamber musician or identifying the spatial location of a recorded sound like an audio engineer. Many recent aural skills textbooks incorporate improvisation, listening for timbre, and even using embodied approaches to understanding music. But our classrooms could fully embrace the approaches to improvisation practiced in music therapy, the kinds of timbre/instrument categorization used by ethnomusicologists, and the embodied approaches to phrase shaping practiced in some private studios. None of these is heavily mediated by knowledge structures.
This proposal does not require that we banish deep and logical thinking from aural skills. In fact, applying music theory knowledge to the interpretation of perceived stimuli is a classic use of working memory, as it involves both storage and manipulation of information and a high degree of control of attentional focus. But so are evaluating the balance and blend of an ensemble, style-specific improvisation, and identifying spatial characteristics of a recorded sound environment. None of these, including the traditional tasks, will work without a solid foundation in truly aural skills—the “perceptual fundamentals.” When we focus exclusively on aural tasks filtered through the lens of music theory, we are likely to be successful in teaching students who already have the ability to imagine sound internally, determine tonic, move at a periodic rate to music, etc. But we will not necessarily build the foundation of perceptual skills in all students that is required for success in the wide variety of classes that build on these skills—including those outside of music theory.

If we fully embrace a responsibility to focus aural skills courses on perceptual fundamentals, then we must move these curricula beyond their reliance on music theory alone. Example 1 demonstrated the extent to which aural skills and music theory are currently linked; Example 15 proposes an alternative model of aural skills instruction that reflects the relationships between perception-focused aural skills and all areas of study within music—indeed, the centrality of aural-perceptual study to every kind of music study. Around the edge of the outer circle are listed a number of subfields of music study. Various aural tasks are then distributed around the diagram. The most fundamentally aural of these—the perceptual fundamentals—are placed nearest the center, suggesting that they are not tied to a specific discipline but rather foundational for all. (For the sake of space, many tasks are presented in abbreviated form; most items that do not include a verb should be understood to imply something similar to “listening for . . .” or “aurally identifying . . .”.) Tasks that require greater degrees of mediating knowledge radiate out from this fundamentally aural center towards the subfield to which they are most relevant.

The tasks appearing in the middle of the diagram, the “perceptual fundamentals,” are those that require almost no mediation through knowledge and that are necessary for further skills acquisition. I would argue that these include attentional control, basic facility with singing, developing “internal hearing” or “audiation,” basic pitch memory, memorization skills such as chunking and building sound-symbol relationships, metric entrainment skills, and identifying repetition and contrast at different temporal scales. No matter what final form this group takes, however, it is crucial that we emphasize them in our teaching.

Given the complex relationships between perception and knowledge, the placement of skills on the diagram should not be considered precise. For example, identifying modulations—a complex task which I have placed in the outer ring near the top of the diagram—involves component skills such as detecting modulations, which may be fundamentally aural; detecting whether they are closely related or not, which might be perceived fairly directly; evaluating possible exact key relationships, which is a far more mediated task; and then perhaps negotiating between memory, this understanding of the modulation, and understanding of notation to come up with a melodic or harmonic dictation, the most mediated task of all. Different disciplinary biases and priorities might also shift these topics around the diagram. For example, a teacher who wishes to decenter European-derived music in their curriculum might not consider learning solfege to be foundational.

The center of the diagram reveals another similarity between music theory and aural skills, which is that both curricula impart skills that are fundamental to virtually all music study. Michael R. Rogers points out that while fundamentals of notation (scales, do–re–mi, roman numerals, and key signatures) are intimately associated with music theory, they “really represent a pre-theory stage” and are “no more a part of the genuine study of music than knowledge of the alphabet, verbs, or commas is a part of the study of literature” (2004, 3). Rogers’s mention of “do–re–mi” seems appropriate: the identification of specific scale degrees need not be considered unique to aural skills classes, as it is important in error detection for conductors, audio producers, and chamber musicians; in sight reading for performing musicians; and in improvisation for music therapists. Similarly, the eye motions used in effective sight reading (Puurtinen 2018) are not
specific to one part of a music curriculum. These might appropriately be considered fundamental
tasks. It is only once one travels further outward in the diagram, towards the field of Music Theory,
that one reaches traditional aural skills tasks that are less “truly aural.”

Curricular Implications

[4.1] How might one put this proposal of centering aural skills on perceptual skills through the lens
of working memory into practice? Is it possible to do so while preserving some version of the
traditional goals of current aural skills pedagogy? This section will sketch out several potential
solutions, proposing two more radical options and a number of other approaches that could be
more easily integrated within current curricular models.

[4.2] The diagram in Example 15 affirms the fact that all music study involves aural skills. When
music education students conduct ensembles, they are exercising aural skills—in the sense of skills
related to the perception and creation of music. When music therapy students play songs for a
client by request, they, too, are exercising this kind of aural skills. What role, then, should the
classes called “aural skills” play in formal music training? I would argue that the answer could go
two different ways. It might argue, on the one hand, for preserving the current role of these classes
as listening skills applied to the music theory curriculum, which might inspire us to create aural
skills classes with alternative curricular focuses. Alternatively, it might advocate for limiting the
aural skills curriculum to truly fundamental skills and moving high-level listening skills applied to
music theory topics into the music theory curriculum.

[4.3] If we opt to leave aural skills curricula unchanged, then we should explicitly acknowledge
that, especially as one gets to more advanced classes, they are focused on listening skills applied to
the music theory curriculum. This admission of redundancy logically implies a need for other,
truly separate “aural skills” courses. After all, if we currently have “Music Theory Aural Skills,”
perhaps we need new high-level courses such as “Conducting Aural Skills,” “Music Therapy Aural
Skills,” “Jazz Aural Skills,” “Chamber Ensemble Aural Skills,” “Audio Production Aural Skills.”
(30) Giving students such options would promote two recent trends in curriculum development: an
increasing value placed on application of material (Schubert 2013, Gawboy 2013, and Duker,
Shaffer, and Stevens 2014) and an increasing value placed on choice and customizability in order to
facilitate access and success for students with diverse musical skills (see, e.g., Covach 2017). On the
other hand, this model has significant drawbacks, perhaps the most obvious of which is the
impractical proliferation of course offerings (and, presumably, facilities and faculty). In addition,
this idea might seem a bit silly. After all, students already practice their conducting aural skills
when they lead rehearsals, their jazz aural skills when they improvise in their ensembles, and their
chamber ensemble aural skills when they meet in their trios and quartets.

[4.4] The contrasting model entails focusing the aural skills curriculum on the most fundamentally
aural skills—possibly reducing the number of required aural skills classes—and relocating high-
level, knowledge-mediated listening skills to specialized classes associated with each subfield. To
an extent, this is already how many subfields already operate. For example, students in music
history classes are commonly given listening quizzes on canonical repertoires and use their newly
acquired knowledge of style to aurally identify likely composers and dates of music they have not
encountered before. Music theory, too, could work in this manner. For example, a quiz on
secondary dominants in a music theory class could include a series of listening-based questions
asking students to identify whether a recorded passage includes a secondary dominant or not
and/or which chord is tonicized. (31) This would, of course, require that we set aside time in music
theory class to work on listening.

[4.5] A likely reaction from some to this proposed change is, “but what material must we give up in
order to work on listening skills in music theory?” Let us consider this response, critically, as an
indication of the extent to which we have used the existence of aural skills classes to curtail
listening opportunities in music theory classes in the interests of covering more “content.” New
theoretic content is wonderful, but if we don’t work in rigorous ways on hearing it, are students
really learning it? If not, what are we really giving up if we in jettison a portion of that content? (32)
It is worth noting that reformulating our aural skills classes might also free up much of the time and energy currently dedicated to traditional sight singing and dictation drills and inspire a diversification of the kinds of aural tasks we ask students to do. No rule exists that says students must perform harmonic dictation in order to “learn” the augmented sixth chord. It might be just as effective and more time-efficient for them to learn to play progressions on an instrument or to memorize a number of representative listening examples that include this chord. (33)

[4.6] These options provide a new perspective on the longstanding debate regarding integration vs. separation of music theory and aural skills classes. (34) An “integrated” approach partly embraces the first curricular option proposed above, effectively linking the mediated aural skills tasks into a closer relationship with music theory. At the same time, it risks overlooking the development of the more fundamentally aural skills, which do not rely on music theory knowledge. A “separated” curriculum has more flexibility to focus on perceptual fundamentals, but to the extent that its aural skills courses incorporate advanced music theory topics, it runs the risk of conflating perceptual fundamentals with skills mediated through music theory. From the perspective of Example 15, then, both models suffer from their inability to distinguish perceptual fundamentals and more mediated skills.

[4.7] Focusing aural skills classes on the perceptual fundamentals would dramatically alter how aural skills classes operate, so it is worth suggesting what kinds of topics might be appropriate to a “Truly Aural Skills” class. These might include the following:

- attentional control
- sing-backs
- solfege
- moving to music at a periodic rate (35)
- determining relationships among levels of pulse, determining tonic and collection (36)
- increasing musical memory capacity by considering music in chunks
- developing and manipulating internal hearing (Klonoski 1998 and Gates 2021)
- practicing memorization
- learning common rhythmic cells
- coordinating with other musicians (particularly in terms of tempo)
- identifying repetition and contrast (37)

Several of these tasks are highly complex and would need to be broken into sub-topics. For example, attentional control can be directed in many different ways, including towards different temporal parts of a passage (Karpinski 2000’s “extractive listening”), towards different contrapuntal strands within a passage (including listening for bass lines) and towards listening for different aspects of sound (evaluating balance and blend, describing dynamic, timbre, or location in space). (38) A fully fledged new approach to aural skills curricula will require years of collaboration for the pedagogy community to develop, but this preliminary list gives some idea of the scope of the truly aural skills that we might investigate further.

[4.8] There are, fortunately, simpler and more immediate options for making aural skills classes more grounded in perception. For example, authors and teachers can incorporate chapters and units on perceptual fundamentals into otherwise traditional aural skills textbooks and courses. Units might address, for example, “developing internal hearing,” “control of attention,” “increasing musical memory,” “responding to music through motion,” etc. One of the most important overlooked topics, even from the perspective of traditional aural skills instruction, is control of attention. In addition to its potential to facilitate better performance on single-stimulus tasks such as melodic dictation and sight reading, control of attention is particularly crucial for listening to bass lines and harmony (since one must focus on specific aspects of a heard sound
within a multiple-voice texture) and hearing modulation (since the introduction of a new diatonic context complicates the task of focusing on and memorizing the music in the original key).

[4.9] Perhaps the most practical and immediate response, however, is this: we can acknowledge our deep-seated preference for logic over perception in aural skills classes and work to minimize its effects; and we can acknowledge that listening tasks are not as well integrated as they should be into music theory classes and work to correct this. In aural skills, instead of harmonic dictation with roman numerals and inversion symbols, we might ask for a bass line dictation with cadences labeled by type and chords of interest circled. A possible learning objective here might be: “by the end of the course, students will be able to reproduce and transcribe bass lines and identify the chords used in cadences.” When we teach melodic dictation for the first time, we might introduce it by working on basic memory and attentional-focus exercises. The learning objective in this case might be: “by the end of the course, students will be able to selectively remember and transcribe either the first half or the second half of a melody that is too long to remember in its entirety.”

In music theory, when we teach primary chords, we might add some aural chord-progression identification questions to our quizzes. The learning objective here might state: “by the end of the course, students will be able to define, construct/identify in notation, and aurally identify primary chords using RN/IS labels.” When we teach counterpoint, we might ask for students to record every exercise on solfege (possible learning objective: “by the end of the course, students will be able to compose and produce, without use of an instrument, phrases of first, second, and third species counterpoint”).

[4.10] Whatever path we choose, there are both challenges and opportunities in this proposal to focus on truly aural skills instead of mechanically applying sight reading and dictation to music theory content. It will be challenging to enact this change. Doing so will require us to explore what we know about perception and the mechanisms that exist for fine-tuning it, and to experiment with different models of instruction. The payoffs will be handsome, however. If we find ways to place our students’ perceptual fundamentals on firmer ground, then we will facilitate success for a greater diversity of students, and aural skills classes can move out of their current state of curricular dependency on music theory and into a new position of prominence as the foundation of all music study.

Timothy Chenee
Department of Music
Utah State University
4015 Old Main Hill
Logan, UT 84322
timothy.chenee@usu.edu

Works Cited


**Footnotes**

1. Some scholars may consider aural skills (and aural skills textbooks) to be a subcategory of music theory (and music theory textbooks). Throughout this article, for reasons that I hope will become clear, I prefer to consider these different categories. I avoid the commonly used term “written music theory,” however, because I will argue that music theory courses are more effective when they include aural (and performance-based) tasks and ways of thinking.

Return to text

2. Klonoski’s article questions the “tacit assumption” that “the sequence of topics typically found in tonal theory texts, normally a highly refined and logical conceptual ordering, also represents the optimal perceptual ordering” (2000). His discussion includes several suggestions on reordering material, but focuses particularly on the importance of auditory imagery as a “perceptual fundamental.”

Return to text

3. Another problem inherited from music theory curricula concerns repertoire and representation: Like music theory curricula, aural skills curricula often focus heavily on common-practice European music at the expense of other repertoires. Duinker and Gauvin 2017 identified eight European males of the common-practice period who were among the top ten composers represented in each of the five major music theory textbooks they examined; seven of these eight are also among the top ten represented among Karpinski 2007’s dictation melodies: J. S. Bach (44 excerpts), Franz Schubert (31), Ludwig von Beethoven (25), Wolfgang Amadeus Mozart (21), Franz Joseph Haydn (16), Robert Schumann (13), and Johannes Brahms (5). Other composers often represented in the first edition of the Karpinski Manual include George Frideric Handel (8), Pyotr Ilyich Tchaikovsky (7), and Béla Bartók (5). (The 2017 second edition does not include an index but does not appear to represent a change in this regard.) This problem extends also to performance skills, since aural skills classes tend to focus on sight-reading. This excludes both repertoires where a written score is not the primary means of dissemination, such as recorded popular music and many traditional repertoires from around the world, and repertoires that are typically conceptualized and communicated in a different notation system.

Return to text

4. One significant exception is Duker and Stevens 2017, which clearly defines action-oriented learning goals. In contrast to my approach, however, Duker and Stevens do not adopt an explicitly cognitive/perceptual perspective.

Return to text

5. Of course, the prioritization of logic and systems over perception by intellectuals stretches back practically to the beginning of recorded music theory. According to Boethius, “it is far greater and nobler to know what someone does than to accomplish oneself what someone else knows, for physical skill obeys like a handmaid while reason rules like a mistress…. How much more admirable, then, is the science of music in apprehending by reason than in accomplishing by work and deed!” (McKinnon 1998, 32).

Return to text

6. Huron 2006 points out that encoding an interval in the brain requires some representation of the pitches themselves—so it is likely that some direct representation of pitch identity, probably scale degree, is foundational to the way the brain encodes music, rather than interval (122–127).

Return to text
7. By this term I mean not identification of isolated chord qualities but rather the recognition of chord relations within a tonal context (usually a phrase or more). While this task is typically framed as “harmonic dictation,” I prefer the term “harmonic identification,” to emphasize the fundamental skill that is relevant to a number of different situations (error detection in ensembles, jazz improvisation, transcription, etc.) rather than the task that we do in aural skills class. Nevertheless, the discussion below will also refer to harmonic dictation.

Return to text

8. Many textbooks also drill bass-line dictation, outer-voice dictation, or even four-voice dictation. Nevertheless, the discussion below will focus on roman numerals and inversion symbols because they are typically the core of how harmony is discussed in these textbooks. For textbooks that require bass-line notation to accompany harmonic dictations, the inversion symbol is redundant; presumably these texts ask for the inversion symbol in order to reinforce the roman numeral/inversion symbol pairs learned in music theory class—another example of how aural skills curricula are based on music theory curricula.

Return to text

9. Throughout this article, I draw on dictations from Karpinski 2017a both because it is popular—Murphy and McConville reported that it was the most-often-used dictation text among their respondents (the two more popular aural skills texts they list are sight singing anthologies with no dictation excerpts; 2017, 211)—and because it explicitly draws on cognitive science. The note “To the Instructor” in Karpinski 2017b states, “The structure and content of this book have been shaped in large part by recent research in music cognition and perception” (xiii). One aspect of the text that may have been influenced by this concern for cognition and perception is that it presents two-part and bass-line dictation before adding RN/IS labels.

Return to text

10. Weinstein, Sumeracki, and Caviglioli 2019 state, “The data point strongly to the conclusion that it is almost impossible to pay attention to more than one thing at the exact same time” (52). This has been applied to musical meter by Justin London, who points out that listeners confronted with polyrhythmic stimuli will either extract a composite pattern or focus on one rhythmic pattern at the expense of the other. In other words, “because the need to maintain a single coherent ground seems to be universal . . . there is no such thing as a polymeter” (2012, 67; italics in original).

Return to text

11. Even if we manage to teach students Gestalt hearing in the context of four-voice, chorale-style dictation excerpts played in the middle register of the piano, it is not clear that this mode of hearing will transfer to more ecologically valid contexts.

Return to text

12. It is true that those experienced with music theory and chord identification can typically deduce stylistic-appropriate harmony from a known bass line, but this probably indicates expertise in music theoretical understanding more than perception.

Return to text

13. If the answer is yes, we might further ask, “is it meaningful enough to justify the amount of time it requires to learn to make this distinction, given other priorities that we must ignore to do so?”

Return to text

14. Karpinski’s text under the heading “Listening for Supertonic Triads” reads in part, “When $\text{4/6}$/$\text{a}$ in the bass supports a $\frac{5}{4}$ chord the result is the subdominant triad—IV or iv. However, when $\text{4/6}$/$\text{a}$ in the bass supports a $\frac{5}{4}$ chord the result is the supertonic triad in first inversion—ii$^6$ or ii$^6$. The differences will be $\frac{1}{4}$/DO or $\frac{2}{4}$/RE in an upper voice and the quality of the chord (major, minor, or diminished)” (Karpinski 2017b, 194). Note that Karpinski works from the bass—an element that is not directly indicated in the chord label—but also emphasizes root and quality.

Return to text
15. Textbooks and curricula that test harmonic hearing in part through outer-voice transcriptions will of course register the accented dissonance of Example 5. The fact that changing an inner voice from scale degree 1 to scale degree 2 alters the RN/IS while the accented dissonance alters only the melody may suggest to students that these are alterations in different aspects of the passage and that the former is more fundamentally harmonic.

Return to text

16. If the root, quality, and inversion of the roman numeral are each accorded their own portion of the grade and graded mechanically, the distinction between ii⁰⁶ and iv will be triply penalized, and even a person who hears the correct function and bass note here may be considered entirely incorrect.

Return to text

17. Many thanks to my colleague at Utah State University, Dr. Cahill Smith, for these recordings.

Return to text

18. This placement of this dictation in Karpinski’s chapter on “Successive Modulation” means that it does not primarily concern distinguishing among V₂, V₃, and vii⁰⁶. Nevertheless, Karpinski does suggest through the answer key’s roman numeral that this chord’s identity is important. The distinction between these three chords is addressed in the pedagogical literature, by the otherwise rather progressive Kleppinger 2017. As I argue above, before devoting time to this distinction, we should first make sure we know why we are asking students to master it; to borrow Kleppinger’s remarks about distinguishing “nationalities” of augmented sixth chords, “does its priority represent the amount of time and reinforcement required to become proficient at mastering and later recalling this skill?” (158).

Return to text

19. The role of kinesthetic imagery in expert musical performance has been well documented (e.g., Lotze 2013). Its role in aural skills instruction, however, is under-studied. While I focus on pianists here, it would make sense for harpists, marimbaists, vibraphone players, and guitarists to have a similar advantage in taking harmonic dictation.

Return to text

20. Concerns about students whose faulty music theory knowledge results in a logically nonsensical label can be found in Stevens 2016 and Jarvis 2015, both of which focus on the distinction between IV and V₁. Jarvis (2015) writes “I am often troubled by the fact that my sophomore aural-skills students confuse V₁ with IV during simple harmonic dictation exercises (e.g., I V₁ I V). As the years go by, it gets increasingly difficult for me to imagine what that experience must be like for them.” This last sentence is a perfect description of a phenomenon called the “curse of knowledge,” where an educated individual (in this case, educated in the systems of harmony typically described in music theory classes) cannot imagine or anticipate the thought processes of a novice that are presumably less mediated by those same knowledge structures.

Return to text

21. The excerpts were presented as short, embedded audio excerpts (Karpinski and Mozart) and an embedded YouTube video of the whole song (Beyoncé). Participants could listen to each as many times as they wished.

Return to text

22. One question asked, “do you think you might have some level of absolute pitch? If so, please briefly describe it. If not, please indicate by typing ‘no.’” Responses were categorized as absolute pitch (AP) if respondents simply answered “yes” or described a high degree of accuracy and immediacy, and as heightened tonal memory (HTM) if respondents described being able to identify only certain tones, certain timbres, etc. These categorizations are based on Ross, Gore, and Marks 2005. Ross, Gore, and Marks prefer “APE” (“ability to perceptually encode”) rather than “AP”; the latter term, however, is simpler in the current context. Participants were also asked to identify all instruments they play and, for each, the number of years they had taken formal lessons.
Primary instruments were determined either by the instrument with the highest number of years of lessons or, for those who neglected to add numbers, the first instrument listed. The average years of formal study was calculated based on the number reported for the primary instrument, since years on a secondary instrument may have overlapped with these; participants who did not indicate their number of years of training were excluded from this average. Because of the anonymous online survey format, it is impossible to verify absolute accuracy of all answers.

23. Notably, within this “Nearly Correct” group, two-thirds (4 out of the 6) of those who indicated that they had completed the course of aural skills training at their institution (or had graduated with a music degree) misidentified this chord as IV. While this sample size is too small to indicate whether this kind of error is likely to be corrected in the course of aural skills training, these data are not promising.

24. If students have been primarily exposed to popular music but their higher-education curriculum focuses on Mozart, Haydn, and Bach, then the kind of exposure necessary for effective statistical learning may not yet have taken place. And while harmonic dictation may be an effective way to determine whether students have effectively accomplished appropriate statistical learning about chord progressions in the style at hand, it is a remarkably inefficient way to expose them to stylistic norms since it typically focuses on a single phrase for ten minutes or more at a time.

25. For instructors who are worried that this suggests dropping roman numerals from listening exercises altogether, such RN/IS-based listening could easily be integrated into music theory classes, as suggested in Section 4 below.


27. For a more detailed examination of recent research on working memory with application to aural skills, see Chenee 2018.

28. Neither the list of subfields around the diagram nor the lists of aural skills within the diagram are intended to be comprehensive; to list every aural skill relevant to every subdiscipline would require a much larger page than is available here. In addition, many of the listed skills could be related to multiple categories; for example, “cueing” is important to both conducting and music education, while listening for timbres is important in different ways to ethnomusicologists, audio engineers, music historians, performers, and conductors. For the most part, I have placed such potential duplicates in just a single category for the sake of space.

29. It is worth noting that this distinction may be (in part) why aural skills classes are often described both as crucial and, perhaps equally often, as irrelevant. To take but two seemingly opposed examples among many, Wright 2016 found that both professional musicians and music students identified “aural” as crucial for their education and careers; Barlow 2016, on the other hand, suggests that “it can very easily be seen as irrelevant” (23). Perhaps the “truly aural skills”—the perceptual fundamentals—are those most important for a music education and career, while the “aural skills” radiating out towards the field of music theory in Example 15 are those that are less obviously relevant to some.

30. Indeed, there are already textbooks devoted to some of these areas, including Jazz Ears: Aural Skills for the Improvising Musician (Mason 1997) and Audio Production and Critical Listening (Corey 2017). The existence of dedicated aural skills textbooks for jazz and audio production might
suggest that practitioners of these two subfields feel particularly overlooked by current aural skills models—and also that they place a high value on aural training.

31. This might also help make sure we are thinking about exactly which skill we are testing. As Kleppinger 2017 points out, traditional grading applied to many high-level dictation tasks may not effectively assess the skill at hand; Kleppinger’s goal in his article is “to inspire introspection about what our aural skills assessment methods actually measure, the expertise we intend for students to gain from this part of their music studies, and the potentially dangerous distance between these two things” (153). Kleppinger focuses entirely on methods of assessment, explicitly leaving aside questions of curriculum change.

32. I do not here intend to denigrate more speculative approaches to music theory, many of which do not require listening for understanding and which I believe should be embraced as part of the liberal arts mission of developing students’ critical thinking abilities and knowledge of musical systems of all kinds. But for obvious curricular reasons, most core undergraduate theory “content” is framed as in some way “useful” in developing listening or performing skills. It is difficult to argue this, however, if we do not explicitly work on developing those skills.

33. Of course, “learn the augmented sixth chord” is a content goal, not a learning objective—but since this article is focused on aural skills courses, I leave the consideration of appropriate music theory learning objectives relevant to this chord for another time.

34. Rogers 2004 describes the “integrated approach” as that which “mixes ear training, analysis, and composition within a single unified course each semester” (16).

35. Dalcroze Eurythmics, which has been applied to contemporary music theory/aural skills instruction by Urista 2016 and Ristow, Thomsen, and Urista 2014, is an excellent source of ideas here.

36. Karpinski 2000 suggests some ideas for understanding this skill and how to develop it (39–48); methods for improving this skill are examined in more detail in Chenette 2020.

37. Though I have not stated these as fully fledged learning goals, they are action-oriented and thus amenable to being phrased as such.

38. Corey 2017 includes instruction and exercises intended to improve this kind of listening for different aspects of sound.

39. Karpinski argues for the importance of this skill, which he calls “extractive listening,” and gives techniques for its improvement (2000, 71–73).

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.

Prepared by Lauren Irschick, Editorial Assistant